

Anonymous Referee #2, 13 May 2026

Report #2

Submitted on 13 May 2026 Anonymous referee #2	
Notes for the submission of interactive comments	
Anonymous: Yes No	
Formal manuscript rating and recommendation to the editor (visible to authors and reviewers only)	
1) Scientific significance Does the manuscript represent a substantial contribution to modelling science within the scope of this journal (substantial new concepts, ideas, or methods)?	Excellent Good Fair Poor
2) Scientific quality Are the scientific approach and applied methods valid? Are the results discussed in an appropriate and balanced way (consideration of related work, including appropriate references)? Do the models, technical advances and/or experiments described have the potential to perform calculations leading to significant scientific results?	Excellent Good Fair Poor
3) Scientific reproducibility To what extent is the modelling science reproducible? Is the description sufficiently complete and precise to allow reproduction of the science by fellow scientists (traceability of results)?	Excellent Good Fair Poor
4) Presentation quality Are the scientific results and conclusions presented in a clear, concise, and well structured way (number and quality of figures/tables, appropriate use of English language)?	Excellent Good Fair Poor
For final publication, the manuscript should be accepted as is accepted subject to technical corrections accepted subject to minor revisions reconsidered after major revisions rejected Please note that this rating only refers to this version of the manuscript!	
Were a revised manuscript to be sent for another round of reviews: I would be willing to review the revised manuscript. I would not be willing to review the revised manuscript.	
Referee Comment: 'Comment on egusphere-2026-551'	

This study presents the implementation of an “urban” tile in the ORCHIDEE land cover and vegetation model for future applications in climate modeling at various spatial scales. The evaluation of energy fluxes is based on the Urban-Plumber database, and shows a nearly systematic improvement in modeling of sensible and latent heat fluxes which demonstrates the added value of the new parameterization. However, some biases persist in ORCHIDEE: it would be relevant to suggest some ways to improve the model and to assess the model’s sensitivity to certain parameters (e.g. roughness and drag coefficient). For the hydrological component, one limitation of the study is the lack of data needed to conduct an objective evaluation. Consequently, this section presents only a sensitivity analysis comparing the different ORCHIDEE configurations, which makes it difficult to critically analyze the results. We would expect to gain a better understanding of the specific characteristics of certain sites that result in a “non-linear” response from the Urban2 version comparing to Urban1.

We thank the reviewer for these suggestions. We agree that the remaining biases require further investigation and that the lack of hydrological observations limits the interpretation of the water-budget results.

For the energy-budget evaluation, we examined the sensitivity of the simulated sensible heat flux to the aerodynamic parameters, including roughness length and the drag coefficient. These diagnostics indicate that the remaining sensible heat biases cannot

be explained primarily by the aerodynamic formulation. Further analysis, detailed in our response to Reviewer 1, instead points to two main limitations: the prescribed soil thermal conductivity and heat capacity at sites with relatively low urban fractions, and an underestimation of latent heat flux at many sites. We have revised the manuscript accordingly and now identify the refinement of the soil thermal properties and evapotranspiration parameterizations as priorities for future model development.

We also agree that the hydrological analysis remains limited because runoff, soil moisture, and the other water-budget components are not observed at the Urban-PLUMBER sites. The results should therefore be interpreted as a sensitivity analysis of the different ORCHIDEE configurations rather than as an objective hydrological validation, but we plan to explore performance at the catchment scale in future work. We have revised the discussion of the apparently non-monotonic response in Urban2 to clarify this limitation and identify possible mechanisms to investigate in future work.

We replaced:

“The Urban2 simulation, which includes a lower level of imperviousness, also leads to an increase in surface runoff (with a corresponding decrease in subsurface runoff), but the magnitude of change is much smaller. Interestingly, at a few stations (e.g., KR-Ochang and SG-TelokKurai), Urban2 even shows a slight increase in subsurface runoff, suggesting a non-linear response of the water balance to moderate imperviousness. Urban2 also results in a decrease in surface runoff at some stations (e.g., FI-Torni and FI-Kumpula).

with:

“The Urban2 simulation, which represents a more moderate reduction in saturated hydraulic conductivity, generally increases surface runoff and decreases subsurface runoff, although the magnitude of these changes is much smaller than in Urban1. However, the response is not consistent across all sites: Urban2 slightly increases subsurface runoff at KR-Ochang and SG-TelokKurai and decreases surface runoff at FI-Torni and FI-Kumpula. These site-dependent changes indicate that the annual water-budget response to a moderate conductivity reduction is not necessarily monotonic. Although saturated hydraulic conductivity is scaled linearly with imperviousness, the resulting water-budget response may be nonlinear because runoff generation and vertical water transfer depend on infiltration thresholds, rainfall intensity, antecedent soil moisture, soil texture, evapotranspiration, and changes in soil-water storage.”

P2, L42

"urban bulk schemes" : add references

Done

P2, L61

"Chancibault et al., 2014": The reference listed here does not appear to be correct; I would suggest Stavropoulos-Laffaille et al. (2018) instead. > DOI: 10.5194/gmd-11-4175-2018

Done

P3, L66-68

You should add a point about the lack of data for process documentation and model evaluation.

We agree that the limited availability of observations is an important constraint on both process understanding and model evaluation. The sentence highlighted by the reviewer was removed when the urban hydrology section was restructured; however, this point is now addressed explicitly at the end of the revised paragraph:

"Progress is also constrained by the limited availability of observations, as urban sites measuring turbulent fluxes rarely provide all water-balance components required to evaluate runoff, irrigation, drainage, and storage changes consistently at the same spatial scale (Jongen et al., 2024)."

P3, 79

Please specify the model name using ORCHIDEE here.

