The efforts made by the authors to address my (too) numerous comments and to improve their manuscript are greatly appreciated, as well as their detailed responses. I find that the paper much improved over the original version. However, I still believe that additional revisions are necessary for it to become more impactful.

Firstly, the paper is still overly long (31 pages without the references and Supplementary Materials). It would benefit from a clearer and more concise organization to better highlight its key developments and findings.
Secondly, certain aspects of the model parameterization and data processing remain unclear, despite their potential impact on the interpretation of the results (while the informations may be included in the paper, they are obscured by the paper's presentation).

Structure of the paper

- Including "QUINCY terrestrial biosphere model" and "simulations" in the title would better define the study's focus and increase the visibility of the authors' model.

- In several places, topically connected sections are disjointed, which makes the reading process more difficult. For instance:
    - l137 and l150 related to GPP and gap-filling
    - l177 and l319: regarding the QUINCY PFT at the three sites
    - l210, l240, l320 and l327: informations on soil optical properties
    - l216-220 and §2.5.3 on the use of the fesc parameterization to calculate SIF with one of the modelling approach
    - in §3.3, l434: sentence related to Figure 7 when the analysis of the figure is provided later…

- The Introduction and Discussion sections can be improved and condensed. The description of the study objectives (from line 93) needs clarifications. QUINCY should be introduced in line 93 already; Given that the authors have mostly followed the approach of Raczka et al. (2019) to represent NPQs, is "how to accounting for it" truly a research question?

- Are all (numerous) figures and tables really necessary? Figure 1 could be removed (or moved to Supplementary Materials). Informations of some tables could be added in the corresponding figures or in the text. For instance the informations on computational efficiency provided in Table 4 could be added directly in the text in §3.1.3.

Clarifications

- Several definitions of the red and far-red spectral ranges co-exist, depending on the measurement / observation characteristics: 680-686 nm vs 687 nm for the red region (American sites vs Finish site), 745-758 nm vs 760 nm (in situ data) vs 740 nm (TROPOSIF) for the far red. It is still not clear whether these distinct spectral samplings are considered or not for the model simulations and comparisons with respect to the observation data.
All the more the presentation of the radiative transfer calculations in QUINCY (§2.4.1, §2.5.3) mentions two large spectral ranges (300-700 nm for the visible and 700-3000 nm for the near-infrared parts). What spectral ranges were considered to prescribe leaf and soil optical properties for simulating SIF in the two narrow spectral windows? Averaged values over 300-700 nm / 700-3000 nm? If so, does it have an impact on the consistency of the model–data comparison?

- There is a misunderstanding regarding the wavelength at which TROPOSIF SIF data are provided with the fitting windows from which they are derived. The authors refer to the product user manual (p11 – description of the variables) to justify the use of "743 nm". The indication of 743 in the variable name is to distinguish the SIF estimates from the two fitting windows. Whatever the fitting window, SIF was scaled at 740 nm (see §3.1 of the same document and the "long_name" attributes of the variables p.12, as well as Guanter et al. (2019)). For the model-data comparison, simulations at 740 nm (instead of 743 nm) would be more relevant (although the impact is likely small).
As I understood, the authors consider different temporal samplings when comparing QUINCY SIF simulations with TROPOSIF data: mid-day values for the simulations and daily averages for TROPOSIF. If so, these different sampling characteristics likely contributes to the observed mismatch in the amplitude of SIF time series between model simulations and observations (with higher SIF values around noon for the simulations). For a more coherent comparison, the authors should consider 1) daily averages both for simulations and observations, and 2) the "daily corrected" SIF estimates from TROPOSIF (SIF_Corr_743 variable) to limit the impact of uneven temporal distribution of observations within a given day.

- The informations on leaf clumping and leaf angle distribution help better understanding the modelling approach. However, "hemispheric" leaf angle distribution is not a common term (no reference is provided); Do the author mean "spherical" distribution?
The range of variation of the clumping index (Figure S1) could be discussed in the Discussion section; is it consistent with values found in the litterature for this type of vegetation?

- Isn't the radiative transfer scheme of mSCOPE also based on a two-stream approach?

