Dear reviewer,

Thank you for the very detailed review and the valuable suggestions for our study. We will soon go through all of them, revising the manuscript accordingly or explaining the reasons otherwise. Meanwhile, we would like to briefly explain the novelty of our study, the choice of the SCOPE model and some assumptions in our study that justify not having used an urban hydrological model currently available to predict ET.

**Novelty and motivation of the study**

The novelty of this study is the solution that combines the temporal dynamic of ET in a vegetated environment with the spatially fragmented land cover in urban environments, providing a less computationally expensive but plausible ET product suitable for most cities in the world. The prediction accuracy (precision and bias) is compatible with the state of the art in urban hydrology models (table 1 and table 2) and potentially more transferable and less demanding.

Another important remark is that the study focuses on modelling with open data available for most medium and large cities of Europe. The intention is to present a model able to generalise for other locations in Berlin and also for other cities based on the promising results shown for the two EC tower locations. The focus of the study is terrestrial ET from plant transpiration and soil evaporation, without considering latent heat fluxes from anthropogenic sources such as car combustion or house heating. It also excludes interception loss. The intention is to develop a method capable of providing ET maps for mitigating the urban heat island effect and droughts by better managing vegetation in cities, which justifies the focus on plant transpiration and soil evaporation.

**Why the choice of the SCOPE model and urban correction**

Models are simplifications by definition. In a fragmented urban environment, it is basically unfeasible to have data to elaborate an ET (or latent heat flux) model that captures all variabilities of the microclimate, land surface, water availability, anthropogenic sources (fixed or mobile) and drainage systems in different temporal and spatial resolution for an
entire city. Therefore, our focus is on soil evaporation and plant transpiration. A sophisticated Soil-Vegetation-Atmosphere Transfer (SVAT) model such as the SCOPE model (Soil Canopy Observation, Photochemistry and Energy fluxes) provides the necessary framework for our application due to the following reasons:

(a) **Capacity to integrate both high-resolution climatological and medium-resolution remote sensing data:** The model provides predictions at a high temporal resolution as it computes the main climatological input parameters required for energy fluxes. But it also incorporates moderate temporal resolution land surface inputs for soil and vegetation such as soil moisture, LAI, vegetation height and many other biophysical and biochemical parameters.

(b) **Sophisticated approach to estimating energy fluxes:** SCOPE calculates the essential elements of the energy fluxes, including LE, H, G, net radiation, soil and canopy temperature, friction velocity, and aerodynamic resistance. It also estimates energy fluxes (LE and H) for soil and vegetation separately and warns when the energy balances cannot be closed for a specific timestamp. There are also options to correct for Monin-Obukhov atmospheric stability and Vcmax for temperature, which is crucial to ET estimation.

(c) **The usability and efficiency of SCOPE compared with urban hydrological models:** The model is divided into different modules, allowing to focus on the model inputs that are important for the target heat flux outputs. The calibration and processing time permits high temporal resolution predictions for many different points in space. The documentation and codes are freely available (https://scope-model.readthedocs.io/en/latest/mSCOPE.html and https://github.com/peiqiyang/mSCOPE).

Furthermore, the strategy of combining hourly SCOPE predictions with high spatial resolution landcover to correct for the homogeneous vegetation assumption is justified as the impervious area (or the spatial domain) is mainly static over a one-year interval. If embedded in the model, the landcover would be primarily redundant and very computationally demanding. This approach allows us to adapt the model easily for different spatial and temporal resolutions, which would be much more complicated if the urban features were used for hourly calculations inside of SCOPE.

- **Why not an urban hydrological model such as SUEWS or UT&C**

Based on the urban models suggested by reviewer #1 and the claim from reviewer #2 that there are more sophisticated urban models to predict ET rather than use SCOPE model corrected for the urban surfaces, we explain below why we think these models are not suitable for our study aims. Notice that this is not a criticism of these model approaches, but an explanation of why the design and model concept do not match our aims to have a transferable and high spatiotemporal model using freely available data to map ET at any location in the city. The arguments are divided into two parts for both approaches, 1) model parameters and data availability, and 2) model accuracy for latent heat fluxes.

**SUEWS model**

**Parameters and data availability:** Despite the intention to reduce the number of variables required for the model compared to the Grimmond and Oke (1991) model as many of those were not routinely (according to Jarvi et al. 2011), there are still many parameters that are difficult to supply for high temporal and spatial applications. For instance, the model authors describe as several model inputs to be important to parametrize for a specific site, including “interception state of the canopy of i_{th} sub
surface”, “wetness state of the soil”, “fraction of irrigated surface area using automatic irrigation”, “fraction of soil without rocks”, “plan area fraction of unirrigated grass”, “population density inside the grid”, “maximum storage capacity of ith soil store”, among others. For example, Rafael et al. (2020) included thermal conductivity, surface albedo and surface emissivity for road, root and walls. In addition, anthropogenic activities such as building heating and traffic can be a constraint to upscale the model. Wald et al. (2016) included energy consumption statistics to estimate anthropogenic heating (QF), the proportion of evergreen and deciduous trees, and albedo and emissivity per surface type. Rafael et al. (2020) state that the availability of measured data is a limitation for applications.

Our modelling approach requires little calibration and uses few variables, being all of them freely available online for our two site locations. While the SCOPE includes more than 60 inputs, our study shows that for ET/QE predictions, ten inputs were enough to have relatively high accuracy. These ten inputs all relied on open-source datasets; the remaining parameters remained constant as default or literature values.