The full name of the ORCHIDEE (Organising Carbon and Hydrology In Dynamic Ecosystems) model has now been specified at its first occurrence.

P3, L85

"...by having a lower albedo"

It is important to note that we are referring here to a relative albedo that takes into account the geometry of the urban canopy, not a surface albedo.

Yes, we changed the sentence in : *"First, cities modify the absorbed shortwave radiation through changes in the effective urban canopy albedo relative to rural areas because of the materials used in cities and the geometry of buildings and streets (Kotopouleas et al., 2021; Qin, 2015)."*

P3, 87

"This lower albedo reflects less solar radiation leading to an increase in net radiation (Rnet), and ultimately in higher surface temperatures"

Some caution is warranted regarding this cause-and-effect relationship: (1) net radiation also increases because of radiative trapping of infrared emissions from surfaces inside the urban canopy, and (2) surface temperatures also rise due to thermal properties of materials.

We have revised the text to clarify that urban net radiation is influenced by both enhanced shortwave absorption and longwave radiative trapping within the urban canopy, while higher surface temperatures also depend on the thermal properties of urban materials: *“Urban geometry also affects longwave exchanges by trapping infrared radiation within the canopy. Together with the high heat capacity and thermal conductivity of urban materials, these radiative effects contribute to increased heat storage and often higher urban surface temperatures.”*

P3, L94

"This is the result of the higher thermal capacity and conductivity of urban materials compared to natural areas..."

The amount of surface area available for heat storage—which is linked to urban geometry—is also a key factor.

Yes, we modified the sentence into:

“This larger heat storage results from both the higher heat capacity and thermal conductivity of urban materials and the greater surface area available for heat exchange and storage within the urban canopy, including roofs, walls, and roads”

P4, L120

"version 2.2 of ORCHIDEE (rev. 8133)."

What does “rev. 8133” mean? Could you clarify and provide the reference?

“Rev. 8133” refers to the specific source-code revision of ORCHIDEE version 2.2 used in this study. This identifier defines the exact code snapshot and is therefore provided to ensure reproducibility. The corresponding repository information and access details are given in the Code and data availability section.

We have clarified the text as follows:

“This study uses ORCHIDEE version 2.2, revision 8133; further details and access information are provided in the Code and data availability section.”

P4, L124

"Hydrological and energy fluxes associated with vegetation ..."

Change by "Water and energy fluxes associated with vegetation and natural soils ..."

"interception" > please clarify (intercepted by vegetation?)

"evaporation" > please clarify (soil evaporation?)

Yes the sentence is now: *“Water and energy fluxes associated with vegetation and natural soils, including transpiration, evaporation of water intercepted by the vegetation canopy, and soil evaporation, ...”*

P5, L125

It is necessary to clarify what you mean by “flux-aggregation approach”

We added: “(calculated separately for each PFT and then aggregated to the grid-cell scale as area-weighted fluxes)” after flux-aggregation approach

Also it is essential to include a figure illustrating the discretization/description of the grid and the soil column for modeling hydrology and thermal processes.

We now include this new figure to present the model:

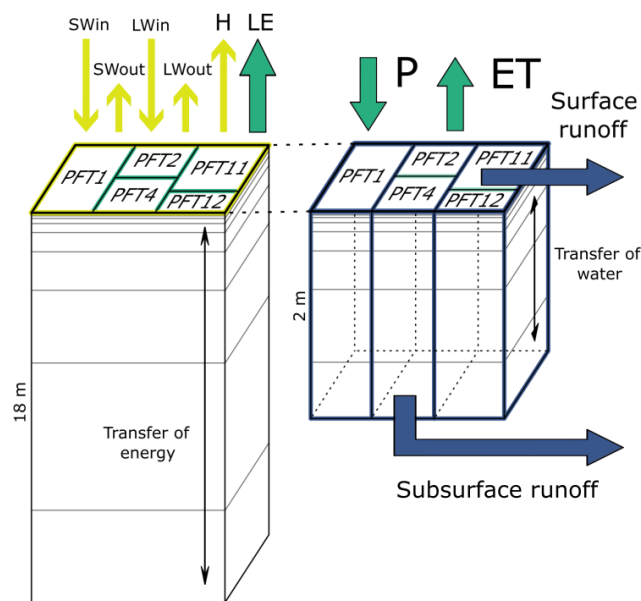


Figure 1: Schematic representation of the energy- and water-budget calculations in ORCHIDEE for a single grid cell, its plant functional types (PFTs), and the three associated soil columns representing bare soil, high vegetation, and low vegetation. Yellow arrows denote fluxes calculated directly at the grid-cell scale, green arrows denote fluxes calculated separately for each PFT and subsequently aggregated to the grid-cell scale, and dark-blue arrows denote fluxes calculated independently for each soil column. For example, sensible heat flux (H) is calculated at the grid-cell scale, evapotranspiration (ET) at the PFT scale, and surface runoff and drainage at the soil-column scale. Each soil column is discretized into 11 hydrological layers extending to a depth of 2 m, while the thermal discretization extends to 18 m. Abbreviations: P, precipitation; ET, evapotranspiration; H, sensible heat flux; SW, shortwave radiation; LW, longwave radiation; PFT, plant functional type.

P5, 144

"First, the model estimates change in surface temperature based on the downwelling and upwelling radiative terms ..."

Requires clarification on how the surface temperature (or the temperature of the first soil layer) is calculated

We have clarified that Tsurf is the grid-cell surface temperature, distinct from the temperature of the uppermost soil layer. It is calculated by implicitly solving the complete surface energy balance, including radiative, turbulent, and conductive heat exchanges. The resulting surface temperature is then used as the upper boundary condition for updating the soil-temperature profile.

