In this paper, the authors parameterize a sustained component of Non-Photochemichal Quenching for boreal evergreen conifer forests within the QUINCY terrestrial biosphere model (TBM), in the view of improving the representation of leaf-level SIF emission. They implement and evaluate 3 modelling approaches to upscale leaf-level SIF fluxes to the canopy scale. Top-of-canopy (TOC) SIF and GPP simulations are compared at three sites against corresponding *in situ* estimates, as well as, for SIF, coarse spatial resolution space-borne SIF retrievals from TROPOMI at two sites.

Because the utilization of space-borne SIF retrievals for improving the parameterisation of TBMs remains challenging due to the large spatial and temporal scales involved, confronting TBM SIF simulations with *in situ* data appears promising for advancing the realism of process-based representations of how light is partitioned between photosynthesis and energy dissipation (fluorescence and NPQ). Although the study addresses current scientific topics on how SIF data can be used to improve our understanding of the dynamics of plant photosynthesis, I feel that the methodology exhibits shortcomings and inaccuracies which question the significance of the outcomes. While the development of SIF observation operators in other TBMs is typically motivated by the goal of improving their representation of GPP dynamics (through the optimization of specific model parameter values using data assimilation), the limited contextualization provided in this study gives the impression that the implementations around SIF simulation conducted here in QUINCY are more of an academic exercise, disconnected from practical applications while the perspective of data assimilation is raised.

In addition, the structure of the manuscript needs to be improved, in particular regarding the presentation of the different couplings between QUINCY and each radiative transfer (RT) approach. In the current version, the flow of information between the variables simulated by QUINCY and those used by each RT is very unclear; what is common across RTs or what is RT-specific should be more clearly presented. The implementation of a sustained component for NPQ, while relevant for the studied ecosystem is questionable and suffers from an insufficiently rigorous calibration.
At the end, I am left with the impression the take-home messages for the TBM community are difficult to grasp.

Although fit for Biogeosciences, major revisions are required before publication. Below is a list of concerns and comments that support this view.

Major concerns

1. The presentation of the coupling between QUINCY and the different modelling approaches is very confusing. The list of variables and parameters of QUINCY that are used by each RT is very not clear. In particular, how are the leaf optical properties (reflectance, transmittance, absorptance) calculated? Is it performed by QUINCY in all cases? How does the leaf nitrogen content translates to leaf chlorophyll content, and how does it impact leaf optical properties, and hence APAR and photosynthetic activity? Do leaf optical properties vary with time and with the canopy depth for all RTs? What is the leaf orientation considered in mSCOPE, and is it consistent with other RTs and with

the one of QUINCY? The spectral dependency of the leaf optical properties and SIF needs to be better explained.
In short, what is common and what is specific to each RT should be presented more clearly. Including a schematic would be beneficial.
Is the 4th modelling approach (§ 2.5.4) really useful? It is acknowledged by the authors as likely unrealistic and the corresponding simulations are largely unexploited in the manuscript. A comparison with the previous implementation (Thum et al., 2017) would have been more relevant.

2. The rationales for the proposed implementation of the sustained component of NPQ for boreal evergreen coniferous forest are questionable. The authors follow the approach of Raczka et al. (2019) for the sustained NPQ model. However, Raczka et al. also developed in parallel a specific parametrisation for the reversible component of NPQ, showing significant impact of how reversible NPQ is represented. The authors use a model that has been calibrated on plant species poorly adapted to the boreal climate that is targeted in this study, and that is then combined to the sustained NPQ model over the whole of the time periods considered. I see a potential risk of partial redundancy in this approach.
The authors could also have considered the reversible model developed in Raczka et al., and assess the impact of their sustained implementation with that of the latter study.
In addition, the presentation of the calibration of the K_{Ns} parameters needs significant improvement.

3. In relation to the previous sentence, I am somewhat surprised that such an advanced TBM as QUINCY is not equipped with an even simple numerical optimization tool to calibrate some model parameter values in a rigorous manner. Only rough estimates of the values of the newly introduced model parameters are provided, without any indication of their associated error and quantification of the corresponding model-data fit. This is a bit problematic, all the more as incorrect quantification of the leaf-level SIF yield can partly contribute to the overestimation of TOC SIF that is observed whatever the RT modelling approach.

