TOPICAL REVIEW • OPEN ACCESS

Cooling hot cities: a systematic and critical review of the numerical modelling literature

To cite this article: E Scott Krayenhoff *et al* 2021 *Environ. Res. Lett.* **16** 053007

View the article online for updates and enhancements.

ENVIRONMENTAL RESEARCH
LETTERS

TOPICAL REVIEW

Cooling hot cities: a systematic and critical review of the numerical modelling literature

E Scott Krayenhoff[1,2,3] 🄳, Ashley M Broadbent[2,3], Lei Zhao[4] 🄳, Matei Georgescu[2,3] 🄳, Ariane Middel[2,5,6] 🄳, James A Voogt[7] 🄳, Alberto Martilli[8], David J Sailor[2,3] 🄳 and Evyatar Erell[9,10]

1   School of Environmental Sciences, University of Guelph, Guelph, ON, Canada
2   Urban Climate Research Center, Arizona State University, Tempe, AZ, United States of America
3   School of Geographical Sciences and Urban Planning, Arizona State University, Tempe, AZ, United States of America
4   Department of Civil and Environmental Engineering, University of Illinois, Urbana-Champaign, IL, United States of America
5   School of Arts, Media and Engineering, Arizona State University, Tempe, AZ, United States of America
6   School of Computing, Informatics, and Decision Systems Engineering, Arizona State University, Tempe, AZ, United States of America
7   Department of Geography, Western University, London, ON, Canada
8   Atmospheric Pollution Division, Department of Environment, CIEMAT, Madrid, Spain
9   J. Blaustein Institutes for Desert Research, Ben-Gurion University of the Negev, Sede Boqer Campus, Sede Boqer 8499000, Israel
10 Department of Geography and Environmental Development, Ben-Gurion University of the Negev, Sede Boqer Campus, Sede Boqer 8499000, Israel

E-mail: skrayenh@uoguelph.ca

Abstract

Infrastructure-based heat reduction strategies can help cities adapt to high temperatures, but simulations of their cooling potential yield widely varying predictions. We systematically review 146 studies from 1987 to 2017 that conduct physically based numerical modelling of urban air temperature reduction resulting from green-blue infrastructure and reflective materials. Studies are grouped into two modelling scales: neighbourhood scale, building-resolving (i.e. microscale); and city scale, neighbourhood-resolving (i.e. mesoscale). Street tree cooling has primarily been assessed at the microscale, whereas mesoscale modelling has favoured reflective roof treatments, which are attributed to model physics limitations at each scale. We develop 25 criteria to assess contextualization and reliability of each study based on metadata reporting and methodological quality, respectively. Studies have shortcomings with respect to neighbourhood characterization, reporting areal coverages of heat mitigation implementations, evaluation of base case simulations, and evaluation of modelled physical processes relevant to heat reduction. To aid comparison among studies, we introduce two metrics: the albedo cooling effectiveness (ACE), and the vegetation cooling effectiveness (VCE). A sub-sample of 47 higher quality studies suggests that high reflectivity coatings or materials offer ≈0.2 °C–0.6 °C cooling per 0.10 neighbourhood albedo increase, and that trees yield ≈0.3 °C cooling per 0.10 canopy cover increase, for afternoon clear-sky summer conditions. VCE of low vegetation and green roofs varies more strongly between studies. Both ACE and VCE exhibit a striking dependence on model choice and model scale, particularly for albedo and roof-level implementations, suggesting that much of the variation of cooling magnitudes between studies may be attributed to model physics representation. We conclude that evaluation of the base case simulation is not a sufficient prerequisite for accurate simulation of heat mitigation strategy cooling. We identify a three-phase framework for assessment of the suitability of a numerical model for a heat mitigation experiment, which emphasizes assessment of urban canopy layer mixing and of the physical processes associated with the heat reduction implementation. Based on our findings, we include recommendations for optimal design and communication of urban heat mitigation simulation studies.

1. Introduction

Urban areas tend to be hotter than their non-developed surrounds, and they are projected to further warm over the 21st century due to global climate change and urban development (Rosenzweig *et al* 2015, Krayenhoff *et al* 2018, Zhao *et al* 2021). Without appropriate adaptation, high temperatures in low and mid-latitude cities are likely to have increasingly adverse effects on health, thermal comfort and energy consumption outcomes during warmer seasons. While these outcomes have multiple causative factors, air temperature is strongly implicated in all of them and has been widely studied.

Intentional modification of urban landscapes can reduce air temperature locally. Common cooling strategies include green infrastructure (Bowler *et al* 2010), bluespace and irrigation (Gunawardena *et al* 2017, Broadbent *et al* 2018a), and use of reflective materials or coatings (Taha *et al* 1988, Krayenhoff and Voogt 2010). Green infrastructure includes street trees, green roofs and facades, parks, lawns and gardens, while bluespace consists of static and dynamic surface waters such as ponds, rivers and lakes. These strategies generate cooling by reflecting solar radiation away from the urban environment or redirecting absorbed solar energy from sensible to latent heat (figure 1).

A priori assessment of potential cooling impacts is helpful to decision makers. Numerous observational studies have reported cooling from urban heat reduction strategies (Bowler *et al* 2010, Aleksandrowicz *et al* 2017, Santamouris *et al* 2017), each reflecting the specific characteristics of the sites monitored and the weather conditions during the measurement period, therefore making generalization difficult. Numerical modelling is generally less demanding than *in situ* assessment of cooling infrastructure impacts and provides ease of experimental control. However, numerical modelling relies, by definition, on imperfect abstractions of urban atmospheric processes, and it is applied at multiple scales and includes widely varying methodological approaches, each of which has received different testing against measured data.

The burgeoning literature focused on numerical simulation of micro-to-regional scale climate effects of urban heat reduction strategies, reported in several recent literature surveys (Krayenhoff and Voogt 2010, Gago *et al* 2013, Santamouris 2014, Aleksandrowicz *et al* 2017, Gunawardena *et al* 2017, Santamouris *et al* 2017, Toparlar *et al* 2017, Pisello *et al* 2018, Tsoka *et al* 2018b, Lai *et al* 2019, Ampatzidis and Kershaw 2020), has so far failed to provide a consistent framework for comparing results systematically. Existing simulation studies vary in terms of spatial and temporal

scale and specificity, background meteorology, neighbourhood context, numerical model construction and application, type and intensity of heat mitigation implementation, and air temperature definition (see table 1 for a summary of degrees of freedom). As a result, the range of simulated cooling magnitudes induced by a given urban heat reduction strategy varies widely; a comprehensive review reports cooling that spans a full two orders of magnitude: <0.1 °C to >5 °C (Santamouris *et al* 2017). Furthermore, key metadata, such as neighbourhood morphology, or the specific air volume or location associated with the reported temperature difference, are often unreported. Consequently, methodological quality becomes uncertain, repeatability becomes challenging, and comparison of cooling effectiveness between studies, and therefore between different geographical contexts, is rendered unfeasible.

In this context, the present review develops the first critical framework for evaluation of simulation studies that assess urban heat reduction (or heat mitigation) strategies. We evaluate both metadata reporting and methodological quality. Beyond provision of appropriate context for interpretation of modelled results, sufficient metadata reporting permits inter-study comparison of the cooling effectiveness of commonly proposed heat reduction strategies. Methodological quality gives an indication of the reliability of the results. Reliability of results is enhanced by appropriate representation of physical processes, suitable model evaluation and calibration, and optimal choices related to model application. Critically, this review is comprehensive and systematic, and it is the first review to undertake an informal meta-analysis to derive typical cooling efficacies for common heat reduction strategies from studies that meet a minimum level of reliability.

An individual heat mitigation simulation study tied to a particular place may be of practical relevance to local planners and policy-makers. However, it must be situated among related studies if it is to inform the broader scientific understanding of the local climate impacts of heat reduction strategies. Hence, we propose that a central aim of urban heat mitigation research, in addition to assessment of local impacts, should be the determination of consensus cooling efficacies for common heat reduction strategies, as well as their dependence on a limited number of factors (e.g. time of day, meteorological conditions; see section 5.4). We further define new indices to aid assessment of the cooling efficacy of albedo-based implementations (albedo cooling effectiveness (ACE)) and vegetation-based strategies (vegetation cooling effectiveness (VCE))—see section 2.5.1.
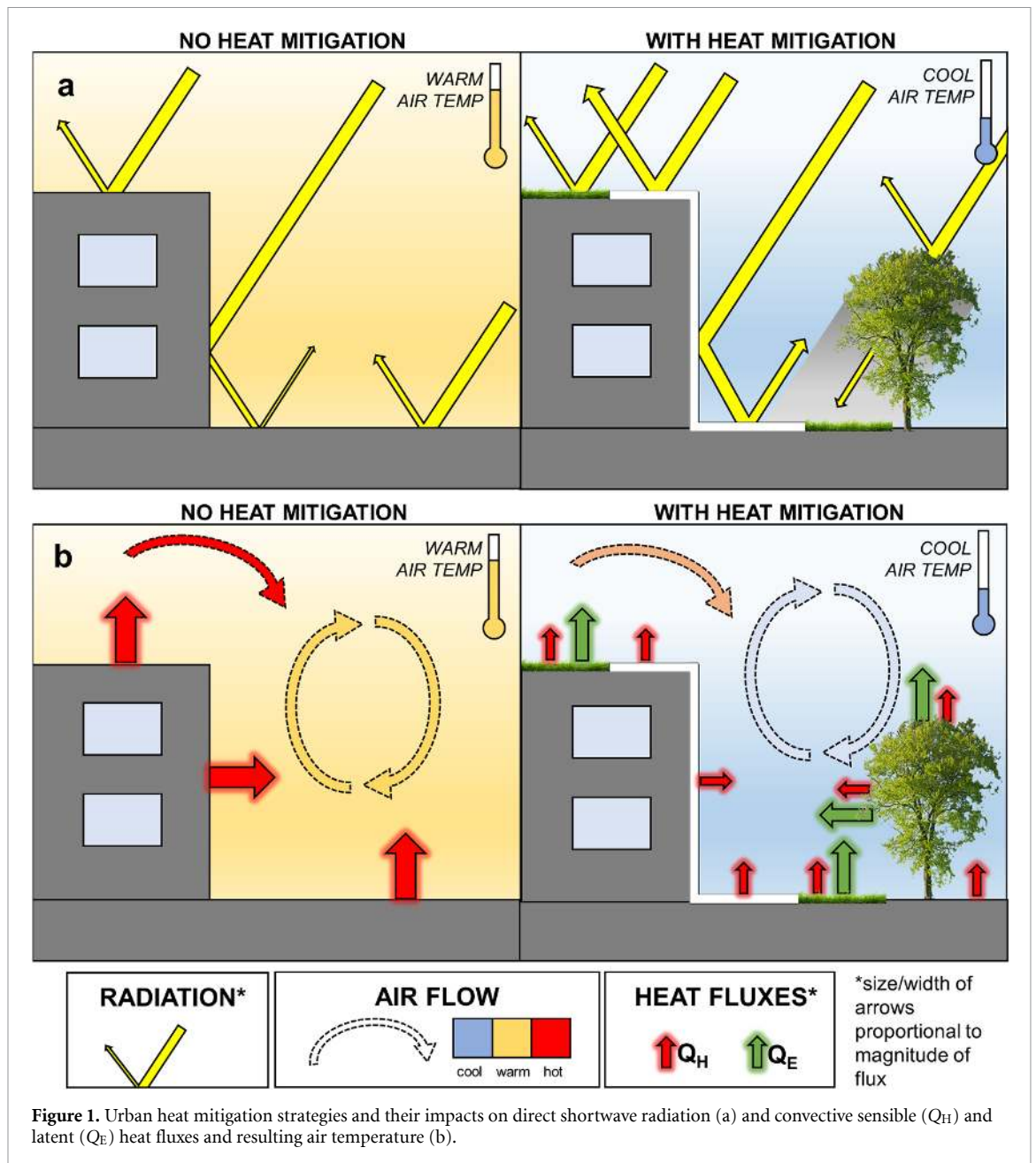
Figure 1. Urban heat mitigation strategies and their impacts on direct shortwave radiation (a) and convective sensible (Q_H) and latent (Q_E) heat fluxes and resulting air temperature (b).

To establish consensus cooling efficacies, urban heat mitigation studies must be reliable and appropriately contextualized so as to be inter-comparable. We define reliability as the trustworthiness of the results, based on a scientifically sound and sufficiently documented modelling methodology. Comparability relies initially on full reporting of context (e.g. meteorology, urban neighbourhood type) and methodological design (e.g. implementation type and intensity), and ultimately on constraint of the many degrees of freedom inherent in urban heat reduction simulation studies (table 1). The local climate zone scheme (Stewart and Oke 2012) represents a similar methodological development in the observational urban heat island literature, aimed at enhancing comparison between measurement sites and studies.

The intent of this systematic review is threefold:

- to assess the urban heat mitigation modelling literature in terms of methodological quality and metadata reporting;
- to derive a state-of-the-art consensus of cooling efficacies associated with implementation of common heat reduction strategies from existing high-quality studies in the literature;
- to recommend key methodological approaches and the reporting of essential metadata for enhanced reliability and comparability of urban heat mitigation studies.

The scope is limited to high albedo materials, photovoltaics and green and blue infrastructure, and to studies that apply numerical modelling; however,

Table 1. Principal degrees of freedom in numerical simulation of air temperature response to urban heat mitigation implementations.

Degree of freedom	Explanation and examples
Context	
Meteorological	Maximum solar angle, mean wind speed, degree of cloudiness, etc
Neighbourhood	Character of the local cover, structure and fabric (e.g. local climate zone)
Tool and application	
Numerical model	Processes represented (scale), and complexity & accuracy of their treatment
Model application	Simulation development, including initial and boundary conditions, model calibration, parameter values and other choices made by researcher
Surface perturbation	
Choice of heat mitigation implementation	Type of surface modification (see supplementary table 2)
Area and intensity of implementation	Spatial coverage and intensity (e.g. degree of albedo or leaf area index increase)
Location of implementation	Roof, walls, ground.
System response	
Vertical specificity	Specific level or layer over which averaging is undertaken (e.g. 2 m air temperature or canopy layer air temperature)
Temporal specificity	Specific time or averaging period corresponding to results (e.g. July 21 diurnal maximum, June–July–August diurnal mean, etc)

the criteria and guidelines herein may inform scale-model and observational work.

2. Methodology

2.1. Defining 'urban heat mitigation'

Infrastructure-based urban heat mitigation can include a broad range of modifications to building envelopes, building and tree configuration, and landscape design. Here, we focus on several commonly investigated heat mitigation strategies that are relevant for both existing and newly-designed neighbourhoods. 'Gray' modifications to neighbourhood configuration, including building morphology and street orientation and engineered shade devices, are excluded from this review in order to constrain the literature sample; although alteration of the built configuration is more difficult in existing neighbourhoods, it is particularly relevant to the design and construction of new neighbourhoods. Thus, this review focuses on non-structural modifications to urban land cover, and/or urban facet properties, that have the intended purpose of cooling local urban climate. All human-wrought changes to urban landscapes—synonymous here with the 'built' series defined by Stewart and Oke (2012)—made with the intention to reduce air temperatures within the urban canopy layer (Oke 1976) or at roof level are considered to be urban heat mitigation implementations for the purposes of this review. These requirements help define the relevant conceptual model and the range of operational tests considered herein (Valiela 2001, Stewart 2011). In many cases, 'urban heat island mitigation' studies qualify, provided that absolute urban cooling is reported in addition to relative cooling (e.g. reduction of urban heat island intensity). However, we do not consider an urban heat island framing to be relevant for discussion of

urban heat mitigation (Martilli *et al* 2020). Urban heat mitigation strategy implementations typically include one or more of the following: street trees, reflective surfaces, short vegetation, permeable surfaces, irrigation, water features, or photovoltaic panels, applied at ground level, on rooftops, or on building walls.

2.2. Literature sample inclusion criteria

The literature sample was drawn from English language, peer-reviewed journal articles published between January 1987 and June 2017. Web of Science search terms with combinations of words related to cities, modelling, mitigation, and cooling and temperature and related searching returned 2414 articles (see supplementary methods (available online at stacks.iop.org/ERL/16/053007/mmedia)). Assessment of these articles by two reviewers was undertaken, qualifying articles for the subsequent review sample when they met all of the following criteria:

- utilizes a physically-based numerical modelling approach;
- employs a model that either yields a steady state solution, or has a prognostic temperature(s)—purely statistical models are excluded;
- simulates effects of urban cover or facet modification (including changes to tree cover) on urban canopy or surface layer air temperature change—studies that exclusively report other temperatures or indices are excluded;
- uses experimental control—studies that do not include an appropriate 'base case' scenario are excluded.

The resulting sample includes 146 articles, employing models that vary from street scale to global scale, from hourly to decadal time scales, and from

past to projected future climates. The articles are divided into two broad categories, namely micro-neighbourhood scale (hereafter 'microscale') models and meso-global scale models (hereafter 'mesoscale'); the former typically resolve buildings but not advection between neighbourhoods or boundary-layer processes (domain size of 100–5000 m, with grid spacing 1–100 m), whereas the latter do not resolve individual buildings but capture the full boundary-layer and city and/or regional scale atmospheric flows (domain size >100 km, with grid spacing 0.5–100 km). The focus here on numerical model results complements reviews of observational results (e.g. Bowler *et al* 2010). Notably, while the critical and quantitative aspects of this review end in 2017, qualitative results from the literature are included up to and including 2020 (e.g. see section 5.3).

