Impact of urban canopy models and external parameters on the modelled urban energy balance in a tropical city

M. Demuzere\textsuperscript{a,b,c}, S. Harshan\textsuperscript{b}, L. Järvi\textsuperscript{d}, M. Roth\textsuperscript{b}, C.S.B. Grimmond\textsuperscript{e}, V. Masson\textsuperscript{f}, K.W. Oleson\textsuperscript{g}, E. Velasco\textsuperscript{h}, H. Wouters\textsuperscript{a}

\textsuperscript{a}KU Leuven, Department of Earth and Environmental Sciences, Celestijnenlaan 200E, 3001 Leuven, Belgium. \textsuperscript{b}Department of Geography, National University Singapore \textsuperscript{c}Laboratory of Hydrology and Water Management, Ghent University, B-9000, Ghent, Belgium \textsuperscript{d}University of Helsinki, Department of Physics, Helsinki, Finland \textsuperscript{e}Department of Meteorology, University of Reading, UK \textsuperscript{f}CNRM, Météo-France and CNRS, Toulouse, France \textsuperscript{g}National Center for Atmospheric Research, Boulder, CO, USA \textsuperscript{h}Singapore-MIT Alliance for Research and Technology (SMART), Center for Environmental Sensing and Modeling (CENSAM), Singapore

*Correspondence to: Celestijnenlaan 200E, 3001 Leuven, Belgium. E-mail: Matthias.demuzere@kuleuven.be

To date, existing urban land surface models (ULSMs) have been mostly evaluated and optimised for mid- and high-latitude cities. For the first time, we provide a comparative evaluation of four ULSMs for a tropical residential neighbourhood in Singapore using directly measured energy balance components. The simulations are performed offline, for an 11-month period, using the bulk scheme TERRA_URB and three models of intermediate complexity (CLM, SURFEX and SUEWS). In addition, information from three different parameter lists are used to quantify the impact (interaction) of (between) external parameter settings and model formulations on the modelled urban energy balance components. Encouragingly, overall results indicate good model performance for most energy balance components and align well with previous findings for mid-latitude regions, suggesting the transferability of these models to (sub)tropical regions. Similar to results from mid-latitude regions, the outgoing longwave radiation and latent heat flux remain the most problematic fluxes. In addition, the various combinations of models and different parameter values suggest that error statistics tend to be dominated more by the choice of the latter than the choice of model. Finally, our intercomparison framework enabled the attribution of common deficiencies in the different model approaches found previously in mid-latitude regions: the simple representation of water intercepted by impervious surfaces leading to a positive bias in the latent heat flux directly after a precipitation event; and the positive bias in modelled outgoing longwave radiation that is partly due to neglecting the radiative interactions of water vapor between the surface and the tower sensor. These findings suggest that future developments in urban climate research should continue the integration of more physically-based processes in urban canopy models, ensure the consistency between the observed and modelled atmospheric properties and focus on the correct representation of urban morphology, water storage and thermal and radiative characteristics.

Key Words: urban canopy models, surface energy balance, tropical residential neighbourhood, water vapor opacity, surface interception distribution, local climate zones

1. Introduction

Cities are hot spots that drive environmental change at multiple scales (Grimm et al. 2008, Georgescu et al. 2014, 2015). As the earth’s climate will change over the coming decades (Stocker et al. 2013), global warming will impact urban areas especially hard being a major threat to the health and well-being of human populations (Watts et al. 2015). The tendency for urban areas to be warmer than their surrounding rural environments (referred to as the urban heat island) is a well established phenomenon and originates from differences in surface energy exchanges over built-up and natural areas (Oke 1982). A range of urban characteristics, including the high thermal admittance of urban materials, dominance of impervious surfaces (and thus reduced natural pervious surfaces) and the presence of three-dimensional geometries, enhance the absorption of incoming short-wave radiation from the sun and storage of heat energy, partition less energy into evapotranspiration, and reduce a city’s ability to cool after sunset thereby warming the atmosphere nearby.

It is now widely accepted by the climate community that dominant processes leading to local and regional urban warming effects need to be incorporated in climate models. Even though some authors argue that the impact of urban areas might be negligible in terms of temperature and precipitation at coarser spatial resolutions (e.g. > 10 km) or depend on the region of interest (Trusilova et al. 2013), others find that e.g. urban expansion implemented at such coarse scales is able to raise near-surface temperatures, not only over the urban areas but also over larger neighbouring areas (Georgescu et al. 2014). Further, the tendency of regional climate modelling towards convection permitting model scales supports the need for a proper representation of the local city climate (Prein et al. 2015, Phelan et al. 2015). This will allow for an improved assessment of local (urban) climate (projections) as well as the potential for investigating various heat-stress mitigation and adaptation strategies (Prein et al. 2015) such as green urban infrastructure (e.g. Bowler et al. 2010, Demuzere et al. 2014)), water sensitive urban design (e.g. Coutts et al. 2013) and changing radiative properties of the built environment such as e.g. ‘cool roofs’ (Oleson et al. 2010a, Georgescu et al. 2014).
Cities and their impacts on atmosphere are included in global, regional and local climate models using urban land surface models (ULSMs). A large number of ULSMs are currently available that vary considerably in complexity: from simple bulk representations of the surface through more recent developments that consider a complete energy balance at various levels within the urban canyon (Best and Grimmond 2015). As the first of its kind, Grimmond et al. (2010) launched the urban land surface model intercomparison project (PILPS-urban) to objectively assess and compare the performance of existing ULSMs. This intercomparison tested a large number of ULSMs (Table I in Best and Grimmond 2015) in offline simulations to evaluate their performance over a light industrial area in Vancouver (British Columbia, Canada) and a suburban area in Melbourne (Australia) (Grimmond et al. 2010, 2011), analysed the representation of the seasonal cycle (Best and Grimmond 2013) and addressed the role of initial conditions and the response to certain atmospheric conditions (Best and Grimmond 2014). This effort helped to identify the dominant physical processes, the level of complexity needed in an application specific context, and parameter requirements. Other ULSMs’ evaluations in online mode (coupled to an atmospheric/climate model) include: a single- and a multi-layer urban parametrisation within the Coupled Ocean-Atmosphere Mesoscale Prediction System for the New York City metropolitan area (Holt and Pullen 2007); various urban canopy schemes (slab, single-layer and multi-layer with and without a building energy model) in the Weather Research and Forecasting/Chemistry model to evaluate the regional climate and air quality of the Yangtze River Delta (China) (Liao et al. 2014), and high-resolution regional climate simulations over Berlin (Germany) with the COSMO-CLM regional climate model coupled to the Town Energy Budget (TEB) model (Trusilova et al. 2013, 2015), the Double Canyon Effect Parametrisation (DCEP) scheme (Schubert and Grossman-Clarke 2012) and TERRA_URB (Wouters et al. 2015, 2016).

In addition to the model physics, the parameters describing the urban surface in terms of land cover, morphology, geometry or radiative and thermal properties play an important role. Ideally site-specific information about building materials are available, but often generalised global values are used. Regional tables such as Jackson et al. (2010) and the ECOCLIMAP data (Chameaux et al. 2005; Faroux et al. 2013) are commonly used by the Community Land Model (Oleson et al. 2008b) or the SURFEX model suite (Masson et al. 2013), respectively. Others have addressed the sensitivity of these parameters via an optimisation approach, e.g. by perturbing a set of selected parameters at each step and evaluating how modelled variables evolve. Loridan et al. (2010) tested the sensitivity of surface energy fluxes to varying input parameter values for the single-layer urban canopy parametrisation used in the Weather Research and Forecasting model, and used this framework to suggest a set of recommended parameter values for three categories of urban areas (Loridan and Grimmond 2012). Song and Wang (2014) coupled a single column model to the single-layer urban canopy model SLUCM to address changing urban morphology, albedo, vegetation fraction and aerodynamic roughness on the growth of the atmospheric boundary layer and the distributions of temperature and humidity in the mixed layer under convective conditions. Their results conclude that changes in land-surface properties (hydrothermal or geometric) have a significant impact on the evolution of the overlapping boundary layer. In addition, Wouters et al. (2016) tested urban canopy parameter value ranges from the Local Climate Zones (Stewart and Oke 2012) in an online simulation over the Belgian territory with COSMO-CLM coupled to TERRA_URB. Their study, amongst other results, reveals that, with respect to surface temperatures, air temperatures and associated urban heat islands, one should prioritise those parameters that are most sensitive: the thermal parameters and the anthropogenic heat emissions.