- With an RMSE of 0.77 units wrt GPP simulations using two QUINCY versions (with / without mSCOPE), the authors conclude that the simulations are similar (l254). Table 2 indicates an

RMSE between simulations and observations of about 0.88 units at a daily scale for the same site. Do I understand correctly that the modelling error is similar to the model data mismatch?

- Quantifying the different components of the error budget between model simulations and observations significantly improved the interpretation of the results. However for SIF, as the main outcome is model overestimation, the slope of the regression line may be the most relevant parameter for this diagnosis. The values could be provided in Table 3.

- Is there a specific reason why NPQs was not considered at FI-Sod? Its accounting would likely be beneficial for some of the analyses, in particular in §3.4.

- The computational efficiency of the radiative transfer models is one of the evaluation criterion. The authors identified limitations in the use of mSCOPE for large scale simulations (§4.2). Later, they discussed the limitations of 1D RT schemes and identified the accounting of the 3D structure of the canopy within QUINCY and RT schemes as a way to improve the realism of the simulations. Wouldn't this in turn increase the computational cost?

<u>Figures and tables</u>

- The diagrams (Figure S2 to Figure S4) really help understanding the different implementations of SIF calculations within SCOPE. However, it is regrettable that the quality of these figures does not match that of the other figures in the article.

- Several of the figures are likely not colorblind friendly (the figures mentioned above as well as Figure 6 or Figure 12 for instance).

- On Figure 3, with the color chosen, mSCOPE and L2SM simulations are hard to distinguish.

- Please homogeneise the notation for "red / far-red" (Figure 3 and others) or " R/FR" (Figure 4).

- Figure S13: It is not clear if the distinct y-axis for the simulations apply to all three sites or not. Why do the simulations at CA-Obs do not cover the same air temperature range as the observations?

- Figure 8 and Figure S16: The months on the x-axis should be provided.

- Figure S12: Different symbols (or colors) should be used for the results with/without NPQs.

- Table 2: I understood that the "upscaled" approach was not considered anymore…

- Table S1: What is the point of indicating "r.u." if no value is provided?

- Is Table S5 really necessary? R^2 and RMSE could be provided in Figure 6 and I do not see what the values of a and b add.

Other comments:

- l15: What does "seasonal development" mean?

- l21: What are those "numerous applications"?

- l41: "The variability in radiative transfer through the canopy": What does that mean?

- l55: The sentence remains unclear.

- l63: Data assimilation studies have been conducted with other models than BEPS (some of the related studies are already cited in the paper).

- l108-l109: What are those satellite missions?

- Title of §2.1: The observations are actually described in §2.2.

- l117: Please correct the link for Ameriflux.

- l134: Do the authors mean "over all three scans" (instead of "over all the observations")?

- l164: Why not present the surface extent evenly for FI-Sod, given the site is considered to assess the impact of spatial averaging (Figure 7)? The presentation of the pixel composition in terms of biomes better fits in the Discussion.

- l198: "ten layers" was already presented in l178.

- l210: "constrained to values larger than 10°": What does that mean?

- l215: "...was caused by the different Farquhard et al. model formulation...". Not clear. Does this relate to something described earlier in the manuscript?

- Section 2.6: Most of the detailed informations could be moved into Supplementary Materials.

- l325: Is the fact that no spinup is needed related to the fact that LAI and leaf N content are prescribed?

- l337-l339: Is that sentence really needed in this "evaluation methodology" section?

- l354: The first sentence should be moved to the introduction of §3.

- Title of §3.1.3: "Performance comparison" instead?

- l420: Why not showing these APAR data on the figure?

- l501: "close to…": But above or below?

- l543: "leaf" optical properties?

- l567-l568: The reference to the figure is missing. This "linear relationship" at FI-Sod does not align well with the fitting curve chosen in Figure 6.

- l586: This uncertainty is consistent with the retrieval error already indicated in the TROPOSIF paper…

- l588: How may site heterogeneity influence the accuracy of TROPOSIF data? Do the authors mean that interpreting the comparison of different datasets at different scales has to be balanced depending on spatial heterogeneity?

- l601: Are the studied sites concerned by such water-stressed conditions?

- l653-l654: "misrepresentation of conifer leaves"… Do the authors refer to their structural description/organisation or their optical properties?

- l660: "The TROPOSIF product from satellite", the formulation is awkward.