2.3. Criteria used to assess literature

Several criteria are developed to evaluate the reliability and comparability of urban heat mitigation studies in the sample. The following four criteria assess metadata reporting, required for contextualization of individual studies and effective comparison between studies.

(1) Site metadata are provided.
(2) Forcing meteorology is characterized.
(3) Heat mitigation implementation metadata are provided.
(4) Air temperature is specified spatially and temporally.

Three subsequent criteria assess appropriateness of the numerical model and its application, and address methodological rigor and subsequent reliability of the modelling results obtained.

(5) Model accurately represents relevant physical processes.
(6) Model evaluation is appropriately targeted and successful.
(7) Model application is sound.

These criteria (described below) are based on the collective experience of the authors, whose training includes building microclimate simulation and atmospheric modelling at micro-to-global scales, largely in North American and European contexts, as well as development of numerical models of urban atmospheres at micro-neighbourhood scales. Each criterion above is sub-divided into multiple sub-criteria to assist scoring, and each study is assigned a grade for each sub-criterion, which represents the degree to which the study fulfils the sub-criterion (table 2; supplementary methods). With the exception of sub-criteria 5c and 5d, grading is based exclusively on information provided in each peer-reviewed article and associated supplementary information;

that is, no other sources are sought and no contact is made with the author(s) at this stage. As such, complete and effective scientific communication is implicitly included as a criterion in the grading scheme (Stewart 2011).

The supplementary methods provide additional detail related to the definition of all criteria and sub-criteria, and table 2 provides an overview of all criteria and sub-criteria and the associated point allotments. There is no universally correct point allotment, and this approach is motivated by a previous critical review of the urban heat island observational literature (Stewart 2011). Since the point allotments are inherently subjective, the total scores for each criterion or set thereof represent value judgments by the authors. We choose point allocations to reflect the relative importance of each sub-criterion with regards to the contextualization and comparability (metadata criteria 1–4) and reliability (methodology criteria 5–7) of the heat mitigation cooling magnitudes reported in the article. Since both reliability and contextualization of the results are essential for meaningful interpretation of modelled cooling magnitudes, they are weighted equally. The total scores for the methodology criteria (criteria 5–7), individually and collectively, are used to assess the suitability of individual studies for inclusion in the informal meta-analysis (see section 2.5). The total scores for metadata criteria (criteria 1–4) are not explicitly used in this work, though associated point allotments in table 2, which are based on considerable deliberation by the authors, are nevertheless provided as additional context and may be useful to the research community in future reviews. The combined scores provide an indication of overall quality of each study.

In brief, the seven criteria are as follows:

(1) Neighbourhood metadata are provided (criterion 1)
The effectiveness of heat mitigation implementations is modulated by the land cover, morphology and materials of the neighbourhood or area surrounding the site. Hence, characterization of the urban neighbourhood provides important context. For example, implementation of high albedo roofs on tall vs. short buildings has been found to strongly modulate pedestrian-level cooling (Botham-Myint *et al* 2015). Suitable metadata include urban structure, cover, and fabric (Oke 2006; supplementary table 1). Alternatively, identification of the local climate zone (Stewart and Oke 2012) provides a confined range of these metadata.

(2) Forcing meteorology is characterized (criterion 2)
Meteorological conditions set the context for the local climate sensitivity to a heat mitigation implementation. Many heat mitigation strategies

Table 2. Percent of articles receiving a passing grade for each criterion and sub-criterion, for numerical modelling studies of heat mitigation implementation effects on air temperature at both the microscale and the mesoscale. A passing grade is defined as receipt of 50% or more of points allotted to the (sub)criterion. Shading indicates failure of the (sub-)criterion.

Criterion	Sub-criterion	Total points allotted	Microscale	Mesoscale
Metadata reporting				
1) Neighbourhood[a]		**2**	**45%**	**20%**
	1a) Land cover fractions	0.5	66%	50%
	1b) Urban structure	0.5	67%	30%
	1c) Thermal parameters	0.5	40%	23%
	1d) Radiative parameters	0.5	59%	58%
2) Meteorology		**2**	**65%**	**67%**
	2a) Origin of meteorological forcing	1	62%	81%
	2b) Characterization of weather	1	72%	44%
3) Heat mitigation strategy		**6**	**96%**	**100%**
	3a) Implementation type	2	96%	100%
	3b) Plan area intensity	2	51%	42%
	3c) Location of implementation	2	96%	100%
4) Temperature specification		**8**	**90%**	**91%**
	4a) Horizontal specification	2	57%	63%
	4b) Vertical specification	2	83%	73%
	4c) Diurnal specification	3	93%	97%
	4d) Temporal averaging	1	99%	30%
Modelling methodology				
5) Model physics		**6**	**76%**	**86%**
	5a) Model scale appropriate	0.5	77%	100%
	5b) Model processes appropriate	1.5	70%	94%
	5c) Completeness of model physics	2	85%	55%
	5d) Evaluation of model physics	2	9%	20%
6) Model evaluation		**6**	**48%**	**58%**
	6a) Quality of evaluation data	1	49%	56%
	6b) Evaluation/mitigation scales match	1	60%	61%
	6c) Small absolute error	1	30%	13%
	6d) High correlation	1	26%	5%
	6e) Subjective assessment	2	39%	69%
7) Model application		**6**	**66%**	**86%**
	7a) Boundary conditions	2	49%	84%
	7b) Model configuration	2	68%	77%
	7c) Initial conditions	2	43%	64%
Total		**36**		

[a] Full marks awarded if local climate zone reported in lieu of sub-criteria.

function by reducing absorption of solar radiation or redirecting absorbed solar radiation to evapotranspiration instead of sensible heat; both processes are influenced by ambient weather conditions—cloud cover, humidity and wind in particular. Choice of meteorological forcing to the model can modulate heat mitigation effectiveness and should therefore be characterized (i.e. description of general conditions, and details about the data source).

(3) Heat mitigation implementation metadata are provided (criterion 3)

Characterization of the type, coverage, distribution, and intensity of a surface modification provides vital context for interpretation of its local climate impacts and permits comparison with related studies.

Supplementary table 2 organizes heat mitigation implementations by scale and process, and details the metadata to report in each case. To enable comparability between studies, it is essential to report the area of surface modified relative to the total plan area. In a study of the cooling impacts of high reflectivity roofs, for example, the area of roofs receiving the albedo treatment relative to the corresponding plan area should be reported in addition to the actual roof albedo increase (see section 2.5.1). Changes to pervious fraction or tree cover fraction should be reported in similar fashion. It is furthermore important to identify the location or facet within the urban environment that is modified. Heat mitigation strategies such as irrigation incorporate a temporal element, which should be described explicitly.

(4) Air temperature is specified spatially and temporally (criterion 4)

Air temperature is the focus of this systematic review; nevertheless, the ensuing discussion is largely relevant to other climate response variables, such as surface temperature.

(4a) Horizontal specification (sub-criterion 4a)

Specification of the area over which air temperature changes are assessed helps contextualize the reported impact of the surface modification. Numerical models that are explicit in three dimensions (e.g. ENVI-met and other computational fluid dynamics (CFD) models; see supplementary figure 1(a)) permit precise selection of the horizontal area for assessment of air temperature impacts due to a heat mitigation implementation. The area selected must be sufficiently constrained so as to fall within the zone of influence of the heat mitigation implementation. Mesoscale numerical models and one-dimensional urban canopy models (e.g. Erell and Williamson 2006, Krayenhoff and Voogt 2010) include implicit horizontal averaging at the local scale. Therefore, microscale horizontal exchanges affecting heat mitigation implementation-induced impacts are implicit or ignored. In these studies, it is sufficient to report the grid size, if applicable, and critically, the subset of grid cells associated with the reported air temperature change; ideally, both spatial mean and variability (e.g. spatial maximum) of impacts are reported.

(4b) Vertical specification (sub-criterion 4b)

Many contemporary models of urban areas vertically resolve the between-building canopy layer as distinct from the above building boundary-layer. Some modelling approaches may resolve several vertical layers below building height (supplementary figures 1(a) and (b)). To enhance comparability between studies, the horizontal location or area in sub-criterion 4a should be specified in the vertical as a layer with specified height and thickness (e.g. the urban canopy layer), or as a level of specified height (e.g. air temperature at 2 m above ground level). Moreover, the model must be capable of accurately representing this level or layer, both in terms of the theory being applied (criterion 5) and evaluation of the model (criterion 6).

(4c & 4d) Temporal specification (sub-criteria 4c & 4d)

The cooling effectiveness of heat mitigation measures varies over the diurnal cycle and with season and meteorological conditions. For example, most heat mitigation strategies provide far more daytime than nighttime cooling (Georgescu 2015, Krayenhoff *et al* 2018). Therefore, it is critical that specific times or temporal ranges corresponding to simulated air temperature changes are reported. Most heat mitigation studies will tend to uncover greater impacts during

fair weather warm season conditions. However, their unintended consequences for other weather conditions and seasons deserve attention (e.g. Krayenhoff and Voogt 2010, Yang and Bou-Zeid 2018). Authors should be clear about the temporal context of their results.

(5) Model represents relevant physical processes (criterion 5)

A fit-for-purpose numerical model represents the primary physical processes that modulate urban canopy air temperature, including key processes introduced by the heat mitigation implementation (figure 1). Moreover, such a model has been evaluated at the process level to ensure that perturbations to the system introduced by heat mitigation strategies will have realistic impacts on modelled energy exchanges and temperatures. Representation of flow and turbulent transport within the urban canopy and exchange with the boundary-layer above are particularly important for air temperature (figure 1). Critically, the ability of a model to represent existing air temperature may not ensure that it is capable of capturing temperature impacts of heat mitigation implementations. Modelled processes expected to respond to the heat mitigation implementation, such as energy partitioning at urban surfaces or radiative interception by trees (figure 1), or related impacts on state variables such as surface temperature, should be evaluated or have been evaluated in a previous peer-reviewed study, e.g. in the original model development study. Authors should explicitly defend their choice of model by relating it to their study objective, both in terms of scale and especially in terms of the degree to which it can represent key processes related to the heat mitigation implementation and its impact on air temperature. Model limitations must be stated and the study results contextualized accordingly.

(6) Model evaluation is appropriately targeted and successful (criterion 6)

Model evaluation and calibration ensure that the model can represent the base case spatio-temporal meteorological variation. The magnitude and variation of simulated air temperature must be assessed at a minimum. An appropriately targeted model evaluation is also conducted at spatial and temporal scales and resolutions that correspond to those of the intended model application. Model output should be compared to high quality measured data within the model domain, whose origin or metadata (where appropriate) are reported. A successful model evaluation demonstrates small model-observation differences, where mean absolute error is similar to, or smaller than, the maximum cooling magnitude attributable to the heat mitigation implementation. Standard model evaluation procedures must be followed and a range of error statistics reported (e.g. Willmott

1982, Moriasi, 2007), including a measure of absolute error (e.g. mean absolute error) and a measure of correlation (e.g. coefficient of determination, R^2, or Willmott's d) at a minimum. Ultimately, the quality of a model evaluation requires some degree of subjective judgment.

(7) Model application is sound (criterion 7)
Proper application of numerical models of the urban atmosphere requires attention to initial conditions, the initial values of state variables, and to boundary conditions, the time-dependent values of variables at domain boundaries. Model spin up, that is, model simulation prior to the period of interest, should be at least one full iteration of the dominant thermal forcing cycle of the system (i.e. one diurnal cycle in most cases) to adjust subsurface temperature profiles. At the microscale one diurnal cycle is typically a minimum. Steady-state microscale models are an exception, and they should apply appropriate convergence criteria. At mesoscales, with representation of the complete surface and atmospheric hydrologic cycle (e.g. the Weather Research and Forecasting, or WRF, model; Skamarock *et al* 2005), several days or weeks are typically required to allow soil moisture to adjust to ambient conditions, depending on climate, geography and season. Model domain size must also be sufficient in both mesoscale and microscale simulations to allow full development of the atmosphere upstream of the zone of heat mitigation implementation. In other words, the zone of heat mitigation impact (sub-criterion 4a) must be outside of the zone of direct influence of the boundary conditions. Ideally, models of urban canopy layer climate are coupled to a boundary-layer or atmospheric model, because large changes to surface properties trigger atmospheric feedbacks, whose impacts on air temperature cooling magnitudes can be substantial (Krayenhoff and Voogt 2010). Finally, numerical models of urban climate and meteorology require numerous choices that serve to configure the model (physics and parameter choices, model options and settings). These should be reported and justified.

2.4. Critical review methodology
Each of the 146 articles in the overall sample defined in section 2.2 is evaluated against the 25 sub-criteria, grouped into the seven criteria, presented in section 2.3, table 2, and the supplementary methods. Metadata sub-criteria contain less subjectivity, and for many articles it is clear whether each item of metadata is reported or not. The methodology sub-criteria contain some degree of subjectivity, because they evaluate the rigour or scientific appropriateness of the modelling methodology rather than the presence or absence of a reported piece of metadata. We design our review methodology accordingly (see supplementary methods).

Assessment of the quality of a reported model evaluation (criterion 6) inherently requires subjective judgment, since many interrelated factors must align in order to achieve a convincing model evaluation. Hence, we allot 1/3 of the model evaluation points to a subjective assessment by the reviewer (criterion 6e). Sub-criteria 5c and 5d entail assessment of the completeness of each model's representation of heat mitigation-relevant physics and the degree and/or quality of their evaluation in all previous peer-reviewed articles, both of which require a level of subjective judgment. To render these latter two sub-criteria as objective as possible, they are evaluated by answering the following two questions:

- 5c: How fully are the physics of the process represented?
- 5d: Has the model representation of the physical process been successfully evaluated (or does it have a strong and valid history in the literature)?

Note that evaluation of model representations is typically conducted in the relevant model development article, and for each model in the sample these original articles and any additional articles undertaking process-level model evaluation (including those that add heat mitigation assessment capabilities) are located and assessed. For all heat mitigation strategies questions 5c and 5d are each answered with respect to two constellations of physical processes that are critical for determination of effects of heat mitigation strategies (e.g. Crank *et al* 2018):

- energy balance partitioning at the 'surface' of the heat mitigation strategy; and
- turbulent transport within the urban surface and canopy layers (which transports heat between the location of the heat mitigation implementation and the location where air temperature cooling is reported, e.g. 2 m).

In the case of street trees, treatment of their impacts on shortwave and longwave radiation exchange are additionally evaluated.

2.5. Informal meta-analysis methodology
Subsequent to the critical review, an informal meta-analysis is undertaken based on those studies from the initial sample that exceed a minimum standard with respect to the methodological criteria. Studies were assessed in terms of whether they exceed a score of 50% on the following six conditions:

(a) the overall methodology (sum of criteria 5, 6 and 7);
(b) criterion 5 (model physics);
(c) criterion 6 (model evaluation/calibration);
(d) criterion 7 (model application);
(e) criterion 5c (completeness of model physics);

(f) criterion 6e (subjective assessment of model evaluation/calibration).

Moreover, plan area fraction of the heat mitigation implementation was required for each study to qualify; authors of studies that did not report this value were contacted, and in many cases the information was provided and the study was included. The final meta-analysis sample is comprised of 47 studies that met four or more of the above conditions, and for which we were able to link air temperature cooling magnitude to the associated plan area of implementation.

2.5.1. Cooling effectiveness
The informal meta-analysis is designed to assess the cooling effectiveness (or efficacy) of individual heat mitigation strategies. Assessment of cooling effectiveness in place of absolute cooling removes a primary degree of freedom that differs between studies, and helps progress toward an objective stated in section 1—constraint of the magnitude of heat mitigation strategy impacts—by enhancing comparability of studies. Cooling effectiveness (CE) is defined here as:

$$CE = -\frac{\Delta T}{\Delta a}, \tag{1}$$

where T is air temperature, and a is a plan area-averaged non-dimensional variable that quantifies the principal change associated with the heat mitigation implementation. Here, we distinguish two cooling efficacies. First, to assess cooling effectiveness of albedo implementations, we follow the definition used by Krayenhoff and Voogt (2010), which we term the Albedo Cooling Effectiveness (ACE):

$$ACE = -\frac{\Delta T}{\Delta \alpha_N} = -\frac{\Delta T}{\Delta \alpha_s \cdot \lambda_s}, \tag{2}$$

where $\Delta \alpha_N$ is the neighbourhood scale, plan area-averaged change in albedo resulting from the implementation of high reflectivity building coatings or pavement, $\Delta \alpha_s$ is the change in albedo of the modified surface, and λ_s is the area of modified surface divided by the corresponding overall horizontal plan area. The ACE value in degrees Celsius represents the cooling obtained from a neighbourhood albedo increase from 0.0 to 1.0, assuming linear temperature responses to albedo changes. Similarly, 10% of the ACE value is the cooling obtained for a neighbourhood-wide albedo increase of 0.10. A recent application of the ACE metric has been used to distinguish those neighbourhood densities and regional climates that enhance effectiveness of roof albedo implementations (Broadbent *et al* 2020a). In effect, the ACE metric accounts for the fact that urban albedo applications may occur not only across different portions of the urban surface, but also that

they may occur with different intensities of albedo increase.