From the above it is clear that off- and online model intercomparisons combined with sensitivity tests related to external parameters are very demanding and also have limitations. Cities are located in vastly different climatic zones and have diverse built-up characteristics (cf. Local Climate Zones, Stewart and Oke 2012). More research is therefore required for (climate) conditions not addressed in previous studies (Best and Grimmond 2015). A recent study by Karsisto et al. (2016a,b) answered this call and tested the performance of three ULSMs for several sites in and around the high-latitude city of Helsinki (Finland): the Community Land Model (CLM, Lawrence et al. 2011), the Surface Urban Energy and Water Balance Scheme (SUEWS, Järvi et al. 2011, 2014) and SURFEX (Masson et al. 2013). The present study extends this effort by evaluating the three above-mentioned ULSMs together with TERRA_URB (Wouters et al. 2015, 2016) over a residential neighbourhood in a tropical city. Even though currently most of the rapidly expanding urban areas are located in (sub)tropical regions (Seto et al. 2012), the total number of (sub)tropical urban climate studies is limited (Roth 2007). Given that the impact of climate (change) on population health and well-being in these regions has not yet been well established (Kjellstrom and McMichael 2013; Caminade et al. 2014), it is key to extend the process-based work that seeks to improve our understanding and representation of the (sub)tropical urban climate processes at play (Roth 2007). Against this background, Singapore’s tropical climate provides a unique testbed against which the ULSMs have not yet been benchmarked: a very humid environment with a very small diurnal temperature range. In addition, the selected observation period includes an exceptionally dry two-months period, allowing for an in-depth exploration of the role of precipitation (and the lack thereof) in a tropical setting.

The objective of the present study is to use directly measured energy balance fluxes (Roth et al. 2016; Velasco et al. 2016) to perform model evaluations using three types of input parameter lists. Evaluations are performed for outgoing short- and long-wave radiation ($K^*$ and $L^*$ respectively), net all-wave radiation ($Q^*$), turbulent sensible and latent heat fluxes ($Q_H$ and $Q_E$ respectively) and the storage heat flux ($\Delta Q_S$). In a first step, the performance of the models is evaluated for the site-specific reference parameter list. Second, we examine the interaction and sensitivity of all model and parameter list combinations. Finally, the sensitivity of the models with respect to impervious water storage and water vapor opacity is assessed. The paper is structured as follows: Section 2 provides a brief description of the urban canopy models, the sensitivity studies and evaluation metrics, followed by the study site, measurement set-up and description of external parameters in Section 3. The base-line model evaluation, results of the inter-parameter list and -model interactions as well as the sensitivity experiments are provided in Section 4. General findings and recommendations for future work are provided in Section 5.

2. Methods

Although most models are part of a numerical weather prediction or regional climate model, for the purpose of the present study they are used in an offline mode. They are therefore forced with atmospheric data observed above the canopy layer, which removes a potential source of error produced by the atmospheric model (see Section 3). For all model simulations, a two week spin-up period is considered to reduce the influence of model initialisation errors. Anthropogenic heat is neglected due to

© 2016 Royal Meteorological Society

Prepared using qjms4.cls
its small contribution to the surface energy balance of the selected site (Quah and Roth 2012).

2.1. Description of urban canopy models

2.1.1. SURFEX

SURFEX combines a range of sub-models able to calculate the exchange of sensible and latent heat, momentum, carbon dioxide and other chemical species, as well as various particles, between the atmosphere and several types of surfaces (Masson et al. 2013). The latter include oceans, inland waters, a large variety of natural land surfaces, and urban areas. Heterogeneity within an area of interest is accounted for by the fractional coverage of each main type (tile) contributing to the total area. Natural tiles are treated by the ISBA (Interaction Soil-Biosphere-Atmosphere) land surface model. Vegetation is described by the original ISBA evapotranspiration model using an externally imposed leaf area index (see also Section 3.2). Urban tiles use the Town Energy Balance (TEB) model, a single-layer urban canopy model (Masson 2000). Here, the energy budgets for roofs, walls, and road surfaces are solved separately for a homogeneous isotropic array of street canyons. The lower boundary conditions for roofs and walls are obtained by prescribing an internal temperature while a zero flux boundary condition is assumed for the road. Although gardens inside street canyons are possible (Lemonsu et al. 2012), in this study vegetated areas are treated as separate tiles with ISBA. The overall structure of TEB is described in more detail in Masson (2000); Masson et al. (2013), while further details on its configuration for the current study can be found in Harsham et al. (2016).

2.1.2. Community Land Model (CLM)

The Community Land Model (CLM) v4.0 (Bonan et al. 2011; Lawrence et al. 2011) is the land surface model of the Community Earth System Model (CESM). In CLM, the land surface heterogeneity is represented by main land units (glaciers, lakes, vegetation, wetland and urban), which are further divided into sub-units. The urban fraction, for example, can consist of roof, sunlit and shaded walls, pervious and impervious canyon floor, while vegetation includes representations of up to 17 possible plant functional types. All biogeophysical processes are independently simulated for each sub-unit using the same atmospheric forcing, with subsequent calculation of surface variables and fluxes by averaging the results for individual sub-units and units weighted by their fractional areas (Oleson et al. 2010b). CLM’s urban parametrisation (CLMU) follows to a large extent the concepts of TEB (Section 2.1.1). In CLMU, liquid and solid precipitation can be intercepted, stored and evaporated from the roof and canyon floor, respectively, while the walls are hydrologically inactive. Recent work by Demuzere et al. (2014a) introduced rainwater tanks, biofiltration systems and urban irrigation, while Buzan et al. (2015) implemented heat stress metrics. These features, however, are not activated in the present study. One of the differences between CLMU and TEB is that in CLMU the roof is coupled to the canyon air properties while in TEB the roof interacts directly with the canopy air aloft (see Demuzere et al. (2013)). A more detailed description of CLMU is available in Oleson et al. (2008a,b, 2010b).

2.1.3. TERRA_URB

TERRA_URB (Wouters et al. 2015, 2016) is the bulk urban land surface scheme of the COSMO-(CLM) model. It represents the variability of ground heat and moisture transport, the turbulent transfer of momentum, heat and moisture, and the surface-atmosphere radiative exchanges in urban areas. TERRA_URB has been extensively evaluated in previous studies (Wouters et al. 2015; Trusilova et al. 2015; Wouters et al. 2016), demonstrating satisfactory skill in reproducing the different urban surface energy balance components and the urban heat island. It has also been used to consider heat-stress scenarios under future climate and urban land-use change scenarios in Belgium within the Climate Report of the Flemish Environmental Agency (Brouwers et al. 2015). The initial release of TERRA_URB features the non-iterative calculation of surface-layer stability functions accounting for the roughness sub-layer (Wouters et al. 2012); the impervious water-storage parametrisation based on a probability density function of water reservoirs (Wouters et al. 2015); the Semi-empirical Urban canopY dependency parametrisation (SURY) (Wouters et al. 2016); and the coupling with the turbulence kinetic energy based surface-layer transfer module of the COSMO-(CLM) model (Doms et al. 2011).

2.1.4. SUEWS

SUEWS (The Surface Urban Energy and Water balance Scheme, Järvi et al. (2011, 2014); Ward et al. (2016)) simulates the surface energy and water balance at the neighbourhood scale. It can be run for multiple grids within a city and each model grid is divided into seven surface types including impervious surfaces (buildings and paved), different vegetated surfaces, bare soil and water. The different surfaces are dynamically connected (e.g. water is allowed to move between them). This study uses version SUEWS_V2016a with an adjusted surface conductance parametrisation particularly suitable for non-irrigated urban surfaces (Ward et al. 2016). The radiative flux components are derived from the incoming shortwave solar radiation using the net all-wave radiation scheme (NARP) (Offerle et al. 2003; Loridan et al. 2011); the storage heat flux the objective hysteresis model, OHM (Grimmond et al. 1991); the latent heat flux uses the Penman-Monteith equation adjusted for urban areas (Grimmond et al. 1991). In contrast to the other models (having $\Delta Q_S$ as the residual), SUEWS has the sensible heat flux as the residual of the energy balance. The model has been evaluated in offline mode against measurements in several cities (Järvi et al. 2011, 2014; Alexander et al. 2016b; Karsisto et al. 2016a,b; Ward et al. 2016) and used to estimate future climate scenarios in connection with local climate zones (Alexander et al. 2016a; Ward and Grimmond 2017). It is also part of the Urban Model Evaluation Predictor (Lindberg et al. 2015). Recent developments include an automatic treatment of reanalysis data to be used to force the model (Kokkonen et al. 2016).