We modified the text as:

“At each time step, ORCHIDEE implicitly solves a single grid-cell surface temperature, T_{surf} , from the surface energy balance, accounting for radiative, turbulent, and conductive heat exchanges. T_{surf} is distinct from the temperature of the uppermost soil layer and provides the upper boundary condition for the implicit soil heat-diffusion scheme. The net radiation is then calculated as:”

P5, 147

" ϵ is the emissivity (1, unitless)"

Why is emissivity equal to 1 ?

In the standard ORCHIDEE configuration used in this study, surface longwave emissivity is fixed to 1 unless a site-specific value is prescribed. We have clarified that this is a model assumption corresponding to a blackbody approximation, rather than a general physical property of land surfaces.

We changed the text in:

“ ϵ is the surface longwave emissivity, fixed to 1 in these simulations, corresponding to a blackbody approximation in the thermal infrared,”

P5, L149-152

"Each vegetation PFT has its own value of albedo (α)"

The albedo should be noted for a leaf consistency with the with the rest of the text and the equations ?

The text on soil albedo needs to be reworded, ex. "For bare soil, the specification of albedo must take into account the spatial variability of reflectance, resulting from composition, moisture, and texture heterogeneities".

We thank the reviewer for pointing this out. We have corrected the notation to α_{leaf} for consistency and replaced the description of bare-soil albedo with:

The use of MODIS-derived bare-soil albedo fields accounts for the spatial variability of soil reflectance resulting from heterogeneities in soil composition, moisture, and texture.

P6, L156

"... weighted by the fractional areas of the PFTs of the grid cell (Equation 3)."

Change by "... weighted by the fractional areas of the PFTs of the grid cell :"

Done

P6, L159

"... of each vegetated PFTs"

Change by "... of each vegetated PFT"

Done

P6, L164

"The latent and sensible heat fluxes are calculated respectively thanks to Eq. (4) and Eq. (5)."

Change by "The sensible and latent heat fluxes are calculated respectively thanks to Eq. (4) and Eq. (5):" to be consistent with the order of equations

Done

P6, 172

"The calculation of C_d (also expressed ..."

If I understand correctly, you should add "for a given PFT"

Done

P7, L183

"... the soil thermal conductivity and the soil heat capacity, and both vary with soil texture and with the water content of each soil layer."

You should remove "and with the water content" because this point is discussed later.

Done

P7, L191-193

"Finally, the soil thermal properties also change with the presence of water inside the soil layers. Eq. (9) shows the representation of the soil heat capacity as a function of the fraction of water inside the layer. Eq. (10) represents the evolution of soil thermal conductivity as a function of the same fraction:"

I would put it simply as "Finally, the soil thermal properties also change according to the soil water content of each soil layer:"

You mention that thermal properties depend on the soil moisture content, but the soil column used for modeling water exchange and heat transfer is not the same depth. How do you deal with this?

We have simplified the introductory sentence as suggested and clarified how soil moisture information is transferred between the hydrological and thermal discretizations. Over the common upper 2 m, volumetric soil moisture calculated by the hydrological scheme is interpolated from hydrological nodes to thermal-layer interfaces, while the saturation degree and layer-integrated water content are transferred directly. Below the hydrological domain, the saturation degree and volumetric water contents are

held equal to those of the deepest hydrological layer, while the layer-integrated water content is adjusted according to the thickness of each thermal layer.

We added to the text:

“The hydrological and thermal discretizations share the upper 2 m of the soil column. Within this common domain, hydrological soil-moisture variables are transferred to the thermal grid, including interpolation of volumetric water content to the thermal-layer interfaces. Below 2 m, the saturation degree and volumetric water contents are held equal to those of the deepest hydrological layer, while the total water content is adjusted for the thickness of each thermal layer.”

P7, L196

"where, λ_s and λ_w are the heat conductivities ..."

Subscripts are capitalized in equations and lowercase in the text.

Done

P8, L215

"Where $K(\theta)$..."

Change by "where $K(\theta)$..."

Done

P9, L225

"Reference saturated hydraulic conductivity (K_{sref}) at the depth $z_{lim} = 0.30\text{m}$, derived from the dominant USDA ..."

What does this “ z_{lim} ” depth refer to?

Explain USDA

U.S. Department of Agriculture

P9, L230

"... represented by a factor $F_{K_{roots}}(z, c)$, applied only"

Could you clarify what is "c" ?

Here, (c) denotes the soil-column index. The factor $F_{K_{roots}}(z, c)$ depends on the type of soil column: it represents the root-induced enhancement of saturated hydraulic conductivity in vegetated soil columns, whereas it is set to 1 for the bare-soil column. We have clarified this notation in the revised manuscript.

P9, L239-240

Check unit of "f" in m^{-1}

We corrected “f = 2 m” into: “f is set to 2 m⁻¹”

P9, L242

"Infiltration is computed using a Green–Ampt type approach"

Add a reference

Done

P10, L250-256

Add the units to the terms E_{total} and βET

Done

P10, 262

"Once the latent heat flux is calculated with Eq. (5) and Eq. (17), ORCHIDEE calculates the different ET, as:"

I suggest "Once the latent heat flux is calculated with Eq. (5) and Eq. (17), ORCHIDEE calculates the different terms contributing to total evapotranspiration:"

Done

Section 2.2.1

This section requires more details to fully understand how the various parameters are defined and calculated/aggregated: Is the albedo a composite (or relative) albedo? How is roughness calculated? ... And here again, a figure would be helpful for understanding.

P11, L288 (and Eq 22)

"the vegetation root factor $FK_{\text{roots}}(z,c)$ with a constant urban factor $K_{\text{facturban}}$..."

To ensure greater consistency in the notation, shouldn't we replace $K_{\text{facturban}}$ with FK_{urban} ?

