Response to Referee #1

We thank both referees for their efforts and constructive comments. Each referee's comments are shown below in ***italics***, followed by our point-by-point responses in **blue** and relevant text in **red**.

Referee #1 Youngryel Ryu (Referee) ryuyr77@gmail.com

Dear Dien and all

I would like to congratulate you for this impressive manuscript. Incredibly comprehensive, in-depth analysis, great attentions to details and robust upscaling approach. I have to admit that I am not the expert in atmospheric transport model, where I didn't make any comment.

The authors developed a biogenic CO_2 balance model which includes GPP, R_{eco}, and NEE. They intended to develop this model for global cities, but actually it is applicable to the global land. The basic idea came from linking SIF and GPP. They developed the slopes between GPP and SIF (CSIF products) across FLUXNET sites. After fine tuning (e.g. crops) in the slopes, they converted CSIF (0.05 degree) to GPP. As urban landscape is heterogeneous, they used very high resolution land cover maps to apply the slopes for the relevant land cover types then aggregated to 0.05 degree. Then the authors developed an R_{eco} model using NN with GPP, Tair and Tsoil. To evaluate the model performance, the authors compiled FLUXNET, INFLUX dataset and urbanVPRM model. Then the authors combined fossil fuel emissions data, XCO_2 data and an atmospheric transport model to tease out the contributions of biogenic CO_2 fluxes in urban CO2 fluxes around the world.

The scope of this manuscript is vast but the authors didn't gloss over important details. Although some parts could be improved further, overall I see this is already too good. Though I would like to make some suggestions for further improvement.

We truly appreciate the recognition and constructive comments from referee #1 Youngryel Ryu and have tried our best to conduct additional analyses and to improve the presentation of this manuscript.

First, evaluate SMUrF NEE directly against FLUXNET data like what you did for GPP and R_{eco} in Fig 5. Good performance in GPP and R_{eco} does not necessarily indicate good performance in NEE which is tiny signal compared to the other two fluxes. The authors reported Fig S10 for NEE evaluation, but I feel it is not enough. It is fine to report rather poor performance in NEE, which is quite well expected as machine learning based NEE (e.g. FLUXCOM) performed poorer than GPP and R_{eco}. It would be an useful point about how to improve SMUrF later.

We agree that NEE evaluation against observations is critical. In the initial manuscript, we have already included a model evaluation of the HOURLY mean biome-specific NEE fluxes against FLUXNET data (Fig. 9 and Sect. 3.2.1). In accordance with the reviewer's point, NEE has a poorer performance compared to GPP and R_{eco}, especially for biomes with less training data (i.e., FLUXNET observations).

Second, the current evaluation focused on diurnal to seasonal scales. Could you provide some discussion on the model performance in interannual to trends? e.g. in case of LA, how NEE varied across dry and wet years? How does NEE/fossil fuel CO_2 varies across dry and wet years?

While the paper would be more informative with additional discussion on the interannual variation/trend in urban NEE, the primary motivation of developing the model is to separate the anthropogenic and biogenic CO_2 signals when interpreting the atmospheric measurements. The column CO_2 retrieved from satellites may be affected by the upstream fluxes over timescales of only O(day). Therefore, we argue that CO_2 fluxes at the diurnal scale matters the most to the downwind atmospheric CO_2 observations than CO_2 fluxes at much longer timescales. For example, the interannual variation/trend in NEE may be a secondary-order effect, superimposed on the urban-rural difference in NEE fluxes that influence the CO_2 background determination. Finally, the paper is already quite lengthy, and considering the main motivation of the manuscript, we are inclined not to touch on NEE fluctuation at the moment, but added relevant text in the future work section for clarification (Sect. 4.2) with changes highlighted in red:

> "Because atmospheric CO_2 concentrations measured from satellites are mainly influenced by the anthropogenic and biogenic CO_2 fluxes a few hours to days ahead of the overpass time, this work focused on presenting and evaluating the diurnal and seasonal CO_2 fluxes. Biogenic CO_2 fluxes over longer timescales, e.g., their interannual variations and long-term trends, may require further investigations. We also hope to examine more cities and different times of the day in future studies to better study the relative biogenic and anthropogenic contributions to XCO_2 anomalies. And, incorporating uncertainties in biogenic fluxes and resultant $XCO_{2.bio}$ is needed for future studies with aims of understanding urban signals especially over the growing seasons."

Third, I would like to recommend adding some discussions for including evaporation in SMUrF, not now but in v2. Your model already has most important components to compute evaporation. One approach would be to use Ball-Berry model to link your GPP, canopy conductance and finally evaporation. I really enjoyed this paper (https://doi.org/10.1073/pnas.2005253117), which stressed the important linkage between irrigation and biogenic CO_2 fluxes in LA. I think SMUrF can track this as well once evaporation module is included.