2.2. Impervious water storage

Large differences in evaporation rates between urban and rural environments suggest a strong impact of urbanisation on the global water and energy cycle (Wouters et al. 2015). Evaporation from engineered pavements (e.g. asphalt and concrete) have long been ignored (Nakayama and Fujita 2010), yet attempts have been made to accurately represent the urban surface water balance, including water storage on impervious surfaces, run-off, evapotranspiration and urban biofiltration and irrigation systems (e.g., Grimmond and Oke 1991; Masson 2000; Wang et al. 2013; Demuzere et al. 2014a; De Ridder et al. 2015; Wouters et al. 2015). This study contributes to these ongoing efforts by testing the representation of water-interception reservoirs (puddles) using the Surface Interception Distribution (SID) approach (Wouters et al. 2015), currently embedded in TERRA_URB. Although this framework was tested for mid-latitude sites, Wouters et al. (2015) revealed that the annual-mean and variability of the surface water balance is very sensitive to these water storage reservoirs. Given that the current study focuses on a tropical site, characterised by

© 2016 Royal Meteorological Society
Prepared using qrms4.cls
an abundance of (intense) precipitation, one can expect an even stronger impact.

The SID approach assumes a linear probability density distribution of water puddles as a function of puddle depth, resulting in the following expression for the evaporative surface fraction (for the full derivation, see Wouters et al. (2015), their section 2.2):

\[ \delta = \delta_m \left( \frac{w}{w_m} \right)^{2/3} \]  

(1)

where \( \delta_m \) is the maximum puddle fraction, \( w \) is the amount of water in the water puddles, and \( w_m \) is the impervious water-storage capacity. Wouters et al. (2015) combined observations from Toulouse (France) and Basel (Switzerland) with model storage capacity. Radiation can be neglected (with water vapor may become even more important than for mid-humidity levels (figure 2); and tropical regions are characterized by high specific aloft (e.g. surface temperatures are typically higher than the air temperature daytime, and in most cases nighttime as well, urban canopy site in Singapore: on the one hand - as for any urban site - they themselves emit infra-red radiation at a lower temperature. We emitted from the surface at high temperatures is absorbed by radiation hitting the sensor at mast height. Infra-red radiation important. One of these interactions is the absorption and emission radiative interactions between the surface and the sensor become located several tens of metres above the urban canopy, for which of offline modelled \( L \) we hypothesise that radiative interactions established by the mixed sites typically identify a large positive daytime bias in \( L \) (according to Appendix B in Wouters et al. 2015), resulting in \( \delta_m = 0.2 \) and \( w_m = 1.31 \) kg m\(^{-2}\). To obtain these numbers, two assumptions are made: 1) the walls are hydrologically inactive (default in CLM/SURFEX) and 2) \( w_m \) is assumed to be identical for roof and road surfaces. In the remainder of this manuscript, this sensitivity run is denoted with ‘SID’.

2.3. The water vapor opacity effect

Previous offline evaluations and comparisons for mid-latitude sites typically identify a large positive daytime bias in \( L \) (e.g. Grimmond et al. 2011; Hénon et al. 2012; Demuzere et al. 2013). Besides other potential explanations (e.g. source areal differences between model and measurement, Järvì et al. (2014)) we hypothesise that radiative interactions established by the mixed air in the urban canopy-layer is disregarded in the offline setups, while being resolved by the atmospheric model in online coupled model setups. Such interactions might lead to a poor performance of offline modelled \( L \) compared to mast sensor observations located several tens of metres above the urban canopy, for which radiative interactions between the surface and the sensor become important. One of these interactions is the absorption and emission by water vapor (hereafter referred to the water vapor opacity effect, WVO), affecting amongst others the upwelling infra-red radiation hitting the sensor at mast height. Infra-red radiation emitted from the surface at high temperatures is absorbed by water vapor molecules. At the same time, the molecules themselves emit infra-red radiation at a lower temperature. We hypothesise that these interactions are also relevant for the current site in Singapore: on the one hand - for any urban site - daytime, and in most cases nighttime as well, urban canopy surface temperatures are typically higher than the air temperature aloft (e.g. Coutts et al. 2016, Wouters et al. 2016), their figure 2); and tropical regions are characterized by high specific humidity levels (Roth et al. 2016), for which radiative interactions with water vapor may become even more important than for mid-latitude sites.

In the absence of hydrometeors, where scattering of infra-red radiation can be neglected (Pielke 2002), the water-vapor opacity effect can be approximated as:

\[ R(z) \approx \epsilon(\Delta P)\sigma T^4_a [1 - \epsilon(q_v, \delta u)] L \]  

(2)

where \( \epsilon \) is the emissivity (taken as the complement of the transmissivity) as a function of the optical path length \( \Delta P = \int_0^z p_q dz \) between the the effective canopy height from where surface radiation originates (\( z_e \)) and the sensor height (\( z_s \)), \( \rho \) is the air density, \( q_v \) is the specific humidity, \( T_a \) is the air temperature and \( \delta u \) is the normal path length between \( z_e \) and \( z_s \). For the current study, \( q_v \) and \( T_a \) are approximated by the forcing values (Section 3.1) and assumed to be a constant throughout the canopy layer. According to Pielke (2002), the infra-red broadband emissivity for water vapour is approximated as follows:

\[ \epsilon(\Delta P) = 0.130 \log_{10}(\Delta P) + 0.54, \text{ for } \log_{10}(\Delta P) > 0. \]  

(3)

with \( \Delta P \) expressed in g cm\(^{-2}\). More details can be found in Appendix A of Wouters et al. (2015).

Since SUEWS directly uses the forcing temperature in order to calculate \( L \) (see Eq. 11 in Lidar et al. 2011), the effect of the water vapor opacity is only tested for CLM, SURFEX and TERRA_URB and denoted by ‘WVO’ in the remainder of the manuscript.

2.4. Evaluation Metrics

Baseline comparison statistics include mean (e.g. \( \bar{X} \)), standard deviation (e.g. \( \sigma_X \)), coefficient of determination \( (r^2) \), mean bias error (MBE), mean absolute error (MAE) and the root mean square error (RMSE) including both its systematic (RMSE\(_s\)) and unsystematic (RMSE\(_u\)) components (e.g. Willmott and Matsuura 2005; Oleson et al. 2008b; Grimmond et al. 2010; Demuzere et al. 2013). In addition, Taylor diagrams (Taylor 2001) are used to provide a simultaneous assessment of each models’ capability to simulate the radiative and turbulent fluxes. Some statistics are normalized by the standard deviation of the observed values (denoted with ‘n’ in front of the statistics’ abbreviation). The significance of the differences between modelled and observed quantities are tested with the Perkins skill score (\( S_{score} \)) (e.g. Perkins et al. 2007; Devis et al. 2013, 2014). This simple metric allows for a comparison across probability density functions (PDF), measuring the common area shared by the modelled and observed PDF. Values range between zero and unity for no and perfect overlap, respectively. In the remainder of this paper we consider PDFs to be significantly different when \( S_{score} \) is <0.8, which is more stringent than the value of 0.7 used in Perkins et al. (2007); Devis et al. (2013, 2014).

3. Site description, measurements and boundary conditions

3.1. Telok Kurau measurement site and observations

The data used to force and evaluate the models were measured over a residential neighbourhood of Singapore, which is a small (716 km\(^2\)), densely populated (5.4 million people in 2013), low-lying island city-state located ~137 km north of the equator. The study area in the Telok Kurau (TK) district has low-rise buildings (2-3 storey row and semi-detached houses and a few 5 storey condominiums) dissected by a network of mostly minor streets. A detailed site survey covering an area within a 500-m radius around the micrometeorological tower (see below) produced the following morphological and land cover parameters representative of TK (Velasco et al. 2013): an average building and tree height of 9.86 and 7.26 m respectively, a surface cover of 85% impervious (39% buildings, 34% gravel/paved, 12% roads) and 15% pervious (11% tree crowns, 4% grass) and an area-averaged height-to-width ratio (H/W) of 0.61 (Table 1). Many houses have small gardens covered with turf grass and most streets are lined with shade trees.
The area corresponds to LCZ 3 or ‘compact low rise’ (Stewart and Oke 2012).