Yes, we replaced it by F_{urb}

P12, Eq 24

The “max” is unnecessary because, by definition, $1 - 0.9f_{\text{imp}}$ is always greater than or equal to 0.1

Done

P12, L302-304

Please provide references.

We modified the sentence into:

“When site-specific information is unavailable, the modeler may prescribe a representative value. For example, the Copernicus Urban Atlas distinguishes discontinuous medium-density urban fabric ($0.30 < f_{imp} \leq 0.50$), discontinuous dense urban fabric ($0.50 < f_{imp} \leq 0.80$), and continuous urban fabric ($f_{imp} > 0.80$; European Environment Agency, 2020).”

European Environment Agency. (2020). Copernicus Land Monitoring Service (CLMS): Urban Atlas Land Cover/Land Use and Street Tree Layer 2012 and 2018—Product User Manual, version 6.3 (6.3.1).

(https://library.land.copernicus.eu/products/Urban_Atlas_Land_Cover-Land_Use_and_Street_Tree_Layer_2012_and_2018_PUM_v6.pdf)

P12, Fig1

Add unit

Done

P12, L313

"... for the grid cells where urban areas cover more than 50% of the grid cell"

I am not sure to understand what is done if the fraction is less than 50%?

For grid cells with an urban fraction of 50% or less, the thermal conductivity and volumetric heat capacity are not modified and retain the values calculated by the standard ORCHIDEE formulation as functions of soil texture and soil moisture. No weighted averaging between natural and urban thermal properties is applied in the current implementation. We have clarified this point in the revised manuscript:

“In grid cells where the urban fraction is below 50%, the standard ORCHIDEE values calculated from soil texture and soil moisture are retained.”

P12, L321

"The thermal conductivity is set to $3.24 \text{ W m}^{-1} \text{ K}^{-1}$ and the heat capacity to $1890 \text{ kJ m}^{-3} \text{ K}^{-1}$ "

In slab-based approaches (as used in SURF, Woutters et al.), thermal properties are adjusted to account for surface density (and can therefore vary depending on the urban typology). How are these properties defined and selected in this study?

In the present implementation, thermal conductivity and volumetric heat capacity are prescribed as fixed effective urban values and do not vary with urban morphology or typology. The values of $3.24 \text{ W m}^{-1} \text{ K}^{-1}$ and $1890 \text{ kJ m}^{-3} \text{ K}^{-1}$ were adopted from the Noah land-surface model. The prescribed thermal conductivity is relatively high, including compared with the effective value of $1.55 \text{ W m}^{-1} \text{ K}^{-1}$ used in SURF, and therefore provides a first-order bulk representation of enhanced heat conduction in urban areas.

However, unlike SURY, these properties are not explicitly derived from surface-area density or other morphological parameters.

We added: *“These values are not adjusted according to urban typology, building morphology, or surface-area density.”*

P13, L343

"The benchmark simulations are the Baresoil and the Lowveget simulations...."

The definition of land cover and land use characteristics is not very clear. Is it only the urban portion of the site that is classified as “lowveg” or “baresoil”? For example, in AU-PRESTON, 23% of the area is covered by trees—are these trees taken into account in the different configurations? Table S1 should be clarified and completed with all characteristics.

We have clarified the treatment of land cover in the different simulations. Only the fraction identified as urban in the Urban-PLUMBER site characteristics is reassigned between experiments. The other observed land-cover fractions, including trees, low vegetation, bare soil, and water, are retained in all configurations. Thus, for example, the 22.5% tree fraction at AU-Preston is represented in every simulation, while its 62% urban fraction is represented as bare soil in Baresoil, as low vegetation in Lowveget, and using the new urban PFT in Urban0, Urban1, and Urban2.

We modified the text to:

“The benchmark simulations are Baresoil and Lowveget. In all simulations, the observed non-urban land-cover fractions of each site, including trees, low vegetation, bare soil, and water, are retained. Only the fraction classified as urban is represented differently among the experiments. In Baresoil, the urban fraction is assigned bare-soil properties, corresponding to the default treatment of urban areas in ORCHIDEE. In Lowveget, the urban fraction is assigned the properties of the dominant low-vegetation type in the surrounding area, either grassland or cropland, providing a non-urban baseline.”

The complete land-cover and land-use fractions for all sites are already provided in Table 1. We have therefore not detailed these values for each simulation in Table S1, as doing so would considerably increase the size of the table, and make it difficult to read.

P13, L351

"For each site, the model is first spun up for 40 years, using repeated cycles of 10 years... "

The data from the Urban-Plumber sites does not cover such long periods of time. How are atmospheric forcings built over long time periods for each site (using which data) ?

As stated in the manuscript, the 40-year spin-up is generated by repeating the 10-year meteorological forcing available for each Urban-PLUMBER site four times. No additional

atmospheric forcing dataset is used. We have slightly rephrased the sentence to make this procedure more explicit.

We changed:

“For each site, the model is first spun up for 40 years by repeating four times the 10-year meteorological forcing, until soil moisture and temperature reach a stable seasonal cycle.”

Into:

“For each site, the model is first spun up for 40 years by repeating four times the 10-year meteorological forcing provided by Urban-PLUMBER, until soil moisture and temperature reach a stable seasonal cycle.”

P15, Table2

Could you add the roughness ?

Roughness is not prescribed as an independent parameter in these simulations. For the urban PFT, the aerodynamic roughness length is diagnosed from the site-specific building height reported in Table 1 using the standard ORCHIDEE formulation described in Eq. (6), also expressed as $1/(r_{ah} \cdot u)$ where u is the wind speed and r_{ah} the aerodynamic resistance, with the ratio between roughness length and PFT height, z_0 , fixed to 0.0625. The resulting drag coefficients are subsequently aggregated at the grid-cell scale according to the PFT fractions. Consequently, roughness varies among sites and cannot be represented by a single value for each simulation. The standard ORCHIDEE aerodynamic treatment is retained for the Baresoil and Lowveget configurations.