Second, to assess cooling effectiveness of vegetation strategies, we define a Vegetation Cooling Effectiveness (VCE) which retains the area fraction of implementation in the denominator but omits the 'intensity' of application:

$$VCE = -\frac{\Delta T}{\lambda_s}, \tag{3}$$

where λ_s is the added surface area of vegetation divided by the associated plan area. Note that a more sophisticated version of equation (3) could include a measure of vegetation 'intensity' such as leaf area index (LAI), corresponding to the $\Delta \alpha_s$ in the ACE metric; however, so few of the studies in our literature sample reported this value that we opt for a simplified definition here. Moreover, air temperature is less likely to respond linearly to increases of LAI than it is to increases of albedo, because added leaf area is more likely to be shaded by existing leaf area and different vegetation covers provide varying evapotranspiration cooling depending on species and environmental context (e.g. Snir *et al* 2016). Transpirative cooling from vegetation further depends on access to soil moisture, which can be highly spatially and temporally variable. For irrigation, a metric similar to VCE could be used, or for certain purposes inclusion of water quantity and/or irrigation timing in such a metric may also be useful (e.g. Daniel *et al* 2018, Broadbent *et al* 2018b).

In equations (1–3), ΔT should be defined for a particular layer or level, ideally at the 2 m level or a vertically-integrated urban canopy layer value (sub-criterion 4b, section 2.3). Furthermore, implicit in these equations is an assumption that air temperature responds linearly to the areal fraction of surface modified, which is appropriate as an initial assumption and furthermore appears to be approximately accurate for roof albedo implementations (Li *et al* 2014). Note that for cooling effectiveness to be meaningful, it is important that horizontal scales of heat mitigation implementations and air temperature response are appropriately related (e.g. sub-criterion 4a).

3. Critical review of urban heat mitigation literature

3.1. Overview of the literature sample
The vast majority of the 146 articles in the literature sample were published after 2006, in the final decade of the 30 year sample (figure 2). The trend is toward greater numbers of articles published per year at both the microscale and mesoscale, probably influenced by the increasing rate of article publication in academia. In addition, the recent surge of urban heat mitigation simulations probably also relates to developments of key modelling tools around the turn of the millennium. ENVI-met
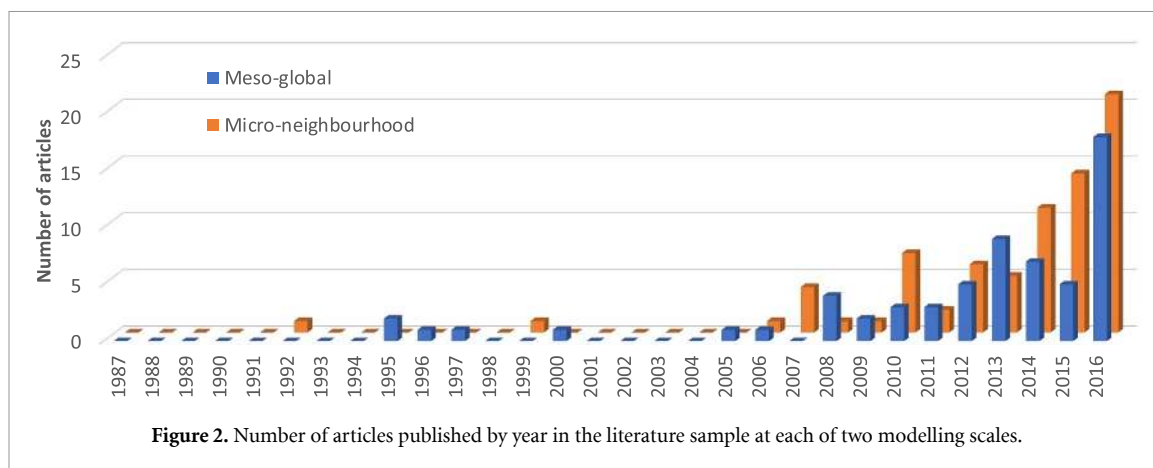
Figure 2. Number of articles published by year in the literature sample at each of two modelling scales.

(Bruse and Fleer 1998) and WRF-urban (Chen *et al* 2011; in particular when applied with the single-layer urban canopy model of Kusaka *et al* 2001, i.e. WRF-SLUCM) are the most commonly-applied models within the sample (figure 3). The development of ENVI-met was first published in 1998, and the first widely-adopted urban canopy models were documented shortly thereafter (Masson 2000, Kusaka *et al* 2001, Martilli *et al* 2002). Moreover, the first version of WRF-SLUCM was released in late 2006, while ENVI-met version 3.1 was released in 2004. The widespread application of both models accounts for close to half of the literature sample (27% of articles use ENVI-met 3.1, and 15% use WRF-SLUCM).

Clear differentiation by model scale is apparent in terms of the types of heat mitigation strategies assessed in the sample (figure 4). Mesoscale models have tended to evaluate albedo strategies and rooftop-based strategies, whereas microscale models have focused preferentially on ground-based applications, especially vegetation and trees. Model physics plays a strong role in this divide. For example, until very recently, street trees have not been integrated into mesoscale implementations of urban canopy models, but their impacts on above-canopy climate have instead been included using the tile approach (Krayenhoff *et al* 2020). This explains the dearth of assessment of street trees at this scale, except via the bulk or tile approach, which cannot represent effects of trees shading impervious surfaces, and other important physical processes related to urban trees (Krayenhoff *et al* 2014, 2015). Conversely, street tree processes that impact the urban canopy layer, such as shading and drag, have been included with some degree of fidelity in ENVI-met and a few other micro-scale models for several years.

We suspect that a bias exists within the research community toward publishing results that include larger cooling magnitudes, related to the particular model and mitigation strategy combinations that are most common (figure 4). For example, ENVI-met typically yields very small street level air temperature cooling as a result of rooftop albedo or vegetation (green roof) applications (Skelhorn *et al* 2014, Middel *et al* 2015, Wang *et al* 2016); WRF-SLUCM, conversely, yields strong '2 m' cooling as a function of rooftop implementations (Georgescu *et al* 2014, Georgescu 2015, Vahmani *et al* 2016), but surprisingly moderate cooling at this atmospheric level from ground level strategies (e.g. ground albedo; Yang *et al* 2016). Model physics formulations appear to play a strong role in both cases (see section 5.1.3 for discussion).

Irrigation, permeable surface, and water bodies received comparatively less attention in the literature sample, and photovoltaic panels have received virtually no physically-based treatment (figure 4). Most 'photovoltaic' treatments in the sample simply entailed changes to rooftop or bulk albedos, and they were recorded as such in figure 4.

3.2. Critical assessment of urban heat mitigation modelling studies

Modelling studies in the sample demonstrate variable performance with respect to the four metadata reporting and three modelling methodology criteria (table 2; figure 5). Articles at both scales typically report sufficient detail related to both heat mitigation type and location (sub-criteria 3a, c) and the timing and vertical level of the reported air temperature (sub-criteria 4b, c), but fewer report plan area intensity of heat mitigation implementations (sub-criterion 3b) or clearly specify the horizontal domain of the urban atmosphere that corresponds to the reported temperature changes (sub-criterion 4a; table 2). Least well reported are site or neighbourhood characteristics, such as the local climate zone and/or information relating to land cover, built structure, or thermal or radiative characteristics (criterion 1), particularly in the case of mesoscale models (e.g. data used in the urban input parameter file). Meteorological forcing (criterion 2) is reasonably well described in the sample, except that mesoscale simulations do not typically describe the prevailing weather conditions during their simulations (sub-criterion 2b).
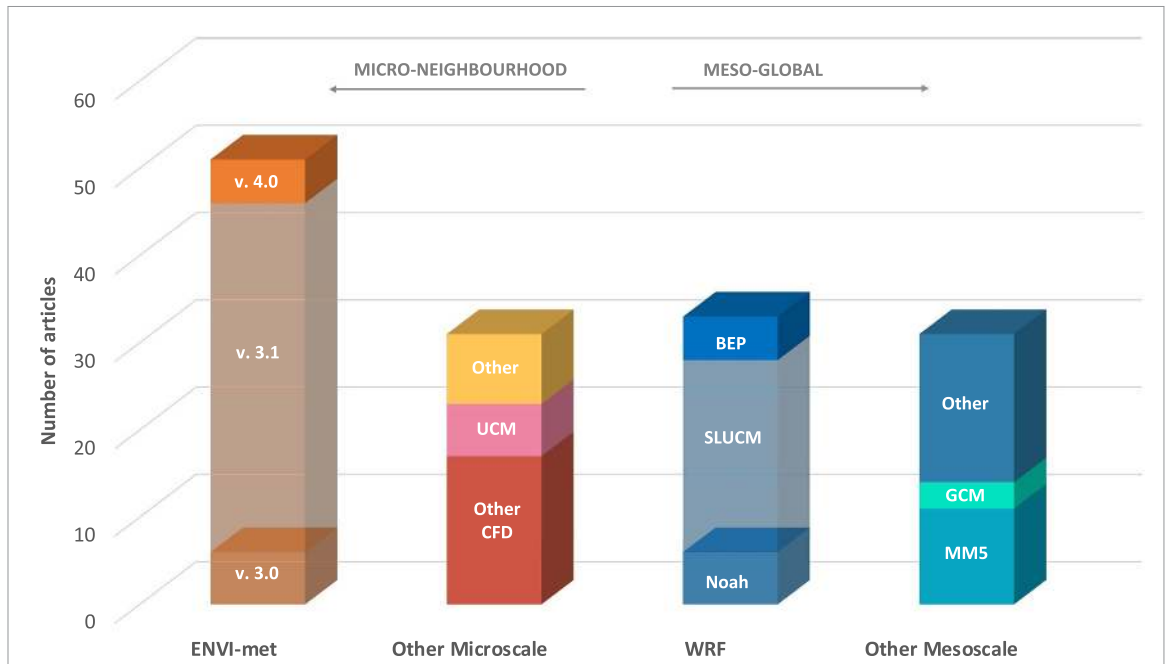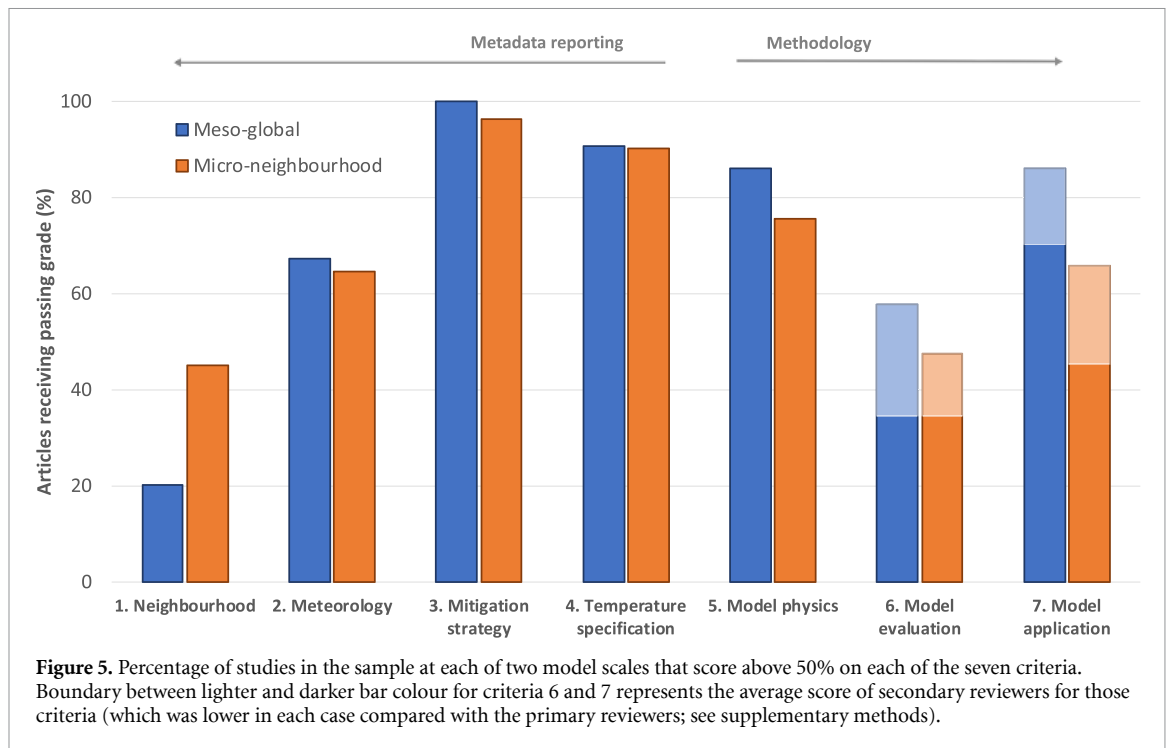
Figure 3. Number of articles in the literature sample by model scale and type. UCM = Urban Canopy model (uncoupled/offline); CFD = computational fluid dynamics; Noah = NCEP, Oregon State, Air force weather agency, NOAA Office of hydrology land surface model; SLUCM = Single-Layer Urban Canopy model (based on Kusaka *et al* 2001); BEP = Building Effect Parameterization (Martilli *et al* 2002); MM5 = fifth-generation Penn State/NCAR mesoscale model; GCM = Global Climate Model.
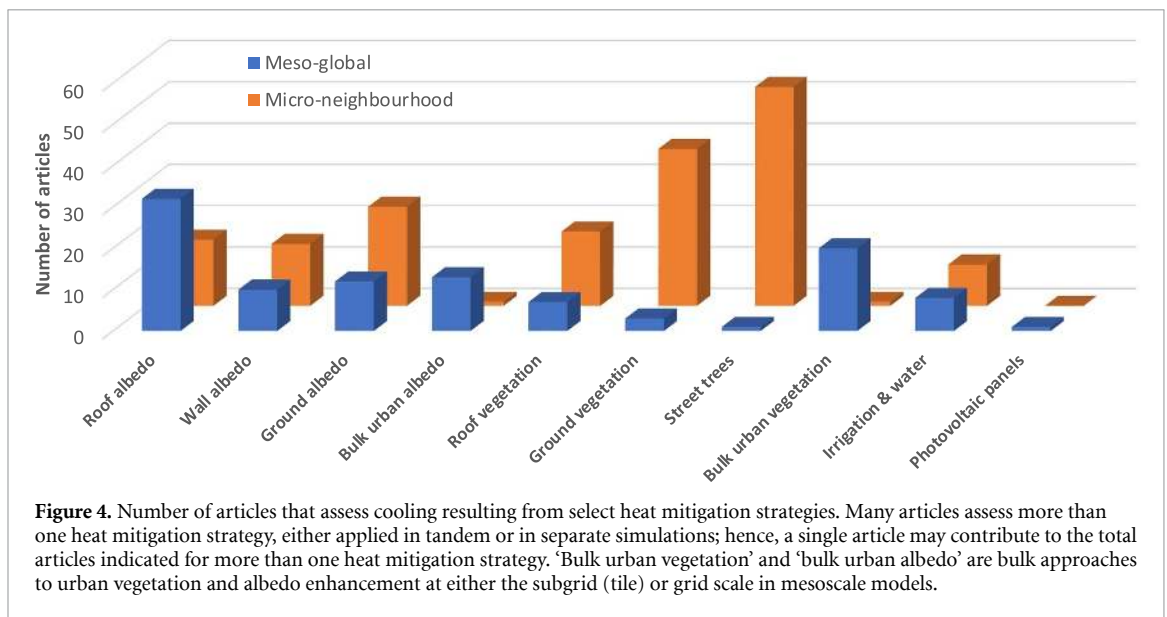
Figure 4. Number of articles that assess cooling resulting from select heat mitigation strategies. Many articles assess more than one heat mitigation strategy, either applied in tandem or in separate simulations; hence, a single article may contribute to the total articles indicated for more than one heat mitigation strategy. 'Bulk urban vegetation' and 'bulk urban albedo' are bulk approaches to urban vegetation and albedo enhancement at either the subgrid (tile) or grid scale in mesoscale models.

At first glance, studies at both scales perform moderately well with regard to model physics and model application, and more poorly in terms of their model evaluation or calibration procedures (figure 5). Of particular note is the small number of studies using a model with substantive evaluation of the physical processes related to the implemented heat mitigation strategy (sub-criterion 5d; table 2). The large discrepancy between completeness and evaluation of model physics (sub-criterion 5c vs. 5d) at the microscale derives from the ENVI-met model, which includes an impressive breadth of

physical processes, most of which have been insufficiently evaluated in the peer-reviewed literature (likely related to the proprietary nature of this software and unavailability of its source code). Few mesoscale models are as complete in their physical representation of processes, partly because of the computational requirements at this scale, but they are somewhat better evaluated at the process level. WRF-SLUCM generally fares better on sub-criteria 5c and 5d, but receives a lower score for physical representation of urban canopy turbulent transport (see section 5.1.3).

Figure 5. Percentage of studies in the sample at each of two model scales that score above 50% on each of the seven criteria. Boundary between lighter and darker bar colour for criteria 6 and 7 represents the average score of secondary reviewers for those criteria (which was lower in each case compared with the primary reviewers; see supplementary methods).

Less encouraging is the performance of many model evaluations in the sample (criterion 6), with very few achieving small errors relative to observational data. Just over half of the sample, on average, used good quality observation data in their evaluation, and evaluated their model at the same scale as they applied it in their heat mitigation assessment (sub-criteria 6a,b; table 2). Moreover, model evaluation was frequently unconvincing, especially at the microscale and with ENVI-met studies in particular (sub-criterion 6e). Model application (criterion 7), including boundary and initial conditions and appropriate model configuration, was conducted reasonably well at the mesoscale. At the microscale, many studies (in particular ENVI-met studies) failed to include a buffer of sufficient size between the model domain boundaries and the location of interest for air temperature mitigation (sub-criterion 7a). Overall, the mesoscale sample of articles passed more methodology criteria, whereas more microscale articles fared better in terms of the metadata reporting criteria (table 2).

