> Dear authors,
>
> I agree with the reviewer of this second round of peer review, who also reviewed the original version, that the quality of the manuscript has clearly improved, but that there are still a number of issues raised by the reviewer that need to be addressed before I can definitively recommend your work for publication. The reviewer is still not fully satisfied with some of your responses and how you have addressed the reviewer's criticisms in the revised version. Your argumentation could be improved and made more stringent, while the length of the text could be reduced by removing sentences that have no real meaning. See below for specific points. Kind regards,
> Nicolas Brüggemann
>
> Associate Editor

Dear editor Nicolas Brüggemann

Thank you for your feedback and for the opportunity to revise our manuscript! We appreciate the acknowledgment of the improvements made so far. We agree with the reviewer's new comments and have addressed them thoroughly in the revised paper. See below the detailed responses to each comment. We have further strengthened our argumentation and ensured our revisions directly address the criticisms. During this process, we also deleted sentences with low value to enhance clarity and shorten the discussion. In addition, the second sentence of 4.1 has been relocated to the beginning of the section, as this proved to be a more logical arrangement, and the final sentence of the first paragraph in 4.3 has undergone minor revisions.

We are committed to refining the manuscript as needed and look forward to any additional guidance.

Sincerely, on behalf of all authors,

Laura Thölix

> Reviewer's comments
>
> L310ff: In contrast to the description and the nice figure 2 in response to my previous comments, the changes made in the text are a bit irritating. Besides from wording issues (‘The details of … are described in detail’, ‘Furthermore’ instead of ‘Therefore’, …) the explanation is unclear. How can you transform data (sap flow records) into ‘estimates of transpiration’ (referring to which area?) by multiplying them with ‘values for sapwood’

> that seem to come from literature (but the references are not given in the literature list)? Revise, and possibly provide the resulting figures in the supplementary.

Thank you for your insightful comments! We agree with your observations and have improved the wording in the relevant section to address the issues mentioned. We've also clarified our method for calculating transpiration and added the missing references to the reference list. In the main analysis we compare the seasonal dynamics of sap flow to modeled transpiration. In addition, we also present a comparison of transpiration per tree where the observed sap flow rate is converted into transpiration. The description of the methods was unclear, especially regarding the conversion into transpiration. The revised methods description reads as follows,

> *First, we compared the sap flow rates and the model estimates of transpiration to analyse seasonal dynamics. Second, the sap flow rates were transformed into estimates of whole tree transpiration by multiplying the rates with species-specific values for sapwood area from the literature, namely 349.7 cm^2 for the birches (Zapater et al., 2013) and 433.8 cm^2 for park trees (Leuzinger et al., 2010).*

The comparison figure of sap flow driven and model estimates of whole-tree transpiration (Figure S5) is referred to in the first paragraph of subsection 3.3.

> L526ff: Also, although better, the discussion of possible 'disagreements' (why not deviations?) is still not very convincing. In particular the impact of a variable trunk water storage should be much too small to be significant. That there is interception is also not a reason for a deviation, except it is explained why the models might do it wrong. Instead, I would mention that the soil water supply in the models is likely different to what real trees experience and perhaps already here mention that the phenology (absolute LAI, temporal and spatial distribution) might be critical to fit measurements to models.

We agree that deviation is more suitable word than disagreement and changed that in the revised MS. We also totally removed the speculation on the trunk water storage and the sentence regarding interception. Instead, we added the suggested topic on model testing and development needs. Now the revised discussion regarding the soil moisture deviation reads, as follows

> *The models had some difficulty in reproducing the temporal dynamics of soil moisture, e.g., the increase following periods of droughts in the urban forest was overestimated in the models. Several factors could lead to discrepancies between model results and observations. Measuring soil moisture is inherently challenging*

(Tarantino et al., 2008; Rasheed et al., 2022). Additionally, the soil water supply as represented in models is likely to differ from actual conditions. This difference can be attributed to factors such as root depth and soil texture, but also to the phenology. In addition, local variations in the precipitation may not be correctly captured in the forcing data. Some part of the precipitation is lost as runoff (Ilvesniemi et al., 2010) or transported to deeper soil layers through preferential flow pathways, while the soil is assumed to be homogeneous in the models. Addressing these issues is essential to improve the accuracy of soil moisture simulations in future model developments.