4. While the temporal dynamics of the simulations of SIF by QUINCY are consistent with *in situ* data (when the sustained component of NPQ is considered), they however show a large positive bias. This is however not the case for GPP nor APAR which agree fairly well with the *in situ* estimates. Those contrasting results seem to indicate that the SIF parameterisations are incorrect, but it remains unclear whether it is related to the equations implemented or inappropriate parameter values. As written above, the coupling between QUINCY canopy module (which simulates leaf-level SIF emission) and the RT schemes is not detailed enough to allow the reader understanding what is going on. The way canopy architecture is represented in QUINCY (leaf clumping and inclination, in addition to LAI) could partly explain those large biases (as acknowledged by the authors in the Discussion) but their is no clear information of those parameterisations (values of leaf clumping and leaf inclination that could be compared to literature values, etc.).
A comparative sensitivity analysis of SIF and GPP could have provided first order information on the key parameters requiring further calibration (like parameters strongly affecting SIF yet marginally impacting GPP).
The authors conclude that their work "pave the way for data assimilation using SIF observations". This has already been undertaken by other groups to correct GPP. Here, given that GPP is already in good agreement with *in situ* data, what could be the benefit of assimilating SIF observations?

5. The paper is a bit long. It contains several redundancies and repetitions that could be avoided. In addition, the chaining of some sections could be re-worked (see comments below) to improve the logical flow with the idea of making the manuscript more concise.

6. The referencing needs improvements.

Title:

7. The title does not reflect the developments that are specific to this study: It is very similar to a previous paper from the authors in which SIF modelling was implemented in QUINCY for the same ecosystems (Thum et al., 2017).

Abstract:

8. L2: "...to satellites" > "... to satellite observation footprints"?

9. L6: What does "alternative" mean in this context?

Introduction:

10. L21: Given the focus is on GPP and SIF, information on the GPP budget of these ecosystems would seem more relevant than biomass.

11. The structure is a bit odd with the introduction of SIF for monitoring boreal coniferous evergreen forests (L25) before more a general presentation of SIF (L41). Same remark for NPQ, a mechanism that is shared across ecosystems, that is presented after its sustained component, which seems to be specific to boreal evergreen conifer forests. The works of Porcar-Castell et al. (2008 and 2011) on sustained NPQ, prior to the work of Adams et al., should have been cited.

12. L37: The sentence about the belowground carbon may be interesting, but it breaks down the chaining between the mention of the impact of climate change and the possible use of TBMs to anticipate the future of ecosystems at high latitudes.

13. L44: Several studies prior to the one of Porcar-Castell et al. (2021) have already pointed out and use satellite SIF data to provide information on CO_2 uptake: In addition to the studies cited in the following sentences, the works of Guanter et al. (2012) and Joiner et al. (2011, 2013) using GOSAT and GOME-2 for instance.

14. There have been also numerous modelling studies that have highlighted the "complexity" of the SIF-GPP relationship depending on the scales considered. The paper of Parazoo et al. (2020) also highlighted model-data discrepancies at US-NR1.

15. L59: "...absorbed", as well as "scattered".

16. L60. Canopy structure also explains TOC SIF anisotropy (Middleton et al. 2015; Joiner et al., 2020; Malenovsky et al., 2021 for instance), commonly referred to as "bidirectional" effects. This makes the use of the term "bi-directional modelling" in L75 ambiguous.

17. L61: The contribution of the soil to TOC SIF *estimates* is very unclear. How is it accounted for in QUINCY?

18. L65: The term "carbon cycle modelling" is ambiguous here, as it is more often associated to global TBMs (those described from L76).

19. L65: In Miller et al. (2005), ChlF is simulated by FluorModLeaf.

20. L68: A proper reference for SAIL should be one of Verhoef et al.

21. L71: "many applications" is too vague.

22. The main reason why the TBM community has implemented SIF models is for assimilating satellite SIF data to improve the simulation of GPP. This is really not reflected here, in particular with respect to the cited papers of Koffi et al. or Bacour et al. It is also really unclear if the perspective of assimilation of SIF data is what is motivating the developments made in QUINCY. In addition, I find it a bit strange that QUINCY and its previous SIF model developments are not presented here, associated with the TBM community.
The various satellite data that have been considered in the cited data assimilation studies could be indicated.
Other SIF modelling works worth mentioning could be: Cui et al. (2020) (https://doi.org/10.1016/j.rse.2019.111373) or Wang et al. (2021) (https://doi.org/10.1016/j.agrformet.2021.108424) with BEPS, or Fan et al. (2025) (https://doi.org/10.1029/2024JG008280).
Associating the work of MacBean et al. (2018) to "computational burden" in relation to the modelling of SIF in TBMs is surprising as they considered a simple SIF-GPP relationship.