Finally, individual heat mitigation modelling studies that are worth emulating in terms of their metadata reporting and/or modelling methodology are selected based on their performance against the seven criteria (see supplementary methods, supplementary results and supplementary table 3). All but one of the papers included in supplementary table 3 fall within the final 4 years of the literature sample, suggesting that both metadata reporting and modelling methodology may have improved over the sample period.

4. Cooling effectiveness of heat mitigation strategies

The fractional area of heat mitigation implementation as well as the intensity of application at the neighbourhood scale strongly control the magnitude of local cooling (see section 2.5.1). To remove this primary degree of freedom, normalized 'cooling effectiveness' metrics ACE and VCE introduced in section 2.5.1 are applied within an informal meta-analysis. Cooling is then more readily compared between studies, and a generalized understanding of the effectiveness of select heat mitigation strategies may be approached. Approximate *a priori* assessment of cooling from potential heat mitigation strategies may ultimately be possible for urban contexts without recourse to labour- and computationally-intensive simulations.

The meta-analysis conducted here is informal for several reasons. First, we do not use effect size variance (i.e. cooling effectiveness variance) to weight the effect size reported by each study, as Bowler *et al* (2010) did for observational studies, due to the substantial variability in methodology and reporting between numerical modelling studies. Moreover, we do not conduct formal assessments of effect size heterogeneity. Second, several known degrees of freedom are not fully controlled for: latitude, local geography, mesoscale/synoptic variability, neighbourhood design and construction, and the height of reported air temperature. In the subsequent analysis, reported results are grouped by season and time of day to approximately control for these factors. The

remaining variation of reported cooling effectiveness from each individual study stems from uncertainty in modelling or reporting of results and/or from sampling across those remaining degrees of freedom (e.g. different cities or neighbourhoods). Where appropriate, a median cooling effectiveness is calculated for each study, and a median of all studies is assessed for select temporal contexts (e.g. summer afternoons).

4.1. Albedo Cooling Effectiveness (ACE)

Roof albedo implementations are one of the most common heat mitigation strategies investigated in our sample, likely due to their ease of modelling implementation. Moreover, they already have a degree of uptake in many jurisdictions. During summer afternoons, ACE ranges widely across the studies in our sample (figure 6(a)). Notably, ENVI-met 3.1 studies have a median ACE of 1.6 °C, or 0.16 °C cooling per 0.10 neighbourhood-scale albedo increase, whereas mesoscale studies have a median ACE of 5.8 °C; moreover, the highest ACE of all ENVI-met studies in our sample is lower than the lowest median ACE of any mesoscale study. This large discrepancy clearly suggests that differences in model physics and/or scale strongly affect modelled cooling magnitudes (see section 5.1.3 for discussion). Mesoscale studies in our sample have a median ACE of 2.2 °C for summer nights, suggesting residual nocturnal cooling is less than half of daytime cooling for roof albedo implementations (figure 6(a)). As expected, ACE derived from seasonal average or wintertime results is smaller than episodic summer ACE, as a result of reduction of solar radiation reaching rooftops due to clouds and larger solar zenith angles. Finally, urban heat mitigation applications are area-limited by definition, and hence advection always serves to reduce their local impact. The difference in ACE between a no-advection case and an advection case (which resembles the median across all mesoscale studies) reported by Krayenhoff and Voogt (2010) suggests that fair weather advection reduces the local impact on air temperature of roof albedo by more than half (figure 6(a), top row).

High albedo surfaces or coatings may also be applied at ground level, for example as 'cool pavement'. Microscale models in our sample, primarily ENVI-met 3.1, derive an ACE of approximately 5.7 °C for ground level albedo applications during summer afternoons, albeit with large variation between studies (figure 6(b)). The only mesoscale study of high albedo pavement in our sample yields an ACE less than half as large. Again, model physics and/or scale likely play a role in this difference (see section 5.1.3). Increases of building wall albedo in conjunction with roof albedo add little cooling compared to roof albedo alone, at both model scales, albeit for only two studies (figure 6(b); note that ACE is calculated here without considering wall area in

equation (2). Two microscale studies increased the albedo of all surfaces, while several mesoscale studies increased the albedo of a bulk (i.e. flat) representation of the urban surface; median ACE across all of these studies is 6.0 °C for summer afternoon conditions (figure 6(b)), which is virtually identical to ACE values from mesoscale roof albedo studies, perhaps because similar model physics are involved in both cases.

4.2. Vegetation Cooling Effectiveness (VCE)

Trees affect pedestrian-level urban climate via multiple physical processes (Oke 1989, Krayenhoff *et al* 2020), including shading, sheltering and evapotranspiration. Most of these processes are represented, in some fashion, in the studies included in our sample. We assess the *VCE* defined in section 2.5.1 across multiple studies that incorporate trees as urban heat mitigation tools; note that VCE cannot be directly compared to ACE values in section 4.1 (see section 2.5.1). Most of these studies were performed at the microscale, particularly by ENVI-met 3.1, and they yield a median summer afternoon VCE of 3.3 °C, or 0.33 °C of cooling per 0.10 increase in tree canopy cover (areal fraction), although select studies find substantially higher values (figure 7(a)). The sole mesoscale study by Loughner *et al* (2012) finds a similar VCE magnitude, as does a recent CFD study (Toparlar *et al* 2018). Nighttime cooling by trees across three modelling studies is approximately 50% of respective daytime cooling for each study (figure 7(a)); however, nighttime results in ENVI-met 3.1 must be interpreted within the context of its very simple representation of urban heat storage (Huttner 2012). Some observational studies report small nighttime cooling from street trees (e.g. Ziter *et al* 2019), in agreement with these simulation results, whereas others find small nighttime warming (e.g. Gillner *et al* 2015), or both cooling and warming depending on time of night (Coutts *et al* 2016).

Vegetation more broadly is typically considered to be an urban heat mitigation strategy, including green roofs and ground-level vegetation such as grass. Most studies considering within-canyon grass in the sample were performed by ENVI-met 3.1. Microscale model-derived VCE for grass varies widely between studies with VCE from about 1 °C to above 10 °C (figure 7(b)). The only mesoscale study that includes within-canyon grass has a VCE of 3.3 °C–8.4 °C, depending on soil moisture. Evapotranspiration and associated thermal climate impacts of grass depend strongly on moisture levels of upper soil layers, and the fact that we do not control for this degree of freedom may explain the large variation in VCE between these studies.

Cooling from green roofs appears to demonstrate a similar dependence on model scale and/or physics as high albedo roofs. That is, summer afternoon green roof VCE from available mesoscale studies
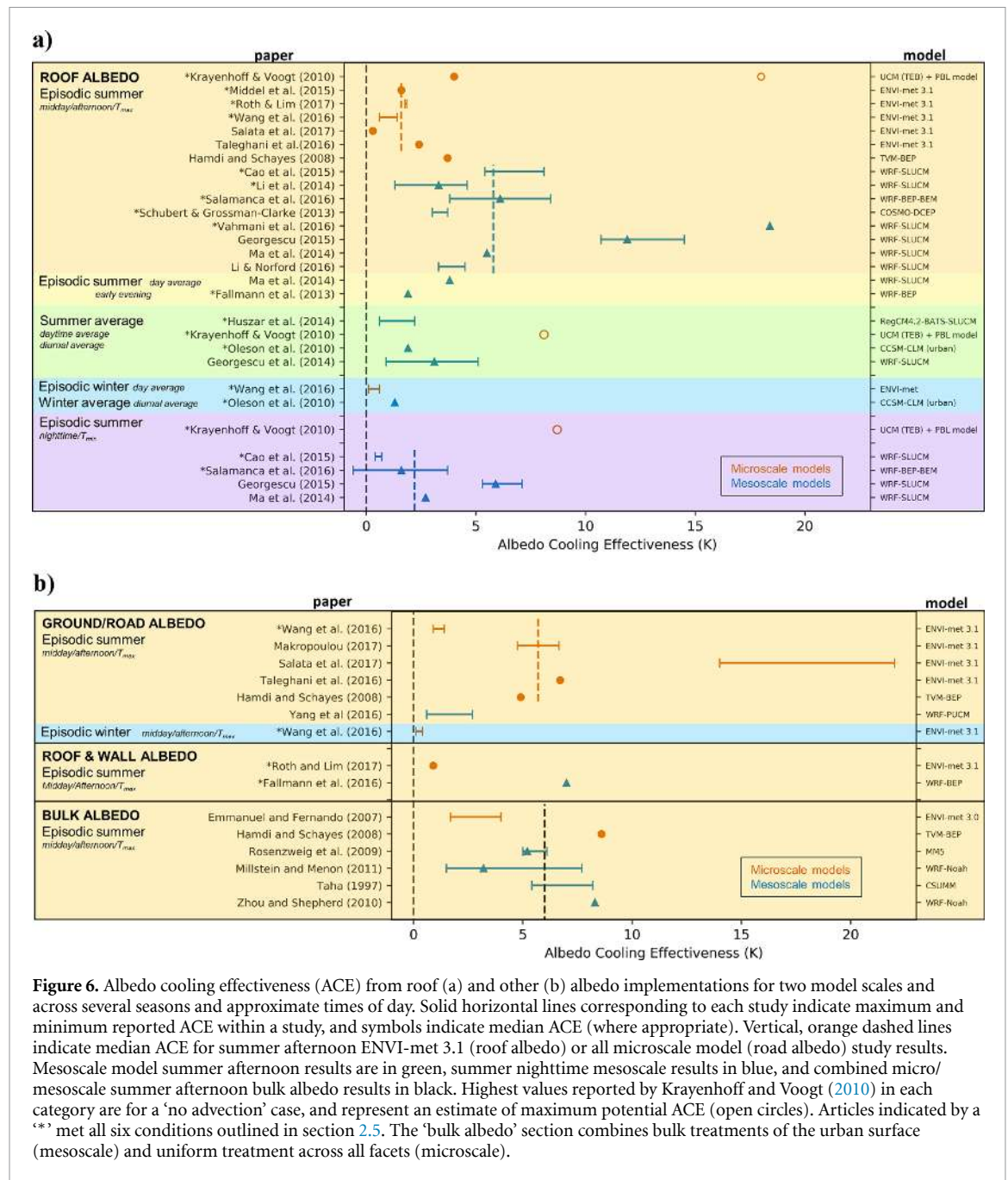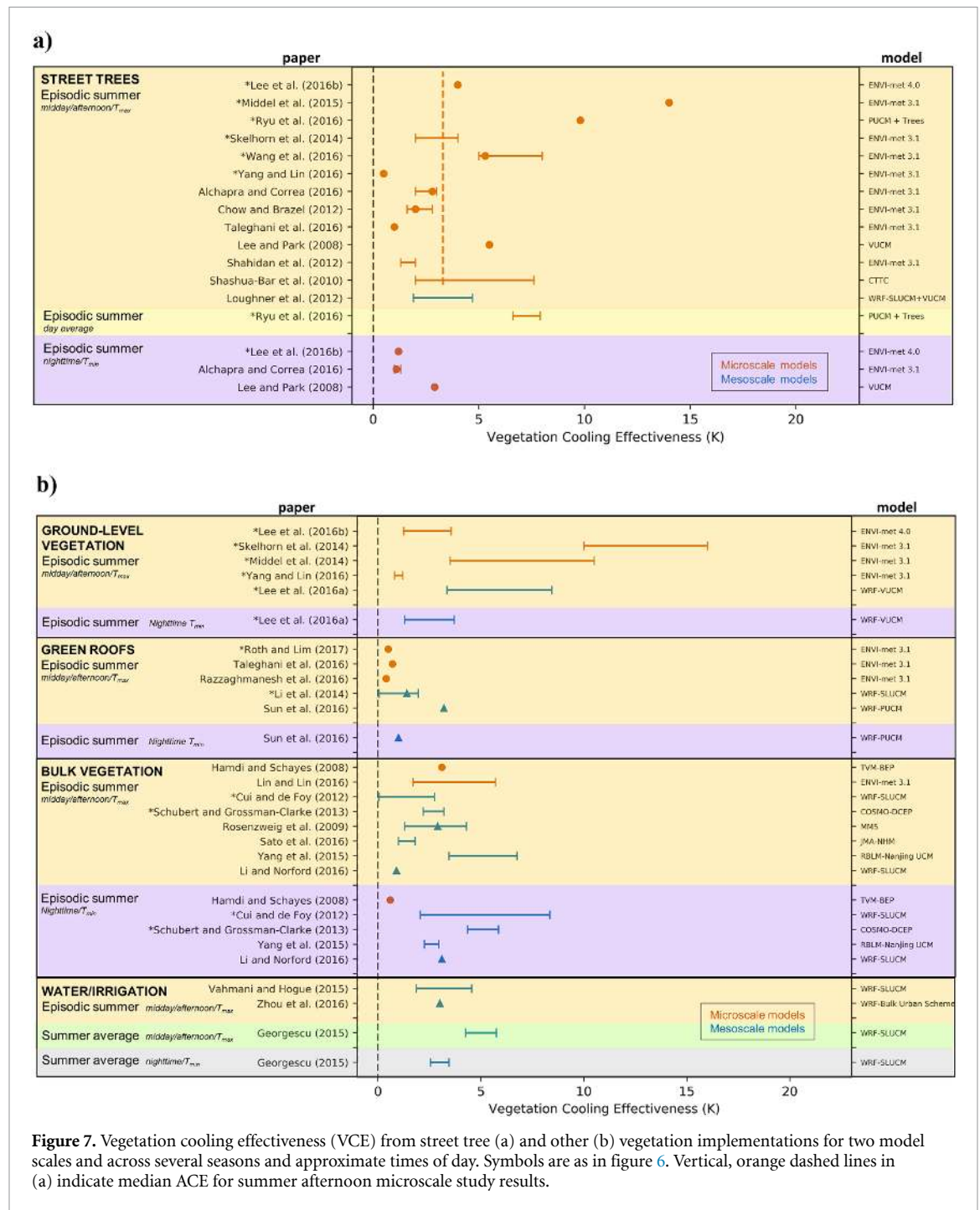
Figure 6. Albedo cooling effectiveness (ACE) from roof (a) and other (b) albedo implementations for two model scales and across several seasons and approximate times of day. Solid horizontal lines corresponding to each study indicate maximum and minimum reported ACE within a study, and symbols indicate median ACE (where appropriate). Vertical, orange dashed lines indicate median ACE for summer afternoon ENVI-met 3.1 (roof albedo) or all microscale model (road albedo) study results. Mesoscale model summer afternoon results are in green, summer nighttime mesoscale results in blue, and combined micro/mesoscale summer afternoon bulk albedo results in black. Highest values reported by Krayenhoff and Voogt (2010) in each category are for a 'no advection' case, and represent an estimate of maximum potential ACE (open circles). Articles indicated by a '*' met all six conditions outlined in section 2.5. The 'bulk albedo' section combines bulk treatments of the urban surface (mesoscale) and uniform treatment across all facets (microscale).

is 1 °C–3 °C, several times higher than the VCE of ≈0.5 °C from microscale ENVI-met 3.1 studies (figure 7(b)). Alternatively, different soil moisture and/or irrigation assumptions between studies may play a role (Li *et al* 2014), particularly given the limited number of studies. In practice, green roofs are typically designed to receive minimal irrigation beyond an initial plant rooting phase. One mesoscale study suggests that green roof cooling is much reduced at night compared to daytime (Sun *et al* 2016); other evidence suggests that green roofs may actually cause nocturnal warming (Scherba *et al* 2011). Overall, available high-quality studies suggest that green roofs yield less canopy level cooling per

area application compared with trees and ground-level vegetation.

Georgescu (2015) models evaporative (i.e. irrigated) roofs, which represent an upper bound of evapotranspirative cooling achievable from green roofs, finding a summertime average VCE of about 5 °C. While few studies of urban irrigation or water bodies remain in the final meta-analysis sample, the two available studies (both at the mesoscale) yield VCE of about 3 °C for summer afternoons, similar to the vegetation implementations discussed previously. Note that the amount and timing of irrigation or the temperature of the water bodies are not controlled for due to insufficient sample size.

Figure 7. Vegetation cooling effectiveness (VCE) from street tree (a) and other (b) vegetation implementations for two model scales and across several seasons and approximate times of day. Symbols are as in figure 6. Vertical, orange dashed lines in (a) indicate median ACE for summer afternoon microscale study results.

Most recent mesoscale modelling of urban vegetation has used the tile approach, where surface-atmosphere energy fluxes from impervious surfaces and from soil-vegetation are computed by independent models and then combined to assess a near-surface (but effectively above-canopy) temperature. While this approach is clearly inadequate for assessment of street trees because it does not represent processes such as tree shading and sheltering of building and roads (Krayenhoff *et al* 2020), it may have more merit for low vegetation such as grass. However, the sub-grid tile approach to vegetation appears to generate more nighttime compared to daytime cooling in mesoscale models (figure 7(b);

Cui and De Foy 2012; Schubert and Grossman-Clarke 2013, Li and Norford 2016), whereas the single study that integrates low vegetation within an urban canopy model cools daytime air temperature more than twice as much as nighttime air temperature (figure 7(b); Lee *et al* 2016a). Recent WRF-SLUCM tile-based modelling also shows predominantly nighttime cooling from urban vegetation (Jacobs *et al* 2018, Imran *et al* 2019), and there is some recent observational support for this result (Hu and Li 2020). Nevertheless, it is unclear if this result is robust or if the tile approach needs to be revisited for representation of ground-level urban vegetation.

5. Discussion

5.1. Key recommendations

Several recommendations arise from the critical portion of this review, particularly related to enhancement of the reliability and comparability (contextualization) of cooling magnitudes reported by heat mitigation studies. The large disagreements in cooling effectiveness between model scales for multiple heat mitigation strategies evident in the informal meta-analysis (section 4) further reinforce the need for more reliable model representations of the physical processes that link heat mitigation implementations to simulated air temperature reductions.