We modified:

“Turbulent transfer of the fluxes between the surface and the atmosphere, is represented thanks to the drag coefficient C_d . The calculation of C_d (also expressed as $1/(r_{ah} \cdot u)$ where u is the wind speed and r_{ah} the aerodynamic resistance) for a given PFT in the standalone version of ORCHIDEE, is calculated with the following equation:

$$C_d = \frac{k^2}{\ln\left(\frac{z_{lev}}{h_{PFT} \cdot z_0}\right)^2} \quad (6)$$

where k is the Von Karman constant; z_{lev} is the height of the first atmospheric layer (10m in standalone simulations); h_{PFT} is the height of the canopy of the considered PFT (fixed) and z_0 is a factor that estimates the roughness height above the canopy and is fixed to 0.0625 (1/16, default value in ORCHIDEE for all PFTs in standalone simulations).”

To:

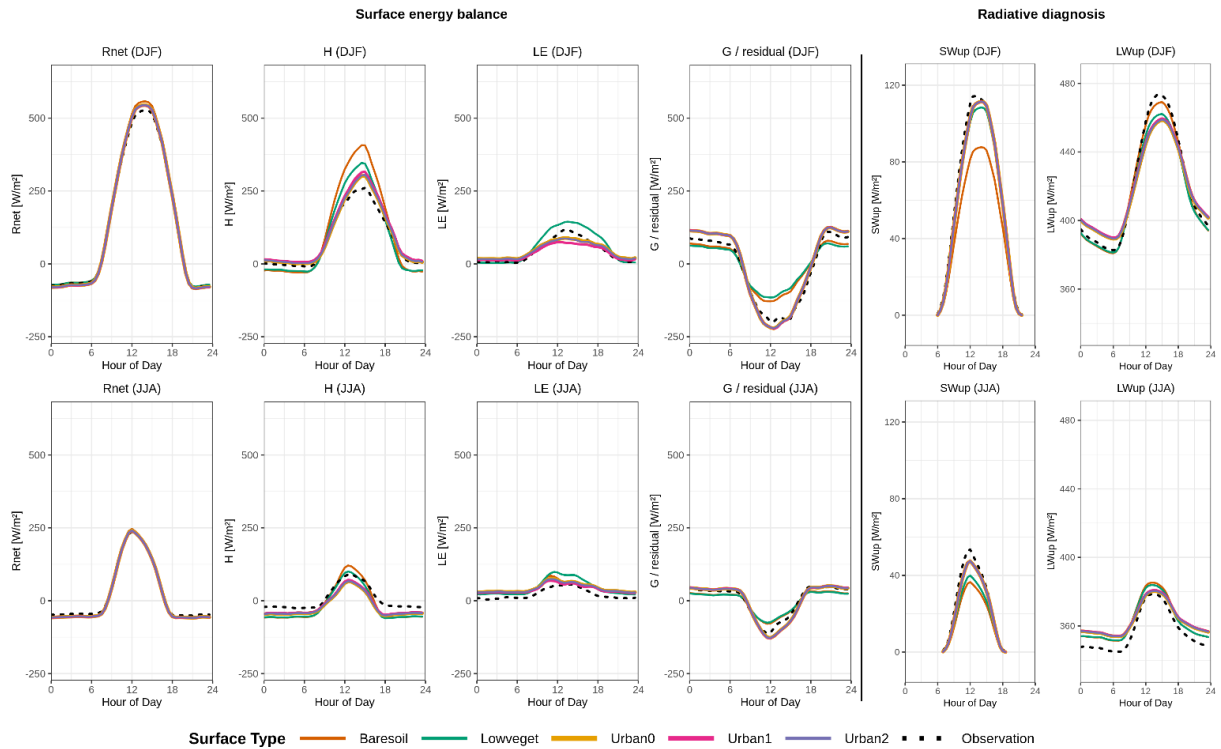
“Turbulent transfer of the fluxes between the surface and the atmosphere, is represented thanks to the drag coefficient C_d . For each PFT in the standalone version of ORCHIDEE, C_d which can also be expressed as $1/(r_{ah} \cdot u)$, where u is the wind speed and r_{ah} the aerodynamic resistance, is calculated as follows:

$$C_d = \frac{k^2}{\ln\left(\frac{z_{lev}}{h_{PFT} \cdot z_0}\right)^2} \quad (6)$$

where k is the Von Karman constant; z_{lev} is the height of the first atmospheric layer (10m in standalone simulations); h_{PFT} is the height of the canopy of the considered PFT (fixed) and z_0 is the dimensionless roughness-length-to-height ratio and is fixed to 0.0625 (1/16, default value in ORCHIDEE for all PFTs in standalone simulations). Thus, the product $h_{PFT} \cdot z_0$ in Eq. (6) corresponds to the aerodynamic roughness length. The aerodynamic roughness length is therefore not prescribed as an additional independent parameter but is diagnosed from the PFT height. For the urban PFT, h_{PFT} corresponds to the prescribed site-specific building height; consequently, its aerodynamic roughness varies among sites. Because ORCHIDEE calculates a single energy budget per grid cell, the PFT-level drag coefficients are subsequently aggregated to the grid-cell scale using the corresponding land-cover fractions.”

Section 4.1

- It seems to me that the Urban-Plumber database provide incoming and upwelling solar and IR radiation terms. It would be useful to compare observed and simulated upwelling S and L to understand the errors noted for Rnet.