L576ff: The additional explanation about uncertainties in the gas exchange of lawns is certainly important. However, wording is not very good and argumentation is inconsistent. For example, if JSBACH has shown to reasonably estimate (gas exchange? Temperature? Water content?) dynamics of lawns, why is it still a problem that needs to be discussed here? In addition, what is a 'soil media' and why does it influence the organic matter? Do you mean that your initialization of soil properties which are important for flux estimations is uncertain? The last sentence in this paragraph ('Therefore, …') is superfluous and only sounds as an excuse for not having an idea where the deviations are coming from.

Thank you for your notes! Upon reviewing the text, we noticed several structural issues and ambiguities ourselves too. First, we have clarified the explanation regarding the model evaluations made in the previous study. Second, we aimed to convey that there are no specific soil properties, such as carbon content, in urban lawns and it is not thus meaningful to generalize lawn TER based on sparse observations. These soils are not in equilibrium with the primary production of the vegetation as something else may have grown there previously, or the soil may have been relocated. Additionally, commercial substrates (~soil media) with high organic matter and nutrient content are often added when establishing lawns affecting the observed TER. Now, we revised the pointed section to read, as follows

In our study, we evaluated the dynamics of GPP on lawns but not the TER, as the momentary measurement of TER are difficult to scale up to a daily level. This is because in open areas such as lawns, changes in radiation can cause significant changes in soil temperature leading up to changes also in TER. This naturally causes uncertainty in the estimated TER and NEE but Trémeau et al. (2024) showed that JSBACH can estimate the seasonal dynamics and absolute level of TER in irrigated and non-irrigated lawns in Helsinki. However, heterotrophic respiration at different lawns depends on the quality and quantity of organic matter in the soil, which in turn

depends on the history of the soil and possible earlier soil amendments such as mulch. Therefore, soil properties and heterotrophic respiration may vary spatially in urban areas without a clear link to vegetation types, as the carbon cycle is rarely in a steady state yet. Without case-specific information on soil carbon pools, model initialisation will be uncertain.

We removed the criticized, last sentence as it clearly added no information to the topic.

> L635ff: It is true that forest soils show stronger response to soil warming than grasslands (Rustad et al. 2001). Probably because of their higher organic matter content. However, you need to consider that higher temperatures will heat up grassland soils much more than well shaded forests. Therefore, the same air temperature rise will probably affect forest soils less than grassland soils. And although you are correct that Meineke et al. indeed found a reduction of aboveground carbon storage under urban warming, this doesn't seem to apply for soils as pointed out by Rustad. Therefore, argumentation need to be better and more carefully formulated. Apart from that, before demanding a more detailed analysis of flux components, authors should specify the uncertainties of the current measurements and perhaps reflect on some quite comprehensive papers on this topic (e.g. Ryan 2023, Zahn et al. 2022, Ueyama et al. 2026).

The pointed section was indeed quite weak and clumsy in places. Rustad et al. has been partly misquoted in relation to grasslands, probably because several authors have revised the text and the quote has moved out of its original context. The whole paragraph has been thoroughly improved with additional references and strengthened arguments. Furthermore, a short discussion on EC general reliability, gap-filling, partitioning, and related issues has been added at the end. During the revision, we also removed some sentences that have no real value. Now it reads as follows,

Increased temperatures are considered to increase soil respiration (Rustad et al., 2000), but a local study showed that increasing soil temperature had less effect than irrigation on heterotrophic respiration in urban tree-covered environments (Karvinen et al., 2024), highlighting the important role of soil moisture also in the north. In subtropical climates warming can also reduce urban tree growth and carbon sequestration (Meineke et al., 2016) but previous article in the hemiboreal study city indicated that temperatures that can locally be considered as extremely high seemed to favour tree photosynthesis (Ahongshangbam et al., 2023).