23. L86: "A 2-stream"... the sentence is not clear.

24. "... pave the way for data assimilation approaches". 1) If this is the main motivation of this work, it should be introduced earlier, 2) so what about some of the works that are cited earlier? Another aspect of data assimilation, in addition to the definition of an observation operator, is the development of a numerical optimization algorithm capable of tuning specific model parameters to minimize the discrepancy between model simulations and observational data. While I view the model implementations conducted in this study very positively, what I have seen regarding the calibration of the model parameters does not meet the standards expected for such a perspective.
It seems to me that those objectives are very much dependent of QUINCY, all the more they have been already undertaken (partly) with other TBMs. The previous works with QUINCY should be introduced before to highlight what is the novelty of this study.

25. The reference to the TROPOMI data that are used in this study should be provided here, given in particular that two datasets exist.

Data and model descriptions

26. §2.1. Do the sites belong to the same ecoclimate class? It is not clear if QUINCY uses plant functional types (PFT) classes for simulations, and if so, what is (are) the PFT(s) considered for those three sites. Do they have the same dominant PFT?

27. L108: About the red and near-infrared regions, it is unclear if the ranges provided here also correspond to the ones used for QUINCY simulations.

28. L112: If there are three vertical scans at CA-Obs, what value is considered in this study?

29. L115. A reference paper for the FI-Sod should be provided if possible.

30. L120: There are two identical references for Magney et al. (2019) (2019a = 2019b). There is no presentation of a specific SIF spectrum from Hyytiälä data in the paper of Magney et al. In addition, later in the manuscript (in §2.6) the authors refer to Fig1 from Magney et al.: All spectra in this figure exhibit different variation ranges than the one shown in Figure 1 of the current study. Clarifications on the source of the SIF emission spectrum considered here, and its availability, should be provided.

31. L123: Which CO_2 fluxes? Are they directly "measured" or "estimated"?

32. Title of §2.2 should mention "SIF". The paper of Köhler et al. (2018) should be cited before the one of Guanter et al. (2021). The description of the spectral characteristics may be too detailed for the (already long) paper.
As indicated in Guanter et al. (2021), the TROPOSIF retrievals are provided at 740 nm (and not 743).
It is unclear whether the authors have used the "instantaneous" retrievals or the "daily averaged" ones. I guess this is the 1st option as they show in Figure 8 daily averages values around noon. This temporal averaging should be indicated here.
What puzzles me, is the spatial scale of 0.5° around the sites that is considered for averaging the satellite data, given the heterogeneity around the flux tower sites. For instance, CA-Obs (analyzed in Figure 8) is surrounded by lakes, with a "big" one about 6 km away from the site. To which extent surfaces different from "boreal evergreen coniferous forests" within the considered satellite footprints can contribute to the averaged values used for model evaluation? Why not considering an averaging at a higher spatial resolution?
Also, why such strict criterion (0.2) on cloud fraction for data selection? A value of 0.8 is used in Guanter et al. (2021). This strongly limits the number of available observations, in particular during winter (which is a specific attention of the study in relation to the impact of sustained NPQ).

33. The structure for the presentation of QUINCY could be improved, in particular by more clearly separating what existed prior to this paper from the developments made in this work. §2.4.1 is included within "§2.4 Leaf level model of chlorophyll fluorescence" but there is no §2.4.2.
The reference to Campbell and Norman (1998) for clumping is a bit vague to understand the parameterisations considered here (the paper of Campbell and Norman is about 300 pages long).
To make the paper more concise, the features/characteristics of the model for boreal evergreen coniferous forests could be provided here. In particular, for the leaf scattering albedo (instead of §2.7). What is the value of leaf clumping? Is a leaf inclination function considered? If so, what is it?

34. The presentation of the various yields and rate constants follows previous works (van der Tol et al., 2014; Raczka et al., 2019); this part could be shortened, possibly moving bits to the "Supplement" section .
I do not understand the formalism considered here: the unit of a "rate constant" is the inverse of time (s^{-1}). If they are unitless, do the authors refer to "relative" rate constants (see Laisk et al., 1997)?
The variable *x* in eq. 9 is referred to as "degree of light saturation" in the reference papers.
L228: The vegetation species for which the relationship between *x* and K_N has been derived in van der Tol et al. (2014) should be clarified (cotton leaves) to highlight one strong assumption of this study (i.e. that a PFT specific parameterisation of "reversible" NPQ is considered, which may not be adapted to boreal conifer ecosystems).
In Table 2, some "variables" are actually "parameters". The statement that they are unitless is incorrect (see for instance Raczka et al. for the case of fluorescence fluxes). The list of parameters and variables is not exhaustive, the value of some parameters is provided in the text (L230 or L250 for instance) while others are missing ($\mathrm{F}_{\mathrm{P}}^{0}$, Φ_{P}^{0}, etc.), which does not permit a comprehensive understanding of the parameterisation employed. Couldn't Table 2 be expanded, while also distinguishing which parameter values come from the literature and from this study?
The values for the parameters *b* and *T* of K_{Ns} are provided in L246 as well as in L251.
The notation for "sustained" is sometimes "s", some other times "sus". This could be homogeneised.