A micrometeorological tower installed in the southwest corner of a grass-covered sports field in the centre of the study area (1°18′51.46″ N, 103°54′40.31″ E; 10 m above sea level) supported various meteorological sensors.

The sensible and latent turbulent energy fluxes were measured using the eddy covariance (EC) technique with a 3-D sonic anemometer/infrared gas analyzer (CSAT3/LI-7500; Campbell Scientific, Logan, Utah / LI-COR Biosciences, Lincoln, NE, USA). The up- and downward short- and longwave fluxes ($K_\uparrow$, $K_\downarrow$, $L_\uparrow$, $L_\downarrow$ respectively) were measured with a 4-component net radiometer (CNR1; Kipp & Zonen, Delft, Holland). Additional instrumentation included a temperature/humidity probe ($T/RH$) (HMP45C; Vaisala, Helsinki, Finland). These sensors were installed at a height of 23.7 m to ensure a sufficient height above the surface roughness to measured spatially representative turbulent fluxes at the neighbourhood scale (e.g. Roth 2000; Velasco and Roth 2010). The flux measurements were processed according and subjected to the usual quality control procedures used in EC work. Finally, a rain gauge (HOBO RG3; Onset Computer Corporation, Bourne, MA, USA) measured rainfall (Precip) near the ground. Observed $\Delta Q_S$ was estimated as the residual from $Q^*$ and all other terms in the energy balance equation. The TK site and EC measurements have been used in a number of recent urban flux studies and are fully described in Velasco et al. (2013) and Roth et al. (2016). All models described in Section 2.1 are forced with the above-mentioned observations of surface pressure (Pa), $K_\downarrow$, $L_\downarrow$ (W m$^{-2}$), $T$ (K), wind speed (m s$^{-1}$), Precip (mm h$^{-1}$) and RH (%).

Because of its equatorial location Singapore has a typical tropical rainforest climate (Köppen classification Af). Temperature is uniformly high throughout the year (long-term annual mean: 27.5 °C) and rainfall abundant (~2340 mm). The diurnal temperature range is relatively small but larger than the mean month-to-month variability (~6-7 °C vs ~2 °C respectively). The climate during the period analysed in the present study (01-06-2013 to 17-04-2014) follows the above mentioned climate normals but with one important exception. An unusual dry period occurred from mid January 2014 to mid March 2014 during which time only 2.2 mm of rainfall was measured on the 8th of February 2014 (the long-term monthly rainfall for February is 160 mm) and temperature (absolute humidity) were slightly above (below) their long-term means for that time of the year (Fig 1).

3.2. External parameters

Each land-surface model (including the urban parametrisation) is supplied with a specific set of global/regional land-cover characteristics. While these datasets are interchangeable between models, each specific model normally uses a ‘native’ database.

SURFEX uses the global 1 km ECOCLIMAP database derived from land cover maps and satellite data (hereafter referred to as MA03) in which each pixel is assigned one of more than 550 land cover types, each associated with a set of parameter values needed by the surface models (Masson et al. 2003). Parameters include the fractional coverage of each main surface type, thermal and radiative characteristics of buildings, walls and roads, and characteristics of the plant functional types (for details, see Masson et al. 2013). The 1 km MA03 grid cell size matches that of the source area of the turbulent flux measurements which during daytime (nighttime) extends to a maximum of 500 (1000) m (Figs. 1 in Velasco et al. 2013; Roth et al. 2016). The MA03 values for the corresponding grid cell is 60% urban, equally divided into a buildings and impervious roads. The remainder of the grid cell consists of broadleaf evergreen tropical trees (20%) and C4 grass (20%). The mean building height is 10 m, with a height-to-width (H/W) ratio of 0.21. Additional details on the total roof, wall and road thickness, thickness of each individual sub-unit layer and radiative and thermal properties are provided in Table 1.

For CLMU, the urban surface properties are taken from Jackson et al. (2010), who provide a global region-specific dataset on urban extent, density, geometry, thermal and radiative characteristics. This is the native generic dataset used for CLMU in global and regional climate studies (e.g. Oleson et al. 2010a, 2013), and is hereafter referred to as JA10. According to this dataset, the urban fraction of the corresponding TK grid cell is ~31% of which 60% is covered by buildings (roofs), 25% by impervious roads and 15% pervious roads, respectively. The vegetation consists mainly out of broadleaf evergreen trees with a smaller fraction of broadleaf deciduous trees, C4 grass and C3 crops. Respective building height and H/W ratio are 30 m and 1.2, reflecting a higher density urban environment than is actually observed (see Table 1).

The reference parameter list (hereafter referred to as REF) is compiled to provide the most realistic description of the Telok Kurau area surrounding the flux tower (Section 3.1 and Table 1). The typical building envelope of the low-residential houses in the area was provided by experts of the National University of Singapore, Department of Architecture. The most common wall material is a double brick layer without any cavity and white plaster on the in- and outside. The roofs generally consist of the following four layers: ceramic tiles, aluminium foil, an air gap and plaster. The typical thickness of these layers and the radiative and thermal characteristics (Table 1) for each of the materials are taken from the materials’ library of the Autodesk Ecotect analysis software (Tools for Sustainability 2012).

In TERRA URB the urban land-cover tile is considered 100% impervious (Wouters et al. 2015). In order to obtain bulk parameter values for roof, wall and road surfaces, the Semi-Empirical Urban canopyY-dependency parametrisation (SURY v1.0) was used to translate facet information from MA03, JA10 and REF to bulk values. Here, the bulk thermal parameter values take into account enhanced ground heat transport and storage due to the increased contact surface expressed by the surface-area-index while radiative parameter values consider multiple-facet radiative numerical experiments for calculating the albedo reduction factor resulting from the radiative trapping by the urban canopy (see sections 2.1.1 and 2.1.2 respectively in Wouters et al. 2016). The resulting bulk values are indicated in brackets in Table 1. These values are slightly higher than those used in previous applications of TERRA URB (Demuzere et al. 2008; De Ridder et al. 2012; Wouters et al. 2015, 2016), but generally fit within the range of LCZ 3 provided in Stewart and Oke (2012).

SUEWS uses the surface cover, morphological (building and tree height) and radiative properties listed in Table 1. Thermal properties for the built-up surfaces are taken into account via the heat storage coefficients used in OHR (Table 1) (Grimmond et al. 1991). For bare soil and vegetated surfaces, the default model values are used (Järvi et al. 2011), except in the calculation of surface conductance where the parameters from (Ward et al. 2016) are used.

4. Results

All models are evaluated using the idealised reference parameter list (REF) (Section 4.1); the impact of other external
parameter values and their interaction with varying ULSMs is described (Section 4.2) and the sensitivity with respect to the impervious water storage and water vapor opacity is assessed (Section 4.3). Due to the similarity between the CLM and SURFEX models, their results are generally discussed together and referred to as CLM/SURFEX.

4.1. Model evaluation using REF parameters

4.1.1. Overall performance

Evaluation metrics using the REF parameter values in Figure 2 are based on 1-hour periods using all (day- and nighttime) data, except for outgoing shortwave radiation $K^\uparrow$ for which only daytime fluxes ($K^\uparrow > 0$ W m$^{-2}$) are used, and all weather conditions. Further details on the statistical values for the simulations using the REF parameter values are provided in the Supplementary Information (Table S1).

Net all-wave radiation ($Q^*$) is well represented by all models, with an $r^2$ and $S_{\text{score}}$ close to 1 and a RMSE varying between 16.6 (SUEWS) and 46.1 W m$^{-2}$ (TERRA_URB). Most models, except SUEWS, tend to underestimate $Q^*$ with a maximum MBE of -17.3 W m$^{-2}$ for TERRA_URB. Although the models are provided with a reference list of external parameters, all models except SUEWS have a RMSEu that is larger than RMSEu indicating issues with the model physics or the parameters used. The underestimation in $Q^*$ for CLM/SURFEX and TERRA_URB is largely driven by an overestimation in outgoing long-wave radiation ($L^\uparrow$). The overestimation of $L^\uparrow$ is further compensated by a slight underestimation of $K^\uparrow$ for all models except TERRA_URB. For $Q^*$, RMSEu is larger than RMSEu for both $K^\uparrow$ and $L^\uparrow$ for all models.