As NEE is a sum of the input and output, a detailed analysis of the flux components would improve our understanding of the role of different weather years and the effect

of increasing temperatures and increasing possibility of extended drought in urban vegetation in northern cities. However, gapfilling of eddy covariance C flux data always requires caution (Vekuri et al., 2023; Zahn et al., 2022), and since fluxes in urban areas include diurnal and otherwise varying amounts of anthropogenic sources (Järvi et al. 2012; Ueyama and Ando, 2016), flux partitioning is challenging. It should also be noted that EC measurements may underestimate some of the component fluxes (Ryan, 2023).

L660: Be careful with your formulation. Phenology might be important to consider in models but not as an initialized value but as a dynamic process that considers the variation of budburst date with spring temperature rise as well as a possible early leaf senescence with increasing drought events. In this context, you might also point out that the drought stress component is not well covered in the model and needs to be better considered (even if you would like to make another paper from this problem).

We agree and changed the formulation from initialization to testing. We also pointed out that the drought issue requires further testing. Now the section reads as follows,

Therefore, it is recommended that the phenological patterns are tested before further use of the models in other cities. Although the drought response in the models appeared reasonable in terms of GPP, the precise description of thresholds and responses to different drought intensities should be further tested, especially to serve future scenarios.

L659: I would like to remark that it would be more specific to talk about different scales between measurements and models, rather than limited observations, that make an evaluation difficult.

We agree. We revised the sentence in the conclusion to be as follows,

However, evaluating absolute levels of net ecosystem exchange in mature trees is hindered by the different scale of observations, which usually focus on single sunlit leaves.

The second sentence of the discussions section was revised, as follows,

However, the measurements were not always realistically represented, but on the other hand, comprehensive measurements at tree and ecosystem level are difficult to collect and are therefore at a different scale to the models.

The issue in the abstract was revised, as follows,

> *However, the validation of absolute level of modelled fluxes proved difficult due to differences in the scale of the observations, particularly for mature trees, and the fact that net ecosystem exchange measurements in urban areas include some anthropogenic emissions*.

References mentioned by the reviewer

Ryan, M. G.: The enduring mystery of differences between eddy covariance and biometric measurements for ecosystem respiration and net carbon storage in forests, New Phytol., 239, 2060-2063, https//doi.org/10.1111/nph.19105, 2023.

Rustad, L., Campbell, J., Marion, G., Norby, R., Mitchell, M., Hartley, A., Cornelissen, J., and Gurevitch, J.: A meta-analysis of the response of soil respiration, net nitrogen mineralization, and aboveground plant growth to experimental ecosystem warming, Oecologia, 126, 543-562, https//doi.org/10.1007/s004420000544, 2001.

Ueyama, M., and Ando, T.: Diurnal, weekly, seasonal, and spatial variabilities in carbon dioxide flux in different urban landscapes in Sakai, Japan, Atmos. Chem. Phys., 16, 14727-14740, https//doi.org/10.5194/acp-16-14727-2016, 2016.

Zahn, E., Bou-Zeid, E., Good, S. P., Katul, G. G., Thomas, C. K., Ghannam, K., Smith, J. A., Chamecki, M., Dias, N. L., Fuentes, J. D., Alfieri, J. G., Kwon, H., Caylor, K. K., Gao, Z., Soderberg, K., Bambach, N. E., Hipps, L. E., Prueger, J. H., and Kustas, W. P.: Direct partitioning of eddy-covariance water and carbon dioxide fluxes into ground and plant components, Agric. Forest Meteorol., 315, 108790, https//doi.org/10.1016/j.agrformet.2021.108790, 2022.