5.1.1. Metadata reporting and study comparability

Sufficient context must be provided both to assess the simulated effectiveness of a heat mitigation implementation and to compare cooling effectiveness between studies. Our sample indicates that, on average, mesoscale modellers have more opportunity to improve metadata reporting than microscale modellers. Based on results from the critical review (table 2; section 3.2), we recommend that particular attention be paid to the reporting of the following metadata.

(a) Characterize the model representation of neighbourhood(s) in which heat mitigation is applied, for example by reporting the local climate zone(s) or built geometry and materials (see supplementary table 1).

(b) Report the ratio of heat mitigation implementation area to plan area, for example the plan area of roofs for cool or green roof implementations, or canopy cover fraction for street tree planting.

(c) Characterize prevailing meteorological conditions during the numerical experiment, for example solar zenith angles or day of year and latitude, mean wind and cloud cover, particularly for shorter term (non-climate scale) simulations.

(d) Spatially contextualize the reported air temperature cooling, for example, indicate the portion of the simulation domain that corresponds to the reported cooling magnitude.

(e) Report the temporal context for the cooling (e.g. diurnal average vs. noontime cooling).

5.1.2. Modelling methodology and reliability of results

A model with well-tested representations of all relevant physical processes, evaluated and applied appropriately to assess the heat mitigation implementation, is required to produce reliable cooling magnitudes. Model application at the microscale, and numerical model development at both scales, would particularly benefit from the following recommendations for improved reliability of results.

(a) Evaluate model representations of physical processes related to both the heat mitigation implementation and the urban canopy layer turbulent transport that governs the air temperature response (e.g. benchmark modelled vertical turbulent exchange of heat in urban canopies, compare measured and modelled street-level shading by trees as a function of leaf area, compare soil moisture dynamics of modelled green roofs to observations).

(b) Conduct an evaluation or calibration of the base case model using high quality measured data, for example, from networks of good quality, calibrated, and appropriately-sited instruments (e.g. Willmott and Matsuura 1995, Rotach *et al* 2005, Oke 2006) distributed among the local climate zones or locations where heat mitigation will be assessed.

(c) Match the scale of model evaluation data with the scale of heat mitigation cooling assessment; for example, through evaluation of building-resolving models with measurements that capture micrometeorological effects (microscale models), and with measurements that capture neighbourhood-scale quantities (e.g. via traverses or other measurements that integrate over scales of hundreds of meters to a few kilometers) in mesoscale models.

(d) Ensure reasonable base case model performance before proceeding, for example by understanding and addressing the root causes of high MAE or RMSE, or low r^2 or d, before conducting heat mitigation simulations.

(e) Use sufficient domain size such that the area over which cooling is assessed (section 2.3.4.1.) is not impacted by domain boundaries, e.g. by testing the influence of a larger buffer between the upwind domain boundary and the zone of cooling assessment (if the buffer is already sufficient, increasing it will not appreciably change results).

Improvement in these areas, in particular, should be priorities in the field of urban heat mitigation modelling. Most critically, the peer-reviewed capability of the applied numerical model to accurately represent turbulent transport of heat in the urban canopy and physical processes associated with the heat mitigation implementation should be demonstrated prior to its application.

5.1.3. Model representations of physical processes relevant to urban heat mitigation

The informal meta-analysis in section 4 suggests that differing representations of urban meteorological processes, especially those that modulate air temperature response to heat mitigation implementations, are primary contributors to the variation of reported cooling magnitudes between studies. The most notable differences appear between microscale
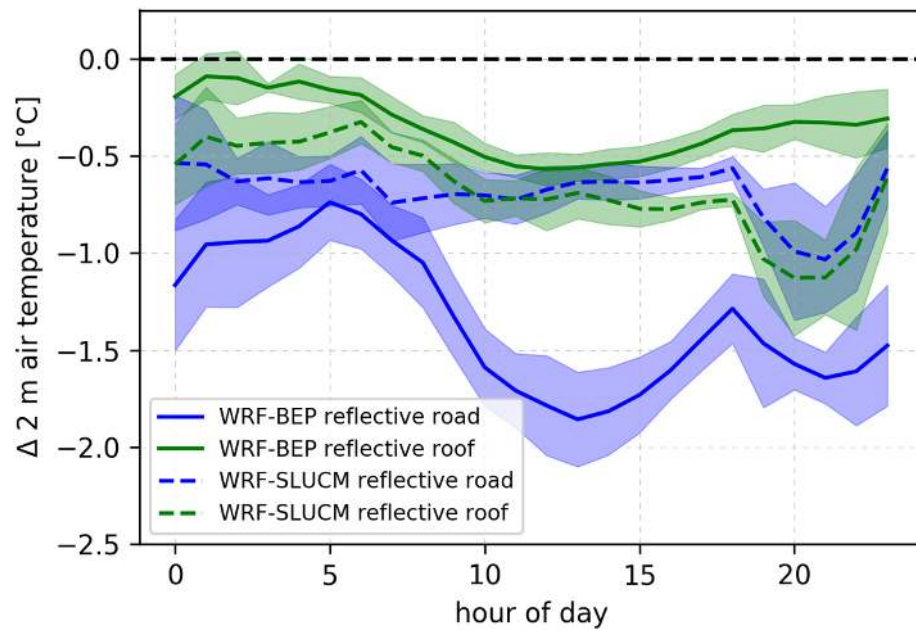
Figure 8. The importance of model physics representations of urban surface layer exchange on air temperature cooling from albedo enhancement. A diurnal composite of cooling during a 2006 heatwave in Phoenix studied by Broadbent *et al* (2020a) is plotted for two urban canopy models (UCMs) embedded in the WRF mesoscale model. For each UCM, the albedo of road and roof are increased in separate simulations to 0.88.

(ENVI-met) and mesoscale models for roof-level implementations (cool and green roofs; figures 6 and 7). Typically, microscale models do not account for local-scale advection or boundary-layer scale vertical mixing, and ENVI-met version 4 (and likely earlier versions) may have additional limitations in terms of accurately simulating vertical mixing in the urban canopy layer (Crank *et al* 2018). Mesoscale models omit or attempt to parameterize the effects of most microscale canopy layer processes and many cannot readily distinguish the pedestrian (2 m) level from the surface layer above roof level (e.g. the standard 2 m temperature reported in most WRF-SLUCM studies does not). Moreover, our own simulations indicate that roof surface temperatures vary strongly between ENVI-met version 4 and WRF coupled with its multi-layer urban canopy model building effect parameterization (BEP) (Martilli *et al* 2002) for identical forcing conditions (not shown). Secondly, underlying assumptions related to roof heat conduction as well as building construction and insulation characteristics will also affect the extent to which roof heat mitigation impacts are partitioned between direct cooling of the atmosphere vs. storage heat into the roof/building, and these may differ substantively by default between ENVI-met and commonly-applied mesoscale models. Hence, model physics differences across one or more key processes may be contributing to the substantive disagreement in cooling effectiveness of rooftop heat mitigation between model scales.

As noted above, the ACE value in degrees Celsius represents the cooling obtained from a neighbourhood albedo increase from 0.0 to 1.0,

assuming linear temperature responses to albedo changes. Similarly, 10% of the ACE value is the cooling obtained for a 0.1 neighbourhood-wide albedo increase, an increase for which the assumption of linearity is more likely to hold. An example of such an increase is a residential neighbourhood with roof plan area of 0.20, where all roofs receive a high albedo coating that increases their albedo by 0.50 (e.g. to 0.60–0.70). The predicted 0.58 °C median air temperature reduction based on the ACE derived from mesoscale modelling studies appears reasonable for this scenario, whereas the 0.16 °C median cooling from ENVI-met 3.1 studies would suggest that roof albedo has negligible impact. Moreover, an ACE of 5.8 °C more closely matches available observational data from area-limited albedo treatments in rural areas that have an ACE of ≈4 °C–9 °C (Rosenfeld *et al* 1995). The necessity for turbulent transport of the cooled air from rooftops to the pedestrian level in an urban area, which is explicitly modelled in the ENVI-met studies but is not present in the rural area reported by Rosenfeld *et al*, would serve to reduce this rural ACE value in an urban context. Most likely, realistic clear sky summer daytime ACE values for high albedo roofs lie somewhere between the values reported by the two model scales, e.g. in the range of 2 °C–6 °C.

Importantly, these model physics-related issues are not limited to differences between model scales. For example, the single-layer (SLUCM) and multi-layer (BEP) urban canopy models, when coupled with the WRF mesoscale model (Chen *et al* 2011), show very different pedestrian-level air temperature

responses to high albedo roads, and to a lesser extent to high albedo roofs, for otherwise identical model input and forcing conditions (figure 8). These differences in cooling between WRF-SLUCM and WRF-BEP probably arise from different parameterizations of urban canopy layer radiation exchange and turbulent mixing between the models, and relatedly, from different approaches to the calculation of a '2 m' air temperature in WRF. BEP represents multiple layers within the urban canopy (supplementary figure 1(b)) and provides source and sink terms to the boundary-layer scheme in WRF, which penetrates to ground level. This approach means that the output 2 m air temperature is calculated by the WRF dynamical core and boundary-layer scheme at a height of approximately 2 m (provided sufficient vertical resolution in the urban canopy is prescribed, as is done for the simulations in figure 8). WRF-SLUCM, conversely, diagnoses a '2 m' air temperature based on the sensible heat flux from the entire urban surface (including roofs), effectively using a bulk flux-gradient relation that does not differentiate between canopy and above-canopy layers (supplementary figure 1(c)), and it is therefore less able to distinguish between roof and road influences on '2 m' air temperature (note, however, that SLUCM does calculate a canopy air temperature which is rarely used, which responds differentially to roof vs. road albedo treatments but does not compare as favourably to measurements in our experience). As a result, the WRF-BEP 2 m air temperature, which is 2 m above the modified roads, demonstrates much higher sensitivity to road albedo than the WRF-SLUCM '2 m' air temperature, which is effectively above roof level (figure 8). The base case configurations of both urban canopy models in WRF have been widely evaluated (e.g. Salamanca *et al* 2018), and they evaluate almost identically against multiple measurement stations in the Phoenix metropolitan area for the current case (supplementary figures 2 and 3). Nevertheless, their representation of physical processes, particularly the physics associated with their coupling to WRF and urban canopy layer mixing, have received less testing, with clear implications for reliable assessment of cooling magnitudes from heat mitigation implementations (figure 8). While more complex model physics representations do not always improve reproduction of base case observations (Grimmond *et al* 2011, Best and Grimmond 2015, Karlicky *et al* 2018), our results suggest that their added complexity is important to consider for accurate assessment of novel forcing or boundary conditions, as in the case of heat mitigation strategy implementation.

Incorrect or inadequate model physics can also persist when air temperature of the base case simulation is the only metric used to assess the suitability of a model for a heat mitigation experiment, as is the case for many heat mitigation model applications (see supplementary figure 4 and supplementary results for an example using ENVI-met). In such applications, apparently successful model evaluation can often be the result of compensating errors. Comparison of simulated and measured air temperature at multiple locations and at multiple points across the diurnal cycle often improves assessment of the base case simulation and its ability to distinguish scale-appropriate spatio-temporal variability of air temperature. However, even a sophisticated model evaluation against high quality air temperature measurements does not on its own indicate that the model is correctly representing the processes relevant to a potential heat mitigation implementation.

Evaluation of the base case model configuration alone is thus not a sufficient prerequisite for simulation of cooling from heat mitigation implementations, principally because the model physics that most strongly impact cooling magnitudes (e.g. a green roof or street tree sub-model) are not often involved in the base case simulation, and therefore accurate simulation of base case air temperatures does not indicate that heat mitigation cooling magnitudes will be accurately simulated. Model physics associated with both canopy layer mixing and the heat mitigation implementation itself must be evaluated (section 2.4), and few models or model combinations that are applied to assess urban heat mitigation cooling have done so (table 2; section 3.2). There are typically three phases of model development and application prior to assessment of heat mitigation strategy-induced cooling with numerical experiments: (a) development and testing of an urban climate or micrometeorological model; (b) development and testing of sub-models that introduce new physical processes required to represent an urban heat mitigation strategy in the urban climate model; and (c) evaluation and calibration of the urban climate model base case simulation prior to conducting heat mitigation experiments (table 3). Each phase should include testing against measured data and/or accepted standards to ensure model reliability. Phase 1 and 2 tests can be undertaken once during urban climate model (or heat mitigation sub-model) development, using datasets from intensive field campaigns for specific sites. These sites do not need to include the exact site where the heat mitigation strategy will be applied, but they ideally include the range of conditions for which the heat mitigation representation will be applied. Model evaluation and calibration during phase 3, alternatively, must be conducted each time the model is applied, using data from the location where the strategy will be implemented.

Our critical review suggests that there is opportunity for improvement in each of the three phases in table 3 (see section 3.2). Critically, testing is most

Table 3. Key physical processes and state variables that are ideally evaluated during the development and application of numerical models designed to quantify cooling from heat mitigation strategy implementations. Guidance is structured according to the typical chronological order of model development and application for heat mitigation assessment: (1) urban climate or micrometeorological model development; (2) heat mitigation strategy sub-model development (e.g. green roof model, street tree model), and (3) base case simulation and associated model evaluation and calibration prior to conducting a heat mitigation experiment.

Heat mitigation strategy	Recommended tests at each model evaluation phase		
	1. Urban climate model development[a]	2. Heat mitigation sub-model development	3. Base case simulation of heat mitigation experiment
Albedo • Reflective roofs • Reflective pavement • Reflective building walls	Vertical transport of heat within and above the urban canopy layer as a function of: • urban density • wind speed • atmospheric stability Urban canopy layer (or ∼2 m) air temperature Sensible heat fluxes and/or surface temperatures of individual surfaces (roofs, walls, roads)	Response of surface temperature and/or sensible heat flux of impervious surfaces to albedo change[b]	Scale appropriate, spatial and temporal variation of urban canopy (∼2 m) air temperature. Minimum recommended resolution of measured temperature data: • Microscale: hourly variation at ∼100 m • Mesoscale: diurnal minimum and maximum between the most common local climate zones in the urban area(s) and surrounds
Street trees		Response of below-tree canopy solar radiation, wind/turbulence and road surface temperature, and of overall latent heat flux to tree cover	
Low vegetation • Green roofs • Ground-level vegetation • *Irrigation*		Response of soil moisture to precipitation and drying periods; response of sensible and latent heat fluxes to soil moisture, vegetation characteristics and vapour pressure deficit[c]	
Irrigation and water bodies		Response of sensible and latent heat fluxes to soil moisture or water temperature; response of soil moisture to irrigation timing and soil type, or water temperature to meteorological forcing	
Photovoltaic panels		Response of panel surface temperature and/or sensible heat flux and power generation to diurnally varying meteorological forcing, with energy exchanges included at active sides of panels[d]	

[a] A greater diversity of tests is recommended, but additional testing varies as a function of model type and scale; tests presented here are important indicators of key physics required to assess canopy layer cooling from heat mitigation strategies.

[b] These tests may also fit more as a part of the model evaluation portion of step 1 (urban climate model development), since no additional sub-model is typically required for albedo-based implementations.

[c] Surface temperature can serve as an alternative to sensible and latent heat fluxes that is simpler to measure, but provides a less nuanced test of the sub-model (i.e. more opportunity for compensating errors). Assessment of modelled fluxes, where possible, is a much stronger test of the model and preferable.

[d] Both sides of photovoltaic panels typically contribute strongly to sensible heat exchange, unless the underside is mounted directly on or very close to the roof (or other host surface)—see Heusinger *et al* (2021). For these cases, models that represent sensible heat (and longwave radiation) exchange on both sides of panels should be used, and they should be able to capture diurnal variation of panel surface temperature with good accuracy.

often weak or absent in phases 1 and 2 (table 2; section 3.2), during which key physical processes that control air temperature response to the implementation of a heat mitigation strategy are assessed. When these phases are skipped, these key physical processes remain untested, potentially contributing to the large diversity of ACE and VCE values in figures 6 and 7. Ideally, this type of evaluation is performed by the developers of the original urban climate model or the developers of associated model extensions related to heat mitigation strategies (e.g. green roof or street tree sub-models that are added to existing urban climate models). ENVI-met presents a particular challenge in this regard, because its code is closed-source, and to date many of its heat mitigation-related processes have received little or no evaluation in the peer-reviewed literature.