We thank the reviewer for the suggestion and agree that an evaporation module is a great function to be implemented in the future. Urban irrigation plays an important role in GPP especially for semiarid urban areas, (e.g., Loridan et al., 2010, Johnson and Belitz 2012, Vahmani and Hogue 2014, Miller et al., 2021). We added the following discussions in Sect. 4.2:

> "Since carbon fluxes are closely tied to the water cycle, anthropogenic moisture input (i.e., urban irrigation) can effectively influence the urban biogenic fluxes, particularly over arid and semi-arid residential areas like Salt Lake City and Los Angeles (Loridan et al., 2010, Johnson and Belitz 2012, Litvak et al., 2017, Miller et al., 2021). Although we currently rely on SIF to pick up potential irrigation effect on urban GPP, it is possible and informative to further relate the water flux exchange to the carbon flux exchange in the future version of SMUrF."

Followings include minor comments:

P4 L5-10: The previous paragraph criticized the limitation of simple R_{eco} model, then this paragraph explained ML for SIF and land surface fluxes. I feel somewhat disconnected from the previous paragraph.

We thank the reviewer for pointing out the disconnection. Our initial thought is to highlight the difficulties and limits in estimating R_{eco}, which is also the motivation of adopting ML technique for R_{eco} estimates. We have now improved the flow of the two paragraphs (with modified text in red):

> "...After all, the complexity of biological and non-biological processes of R_{eco} and the lack of mechanistic understanding of how biotic and abiotic factors affect R_{eco} make the mechanistic modeling quite challenging. Given the complexity in modeling R_{eco}, machine learning (ML) techniques will be adopted in this study.
>
> ML approaches have been increasingly applied in many disciplines including ecosystem modeling to help answer complicated, entangled problems through extracting patterns from data streams for predictions and generalizations..."

P9: pure temperature -> revise

We have now changed "pure temperatures and GPP observations" to "direct temperatures and GP observations".

P10 L30: I feel the assumption for no correlation between GPP and R_{eco} is overly simplified. SMUrF model structure indicates GPP is a forcing to R_{eco} (P6 L16).

We agree that the assumption for no correlation is overly simplified and have added a correlation term when calculating the uncertainties in NEE in the code.

P13 L6: What's GEE? Isn't it GPP?

GEE is the gross ecosystem exchange and closely related to GPP. GEE is often used by researchers working with eddy covariance observations, while GPP is used more often by ecologists. We define GEE = -GPP, so that photosynthetic uptake represents removal from the atmosphere (Fig. 6).

P18 L12: what is QF?

QF stands for quality flag, a measure provided by the OCO-2 XCO_2 retrieval to indicate data quality. We have now modified that sentence.

P20 L3: spatial SIF -> revise

We have now changed "spatial SIF over cities" to "SIF retrieval with a broader spatial coverage over cities".

P20 L10-22: It is worth discussing complex SIF-GPP relationships reported in recent literature. Consistent, linear relationship disappears in some cases. https://doi.org/10.1016/j.rse.2018.07.008 https://doi.org/10.1002/2017JG004180 https://doi.org/10.1038/s41598-018-32602-z

We thank the reviewer for sharing these relevant publications and have added some text on the nonlinearity of GPP-SIF relationship in the discussion section (Sect. 4.2):

> "Second, GPP within SMUrF is currently estimated as a linear function of SIF, using a set of constant biome-specific linear slopes (α) without considering temporal or inter-site variations. The adoption of SIF has dramatically benefited and simplified the GPP calculation, as no extra satellite indices or impervious fractions need to be plugged in. However, previous research based on ground-based SIF measurements (Miao et al., 2018; Wohlfahrt et al., 2018; Yang et al., 2018) revealed the GPP-SIF relation deviated from linearity at the sub-diurnal scale, under unstable light conditions, or heat stress. While SIF and absorbed PAR are linearly related, the GPP-SIF relationship can deviate from linearity due to complex LUE:SIF yield relationships in light-saturating vs. light-limiting regimes (Miao et al., 2018). Thus, considering additional environmental factors related to the modeling of light use efficiency—e.g., relative humidity, cloudiness, and growth stage of crops, could improve the SIF-based GPP estimates (Yang et al., 2018). Although the nonlinear GPP-SIF relationship was not explicitly accounted for in this first iteration of SMUrF, our estimated flux uncertainties against dozens of flux tower sites implicitly account for the overall potential error associated with the linear assumption. Nevertheless, we anticipate future efforts to add more degree of freedoms in the estimate of GPP-SIF relation."

Again, this is a great manuscript. I really enjoyed reading it, and also learned a lot.
Thanks- Youngryel

Thank you, Youngryel. – Dien on behalf of the team