Model errors for the turbulent fluxes $Q_H$, and especially $Q_E$, are larger than the radiative flux component errors. For $Q_H$, the $S_{\text{score}}$ of all models is larger than 0.8 while only SUEWS reaches this $S_{\text{score}}$ threshold for $Q_E$. (Table S1 and Figure 2). Apart from TERRA_URB, all models have a positive bias for $Q_H$, up to 9.3 W m$^{-2}$ for CLM, but for all models, RMSEu is larger than RMSEu. Model performance is poorest for $Q_E$ with a maximum $r^2$ of 0.6 (for SUEWS) and RMSE of 41.7 W m$^{-2}$ (SURFEX). The overall magnitude of $Q_E$ is too low for all models, but the RMSEu is always larger than RMSEu, except for SUEWS.

Finally, the storage heat flux $\Delta Q_S$ is best modelled by SUEWS which is surprising since SUEWS uses default values for OHM, not specifically tailored towards the Telok Kurau site (Table S1 and Figure 2). It is the only model that has a lower systematic than unsystematic RMSE and a $S_{\text{score}}$ above the threshold of 0.8. The other models generally have a too low $\Delta Q_S$ and $\sigma$ compared to the observations with a RMSE up to 64 W m$^{-2}$ for TERRA_URB and a negative bias up to -7.1 W m$^{-2}$ for CLM/SURFEX.
except for $Q_L$ and $E$ during the day too much heat is transported away from the surface (via evaporation or stored in the ground), being strongly positive during the day and negative during the night, especially for CLM/SURFEX and TERRA_URB (see also Section 4.1.2). This is especially true for impervious surfaces (buildings) (see also Jarvis et al. 2011).

### 4.1.2. Performance during day-nighttime and specific weather conditions

After stratifying these result for day-nighttime and distinct weather conditions, the largest biases generally occur at nighttime, except for $Q_E$ and $L^\uparrow$ (Fig. 3). For the latter, the large RMSE errors for CLM/SURFEX and TERRA_URB described in Section 4.1.1 are mainly due to high daytime RMSE values. In contrast, SUEWS performs better with nRMSE = 0.4 W m$^{-2}$ and similar nRMSE$_a$ and nRMSE$_b$ values. The nRMSE of the latent heat flux is slightly higher during the day than at night, especially for CLM/SURFEX. For all models and fluxes, nighttime nRMSE$_a$ is larger than nRMSE$_a$ except for $L^\uparrow$ of SUEWS.

Table 1. Overview of surface cover, morphological, radiative and thermal characteristics describing the residential area of Telok Kurau. REF - reference namelist (measured values), MA03 - based on ECOCLIMAP database used in SURFEX, JA10 - generic dataset used in CLMU; $H/W$ - canyon height-to-width ratio, $w_{ref}$ - maximum water storage on impervious surfaces (roof and impervious road). $\alpha$, $\varepsilon$, $C_v$, and $\lambda$ refer to albedo, emissivity, volumetric heat capacity and thermal conductivity respectively. Values in brackets are the bulk values used for TERRA_URB that are derived from urban-canopy parameters using the Semi-Empirical Urban canopY parametrization (SURY) (Wouters et al. 2016).

<table>
<thead>
<tr>
<th>Parameter</th>
<th>REF</th>
<th>MA03</th>
<th>JA10</th>
</tr>
</thead>
<tbody>
<tr>
<td>Urban fraction [%]</td>
<td>0.85</td>
<td>0.6</td>
<td>0.31</td>
</tr>
<tr>
<td>Buildings</td>
<td>0.39</td>
<td>0.3</td>
<td>0.19</td>
</tr>
<tr>
<td>Pervious roads</td>
<td>-</td>
<td>-</td>
<td>0.04</td>
</tr>
<tr>
<td>Impervious roads</td>
<td>0.46</td>
<td>0.3</td>
<td>0.08</td>
</tr>
<tr>
<td>Vegetation fraction [%]</td>
<td>0.15</td>
<td>0.4</td>
<td>0.69</td>
</tr>
<tr>
<td>Broadleaf evergreen tropical tree</td>
<td>0.11</td>
<td>0.2</td>
<td>0.37</td>
</tr>
<tr>
<td>Broadleaf deciduous tropical tree</td>
<td>-</td>
<td>-</td>
<td>0.09</td>
</tr>
<tr>
<td>C4 grass</td>
<td>0.04</td>
<td>0.2</td>
<td>0.17</td>
</tr>
<tr>
<td>C3 crop</td>
<td>-</td>
<td>-</td>
<td>0.06</td>
</tr>
<tr>
<td>Height of roof [m]</td>
<td>9.86</td>
<td>10</td>
<td>30</td>
</tr>
<tr>
<td>$H/W$ [unitless]</td>
<td>0.61</td>
<td>0.21</td>
<td>1.2</td>
</tr>
<tr>
<td>$w_{ref}$ [kg m$^{-2}$]</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Total thickness [m]</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Roof</td>
<td>0.1356</td>
<td>0.55</td>
<td>0.05</td>
</tr>
<tr>
<td>Wall</td>
<td>0.13</td>
<td>0.195</td>
<td>0.34</td>
</tr>
<tr>
<td>Impervious road</td>
<td>1.15</td>
<td>1.15</td>
<td>0.05</td>
</tr>
<tr>
<td>Thickness of model layers [m]</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Roof</td>
<td>0.05, 0.0006, 0.0075, 0.01</td>
<td>0.05, 0.04, 0.1</td>
<td>0.0033$^b$</td>
</tr>
<tr>
<td>Wall</td>
<td>0.01, 0.11, 0.01</td>
<td>0.02, 0.13, 0.05</td>
<td>0.023$^b$</td>
</tr>
<tr>
<td>Impervious road</td>
<td>0.05, 0.1, 1</td>
<td>0.05, 0.1, 1</td>
<td>0.025 $\times \exp[0.5 \times (i - 0.5)] - 1$$^c$</td>
</tr>
<tr>
<td>$\alpha$ [unitless]</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bulk</td>
<td>(0.21)</td>
<td>(0.13)</td>
<td>(0.21)</td>
</tr>
<tr>
<td>Roof</td>
<td>0.15</td>
<td>0.15</td>
<td>0.28</td>
</tr>
<tr>
<td>Wall</td>
<td>0.647</td>
<td>0.25</td>
<td>0.25</td>
</tr>
<tr>
<td>Impervious road</td>
<td>0.08</td>
<td>0.08</td>
<td>0.13</td>
</tr>
<tr>
<td>$\varepsilon$ [unitless]</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bulk</td>
<td>(0.92)</td>
<td>(0.91)</td>
<td>(0.87)</td>
</tr>
<tr>
<td>Roof</td>
<td>0.9</td>
<td>0.9</td>
<td>0.82</td>
</tr>
<tr>
<td>Wall</td>
<td>0.9</td>
<td>0.85</td>
<td>0.88</td>
</tr>
<tr>
<td>Impervious road</td>
<td>0.93</td>
<td>0.94</td>
<td>0.91</td>
</tr>
<tr>
<td>$C_v$ [MJ m$^{-3}$ K$^{-1}$]$^d$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bulk</td>
<td>(2.65)</td>
<td>(1.38)</td>
<td>(2.39)</td>
</tr>
<tr>
<td>Roof</td>
<td>1.31, 2.37, 0.0012, 1.12</td>
<td>2.11, 2.8, 2.9</td>
<td>1.0</td>
</tr>
<tr>
<td>Wall</td>
<td>1.12, 1.4, 1.12</td>
<td>1.55, 1.55, 2.9</td>
<td>1.01</td>
</tr>
<tr>
<td>Impervious road</td>
<td>2.1, 1.8, 1.5</td>
<td>1.94, 1.28, 1.28</td>
<td>2.06, 1.71</td>
</tr>
<tr>
<td>$\lambda$ [W m$^{-1}$ K$^{-1}$]$^d$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bulk</td>
<td>(0.91)</td>
<td>(0.39)</td>
<td>(2.23)</td>
</tr>
<tr>
<td>Roof</td>
<td>0.93, 2.1, 0.026, 0.53</td>
<td>1.51, 0.08, 0.05</td>
<td>0.39</td>
</tr>
<tr>
<td>Wall</td>
<td>0.53, 0.79, 0.53</td>
<td>0.93, 0.93, 0.05</td>
<td>1.38</td>
</tr>
<tr>
<td>Impervious road</td>
<td>2.1, 1.8, 1.5</td>
<td>1.94, 1.28, 1.28</td>
<td>1.67, 0.56</td>
</tr>
</tbody>
</table>

$^a$ The Semi-Empirical Urban canopY parametrization (SURY) is freely available at [https://github.com/hendrikwouts/sury](https://github.com/hendrikwouts/sury).