Thank you for this suggestion. We have added the observed and simulated upwelling shortwave and longwave radiation components to Fig. 2. The incoming shortwave and longwave radiation terms are prescribed from the Urban-PLUMBER atmospheric forcing and are therefore identical across simulations, so they are not shown.

The figure displays the new simulations accounting for the soil texture information.

The radiative decomposition shows that Baresoil substantially underestimates upwelling shortwave radiation, while the urban configurations are closer to the observations because of their modified albedo. The urban configurations nevertheless slightly underestimate upwelling longwave radiation around midday. The remaining positive bias in Rnet therefore results rather from an underestimation of emitted longwave radiation.

We changed the text in consequence:

“However, looking at upwelling shortwave (SWup) and longwave (LWup) components reveals compensating differences among the configurations. The grid-cell albedo increases from 0.110 in Baresoil and 0.131 in Lowveget to 0.141 in Urban0, Urban1, and Urban2. Consequently, the urban configurations reflect more shortwave radiation than the benchmark simulations: Baresoil produces the lowest SWup, Lowveget intermediate values, and the three urban configurations the highest values. The urban configurations also modify the diurnal cycle of upwelling longwave radiation. Relative to Baresoil and Lowveget, they emit less longwave radiation around midday and more during the night, particularly in summer. This pattern indicates a reduced diurnal amplitude of surface

radiative temperature and is consistent with the greater effective thermal inertia resulting from the changes in heat capacity and thermal conductivity.”

- Why is the albedo prescribed to 0.141 when it is informed equal to 0.151 in the Urban-Plumber database (Table 1) ?

The value of 0.151 reported in Table 1 corresponds to the albedo assigned to the urban PFT. However, R_{net} is calculated using the grid-cell albedo, obtained by weighting the albedos of all PFTs by their respective land-cover fractions. At AU-Preston, the grid cell also contains tree, low-vegetation, and bare-soil fractions of 0.225, 0.15, and 0.005, respectively. The resulting grid-cell albedo is therefore 0.141. We have revised the text to clarify that 0.141 refers to the grid-cell albedo, whereas 0.151 is the albedo prescribed to the urban PFT:

“The grid-cell albedo increases from 0.110 in Baresoil and 0.131 in Lowveget to 0.141 in Urban0, Urban1, and Urban2. The albedo values discussed here correspond to the effective grid-cell albedo used to calculate net radiation, rather than to the albedo assigned to an individual PFT. The grid-cell albedo is calculated as the land-cover-fraction-weighted mean of the PFT-level albedos.”

- Even though the heat storage flux G is not available in the Urban-Plumber database, it would be possible to compare the “residual term” of the energy budget (i.e., $R_{net} - H - LE$), given that there may be an additional anthropogenic effect (Fig 2). Is G calculated as the residual term in ORCHIDEE ?

Thank you for this suggestion. We have added the observational energy-balance residual to Fig. 2 for a qualitative comparison with the ground heat flux simulated by ORCHIDEE. Because the ORCHIDEE sign convention defines G as negative when heat is transferred into the substrate, we plot the observational residual as $H+LE-R_{net}$, i.e. the negative of the conventional residual $R_{net}-H-LE$.

In ORCHIDEE, G is not calculated as a residual of the surface energy budget. It is explicitly computed from the temperature gradient between the surface and the upper soil or urban-substrate layer, as described in Eq. (7), and provides the upper boundary condition for the soil heat-diffusion equation. We have clarified this distinction in the revised manuscript and discuss the comparison as follows:

“During summer, Baresoil and Lowveget underestimate the magnitude of daytime heat storage, as their simulated G values are less negative than the observational energy-balance residual. The urban configurations reproduce the observed daytime magnitude more closely, although their slightly more negative values indicate an overestimation of heat storage. At night, the urban configurations also agree more closely with the observational residual, but their more positive values suggest that nocturnal heat release is overestimated. This comparison remains qualitative because the colored curves represent the ground heat flux explicitly calculated by ORCHIDEE (eq. 7),

whereas the black curve is an energy-balance residual rather than a direct observation of G. The observational residual may additionally include contributions from unaccounted anthropogenic heat and net horizontal advective transport, as well as errors associated with measurement uncertainty.”

- You suggest that the negative bias in winter for H could be related to the anthropogenic fluxes that are not considered in ORCHIDEE: while this may certainly play a role, in this case the biases appear to be equivalent between H (negative) and LE (positive), which suggests that it is the competition between the two fluxes that is not modeled correctly.

We change the sentence into:

“The comparable negative bias in H and positive bias in LE suggests that the main issue lies in the simulated partitioning of available energy between sensible and latent heat fluxes.”

P19, L441

"...due to the release of stored heat accumulated during the day."

You could add "and the persistence of heat transfer by convection at night"

Done.

P19, Fig3

Could you clarify whether the data in the boxplots represent the daily bias for all modeled days and all sites, or an average bias per site (i.e., 20 values per boxplot)?

Thank you for requesting this clarification. The boxplots do not pool the daily biases from all sites. For each site, simulation, season, and flux, we first calculate the observed and simulated daily maximum or minimum using only matching model–observation timestamps. The daily simulated-minus-observed differences are then averaged over all valid days at that site, yielding one site-level mean bias. Each boxplot therefore contains 21 values (There are 20 sites but Minneapolis has two datasets depending on the wind directions), one per site.

We have clarified this procedure in the text and in the caption of Fig. 3. We retained the site-level representation because it gives equal weight to each site, irrespective of the duration of its observational record. Pooling all daily values would give greater weight to sites with longer records and would represent variability among site-days rather than consistency of model performance across sites. In addition, consecutive daily biases from the same site are not statistically independent.