While rigorous evaluation of model physics in phases 1 and 2 is challenging, without it heat mitigation simulation results are not reliable and have limited utility. Critical to reliable simulation of heat reduction is evaluation of surface state variables and fluxes directly modulated by each heat mitigation strategy (i.e. phase 2 in table 3). For example, assessment of the temporal evolution of green roof soil moisture and heat fluxes (Heusinger and Weber 2017, Heusinger *et al* 2018), temperatures and heat fluxes from reflective materials (Rosado *et al* 2014, Kim *et al* 2020, Middel *et al* 2020), temperatures of photovoltaic panels (Broadbent *et al* 2019b, Heusinger *et al* 2021), and temperatures of the road and associated radiation fluxes underneath street trees (Krayenhoff *et al* 2020), present realistic opportunities for more rigorous assessment of heat mitigation-related model physics. Furthermore, phase 1 assessment of simulated vertical turbulent (and dispersive) transport in the urban canopy layer and roughness sublayer as a function of ambient wind, stability and urban geometry in a standardized way, in comparison with urban large-eddy simulation results for example (Giometto *et al* 2016, Nazarian *et al* 2020), would promote more consistent and realistic ventilation of the street-level environment, as well as improved transport of roof-level air temperature reductions to street-level. If model physics are appropriately evaluated during model development phases 1 and 2, then appropriate model evaluation against only air temperature during a heat mitigation study is more justifiable (i.e. phase 3). During phase 3, evaluation of the ability of models to assess intra-urban air temperature variation at a model-appropriate scale should be undertaken (table 3). Salamanca *et al* (2011) present an example of intra-urban (i.e. inter-neighbourhood or inter-local climate zone) evaluation/calibration of a mesoscale model, while Roth and Lim (2017) evaluate a microscale model with intra-neighbourhood (i.e. ∼100 m) scale observations. A key challenge more generally is the limited availability of observational data at the appropriate

temporal and spatial scales for model evaluations in all three phases, although sufficient data currently exists within the global scientific community for most, if not all, of the recommendations in table 3. Development of open source databases containing high quality measurements for evaluation of numerical model representations of urban canopy processes and processes related to heat mitigation implementations is recommended, and will need to properly recognize the substantive efforts undertaken by the scientists involved in collecting these observations.

5.1.4. Combining heat mitigation strategies
Several studies in our sample, particularly those applying ENVI-met, introduce multiple heat mitigation strategies into a single heat mitigation simulation experiment. For example, trees are planted, grass is added, and pavement albedo is increased. While such studies may be helpful for optimizing the design of particular street or neighbourhood, they are less useful for approaching a generalized understanding of urban heat mitigation effectiveness. Because these studies do not isolate individual mitigation strategies they were not included in the informal meta-analysis in section 4. Moreover, because the individual effect of a heat mitigation strategy is not accounted for, place-based prioritization of what works best within a geographical context cannot occur (Georgescu *et al* 2014, Krayenhoff *et al* 2018). If computational resources are available, we recommend assessment of each heat mitigation strategy in isolation in addition to the combined simulation, permitting assessment of cooling magnitude and effectiveness for each strategy independently. This approach also permits assessment of interactions between heat mitigation strategies, which have not been fully explored in the literature. Evidence from Australia and the U.S. suggests these interactions may be largely antagonistic during daytime for both heat wave days (Jacobs *et al* 2018)) and the summer season (Krayenhoff *et al* 2018). Examples of assessment of interactions between urban climate processes with the Stein and Alpert (1993) factor separation method include Martilli (2002), Ryu and Baik (2012) and Krayenhoff *et al* (2018).

5.2. Limitations of this review
There are several limitations to the systematic and critical review and informal meta-analysis portions of this work. In terms of the definition of our sample, we limit our literature sample to one database: Web of Science, and we exclude articles in languages other than English. The critical portion of this review relies on definition of 25 sub-criteria, which may not be fully comprehensive. In our view, the most substantial limitations arise at the informal meta-analysis stage; for most heat mitigation strategies, there are insufficient studies exhibiting satisfactory methodology to properly assess a central tendency of cooling

effectiveness. Second, the preponderance of WRF-SLUCM and especially ENVI-met 3.1 in the meta-analysis sample may bias results, given the likelihood that there is substantial opportunity for improvement in the representations of physical processes of relevance to certain heat mitigation strategies in these models (see section 5.1.3). Moreover, available studies are unlikely to provide distributions of cooling magnitudes that are representative of the range of potential synoptic and mesoscale climates and local climate zones in which heat mitigation strategies may be applied. Many degrees of freedom are unaccounted for in the informal meta-analysis, such as weather conditions, leaf area density, simulated air temperature height—contributing to its 'informal' nature. Importantly, informal meta-analysis results focus on episodic summer scenarios when heat mitigation implementations are likely to have the greatest benefits, whereas guidance for practitioners (e.g. urban planners) should include assessment of impacts throughout the whole year (Krayenhoff and Voogt 2010, Yang and Bou-Zeid 2018).

5.3. Ranking effectiveness of heat mitigation strategies

Sample sizes of high-quality studies for each heat mitigation strategy in section 4 can provide initial guidance related to effectiveness of heat reduction options, particularly in combination with recent evidence from the literature. In addition to limited sample size, assessment of consensus efficacies of different heat mitigation strategies is complicated by the strong modulation of heat mitigation effectiveness by meteorological and neighbourhood contexts, even for clear sky, low wind, afternoon conditions. For example, low vegetation will tend to be more effective in dry conditions than moist atmospheric conditions, and rooftop strategies are likely to substantively influence the pedestrian level for lowrise neighbourhoods, but less so for neighbourhoods with taller buildings (Botham-Myint *et al* 2015).

Reflective surfaces and coatings have been widely studied across both model scales. Our meta-analysis results yield ACE values for roofs of 1.6 °C–5.8 °C, where the upper limit is in agreement with a previous review (Krayenhoff and Voogt 2010). However, our meta-analysis does not permit clear ranking of roof versus street-level albedo in terms of air temperature reduction effectiveness, perhaps due to the current status of canopy layer turbulent mixing parameterizations (see section 5.1.3). Santamouris *et al* (2017) report ACE values that generally agree with our meta-analysis results (2.3–6.2 for roofs, and 2.7–9.5 for pavement), and they provide some indication that reflective pavement may provide more pedestrian level cooling than reflective roofs. Our current Phoenix-based mesoscale simulation results, using a multi-layer urban canopy model with very detailed resolution of the urban canopy, suggest that

reflective pavement (ACE ≈11) may be about twice as effective as reflective roofs (ACE ≈4–5) at lowering pedestrian-level air temperature, even for lowrise neighbourhoods (figure 8; supplementary figure 5). However, recent simulation results find street-level albedo ACE of ∼2 °C–3 °C (Mohegh *et al* 2017; Tsoka *et al* 2018a), and recent measurements from California cities yield roof albedo ACE of 2.5 °C–18.4 °C (Mohegh *et al* 2018). Few studies have examined reflective building walls in isolation, but recent studies indicate that they are substantially less effective at cooling air temperature compared to reflective roofs, at least during daytime (Zhang *et al* 2019), and their impacts more strongly depend on factors such as urban density (Nazarian *et al* 2019). Therefore, the available evidence does not clearly distinguish between the cooling effectiveness of high albedo implementations when applied on ground level surfaces vs. roofs (when applied on shorter, e.g. 1–2 storey buildings), but does suggest that high albedo building walls are less effective. Importantly, high ground and wall albedos can have substantial negative consequences for pedestrians, drivers, and buildings (Yaghoobian *et al* 2010, Erell *et al* 2014, Schrijvers *et al* 2016), which reflective rooftops typically avoid. Moreover, elevated ground-level albedos are difficult to maintain in practice and do not combine well with trees, while tree cover generally provides large radiative benefits to pedestrians instead of the potential disbenefits associated with reflective pavement.

Urban vegetation has been widely studied; however, unlike reflective coatings, which are applied to engineered materials, model representation of urban vegetation typically requires complex soil-vegetation schemes that must be combined with 'urban' models of building and road interactions with the atmosphere. Moreover, cooling from urban vegetation, particularly low vegetation, is highly dependent on soil moisture, and often LAI as well, both of which can exhibit strong variation spatiotemporally and between studies. Therefore, ranking vegetation strategy effectiveness based on the diversity of existing scenarios in the literature is a complex task. Nevertheless, our informal meta-analysis suggests that green roofs exhibit less cooling effectiveness compared to ground-level vegetation and street trees (section 4.2). Likewise, a recent review suggests that trees provide more cooling than green roofs, and that low vegetation at ground-level, such as grass, may fall somewhere in between (Santamouris *et al* 2017). De Munck *et al* (2018) also find that ground-level vegetation is more effective than green roofs for street-level cooling. There is observational evidence that trees provide more cooling relative to ground-level vegetation during daytime (Shashua-Bar *et al* 2009) but the opposite at night, that the combination of low and high vegetation is additive, and that green roofs can provide cooling in some cases (Bowler *et al* 2010). If

we assume that the contrast between urban parks and their built-up surroundings provides a reasonable estimate of urban cooling from unirrigated vegetation, the meta-analysis by Bowler *et al* (2010) suggests a minimum daytime VCE value of approximately 1 °C (range from 0 °C to 2 °C–3 °C) for urban vegetation. Their value represents a minimum VCE because it assumes nearby urban areas that serve as the urban reference for the park cooling signal do not have any vegetation ($\lambda_s = 0$), and that urban parks are devoid of impervious surfaces ($\lambda_s = 1$), conditions that are unlikely to be fully satisfied. This result from Bowler *et al* (2010) is in broad agreement with, or slightly lower than, the simulated VCE ranges reported here for trees (3.3 °C; figure 7(a)) and low/bulk vegetation (~3 °C; figure 7(b)).

Urban irrigation has received less attention compared with vegetation and reflective surfaces (Gao and Santamouris 2019). Both irrigation and water bodies have the potential to cool the daytime urban environment, and available evidence indicates that irrigation may be particularly effective (Vahmani and Hogue 2015, Broadbent *et al* 2018a). However, recent simulations suggest that irrigation is less effective than trees (Broadbent *et al* 2019a), and that it increases nocturnal temperatures (Vahmani and Ban-Weiss 2016), as do water bodies during mid-summer (Hu and Li 2020) and late summer (Theeuwes *et al* 2013, Steeneveld *et al* 2014), by virtue of their high heat capacity. Simulation results also demonstrate the dependence of urban irrigation-induced cooling magnitude on pervious fraction of urban neighbourhoods (Vahmani and Hogue 2015), and recent modelling work suggests pavement watering as a particularly effective heat mitigation strategy in densely built urban areas that may lack pervious area (Daniel *et al* 2018). Importantly, air temperature cooling from extensive evaporation also increases humidity, and therefore does not necessarily improve thermal comfort or reduce the thermal load on air conditioning systems. Vegetation strategies in many climates will not be effective without irrigation, and hence these two strategies are often strongly linked. For example, numerical modelling suggests that extensive green roofs offer little to no cooling without sufficient soil moisture (Li *et al* 2014, Heusinger *et al* 2018), which depends on irrigation in drier locations or seasons.

Overall, there is partial agreement that ground-level strategies provide more street-level air temperature cooling, particularly for greening strategies. Comparison of reflective coatings with vegetation and irrigation strategies is complicated by the variation of cooling effectiveness of the latter with LAI and especially soil moisture and/or irrigation amount and timing. Our informal meta-analysis does suggest that their cooling potential is of the same order of magnitude (e.g. assuming albedo treatments increase local surface albedo by 0.50 on average, median ACE and VCE in figures 6 and 7 are approximately equivalent). Observational evidence indicates that greening strategies are more effective than water bodies during mid-summer for a city in a temperate climate (Hu and Li 2020). Although not the subject of the current review, vegetation and irrigation strategies applied below roof level within the urban canopy have the advantage of improving rather than worsening the radiation environment during hot conditions (e.g. Daniel *et al* 2018), unlike reflective surfaces. Trees are particularly effective in this regard (Coutts *et al* 2016, Middel and Krayenhoff 2019).

5.4. Application of ACE and VCE metrics to heat reduction decision-making

Decision making by urban policy makers and city planners should start by establishing appropriate objectives that can contribute to community well-being, resilience and sustainability (Sailor *et al* 2016). These may include improving pedestrian thermal comfort, reducing energy demand in buildings, improving building resilience indoors in the absence of air conditioning or improving urban air quality. Air temperature can be an important factor in all of these, but a reduction in air temperature does not necessarily indicate that the strategy applied is successful, nor is the lack of a temperature effect a definite indicator that a proposed strategy is not effective. The following examples illustrate this point.

(a) Daytime air temperature beneath a fabric shade canopy may be slightly higher than the temperature in an adjacent unshaded space (because of reduced aerodynamic mixing), yet thermal comfort is enhanced substantially due to the reduction in the radiant load (Shashua-Bar *et al* 2011).

(b) Replacing a concrete pavement with grass is unlikely to result in a substantial reduction in near-surface air temperature, yet it can provide a substantial reduction in mean radiant temperature adjacent to it, and thus a real improvement in thermal comfort to pedestrians (Snir *et al* 2016, Middel and Krayenhoff 2019).

(c) Urban densification will almost certainly intensify the nocturnal urban heat island, yet total annual energy consumption for air conditioning and heating buildings may be reduced, even in relatively warm climates with mild winters and warm summers (Erell and Kalman 2015). The balance will be determined not only by climate, but by the thermal characteristics of the building in question and by the effect of building massing on mutual shading and the potential for natural ventilation.

(d) High albedo pavements and wall surfaces will reduce daytime air temperature in urban street canyons, yet pedestrian thermal comfort may be adversely affected due to an increase in the radiant load (Erell *et al* 2014, Schrijvers *et al* 2016),

and the increased radiant load on windows and other glazed surfaces may increase building cooling demand (Yaghoobian *et al* 2010, Yaghoobian and Kleissl 2012).

Consequently, air temperature reduction *per se* should not be the be-all and end-all of strategies to mitigate urban heat. Rather, it is important to be able to account for the effect of different mitigation strategies on air temperature, because it is an essential input for assessing metrics relevant to each of the objectives proposed above, if it is provided at the appropriate spatial and temporal scales. Because planners and engineers outside the research community will not generally have access to appropriate modelling tools and the expertise required to use them, this information should be generated by the urban climate modelling community, communicated in formats that are accessible and useful (Mills *et al* 2010).

What then is the practical benefit of the consensus values of ACE and VCE metrics generated in this study after rigorous evaluation of the existing peer-reviewed literature? They may be used as an initial guideline for assessing potential outcomes for a given heat mitigation strategy at the neighbourhood scale, helping to avoid unrealistic expectations that may be fostered by proponents of particular design interventions.

For albedo strategies, the temperature reduction achievable is:

$$\Delta T = -\Delta \alpha_s \cdot \lambda_s \cdot \text{ACE}, \tag{4}$$

where $\Delta \alpha_s$ is the change in albedo of the modified surface (i.e. mean albedo of original roofs or pavement subtracted from albedo of proposed reflective roofs or pavement), and λ_s is the area of modified surface divided by the corresponding overall plan area (e.g. the area fraction of roof or road to be modified, derived from aerial imagery, for example). Note that $\Delta \alpha_s$ will reduce over time for many high albedo coatings due to weathering.

VCE may be applied analogously to assess temperature reduction potential of vegetation applications:

$$\Delta T = -\lambda_s \cdot \text{VCE}. \tag{5}$$

However, for low vegetation and green roofs, temperature reduction calculated based on VCE is likely to be applicable only for soil moisture conditions that closely resemble those in the studies used to obtain VCE. An initial objective should therefore be assessment of VCE for several different green infrastructure implementations with sufficient root zone soil moisture and therefore minimized stomatal resistance (e.g. irrigated vegetation). Trees have deeper roots and provide cooling additionally via shade, and therefore their VCE is less likely to be strongly influenced by soil moisture.

Great caution should be exercised when using these metrics because they are affected by several factors, including neighbourhood characteristics and ambient meteorology. Thus, different ACE and VCE values may be obtained for different seasons/latitudes, for compact and/or high rise neighbourhoods vs. dispersed and/or low rise neighbourhoods, and for different times of day (as in figures 6 and 7). Cooling from addition of trees may vary nonlinearly with existing tree canopy cover over certain canopy cover ranges (Ziter *et al* 2019), indicating that VCE may need to be assessed a function of existing canopy cover. Given the large potential variation along several degrees of freedom, worthy initial objectives include assessment of ACE and VCE for each heat mitigation implementation for hot conditions in neighbourhood types that are most inhabited worldwide, namely, for clear sky, light wind, summertime afternoon conditions in open and compact residential neighbourhoods, and for both mid-latitude and (sub)tropical cities. Any reported ACE or VCE magnitudes should be clearly contextualized with metadata describing their meteorological, neighbourhood, and temporal context. Importantly, equations (4) and (5) are applicable at the neighbourhood scale ($\sim$500 m and larger), and the cooling impacts of any such heat reduction application across a neighbourhood will influence downwind areas and also be subject to dilution at upwind edges (e.g. Broadbent *et al* 2020a).

6. Conclusions

We systematically and critically review 146 studies from the past three decades that conduct physically-based numerical modelling of air temperature reductions resulting from implementation of infrastructure-based urban heat reduction strategies. Studies are grouped into two modelling scales for critical assessment: microscale (e.g. ENVI-met), and mesoscale (e.g. WRF-SLUCM). Street tree cooling has primarily been assessed at the microscale, whereas mesoscale modelling has favoured reflective roof treatments, which are attributed to model physics limitations at each scale. We develop 25 sub-criteria, grouped into seven criteria, that assess contextualization and reliability of each study based on metadata reporting and methodological quality, respectively. Metadata criteria include reporting of neighbourhood characteristics and meteorological conditions, heat mitigation strategy application, and spatio-temporal specification of the affected air temperature. Methodological criteria assess model physics, model evaluation, and model application. Studies most often fail to adequately characterize the neighbourhood(s) in which heat mitigation was applied, particularly at the mesoscale, or perform a suitable evaluation of the base case model configuration applied to assess cooling, particularly at the

microscale. Specific sub-criteria that most require improvement include reporting the areal coverage of the heat mitigation implementation, and evaluation of model physics representations of both urban canopy layer turbulent transport and the heat mitigation implementations themselves. Example heat reduction modelling studies that exemplify good metadata reporting and/or modelling methodology are selected based on their performance against the seven criteria.