$^b$ The 15 roof and wall layers have a constant thickness.

$^c$ The impervious road is discretized into fifteen layers (denoted by $i$) with an exponentially increasing node depth and thus thickness (see Eq. 4.8 in Oleson et al. 2010).

$^d$ SUEWS uses the Objective Hysteresis Model (Grimmond et al. 1991). The coefficients used in the current study are $\alpha_1$ = -0.719 (0.238), $\alpha_2$ = -0.194 (0.427) and $\alpha_3$ = -36.6 (-16.7) for impervious surfaces (buildings) (see also Jarvis et al. 2011).
SUEWS. It is important to note here that measurement errors are generally larger at night (Best and Grimmond 2015).

Time series of daily mean MBE for the 11-month period for all models and fluxes are provided in Fig. 4. The similarity between CLM and SURFEX causes their time series (red and blue lines) to overlap most of the time. Overall, the skill of all models have a strong temporal variability, although some model/flux combinations tend to maintain similar skill throughout. TERRA_URB for example closely follows CLM/SURFEX for $L^*$ while SUEWS often tends to have a very different behaviour. At the start of the dry spell in mid-January, MBE for $L^*$ drops for SUEWS while it slightly increases for all other models. Simultaneously, SUEWS’ bias for $Q^e$ increases, while it remains at the same level for the other models. Finally, SUEWS has a more strong positive bias during this dry period, which is accompanied by and overestimation of $\Delta Q_S$. The prolonged period without precipitation is also accompanied by an observed decrease (increase) in latent (sensible) heat flux. All models are able to capture the slope of this decrease/increase, although the absolute magnitude is underestimated, especially for the latent heat flux.

To further evaluate the model performance under specific weather conditions, the observation period is stratified into a two-month dry period (between 15/01/2014 - 15/03/2014) and wet period (15/11/2013 - 14/01/2014) (Fig. 5). Notwithstanding a few exceptions, all models perform better during the wet than the dry conditions. This is especially the case for SUEWS, except for $L^*$. All other models have a smaller difference between the dry and wet period for $K^*$ and $L^*$. This is also reflected in the results for $Q^e$. The inter-quartile range (IQR) error variability for SUEWS is a lot smaller for $L^*$ and $Q^e$, spanning an error range of approximately 10 W m$^{-2}$, while this is always greater than 30 W m$^{-2}$ for the other models.

TERRA_URB and SUEWS have the best performance for $Q_H$ for the dry and wet period, with a MBE decreasing from -7.8 and -5.8 W m$^{-2}$ to -0.4 and 3.7 W m$^{-2}$, respectively. For the latent heat flux, the overall negative MBE remains, but is stronger during the dry period compared to the wet period, even though the magnitude of the flux is smaller during the dry period. This deterioration also occurs in the RMSE components: all models (except SUEWS) have a larger RMSE$_{\Delta Q_S}$ than RMSE$_{\Delta Q_H}$, which is no longer the case during the dry period. The performance of $\Delta Q_S$ for SUEWS improves significantly during the wet period compared to the dry period. Whereas, CLM/SURFEX and TERRA_URB have a slight decrease in MBE, but their $S_{\text{score}}$ is better for the wet than the dry period (not shown).

4.1.3. Performance following a precipitation event

The ability of the models to predict the surface-atmosphere exchange following precipitation is investigated using the nMBE as a function of time (hours) since a precipitation event (Fig. 6). In total, 72 rainfall events (irrespective of intensity) followed by at least 24 hours without precipitation occurred.

For $K^*$, $Q^e$, $Q_H$ and $\Delta Q_S$, the bias is mostly independent of the time since a precipitation event, with an almost constant small positive or negative bias depending on the model or flux of interest (not shown). For $L^*$ and $Q_E$, SUEWS has an almost constant positive and negative bias, respectively, irrespective of the hours since the precipitation event. In contrast for TERRA_URB, CLM and SURFEX, the nMBE for $L^*$ more than doubles when comparing their skill directly after, and 24 hours after, a rainfall event. For $Q_E$, SURFEX and CLM behave similarly, with a peak positive bias one hour after a precipitation event, reducing to almost 0 W m$^{-2}$ after 6 hours and further decreasing to a negative normalized bias of -0.6 W m$^{-2}$ after 24 hours. Stratifying by accumulated rainfall totals (low, medium, high) in the 24 hours leading up to each precipitation event led to a similar relative magnitude and sign of the biases as those in Fig. 6 (not shown).

4.2. Inter-parameter list and model variability

The overall influence of using different parameter lists for each model is summarised in Figure 7. Similar to the evaluation of the models using the REF parameter list, CLM and SURFEX behave almost identically when using MA03 and JA10 and they are discussed together. For CLM/SURFEX, altering the parameter list settings to JA10 has the largest impact on $L^*$ and $Q_E$. For $L^*$, biases in normalized RMSE and standard deviation are reduced with no change in the correlation coefficient. For $Q_E$, the results deteriorate with an increase in nRMSE. This is accompanied by an improvement in correlation and deterioration in variance. Other surface energy balance components are less influenced by changing the parameter list. The results using MA03 are generally within the range of results obtained using REF and JA10.

TERRA_URB has a similar behaviour as the single layer canopy models CLM/SURFEX. The error for $L^*$ decreases when using the JA10 parameter list instead of REF, while the correlation remains constant and the standard deviation improves. Simultaneously, the skill for $Q_E$ deteriorates, while the underestimated variability in REF is strongly overestimated using JA10. Since the SUEWS model used its own thermal parameters which were not changed between the simulations, this model has the smallest sensitivity to changes in the parameter list, except for the latent heat flux. For the latter, all skill scores deteriorate with a more than doubling of the nRMSE and a large overestimation of the standard deviation when using JA10. For TERRA_URB and
SUEWS the results for the MA03 parameter list are again between those from REF and JA10 (Fig. 7).

To further investigate the parameter list versus inter-model performance, the median and inter-quartile range (IQR) of the hourly biases are analysed (Fig. 8) individually for each model and parameter list used. For $K^\uparrow$ (daytime only), all model and parameter list combinations (except JA10 and TERRA_URB) indicate an underestimation, both in median and IQR (see Fig. 8a and b). When using REF, the IQR is smaller than for all other settings (6 W m$^{-2}$ compared to 17 W m$^{-2}$, respectively). The underestimation of $K^\uparrow$ contrasts to an overall daytime (nighttime) overestimation (underestimation) in $L^\uparrow$. A large difference between the interquartile ranges can also be noted. During the day it reaches 40 W m$^{-2}$ for the simulations using the REF parameter list (Fig. 8a), while it is only 10.6 W m$^{-2}$ for the SUEWS model using different parameter lists (Fig. 8b). At night, the differences are smaller with the largest (smallest) median bias and IQR for the JA10 parameter list (SUEWS model) (Fig. 8c and d). The results for both $K^\uparrow$ and $L^\uparrow$ are also reflected in $Q^*$. During the day, most parameter list and model combinations underestimate $Q^*$, while this is opposite during the night. Again, the IQR is smallest for JA10 and SUEWS during day-and-nighttime.

While the biases for the radiation fluxes are consistent between parameter list and model groups, this is different for daytime turbulent heat fluxes. The overestimation in $Q_H$ using the REF namelist is converted to an underestimation using JA10. This is compensated by a strong increase in bias and IQR for the latent heat flux when moving from REF to JA10 (Fig. 8a). The results for each model using different parameter lists are more consistent. All models underestimate $Q_H$ while $Q_E$ biases are similar but slightly overestimated. During the night, the biases are generally negative and small with a narrow distribution. Best results for nighttime $Q_H$ are obtained with TERRA_URB, while for $Q_E$ the bias distribution is almost identical for all models (Fig. 8c and d). Finally, the daytime storage heat flux has a similar bias behaviour across parameter lists and models, generally being underestimated, except for SUEWS. During the night, $\Delta Q_S$ is most problematic with a median overestimation of at least 18.5 W m$^{-2}$ (SUEWS). IQR is also larger compared to all other fluxes, ranging up to 24 W m$^{-2}$ (SUEWS).