We added:

“For each site, simulation, season, and flux, the daily simulated-minus-observed differences are calculated using only matching model–observation timestamps and are

then averaged over all valid days. Each site therefore contributes one value to each boxplot, so that all sites are weighted equally regardless of the length of their observational record. This approach evaluates model performance during both the daytime and nighttime phases of the diurnal cycle.”

L21, L485

"...except for CA-Sunset, UK-Swindon, and FI-Kumpula, where performance is slightly degraded"

We notice in Fig5 a significant drop in H score for the Phoenix site, which you don't mention in the text. Could you explain ?

Thank you for pointing this out. US-WestPhoenix indeed exhibits a more pronounced degradation in sensible-heat-flux performance than the other sites mentioned in the text. The H MAE increases from 28.7 Wm⁻² in Baresoil to 37.2 Wm⁻² in Urban1. This degradation is common to all three urban configurations, with MAEs of 38.1, 37.2, and 37.8 Wm⁻² for Urban0, Urban1, and Urban2, respectively. It is therefore not primarily caused by the Urban1 imperviousness formulation.

Urban configurations substantially reduce the mean negative R_{net} bias at this site, from -18.9 in Baresoil to -1.0 W m⁻² in Urban1. This improvement results from reductions in both upwelling radiative components. Based on the mean summer daytime cycle, Urban1 decreases SW_{up} by approximately 9 W m⁻² relative to Baresoil, consistent with the decrease in grid-cell albedo from approximately 0.206 to 0.190. It also decreases LW_{up} by approximately 11 W m⁻², consistent with its lower daytime radiative surface temperature. Together, these changes increase daytime R_{net} by approximately 20.9 W m⁻², which most of this additional available energy is transferred to sensible heat.

US-WestPhoenix has an urban fraction of 48%, slightly below the 50% threshold used to activate the prescribed urban thermal conductivity and heat capacity. The urban configurations therefore modify the radiative and aerodynamic properties without activating the corresponding increase in substrate thermal inertia. This may contribute to the insufficient daytime heat storage and excessive sensible heat simulated at this site. We have added US-WestPhoenix to the text and clarified this site-specific behaviour.

We modified the text to:

“For sensible heat, improvements are also seen at nearly all sites, except for CA-Sunset, US-WestPhoenix, UK-Swindon, and FI-Kumpula, where performance is slightly degraded. Detailed analysis suggests that these degradations arise from site-specific changes in energy partitioning. At FI-Kumpula and US-WestPhoenix, improvements in net radiation are not accompanied by sufficient partitioning toward latent heat and/or heat storage, resulting in increased sensible heat flux. Their urban fractions (46% and

48%, respectively) lie just below the 50% threshold used to activate the prescribed urban thermal conductivity and heat capacity. Consequently, the urban configurations modify the radiative and aerodynamic properties at these sites without activating the corresponding increase in substrate thermal inertia, which may contribute to insufficient daytime heat storage and the degradation in sensible-heat-flux performance.”

Section 4.3

- Fig6 : Wouldn't it make more sense to present the different components (drainage, runoff, evap) as percentages of total precipitation (with the total amount of precipitation indicated in the fig)? It seems to me that this would make comparison easier, and it would allow for a single bar (combining the three components) for each experiment (and each site).

Thank you for this helpful suggestion. We agree that expressing the water-balance components relative to total precipitation facilitates comparison among experiments and sites. We have therefore revised Fig. 6 to show evapotranspiration, surface runoff, and subsurface runoff as percentages of annual precipitation. The three components are now combined into a single stacked bar for each experiment at each site, and the mean annual precipitation amount is indicated above each panel. The figure caption and corresponding discussion in Section 4.3 have been updated accordingly.



- The fact that this section is based solely on a sensitivity analysis comparing the ORCHIDEE experiments, without any objective evaluation, remains a significant limitation of this study.

We agree that the absence of an objective evaluation of the simulated hydrological fluxes is an important limitation of this study. The Urban-PLUMBER dataset provides latent heat flux observations, allowing the evapotranspiration response to be evaluated in the preceding section, but it does not provide runoff, drainage, or soil-moisture observations at the investigated sites. Consequently, the analysis presented in this section should be interpreted as a sensitivity assessment of how the different ORCHIDEE configurations partition the water balance, rather than as a validation of the simulated runoff and drainage.

This observational limitation is not specific to our study. Jongen et al. (2024), in their evaluation of the water balance in 19 Urban-PLUMBER models, were also limited by the unavailability of these observations, and where still able to show large inter-model spread in runoff, showing the values produced by ORCHIDEE and its different configuration allow us to place our results in this spread.

We have revised the manuscript to state this limitation more explicitly and to avoid interpreting the simulated changes as evidence of hydrological performance. An objective evaluation would require spatially distributed simulations over instrumented urban catchments, including river routing, followed by comparison with observed streamflow. Producing such simulations also requires appropriate high-resolution meteorological forcing (because urban catchments are of small size), which remains challenging and is beyond the scope of the present study, but is a priority for future work.

In the introduction:

“Despite these advances, recent intercomparison work shows that urban water-balance representation remains uncertain in ULSMs, especially regarding water-balance closure, storage dynamics, and runoff parameterization (Jongen et al., 2024). Progress is also constrained by observations, because urban sites providing turbulent fluxes do not provide all water-balance terms needed to evaluate runoff, irrigation, drainage, and storage changes consistently at the same spatial scale (Jongen et al., 2024).”

In the 4.3 Annual water budget:

“Because runoff and subsurface-runoff observations are unavailable at these sites, this analysis should be interpreted as a sensitivity assessment rather than a validation of hydrological performance. Evaluating the simulated runoff would require spatially distributed simulations over instrumented urban catchments and comparison with observed streamflow, which is beyond the scope of this study.”

- Couldn't a surface water retention basin be easily defined in ORCHIDEE “urbanized”?