Forty-seven studies that exhibit higher methodological quality were identified for inclusion in an informal meta-analysis of cooling effectiveness for select heat mitigation strategies. We define the Albedo Cooling Effectiveness (ACE) and Vegetation Cooling Effectiveness (VCE) metrics to control for a primary degree of freedom that generates variation of air temperature cooling magnitude between studies: the plan area fraction or intensity of the heat mitigation implementation. High reflectivity coatings or materials offer $\approx$0.2 °C–0.6 °C cooling per 0.10 neighbourhood albedo increase (i.e. ACE $\approx$2 °C– 6 °C), and trees yield $\approx$0.3 °C cooling per 0.10 canopy cover increase (i.e. VCE $\approx$3 °C), for afternoon clear-sky summer conditions. VCE of low vegetation and green roofs varies more strongly between studies. As expected, *ACE* and VCE are generally smaller during nighttime and during winter, indicating that these strategies generate more cooling for daytime clear sky summer conditions. When combined with evidence from the literature, our informal meta-analysis suggests that: (a) albedo and vegetation strategies offer broadly similar magnitudes of air temperature reduction per area application; (b) ground-level albedo implementations may offer somewhat more pedestrian-level air temperature reduction than roof-level albedo per neighbourhood average albedo increase; and (c) trees and ground-level vegetation offer somewhat more pedestrian-level cooling per area application compared with green roofs. More high-quality studies are required to assess and refine these preliminary conclusions, and future meta-analyses are recommended, particularly based on observations (e.g. Bowler *et al* 2010). Results from vegetation strategies (i.e. VCE) are complicated by the variation of soil moisture between studies, and other factors such as plant species and LAI, whereas the ACE metric accounts for the variation in the magnitude of albedo increase between studies. Although available high-quality simulation studies offer no clear 'winner' in terms of air temperature cooling effectiveness, this is not unexpected since different heat mitigation strategies will be more adapted to different climates. Importantly, albedo and vegetation strategies have numerous additional benefits and disbenefits that must be factored into decision making; for example, vegetation can increase humidity and associated disbenefits, while trees reduce solar radiation exposure and benefit pedestrians and buildings during hot weather, and high

ground-level albedo can radiatively impact pedestrians and drivers, rendering large albedo increases more feasible on rooftops.

The most striking finding from the informal meta-analysis, however, is the strong apparent dependence of ACE and VCE on model scale, particularly for roof albedo and green roof implementations. We present further evidence demonstrating that ACE (and likely VCE) can also vary strongly between models at the same scale for otherwise identical simulation scenarios. These results suggest that current model physics representations of heat mitigation strategies play a large role in the variation of cooling magnitudes between studies. Importantly, evaluation of an urban climate or micrometeorological model's ability to reproduce observed air temperatures does not indicate that the model is suitable for a heat mitigation experiment, because the physics associated with the heat mitigation strategy (e.g. the green roof or street tree sub-model) is not evaluated in the base case simulation. We conclude that evaluation of the base case model configuration is not a sufficient prerequisite for simulation of cooling from heat mitigation implementations.

In response, we identify a three-phase framework for assessment of the suitability of a numerical model for a heat mitigation experiment. Phase 1 comprises development and evaluation of the urban climate or micrometeorological model, including testing model accuracy with respect to vertical convective transport of heat within and above the urban canopy. Phase 2 involves development and testing of the modelled physical processes associated with the heat mitigation implementation (e.g. a green roof or street tree sub-model). Phases 1 and 2 will typically occur prior to a heat mitigation experiment. Phase 3 occurs as part of a heat mitigation experiment and comprises evaluation and calibration of the base case model against air temperature measurements at spatial and temporal scales of relevance to the intended numerical experiment. While there is opportunity to improve model testing at all three phases, it is the testing portion of phase 2, that is, assessment of the model representation of the heat mitigation strategy against measured data, that is least often performed yet critical for accurate simulation of heat mitigation cooling. Moreover, phase 1 assessment of vertical heat transport is also critical for accurate simulation of heat mitigation cooling yet seldom undertaken in rigorous fashion. Rigorous evaluation of model physics in phases 1 and 2 underpins the reliability and utility of heat mitigation simulation results, but is hindered by the limited availability of observational data at appropriate temporal and spatial scales. Development of standardized measurement databases for evaluation of urban canopy processes and processes related to common heat mitigation implementations is recommended (see section 5.1.3).

Although model physics currently influences numerical assessment of heat mitigation strategy cooling effectiveness, many other factors modulate actual cooling effectiveness. For example, different heat reduction strategies will be more adapted to different synoptic scale climates or neighbourhood configurations. Moreover, several strategies have climate-relevant impacts that are not fully accounted for by their effects on air temperature, for example shading by street trees, or the building insulation benefits of green roofs. Ultimately, the underlying objective of infrastructure-based heat mitigation strategies is not to reduce air temperature alone, but rather to provide environments that are healthier and more thermally comfortable and efficient. Each heat mitigation strategy additionally has numerous non-climatic benefits and/or drawbacks related to aesthetics, habitat, function, hydrology, cost, historical context, health and so on, which must be factored into choices among different strategies. Implementation of heat reduction strategies must ultimately be place-based decisions that account for the unique context of each city, neighbourhood or street, and heat mitigation strategies, and combinations thereof, should be chosen in order to maximize co-benefits and minimize potential negative consequences. Moreover, near-term implementation of heat adaptation strategies should anticipate future urban climates, since large increases in heat exposure (Broadbent *et al* 2020b) and impacts on liveability (Kusaka *et al* 2012) are projected for scenarios without adaptation, and some adaptation options take time (e.g. street tree maturation). While the infrastructure-based air temperature reduction strategies addressed in this work are critical for decreasing the future heat burden in cities, maintenance of current warm season liveability and health outcomes in cities globally additionally requires substantive greenhouse gas emissions reductions (Krayenhoff *et al* 2018) and street-scale strategies, such as provision of shade, that help target heat exposure more comprehensively (Middel and Krayenhoff 2019).

Data availability statement

The data that support the findings of this study are available upon reasonable request from the authors.

Acknowledgments

ORCID iDs

E Scott Krayenhoff ⬤ https://orcid.org/0000-0002-4776-4353
Lei Zhao ⬤ https://orcid.org/0000-0002-6481-3786
Matei Georgescu ⬤ https://orcid.org/0000-0001-7321-2483
Ariane Middel ⬤ https://orcid.org/0000-0002-1565-095X
James A Voogt ⬤ https://orcid.org/0000-0002-7975-0052
David J Sailor ⬤ https://orcid.org/0000-0003-1720-8214

References

Alchapar N L and Correa E N 2016 The use of reflective materials as a strategy for urban cooling in an arid "OASIS" city *Sustain. Cities Soc.* **27** 1–14

Aleksandrowicz O, Vuckovic M, Kiesel K and Mahdavi A 2017 Current trends in urban heat island mitigation research: observations based on a comprehensive research repository *Urban Clim.* **21** 1–26

Ampatzidis P and Kershaw T 2020 A review of the impact of blue space on the urban microclimate *Sci. Total Environ.* **730** 139068

Best M J and Grimmond C S B 2015 Key conclusions of the first international urban land surface model comparison project *Bull. Am. Meteorol. Soc.* **96** 805–19

Botham-Myint D, Recktenwald G W and Sailor D 2015 Thermal footprint effect of rooftop urban cooling strategies *Urban Clim.* **14** 268–77

Bowler D E, Buyung-Ali L, Knight T M and Pullin A S 2010 Urban greening to cool towns and cities: a systematic review of the empirical evidence *Landsc. Urban Plan.* **97** 147–55

Broadbent A M, Coutts A M, Nice K A, Demuzere M, Krayenhoff E S, Tapper N J and Wouters H 2019a The air-temperature response to green/blue-infrastructure evaluation tool (TARGET v1. 0): an efficient and user-friendly model of city cooling *Geosci. Model Dev.* **12** 785–803

Broadbent A M, Coutts A M, Tapper N J and Demuzere M 2018b The cooling effect of irrigation on urban microclimate during heatwave conditions *Urban Clim.* **23** 309–29

Broadbent A M, Coutts A M, Tapper N J, Demuzere M and Beringer J 2018a The microscale cooling effects of water sensitive urban design and irrigation in a suburban environment *Theor. Appl. Climatol.* **134** 1–23

Broadbent A M, Krayenhoff E S and Georgescu M 2020a Efficacy of cool roofs at reducing pedestrian-level air temperature during projected 21st century heatwaves in Atlanta, Detroit, and Phoenix (USA) *Environ. Res. Lett.* **15** 084007

Broadbent A M, Krayenhoff E S and Georgescu M 2020b The motley drivers of heat and cold exposure in 21st century US cities *Proc. Natl Acad. Sci.* **117** 21108–17

Broadbent A M, Krayenhoff E S, Georgescu M and Sailor D J 2019b The observed effects of utility-scale photovoltaics on

near-surface air temperature and energy balance *J. Appl. Meteorol. Climatol.* **58** 989–1006

Bruse M and Fleer H 1998 Simulating surface–plant–air interactions inside urban environments with a three dimensional numerical model *Environ. Modelling Softw.* **13** 373–84

Cao M, Rosado P, Lin Z, Levinson R and Millstein D 2015 Cool roofs in Guangzhou, China: outdoor air temperature reductions during heat waves and typical summer conditions *Environ. Sci. Technol.* **49** 14672–9

Chen F, Kusaka H, Bornstein R, Ching J, Grimmond C S B, Grossman-Clarke S and Sailor D 2011 The integrated WRF/urban modelling system: development, evaluation, and applications to urban environmental problems *Int. J. Climatol.* **31** 273–88

Chow W T and Brazel A J 2012 Assessing xeriscaping as a sustainable heat island mitigation approach for a desert city *Build. Environ.* **47** 170–81

Coutts A M, White E C, Tapper N J, Beringer J and Livesley S J 2016 Temperature and human thermal comfort effects of street trees across three contrasting street canyon environments *Theor. Appl. Climatol.* **124** 55–68

Crank P J, Sailor D J, Ban-Weiss G and Taleghani M 2018 Evaluating the ENVI-met microscale model for suitability in analysis of targeted urban heat mitigation strategies *Urban Clim.* **26** 188–97

Cui Y Y and De Foy B 2012 Seasonal variations of the urban heat island at the surface and the near-surface and reductions due to urban vegetation in Mexico City *J. Appl. Meteorol. Climatol.* **51** 855–68

Daniel M, Lemonsu A and Viguie V 2018 Role of watering practices in large-scale urban planning strategies to face the heat-wave risk in future climate *Urban Clim.* **23** 287–308

De Munck C, Lemonsu A, Masson V, Le Bras J and Bonhomme M 2018 Evaluating the impacts of greening scenarios on thermal comfort and energy and water consumptions for adapting Paris city to climate change *Urban Clim.* **23** 260–86

Emmanuel R and Fernando H J 2007 Urban heat islands in humid and arid climates: role of urban form and thermal properties in Colombo, Sri Lanka and Phoenix, USA *Clim. Res.* **34** 241–51

Erell E and Kalman Y 2015 Impact of increasing the depth of urban street canyons on building heating and cooling loads Tel Aviv, Israel *ICUC9—The 9th Int. Conf. Urban Climate* (*Toulouse, France, 20–24 July 2015*)

Erell E, Pearlmutter D, Boneh D and Kutiel P B 2014 Effect of high-albedo materials on pedestrian heat stress in urban street canyons *Urban Clim.* **10** 367–86

Erell E and Williamson T 2006 Simulating air temperature in an urban street canyon in all weather conditions using measured data at a reference meteorological station *Int. J. Climatol.* **26** 1671–94

Fallmann J, Emeis S and Suppan P 2013 Mitigation of urban heat stress—a modelling case study for the area of Stuttgart *DIE ERDE–J. Geogr. Soc. Berlin* **144** 202–16

Fallmann J, Forkel R and Emeis S 2016 Secondary effects of urban heat island mitigation measures on air quality *Atmos. Environ.* **125** 199–211

Gago E J, Roldan J, Pacheco-Torres R and Ordóñez J 2013 The city and urban heat islands: a review of strategies to mitigate adverse effects *Renew. Sustain. Energy Rev.* **25** 749–58

Gao K and Santamouris M 2019 The use of water irrigation to mitigate ambient overheating in the built environment: recent progress *Build. Environ.* **164** 106346

Georgescu M 2015 Challenges associated with adaptation to future urban expansion *J. Clim.* **28** 2544–63

Georgescu M, Morefield P E, Bierwagen B G and Weaver C P 2014 Urban adaptation can roll back warming of emerging megapolitan regions *Proc. Natl Acad. Sci.* **111** 2909–14

Gillner S, Vogt J, Tharang A, Dettmann S and Roloff A 2015 Role of street trees in mitigating effects of heat and drought at highly sealed urban sites *Landsc. Urban Plan.* **143** 33–42

Giometto M G, Christen A, Meneveau C, Fang J, Krafczyk M and Parlange M B 2016 Spatial characteristics of roughness sublayer mean flow and turbulence over a realistic urban surface *Bound.-Layer Meteorol.* **160** 425–52

Grimmond C S B, Blackett M, Best M J, Baik J J, Belcher S E, Beringer J and Dandou A 2011 Initial results from Phase 2 of the international urban energy balance model comparison *Int. J. Climatol.* **31** 244–72

Gunawardena K R, Wells M J and Kershaw T 2017 Utilising green and bluespace to mitigate urban heat island intensity *Sci. Total Environ.* **584–5** 1040–55

Hamdi R and Schayes G 2008 Sensitivity study of the urban heat island intensity to urban characteristics *Int. J. Climatol.* **28** 973–82

Heusinger J, Broadbent A M, Krayenhoff E S and Weber S 2021 Adaptation of a photovoltaic energy balance model for rooftop applications. *Build. Environ.* **192** 107628

Heusinger J, Sailor D J and Weber S 2018 Modeling the reduction of urban excess heat by green roofs with respect to different irrigation scenarios *Build. Environ.* **131** 174–83

Heusinger J and Weber S 2017 Surface energy balance of an extensive green roof as quantified by full year eddy-covariance measurements *Sci. Total Environ.* **577** 220–30

Hu L and Li Q 2020 Greenspace, bluespace, and their interactive influence on urban thermal environments *Environ. Res. Lett.* **15** 034041

Huszar P, Halenka T, Belda M, Zak M, Sindelarova K and Miksovsky J 2014 Regional climate model assessment of the urban land-surface forcing over central Europe *Atmos. Chem. Phys.* **14** 12393–413

Huttner S 2012 *Further Development and Application of the 3D Microclimate Simulation ENVI-met* vol 147 (Mainz: Johannes Gutenberg-Universitat)

Imran H M, Kala J, Ng A W M and Muthukumaran S 2019 Effectiveness of vegetated patches as Green Infrastructure in mitigating Urban Heat Island effects during a heatwave event in the city of Melbourne *Weather Clim. Extremes* **25** 100217

Jacobs S J, Gallant A J, Tapper N J and Li D 2018 Use of cool roofs and vegetation to mitigate urban heat and improve human thermal stress in Melbourne, Australia *J. Appl. Meteorol. Climatol.* **57** 1747–64

Karlický J, Huszár P, Halenka T, Belda M, Žák M, Pišoft P and Mikšovský J 2018 Multi-model comparison of urban heat island modelling approaches *Atmos. Chem. Phys.* **18** 10655–74

Kim G A, Song B G and Park K H 2020 Long-term monitoring for comparison of seasonal effects on cool roofs in humid subtropical climates *Energy Build.* **206** 109572

Krayenhoff E S et al 2020 A multi-layer urban canopy meteorological model with trees (BEP-tree): street tree impacts on pedestrian-level climate *Urban Clim.* **32** 100590

Krayenhoff E S, Christen A, Martilli A and Oke T R 2014 A multi-layer radiation model for urban neighbourhoods with trees *Bound.-Layer Meteorol.* **151** 139–78

Krayenhoff E S, Moustaoui M, Broadbent A M, Gupta V and Georgescu M 2018 Diurnal interaction between urban expansion, climate change and adaptation in US cities *Nat. Clim. Change* **8** 1097

Krayenhoff E S, Santiago J L, Martilli A, Christen A and Oke T R 2015 Parametrization of drag and turbulence for urban neighbourhoods with trees *Bound.-Layer Meteorol.* **156** 157–89

Krayenhoff E S and Voogt J A 2007 A microscale three-dimensional urban energy balance model for studying surface temperatures *Bound.-Layer Meteorol.* **123** 433–61

Krayenhoff E S and Voogt J A 2010 Impacts of urban albedo increase on local air temperature at daily–annual time scales: model results and synthesis of previous work *J. Appl. Meteorol. Climatol.* **49** 1634–48

Kusaka H, Hara M and Takane Y 2012 Urban climate projection by the WRF model at 3-km horizontal grid increment:

dynamical downscaling and predicting heat stress in the 2070's August for Tokyo, Osaka, and Nagoya metropolises *J. Meteorol. Soc. Japan* **90** 47–63

Kusaka H, Kondo H, Kikegawa Y and Kimura F 2001 A simple single-layer urban canopy model for atmospheric models: comparison with multi-layer and slab models *Bound.-Layer Meteorol.* **101** 329–58

Lai D, Liu W, Gan T, Liu K and Chen Q 2019 A review of mitigating strategies to improve the thermal environment and thermal comfort in urban outdoor spaces *Sci. Total Environ.* **661** 337–53

Lee H, Mayer H and Chen L 2016b Contribution of trees and grasslands to the mitigation of human heat stress in a residential district of Freiburg, Southwest Germany *Landsc. Urban Plan.* **148** 37–50