**Accuracy for latent heat flux:** The model provides many outputs for urban environments, but the estimations of latent heat flux (QE) may be the most critical ones. The accuracy for QE is very low in some cases, especially for densely urbanised sites. Rafael et al. (2020) applied the model in two locations in Portugal. They concluded that the better performance of QE in suburban areas compared to dense urban areas is broadly consistent with previous studies (Ward et al., 2016). Wald et al. (2016) which also study two areas with different urbanisation made a similar statement saying that the results for the UK sites seem to be broadly consistent with previous studies, citing that QE is underestimated at both sites in Helsinki, although performance is generally better at the suburban site than the city-centre site (Karsisto et al., 2015).

Table 1 - SUEWS models accuracy for urban and suburban sites according to the literature

<table>
<thead>
<tr>
<th>Publication</th>
<th>Suburban site metrics</th>
<th>Urban (denser) site metrics</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rafael et al., 2020</td>
<td>r* 0.61 and MBE 0.27</td>
<td>r* 0.13 and -12 (Porto)</td>
</tr>
<tr>
<td></td>
<td>(Aveiro)</td>
<td></td>
</tr>
<tr>
<td>Ward et al., 2016</td>
<td>R² 0.72 and MBE 8.9</td>
<td>R² 0.25 and -6.4 (London)</td>
</tr>
<tr>
<td></td>
<td>(Swindon)</td>
<td></td>
</tr>
<tr>
<td>Karsisto et al., 2015</td>
<td>r* 0.79 and MBE -23.8</td>
<td>r* 0.44 and MBE -23.9</td>
</tr>
<tr>
<td></td>
<td>(Kumpula**)</td>
<td>(Torni**)</td>
</tr>
</tbody>
</table>

* reported correlation r rather than R². ** reported the best accuracy from summertime

Our modelling approach also presents better accuracy for the suburban site (R² 0.82 against 0.47). Therefore for us, there is no justification to use a sophisticated but demanding urban model that performs worse in an urban environment. Ward et al. (2016) also suggest that future model development should allow some evaporation to occur from paved and built surfaces (other than evaporation of intercepted water), which is similar to our simplification to consider no ET from 100% impervious surfaces using the correction factor (excluding interception).
UT&C model:

**Parameters and data availability:** The model is based on the infinite urban canyon approximation (Meili et al., 2020) and is very dependent on urban geometry such as canyon height, canyon width, roof width and street directions. UT&C accounts for anthropogenic heat flux, which is added to the canyon air, assuming that heat emissions mainly occur within the urban canyon. Hence, anthropogenic heat emissions caused by air conditioning, car exhaust, industry, human metabolism, or any other anthropogenic heat source need to be estimated prior to simulation. In a city like Berlin, the assumption of canyons and the requirements to specify a precise spatial configuration is prohibitive to model ET at the spatial and temporal resolution desired or mapping ET for the entire city. According to the 81-page supplementary material, the parameters used for the model validation in Singapore (SG), Melbourne (MB), and Phoenix (PH), includes the distance of the wall to the tree trunk (m), albedo and emissivity of the wall, volumetric heat capacity of impervious, roof and wall, the thickness of wall and roof layers. These variables are possible for experimental models at a very reduced scale to have some insights and increase the knowledge of the potential factors affecting heat flux in the urban environment. However, they are unreasonable in a transferable and generalisable model to be applied in real-life cases, especially with the aim of mapping ET at a high resolution for an entire city.

To reiterate, our modelling approach requires little calibration and uses few variables, with all of them being freely available online. There are infinite variables and interactions that a model can explore, but a transferable model (empirical or physical-based) requires a source of parsimony. Therefore, the model to fit our demand should capture the essential sources of ET that, in our assumption, is from pervious soil and vegetation.

**Accuracy for Latent heat flux:** Although the UT&C model is a very sophisticated and detailed model for urban environments (i.e. urban canyon design), the model’s accuracy is not so different from SUEWS models. The reported $R^2$ range from 0.50 to 0.62, which is not that high as the model was developed and calibrated based on this study (see table below).

Table 2 - UT&C models accuracy for three cities compared to the literature

<table>
<thead>
<tr>
<th>City</th>
<th>Metrics from UT&amp;C models</th>
<th>Literature using other models</th>
</tr>
</thead>
<tbody>
<tr>
<td>Singapore</td>
<td>$R^2$ 0.60 and MBE -1.2</td>
<td>$R^2$ 0.30-0.61 and MBE -10.8</td>
</tr>
<tr>
<td>Melbourne</td>
<td>$R^2$ 0.62 and MBE 1.9</td>
<td>$R^2$ 0.45 and MBE -0.8</td>
</tr>
<tr>
<td>Phoenix</td>
<td>$R^2$ 0.50 and MBE 4.1</td>
<td>-</td>
</tr>
</tbody>
</table>

Our modelling approach has better accuracy for one site and the other similar to reported for Phoenix ($R^2$ 0.82 against 0.47). Furthermore, our combined approach can be transferred to any area of the city without issues collecting the necessary data inputs or without the need to define the canyon limits and how individual urban features relate to the canyon.

We hope that our general explanation provided above can justify most of our model
approach choices. Although we acknowledge that anthropogenic sources contribute to latent heat fluxes, the primary sources of terrestrial ET are still coming from plant transpiration and soil evaporation. Climatological conditions are the main drivers of ET, which present a high temporal dynamic. On the other hand, fragmented urban landcover and impervious surfaces are the main constraints of ET released to the atmosphere. Therefore, a model to predict urban ET accurately requires high-temporal and -spatial resolutions. Still, processing all time-space interactions is demanding and currently unfeasible for the resolution needed for our plans. Based on these assumptions, we propose this two-stage model approach (SCOPE plus correction) to capture most of the spatiotemporal variability of ET without making the model overly complex.

Best regards,

Alby Duarte Rocha (on behalf of my co-authors)

References used in the justification