A surface water retention reservoir could, in principle, be introduced into the urban version of ORCHIDEE. However, its implementation would require additional assumptions regarding, for example, storage capacity, the fraction of impervious area connected to the reservoir, infiltration and evaporation losses, and the rate at which stored water is released. The available observations at the Urban-PLUMBER sites do not allow these parameters to be constrained or the resulting model behaviour to be evaluated. For this first urban implementation in ORCHIDEE, we therefore chose to remain parsimonious and to focus on representing the effects of imperviousness on infiltration, runoff, and drainage. A surface water retention reservoir could be considered in future developments if catchment-scale evaluations indicate that such a process is necessary to reproduce observed urban hydrological behavior.

We added a mention of rainfall interception by urban structures to the conclusion.

“Additional developments, such as spatially and temporally varying urban parameters (e.g., from WUDAPT), interception of rainfall on urban structures, refinement of thermal properties, and the inclusion of anthropogenic heat fluxes, will further enhance the performances of this scheme.”

Discussion

- The points raised in the “Discussion” section are interesting but should be more clearly contextualized in relation to the scientific literature and the approaches used in other climate models applied to similar applications (as it is done for thermal properties).

We thank the reviewer for pointing this out. We added to the manuscript:

«Despite the implemented urban developments, the simulations exhibit persistent seasonal biases. Most notably, during summer, urban simulations show an overestimation of daily sensible heat flux and an underestimated latent heat flux, indicating that too much of the available energy is partitioned into sensible rather than latent heat. This high-sensible-heat, low-latent-heat bias is a recognized limitation of urban land-surface models and has been associated with the omission or insufficient representation of urban vegetation and water availability (Best and Grimmond, 2015). Indeed, latent heat flux remains underestimated by many current urban models (Lipson et al., 2024). In winter, the daytime bias reverses, with an underestimated sensible heat and an overestimated latent heat. The absence of anthropogenic heat in the current scheme may contribute to the winter underestimation of sensible heat, as its representation has been identified as important for simulating wintertime urban energy budgets (Hertwig et al., 2020; Jin et al., 2021; Karsisto et al., 2016). Future work will test whether prescribing anthropogenic heat fluxes from observation-based gridded datasets can reduce this bias in kilometer-scale simulations. On the contrary, the winter overestimated latent heat is not consistently reported across other models (Lipson et al., 2024) and may therefore reflect ORCHIDEE-specific seasonal controls on evaporation and water availability. At night, sensible heat also remains underestimated in winter, while latent heat exhibits smaller positive biases in both seasons. The bias pattern displayed by this new urban representation therefore suggests that evaporation is insufficient in summer, when urban vegetation and irrigation are most active, but insufficiently constrained in winter and at night. A potential development for ORCHIDEE would therefore be to embed an active vegetated fraction, including seasonal phenology, directly within the urban PFT. This would allow urban vegetation, and potentially irrigation, to enhance transpiration during the summer while limiting vegetation-related evaporation during winter.»

And:

«A more refined representation of urban hydrology could also help address the above biases. Urban land surface models use a range of runoff strategies, from routing a fixed fraction of rainfall directly to runoff to more process-based formulations accounting for infiltration capacity, soil saturation, surface-water storage, and transfers between pervious and impervious tiles. The simpler formulations may be insensitive to rainfall intensity, antecedent soil moisture, and site characteristics (Jongen et al., 2024). The

present study introduces a novel physically based representation of imperviousness by linking it to saturated hydraulic conductivity, thereby affecting infiltration, subsurface water transfer, and runoff generation. This development directly addresses the weak sensitivity of runoff to imperviousness found in 7 of the 18 models evaluated by Jongen et al. (2026). Nevertheless, Jongen et al. (2026) also found that 10 models omitted at least one major runoff-generation mechanism. Future developments should therefore represent surface ponding and interception storage, irrigation, drainage connectivity, and water transfers between urban and vegetated tiles. These processes should ultimately be evaluated against hydrological observations, because the present sensitivity analysis alone cannot establish whether the simulated water partitioning is realistic. Importantly, advancing the hydrological realism of the model should not be guided by latent heat performance alone, but by a comprehensive evaluation of the full water balance (Jongen et al., 2024). However, such evaluation is currently limited by a lack of observational data for key hydrological variables (e.g., runoff and soil moisture) at the urban flux tower sites used in this study (Jongen et al., 2026, 2024). This underscores the need for coordinated efforts to collect hydrological data in urban environments to support model development and validation.»

P25, L552-557

"Our current thermal conductivity value, derived from the NOAH land surface model (He et al., 2023) and relatively high compared to values used in CLM (Lawrence et al., 2019) for instance, may require further refinement. Our current thermal conductivity value ($3.24 \text{ W m}^{-1} \text{ K}^{-1}$), taken from the NOAH land-surface model, is substantially higher than values commonly used in CLM for example, $0.767 \text{ W m}^{-1} \text{ K}^{-1}$ in its original urban implementation (Loridan and Grimmond, 2012) or $1.55 \text{ W m}^{-1} \text{ K}^{-1}$ in the later SURY configuration (Wouters et al., 2016), and may therefore require further refinement."

There's a problem with repeated sentences here. You could simplify: "Our current thermal conductivity value ($3.24 \text{ W m}^{-1} \text{ K}^{-1}$), taken from the NOAH land-surface model (He et al., 2023), is substantially higher than values commonly used in CLM for example, $0.767 \text{ W m}^{-1} \text{ K}^{-1}$ in its original urban implementation (Loridan and Grimmond, 2012) or $1.55 \text{ W m}^{-1} \text{ K}^{-1}$ in the later SURY configuration (Wouters et al., 2016), and may therefore require further refinement."

Done, thank you.