Figure 9 confirms that generally the best model results are obtained for $Q^*$, $K^\uparrow$ and $\Delta Q_S$, and the worst for $L^\uparrow$ and $Q_E$ with nMBE values larger than 1. Especially for $Q_E$ (and to a lesser extent for $Q_H$ and $L^\uparrow$) it is striking that the sign and magnitude of the error depend more on the choice of parameter values rather than choice of model itself. For example, the model results using JA10 strongly overestimate the latent heat flux, while an underestimation is observed when using REF. Here, the best results are obtained for the models using the MA03 parameter list. SUEWS-JA10 ranks best in terms of $K^\uparrow$ but simultaneously performs worst for $Q_H$ and $Q_E$. CLM-MA03 ranks best for $Q_H$ and $Q_E$ but performs poorly for $K^\uparrow$, $L^\uparrow$ and $\Delta Q_S$. In addition, rank numbers for CLM and SURFEX are in most cases only one number apart from each other, which again confirms

![Figure 3](image-url)
Figure 4. Time series of daily mean bias (W m$^{-2}$) for four models (indicated by different colours) and six fluxes (panels a to f) for the full period using the REF parameter list. Blue bars in top panels are daily precipitation (mm day$^{-1}$). Blue and red shaded areas indicate the selected wet and dry period respectively. A five day moving average is applied to all time series for clarity. Note different y-axis ranges for different fluxes. Breaks in time series are time periods with missing observations.

Figure 5. Error distribution statistics for the dry (red) and wet (blue) periods for all models using the REF parameter list. Mean values (corresponding to the MBE) and inter-quartile ranges are depicted by dark coloured triangles and light coloured boxes, respectively. For $K^+$, only daytime fluxes are used.
their similarity noted above (when using the same parameter list settings). Finally, the model, parameter list and their averages do not necessarily perform better than their individual counterparts, although this result is very much dependent on the flux of interest. The simple model averaging solution (same weight for each member, MmPl, multi-model&Parameter list) provides average rankings from 6 for \( Q^* \) to rank 11 for \( L^\uparrow \) and \( Q_H \). Additional tests using the reliability ensemble averaging (Miao et al. 2014), which uses a weighted average of the ensemble members based on the reliability of its members, did not result in a better ensemble performance (not shown).

### 4.3. Sensitivity to the treatment of impervious water storage and water vapor opacity

Given the similarity between CLM and SURFEX, the surface interception distribution (SID) framework described in Section 2.2 is illustrated for CLM only (note that the SID formulation is already included by default in TERRA_URB). From Figure 10 it is clear that the alternative representation of water puddles on the impervious surface has a strong impact on the modelled \( Q_E \). Both the peak overestimation until 6 hours after such precipitation event as well as the underestimation in the last 6 hours of this 24-hour period is strongly reduced (compare Fig. 10 with Fig. 6). In addition, the SID approach positively impacts the full-period error statistics for \( L^\uparrow \), \( L^\uparrow \), \( Q_H \) and \( Q_E \). At the same time, there is a trade-off in skill with a slightly worse performance for \( Q^* \) and \( \Delta Q_S \) (Table S2).

The sensitivity of the modelled \( L^\uparrow \) on the effect of water vapor opacity (WVO) is tested for CLM/SURFEX and TERRA_URB. The error statistics for \( L^\uparrow \) (Table S3) compared to the default REF simulations’ error statistics (Table S1) are improved. For all models, the \( S^\text{score} \) is larger than 0.9 which is now in line with the default SUEWS model performance. The RMSE is reduced from \( \sim 30 \text{ W m}^{-2} \) to \( \sim 5 \text{ W m}^{-2} \). The large gap between the systematic and unsystematic component of the RMSE in the default setting is now almost negligible for the WVO simulations.

As the water vapor opacity effect depends on atmospheric humidity levels, the WVO results are tested for the dry and wet periods discussed in Section 4.1.2. During the dry period, the nighttime negative bias is completely removed, while the strong daytime overestimation (\( \sim 80 \text{ W m}^{-2} \)) becomes a small underestimation (\( \sim 10 \text{ W m}^{-2} \)) (Fig. 11a). For the wet period, the WVO correction results in a complete removal of the bias throughout the day (Fig. 11b).

Finally, as the skill of the models in representing \( L^\uparrow \) is a function of hours since a precipitation event (Figure 6), the combined effects of SID and WVO are illustrated for CLM in Figure S1. Without taking into account WVO, there is no systematic improvement in modelled \( L^\uparrow \). Introducing WVO has the largest impact and supports the results described above: the modelled \( L^\uparrow \) now closely follows the observed \( L^\uparrow \) dynamics in the 24 hours after the selected rainfall events by removing the positive biases at the beginning and end of this 24-hour period. Outside of
these 24-hour rainfall periods, the SID approach only has a minor effect on the evaporation.

5. Discussion and Conclusions

The present study provides the first comparative offline evaluation for a tropical residential neighbourhood (Telok Kurau, Singapore) using four urban land surface models. These ULSMs include the bulk scheme TERRA_URB and three models of intermediate complexity, viz. CLM, SURFEX and SUEWS. All simulations are performed using three different external parameter lists which include the global region-specific parameters.

The ULSMs under investigation have been evaluated extensively in mid- and high-latitude cities, and are therefore potentially optimised to these regions. Encouragingly, our results using the best available external parameters (REF) align well with previous findings. For example, the second phase of the PILPS-urban project (targeting a suburban site in Melbourne (Australia) characterized by a temperate oceanic climate) found that $L^\uparrow$ is overall not as well modelled as $K^\uparrow$ and that $Q^*$ is modelled better than either $K^\uparrow$ or $L^\uparrow$ (Grimmond et al. 2011). This is true also for the current model evaluation over Telok Kurau. In addition, $L^\uparrow$ and $Q_E$ are identified as the most problematic fluxes, which is again in agreement with the findings of the urban intercomparison project PILPS-Urban (Grimmond et al. 2011; Best and Grimmond 2015).

The performance of varying parameter list and model combinations largely depends on the respective combination and flux of interest. The outgoing longwave and latent heat flux are the most sensitive to changes in the parameter list, but with some exceptions. Since $L^\uparrow$ in SUEWS depends on the forcing temperature $T$, the use of a different parameter list has almost no impact. Whereas for SUEWS’ $Q_E$, the impact of the external parameters is largest (compared to the other models). For all
models, using the JA10 parameter list results in strong negative (positive) bias for $Q_H$ ($Q_E$). The opposite occurs when using the REF parameter list, while using MA03 provides results which are between the former two. The most significant difference between these parameter lists is the amount of urban fraction, which is 30.8, 60 and 85% for JA10, MA03 and REF, respectively. The present results suggest a significant impact of this value, but similar to findings in Loridan and Grimmond (2012), the use of a site-specific urban fraction does not always yield the best model result for $Q_H$ and $Q_E$.

Combining all model and parameter list performances suggests that the error statistics tend to be more dominated by the choice of external parameter values than the choice of model (structure, parametrisations, etc.). For example, the variability between all models driven by one parameter list is often smaller than the variability of one model driven by different parameter lists. Yet, the multi-parameter list and/or multi-model averages do not necessarily outperform each other or the individual realisations, a result, however, that very much depends on the flux of interest. If the focus is on a robust representation of the surface energy balance at an aggregated neighbourhood scale, a simple representation (with a limited number of parameters) such as TERRA_URB may be sufficient. Such a scheme is also advantageous in terms of a lower computational cost; bulk parameters are determined before-hand based on detailed radiation model studies, hence avoiding the radiation calculation during the simulation. But as stated in Best and Grimmond (2015), such representations might not have the physical requirements for more advanced applications such as street-level heat stress studies that benefit from detailed in-canyon radiation information (Buzan et al. 2015).