Lee S H, Lee H, Park S B, Woo J W, Lee D I and Baik J J 2016a Impacts of in-canyon vegetation and canyon aspect ratio on the thermal environment of street canyons: numerical investigation using a coupled WRF-VUCM model *Q. J. R. Meteorol. Soc.* **142** 2562–78

Lee S H and Park S U 2008 A vegetated urban canopy model for meteorological and environmental modelling *Bound.-Layer Meteorol.* **126** 73–102

Li D, Bou-Zeid E and Oppenheimer M 2014 The effectiveness of cool and green roofs as urban heat island mitigation strategies *Environ. Res. Lett.* **9** 055002

Li X X and Norford L K 2016 Evaluation of cool roof and vegetations in mitigating urban heat island in a tropical city, Singapore *Urban Clim.* **16** 59–74

Lin B S and Lin C T 2016 Preliminary study of the influence of the spatial arrangement of urban parks on local temperature reduction *Urban For. Urban Greening* **20** 348–57

Loughner C P, Allen D J, Zhang D L, Pickering K E, Dickerson R R and Landry L 2012 Roles of urban tree canopy and buildings in urban heat island effects: parameterization and preliminary results *J. Appl. Meteorol. Climatol.* **51** 1775–93

Ma H, Shao H and Song J 2014 Modeling the relative roles of the foehn wind and urban expansion in the 2002 Beijing heat wave and possible mitigation by high reflective roofs *Meteorol. Atmos. Phys.* **123** 105–14

Makropoulou M 2017 Microclimate improvement of inner-city urban areas in a Mediterranean Coastal City *Sustainability* **9** 882

Martilli A 2002 Numerical study of urban impact on boundary layer structure: sensitivity to wind speed, urban morphology, and rural soil moisture *J. Appl. Meteorol.* **41** 1247–66

Martilli A, Clappier A and Rotach M W 2002 An urban surface exchange parameterisation for mesoscale models *Bound.-Layer Meteorol.* **104** 261–304

Martilli A, Krayenhoff E S and Nazarian N 2020 Is the urban heat Island intensity relevant for heat mitigation studies? *Urban Clim.* **31** 100541

Masson V 2000 A physically-based scheme for the urban energy budget in atmospheric models *Bound.-Layer Meteorol.* **94** 357–97

Middel A, Chhetri N and Quay R 2015 Urban forestry and cool roofs: assessment of heat mitigation strategies in Phoenix residential neighborhoods *Urban For. Urban Greening* **14** 178–86

Middel A, Häb K, Brazel A J, Martin C A and Guhathakurta S 2014 Impact of urban form and design on mid-afternoon microclimate in phoenix local climate zones *Landsc. Urban Plan.* **122** 16–28

Middel A and Krayenhoff E S 2019 Micrometeorological determinants of pedestrian thermal exposure during record-breaking heat in Tempe, Arizona: introducing the MaRTy observational platform *Sci. Total Environ.* **687** 137–51

Middel A, Turner V K, Schneider F A, Zhang Y and Stiller M 2020 Solar reflective pavement—a policy panacea to heat mitigation? *Environ. Res. Lett.* **15** 064016

Mills G, Cleugh H, Emmanuel R, Endlicher W, Erell E, McGranahan G, Ng E, Nickson A, Rosenthal J and Steemer K 2010 Climate information for improved planning and management of mega cities (needs perspective) *Proc. Environ. Sci.* **1** 228–46

Millstein D and Menon S 2011 Regional climate consequences of large-scale cool roof and photovoltaic array deployment *Environ. Res. Lett.* **6** 034001

Mohegh A, Levinson R, Taha H, Gilbert H, Zhang J, Li Y, Tang T and Ban-Weiss G A 2018 Observational evidence of neighborhood scale reductions in air temperature associated with increases in roof albedo *Climate* **6** 98

Mohegh A, Rosado P, Jin L, Millstein D, Levinson R and Ban-Weiss G 2017 Modeling the climate impacts of deploying solar reflective cool pavements in California cities *J. Geophys. Res. Atmos.* **122** 6798–817

Moriasi D N, Arnold J G, Van Liew M W, Bingner R L, Harmel R D and Veith T L 2007 Model evaluation guidelines for systematic quantification of accuracy in watershed simulations *Trans. ASABE* **50** 885–900

Nazarian N, Dumas N, Kleissl J and Norford L 2019 Effectiveness of cool walls on cooling load and urban temperature in a tropical climate *Energy Build.* **187** 144–62

Nazarian N, Krayenhoff E S and Martilli A 2020 A one-dimensional model of turbulent flow through "urban" canopies (MLUCM v2. 0): updates based on large-eddy simulation *Geosci. Model Dev.* **13** 937–53

Oke T R 1976 The distinction between canopy and boundary-layer urban heat islands *Atmosphere* **14** 268–77

Oke T R 1982 The energetic basis of the urban heat island *Q. J. R. Meteorol. Soc.* **108** 1–24

Oke T R 1989 The micrometeorology of the urban forest *Phil. Trans. R. Soc.* B **324** 335–49

Oke T R 2006 Instruments and observing methods: report No. 81: initial guidance to obtain representative meteorological observations at urban sites *World Meteorological Organization, WMO/TD (1250)*

Oleson K W, Bonan G B and Feddema J 2010 Effects of white roofs on urban temperature in a global climate model *Geophys. Res. Lett.* **37**

Pisello A L, Saliari M, Vasilakopoulou K, Hadad S and Santamouris M 2018 Facing the urban overheating: recent developments. Mitigation potential and sensitivity of the main technologies *Wiley Interdiscip. Rev. Energy Environ.* **7** e294

Razzaghmanesh M, Beecham S and Salemi T 2016 The role of green roofs in mitigating Urban Heat Island effects in the metropolitan area of Adelaide, South Australia *Urban For. Urban Greening* **15** 89–102

Rosado P J, Faulkner D, Sullivan D P and Levinson R 2014 Measured temperature reductions and energy savings from a cool tile roof on a central California home *Energy Build.* **80** 57–71

Rosenfeld A H, Akbari H, Bretz S, Fishman B L, Kurn D M, Sailor D and Taha H 1995 Mitigation of urban heat islands: materials, utility programs, updates *Energy Build.* **22** 255–65

Rosenzweig C, Solecki W D, Parshall L, Lynn B, Cox J, Goldberg R and Dunstan F 2009 Mitigating New York City's heat island: integrating stakeholder perspectives and scientific evaluation *Bull. Am. Meteorol. Soc.* **90** 1297–312

Rosenzweig C, Solecki W, Romero-Lankao P, Mehrotra S, Dhakal S, Bowman T and Ali Ibrahim S 2015 ARC3.2 summary for city leaders *Climate Change and Cities: Second Assessment Report of the Urban Climate Change Research Network* ed C Rosenzweig, W Solecki, P Romero-Lankao, S Mehrotra, S Dhakal and S Ali Ibrahim (Cambridge: Cambridge University Press)

Rotach M W, Vogt R, Bernhofer C, Batchvarova E, Christen A, Clappier A and Mitev V 2005 BUBBLE—an urban boundary layer meteorology project *Theor. Appl. Climatol.* **81** 231–61

Roth M and Lim V H 2017 Evaluation of canopy-layer air and mean radiant temperature simulations by a microclimate model over a tropical residential neighbourhood *Build. Environ.* **112** 177–89

Ryu Y H and Baik J J 2012 Quantitative analysis of factors contributing to urban heat island intensity *J. Appl. Meteorol. Climatol.* **51** 842–54

Ryu Y H, Bou-Zeid E, Wang Z H and Smith J A 2016 Realistic representation of trees in an urban canopy model *Bound.-Layer Meteorol.* **159** 193–220

Sailor D, Shepherd M, Sheridan S, Stone B, Kalkstein L, Russell A, Vargo J and Andersen T 2016 Improving heat-related health outcomes in an urban environment with science-based policy *Sustainability* **8** 1015

Salamanca F, Georgescu M, Mahalov A, Moustaoui M and Martilli A 2016 Citywide impacts of cool roof and rooftop solar photovoltaic deployment on near-surface air temperature and cooling energy demand *Bound.-Layer Meteorol.* **161** 203–21

Salamanca F, Martilli A, Tewari M and Chen F 2011 A study of the urban boundary layer using different urban parameterizations and high-resolution urban canopy parameters with WRF *J. Appl. Meteorol. Climatol.* **50** 1107–28

Salamanca F, Zhang Y, Barlage M, Chen F, Mahalov A and Miao S 2018 Evaluation of the WRF-urban modeling system coupled to Noah and Noah-MP land surface models over a semiarid urban environment *J. Geophys. Res. Atmos.* **123** 2387–408

Salata F, Golasi I, Petitti D, de Lieto Vollaro E, Coppi M and de Lieto Vollaro A 2017 Relating microclimate, human thermal comfort and health during heat waves: an analysis of heat island mitigation strategies through a case study in an urban outdoor environment *Sustain. Cities Soc.* **30** 79–96

Santamouris M 2014 Cooling the cities–a review of reflective and green roof mitigation technologies to fight heat island and improve comfort in urban environments *Sol. Energy* **103** 682–703

Santamouris M, Ding L, Fiorito F, Oldfield P, Osmond P, Paolini R, Prasad D and Synnefa A 2017 Passive and active cooling for the outdoor built environment–Analysis and assessment of the cooling potential of mitigation technologies using performance data from 220 large scale projects *Sol. Energy* **154** 14–33

Sato Y, Higuchi A, Takami A, Murakami A, Masutomi Y, Tsuchiya K, Goto D and Nakajima T 2016 Regional variability in the impacts of future land use on summertime temperatures in Kanto region, the Japanese megacity *Urban For. Urban Greening* **20** 43–55

Scherba A, Sailor D J, Rosenstiel T N and Wamser C C 2011 Modeling impacts of roof reflectivity, integrated photovoltaic panels and green roof systems on sensible heat flux into the urban environment *Build. Environ.* **46** 2542–51

Schrijvers P, Jonker H, de Roode S and Kenjereš S 2016 The effect of using a high-albedo material on the universal temperature climate index within a street canyon *Urban Clim.* **17** 284–303

Schubert S and Grossman-Clarke S 2013 The influence of green areas and roof albedos on air temperatures during extreme heat events in Berlin, Germany *Meteorol. Z.* **22** 131–43

Shahidan M F, Jones P J, Gwilliam J and Salleh E 2012 An evaluation of outdoor and building environment cooling achieved through combination modification of trees with ground materials *Build. Environ.* **58** 245–57

Shashua-Bar L and Hoffman M E 2000 Vegetation as a climatic component in the design of an urban street: an empirical model for predicting the cooling effect of urban green areas with trees *Energy Build.* **31** 221–35

Shashua-Bar L, Pearlmutter D and Erell E 2009 The cooling efficiency of urban landscape strategies in a hot dry climate *Landsc. Urban Plan.* **92** 179–86

Shashua-Bar L, Pearlmutter D and Erell E 2011 The influence of trees and grass on outdoor thermal comfort in a hot-arid environment *Int. J. Climatol.* **31** 1498–506

Shashua-Bar L, Potchter O, Bitan A, Boltansky D and Yaakov Y 2010 Microclimate modelling of street tree species effects within the varied urban morphology in the Mediterranean city of Tel Aviv, Israel *Int. J. Climatol.* **30** 44–57

Skamarock W C, Klemp J B, Dudhia J, Gill D O, Barker D M, Wang W and Powers J G 2005 *A description of the advanced research WRF version 2* (No. NCAR/TN-468+ STR) (Boulder, CO: National Center For Atmospheric Research)

Skelhorn C, Lindley S and Levermore G 2014 The impact of vegetation types on air and surface temperatures in a temperate city: a fine scale assessment in Manchester, UK *Landsc. Urban Plan.* **121** 129–40

Snir K, Pearlmutter D and Erell E 2016 The moderating effect of water-efficient ground cover vegetation on pedestrian thermal stress *Landsc. Urban Plan.* **152** 1–12

Steeneveld G J, Koopmans S, Heusinkveld B G and Theeuwes N E 2014 Refreshing the role of open water surfaces on mitigating the maximum urban heat island effect *Landsc. Urban Plan.* **121** 92–96

Stein U and Alpert P 1993 Factor separation in numerical simulations *J. Atmos. Sci.* **50** 2107–15

Stewart I D 2011 A systematic review and scientific critique of methodology in modern urban heat island literature *Int. J. Climatol.* **31** 200–17

Stewart I D and Oke T R 2012 Local climate zones for urban temperature studies *Bull. Am. Meteorol. Soc.* **93** 1879–900

Sun T, Grimmond C S B and Ni G H 2016 How do green roofs mitigate urban thermal stress under heat waves? *J. Geophys. Res. Atmos.* **121** 5320–35

Taha H 1997 Modeling the impacts of large-scale albedo changes on ozone air quality in the South Coast Air Basin *Atmos. Environ.* **31** 1667–76

Taha H, Akbari H, Rosenfeld A and Huang J 1988 Residential cooling loads and the urban heat island—the effects of albedo *Build. Environ.* **23** 271–83

Taleghani M, Sailor D and Ban-Weiss G A 2016 Micrometeorological simulations to predict the impacts of heat mitigation strategies on pedestrian thermal comfort in a Los Angeles neighborhood *Environ. Res. Lett.* **11** 024003

Theeuwes N E, Solcerová A and Steeneveld G J 2013 Modeling the influence of open water surfaces on the summertime temperature and thermal comfort in the city *J. Geophys. Res. Atmos.* **118** 8881–96

Toparlar Y, Blocken B, Maiheu B V and van Heijst G J F 2018 The effect of an urban park on the microclimate in its vicinity: a case study for Antwerp, Belgium *Int. J. Climatol.* **38** e303–22

Toparlar Y, Blocken B, Maiheu B and Van Heijst G J F 2017 A review on the CFD analysis of urban microclimate *Renew. Sustain. Energy Rev.* **80** 1613–40

Tsoka S, Theodosiou T, Tsikaloudaki K and Flourentzou F 2018a Modeling the performance of cool pavements and the effect of their aging on outdoor surface and air temperatures *Sustain. Cities Soc.* **42** 276–88

Tsoka S, Tsikaloudaki A and Theodosiou T 2018b Analyzing the ENVI-met microclimate model's performance and assessing cool materials and urban vegetation applications–a review *Sustain. Cities Soc.* **43** 55–76

Vahmani P and Ban-Weiss G 2016 Climatic consequences of adopting drought-tolerant vegetation over Los Angeles as a response to California drought *Geophys. Res. Lett.* **43** 8240–9

Vahmani P and Hogue T S 2015 Urban irrigation effects on WRF-UCM summertime forecast skill over the Los Angeles metropolitan area *J. Geophys. Res. Atmos.* **120** 9869–81

Vahmani P, Sun F, Hall A and Ban-Weiss G 2016 Investigating the climate impacts of urbanization and the potential for cool roofs to counter future climate change in Southern California *Environ. Res. Lett.* **11** 124027

Valiela I 2001 *Doing Science: Design, Analysis, and Communication of Scientific Research* (Oxford: Oxford University Press)

Wang Y, Berardi U and Akbari H 2016 Comparing the effects of urban heat island mitigation strategies for Toronto, Canada *Energy Build.* **114** 2–19

Willmott C J 1982 Some comments on the evaluation of model performance *Bull. Am. Meteorol. Soc.* **63** 1309–13

Willmott C J and Matsuura K 1995 Smart interpolation of annually averaged air temperature in the United States *J. Appl. Meteorol.* **34** 2577–86

Yaghoobian N and Kleissl J 2012 Effect of reflective pavements on building energy use *Urban Clim.* **2** 25–42

Yaghoobian N, Kleissl J and Krayenhoff E S 2010 Modeling the thermal effects of artificial turf on the urban environment *J. Appl. Meteorol. Climatol.* **49** 332–45

Yang J and Bou-Zeid E 2018 Should cities embrace their heat islands as shields from extreme cold? *J. Appl. Meteorol. Climatol.* **57** 1309–20

Yang J, Liu H, Sun J, Zhu Y, Wang X, Xiong Z and Jiang W 2015 Further development of the regional boundary layer model to study the impacts of greenery on the urban thermal environment *J. Appl. Meteorol. Climatol.* **54** 137–52

Yang J, Wang Z H, Kaloush K E and Dylla H 2016 Effect of pavement thermal properties on mitigating urban heat islands: a multi-scale modeling case study in Phoenix *Build. Environ.* **108** 110–21

Yang S R and Lin T P 2016 An integrated outdoor spaces design procedure to relieve heat stress in hot and humid regions *Build. Environ.* **99** 149–60

Zhang J, Li Y, Tao W, Liu J, Levinson R, Mohegh A and Ban-Weiss G 2019 Investigating the urban air quality effects of cool walls and cool roofs in Southern California *Environ. Sci. Technol.* **53** 7532–42

Zhao L, Oleson K, Bou-Zeid E, Krayenhoff E S, Bray A, Zhu Q, Zheng Z, Chen C and Oppenheimer M 2021 Global multi-model projections of local urban climates *Nat. Clim. Change* (https://doi.org/10.1038/s41558-020-00958-8)

Zhou X, Ooka R, Chen H, Kawamoto Y and Kikumoto H 2016 impacts of inland water area changes on the local climate of Wuhan, China *Indoor Built Environ.* **25** 296–313

Zhou Y and Shepherd J M 2010 Atlanta's urban heat island under extreme heat conditions and potential mitigation strategies *Nat. Hazards* **52** 639–68

Ziter C D, Pedersen E J, Kucharik C J and Turner M G 2019 Scale-dependent interactions between tree canopy cover and impervious surfaces reduce daytime urban heat during summer *Proc. Natl Acad. Sci.* **116** 7575–80