The humid tropical setting of this residential site together with the exceptional two-month dry period allowed for a more in-depth evaluation of the models’ performances during specific weather conditions. First, results vary across models and fluxes considered, but overall their skill deteriorates during dry compared to wet conditions, as was also found by Harshan et al. (2016). While this information is valuable in itself, it might have implications for
e.g. urban heat island and heat stress studies. Fischer and Schär (2010) and Oleson et al. (2015) clearly pointed out that additional heat exacerbates heat stress, morbidity and heat-related mortality, shown to be higher in urban environments during heat waves. Such episodes manifest themselves during heat-wave periods that are generally characterized by dry conditions. An inadequate representation of the energy balance components during such events might lead to a misrepresentation of surface and canyon air temperatures (as well as e.g. humidity) in turn leading to incorrect heat stress metrics (e.g., Buzan et al. 2015).

Second, all models reproduce the observed decrease in $Q_E$ during the exceptional dry period. Yet, the magnitude of the modelled latent heat flux is underestimated (stronger during the dry period compared to the wet period). When representing the modelled latent heat flux skill as a function of dry hours since a set of precipitation events, SURFEX and CLM show an interesting behaviour: both models overestimate the latent heat flux until 6 hours after these events followed by an underestimation in $Q_E$. This ‘peak’ in the first six hours resembles the peak of the ‘PTEB’ simulation displayed in Fig. 5 of Wouters et al. (2015). To address this, the surface interception distribution (SID) approach assuming a depth distribution of water reservoirs on the impervious surface was proposed. This is illustrated in the current study by increasing the maximum water ponding depth $w_m$ from 1 to 1.31 kg m$^{-2}$ and adding a maximal evaporating surface fraction parameter ($\delta_m=0.12$), as was estimated by Wouters et al. (2015). Our results indicate that this framework is both able to alleviate the evaporation peak in the first 6 hours after rainfall, as well as the error statistics over the full period. Although the SID and storage parameters $w_m$ and $\delta_m$ were derived and evaluated by Wouters et al. (2015) over Toulouse (France) and Basel (Switzerland) respectively, the values are shown to be a good first approximation for tropical Singapore. Given this approach provides a more physical basis for the maximum water storage $w_m$ compared to the currently used arbitrary constant of 1 kg m$^{-2}$, it is advised to integrate and further develop the SID framework for improving the performance of numerical meteorological and climate models (Wang et al. 2013; Prein et al. 2015).

This finding contributes to the ongoing efforts in improving the often inadequate representation of the urban water budget and the latent heat in ULSMs (Grimmond et al. 2010, 2011). Consequently, this result supports a better assessment of urban adaptation strategies such as e.g. climate- and water-sensitive urban design and green urban infrastructure (Starke et al. 2010; Coutts et al. 2013; Demuzere et al. 2014a,b); and is relevant for improving the performance of numerical meteorological and climate models (Wang et al. 2013; Prein et al. 2015).
for tropical (humid) regions, such advancements are critical for a better assessment of e.g. the two-way interaction between urbanisation and the initiation of thunderstorms and their socio-economic consequences (Haberlie et al. 2015; Thiery et al. 2015, 2016).

Third, $L^\uparrow$ results indicate a distinct difference between SUEWS and the three other models. CLM/SURFEX and TERRA_URB are not only characterised by a large (positive) bias in $L^\uparrow$ during the day turning negative at night but also an increasing bias in $L^\uparrow$ during the dry period. In contrast, SUEWS has a small overall bias in $L^\uparrow$ which decreases during the dry period. A reason for this different behaviour can be found in the way the models calculate $L^\uparrow$. In SUEWS, the radiative flux components are derived from the incoming shortwave solar radiation using the net all-wave radiation scheme (NARP) (Loridan et al. 2011). In the latter, $L^\uparrow$ depends on the forcing temperature $T$ and a correction factor that takes into account the differences between the radiative temperature of the surface and $T$. This is clearly different from the other models, in which emitted longwave radiation is a direct function of modelled surface temperature and emissivity. When these models are evaluated in an offline setting, we hypothesise that neglecting the absorption and emission by water vapor (water vapor opacity effect, WVO) leads to an (artificial) poor representation of modelled $L^\uparrow$ compared to observations from a micrometeorological tower that do register these radiative interactions. This is illustrated in the current study by applying the WVO framework developed by Wouters et al. (2015) to TERRA_URB and CLM/SURFEX. While the REF-driven baseline simulations had a RMSE > 30 W m$^{-2}$ and $S_{scores}$ below 0.7, the simulated $L^\uparrow$ corrected for WVO reduces the RMSE to ~5 W m$^{-2}$ and $S_{scores}$ > 0.9, in line with the results obtained for SUEWS.

All of the above underlines the need to continue our efforts in developing and evaluating ULSMs which can ultimately support the development of urban climate adaptation strategies for (sub)tropical regions (and beyond). When model developments are tested offline using observational datasets, one needs to ensure that there is no mismatch between what is actually measured and modelled (cf. water vapor opacity effect). In addition, future model developments should not only focus at integrating more physically-based characteristics in urban canopy models but also on the correct representation of urban morphology and thermal and radiative characteristics. In terms of urban characteristics, the correct representation of the urban extent, and more specifically the impervious fraction, should be a primary concern when studying the urban impact on the atmosphere at the local, regional or global level. For example, Schneider et al. (2010) note an order of magnitude difference between the global urban extent (expressed in km$^2$) derived from different global products. In addition, Nordbo et al. (2015) emphasizes how, within the the urban extent, also the amount of vegetation plays a key role in urban climate simulations. In this respect, the ‘World Urban Database and Access Portal Tools’ (WUDAPT) framework (Bechtel et al. 2015; See et al. 2015) is likely a promising tool. Herein, consistent data are collected at various stages, with level 0 being the Local Climate Zones, LCZs, (Stewart and Oke 2012) and higher level products providing more specific parameters about urban form (canyon height-to-width ratio, building/canyon height), built materials and function. Since the LCZ classifier uses Landsat 8 red, near infrared and thermal bands, available at 30 m horizontal resolution, the resulting products can be expected to provide a detailed, globally available consistent and comprehensive dataset on the urban landscape with respect to its canopy layer climate (Bechtel et al. 2015; Alexander et al. 2016a).

Finally, future ULSMs evaluations should not only continue for different background climates but also for specific weather conditions within these climates (e.g. Ward et al. 2016). Where possible, the evaluation procedure should aim at an extended multi-variable model approach. Here, the use of e.g. surface temperature measurements either from infrared thermometers (e.g. Harshan et al. 2016) or thermal infrared satellite data (e.g. Xu et al. 2008; Parlow et al. 2014; Rayner et al. 2014; Zhao et al. 2014; Wouters et al. 2016) and soil moisture profiles sampled from vegetated or bare soil fractions in an urban environment (e.g. Demuzere et al. 2014a) might provide more information.

---

**Figure 11:** Observed and modelled ensemble mean diurnal variation of $L^\uparrow$ for (a) the dry period and (b) the wet period. Full (dashed) lines refer the default model output without (with) accounting for the water vapor opacity framework described in Section 2.3. Note that the WVO framework is not relevant for the SUEWS model, but its default output has nevertheless been added for completeness. Because of the similarity between CLM and SURFEX, they are plotted together.
about the drivers of error in the modelled surface energy balance components.

Acknowledgement

This work has been funded by the Flemish regional government through a FWO (Fund for Scientific Research) post-doctoral position and FWO mobility grant, respectively. The model simulations used resources provided by the VSC - Flemish Supercomputer centre, which is funded by the Hercules foundation and the Flemish Government (EWI). We also thank the Academy of Finland (Project numbers: 138328, 1127756, ICOS-Finland 281255 and 263149). NCAR is sponsored by the National Science Foundation (NSF). This work also received funding from the Belgian Science Policy Office through its Science for a Sustainable Development Programme under contracts SD/CS/041/MACCBET and BR/143/A2/CORDEX.be. The flx tower operation has been supported by the National Research Foundation Singapore through the Singapore MIT Alliance for Research and Technologys CENSAM research program and the National University of Singapore (research grant R-109-000-091-112). Support from the Met Office/Newton Fund CSSP-China is also acknowledged. Finally, we would like to thank three anonymous reviewers for the insightful and useful comments on an earlier version of this paper.

References


Faroux S, Kaptul Tchuenté A, Roujean J, Masson V, Martin E, Le Moigne P. 2013. ECOCCLIMAP-II: Europe: a twofold database of ecosystems and


© 2016 Royal Meteorological Society