35. The presentation of the calibration of K_{Ns} needs improvement.
L242, Raczka et al. also accounted for the same temperature acclimation in their parameterisation of sustained NPQ; however, they complemented this with the parallel development of a specific reversible component. It is not clear to me why this is not considered in this study, all the more this could possibly result in a form of "double counting" of NPQ when K_{Ns} is not zero.
I do not understand the justification about the use of only the "SIF observations" at US-NR1 for the parameter calibration, also given there is no attempt in this paper to compare the K_{Ns} obtained in this paper to that of Raczka et al. It is very unclear what are those "SIF observations"; I guess this must relate to some PAM measurements, but they are not presented in the "Data section" for the corresponding sites.
Is the parameter calibration at US-NR1 performed over the whole of the time series or by considering specific time periods?

1. L173: The reference "NOA" is a bit odd. Other references to FLUXNET would be more appropriate.

36. What is "1/C°" in eq 5?

37. Doesn't eq 12 depends on the layer / LAI?

38. The coupling of the canopy module of QUINCY with the different RT approaches is very confusing. It is not clear how eq 12 is exploited by each of the RT approach.
It is unclear whether leaf-level SIF emission is considered as a diffuse flux in all cases, and whether mSCOPE represents a directional component of TOC SIF compared to the other RTs: the observation geometry (nadir) is detailed for mSCOPE but not for L2SM and LZ. Do that mean that for L2SM and LZ, the whole upper hemisphere contributes to the simulated TOC SIF? If so, is the approach adapted for a comparison against nadir viewing instruments?

Why such distinction between mSCOPE and the other models? There is a QUINCY-mSCOPE version but no QUINCY-L2SM or QUINCY-LZ, all the more it finally appears that L2SM is the preferred option by the authors.

The spectral dependency of the leaf optical properties and SIF needs to be better explained; it is absent from all the equations. L347: the specific equations should be provided.
What do the "red" and "far red" ranges correspond to? The source of the SIF emission spectrum is unclear (see previous comment) and its application in each case is not detailed enough (what is done for the simulations at 740 nm?).
In eq 21, is $SIF_{internal}$ somewhat related to SIF_{cl}? If not, how is it defined? What is E in eq 21? E_{leaf}?

Eq 22 is a very generic equation, which does not seem specific to the LZ approach. Where is the spectral dependency in eq 22, as LZ is used to calculate TOC SIF in the red and far red ranges?
Is the “upscaling only” implementation really useful, as it is not realistic and little discussed later in the paper?

39. Title of §2.7 could be clarified, as for instance "site scale flux simulation protocol".
It is unclear whether some kind of spin-up is performed at the different sites.
L358-L360: same remark as before on the way the values for soil albedo and LAI are prescribed. Is the soil albedo the same for the three sites?

40. A section dedicated on the description of how the models are evaluated is missing. It could detail typical evaluation metrics considered here (R^2, RMSE, etc.). Because it seems that the model bias dominates the error budget, RMSE could also be decomposed into systematic vs random components.
The section could also include how the escape factor is calculated for each RT approach as well as information about the computational efficiency. It appears that the latter criterion is the one the authors ultimately considered for the selection of the optimal RT scheme, but there is no information/number to allow assessing how much “faster” is L2SM compared to mSCOPE. The corresponding analysis should be provided somewhere in the result section.
Finally, this “evaluation methodology” section could also include the description of the specific processing of the data (monthly averaging and calculation of the midday values, etc.), that do not fit in the Result section (L368).

Results:

41. Because the modelling of sustained NPQ is presented before the RT approaches in the Methods section, I would find it more logical to present the corresponding results on the impact of the accounting of K_{NPQs} on the simulations first.

42. Section 3.1 is way too long. It could be separated into specific subsections (for instance by splitting the evaluation of the escape factor estimates) and redundancies should be avoided. In addition, several discussion of the results would rather fit in a “Discussion” section (ex: L390-L395, L397-L399, L406-L409, etc.) or in the “Methods” section (L406, L420, L433, etc.).

43. A figure synthetising the model-data (dis)agreement over all sites for all RT schemes is missing.

44. L380: I do not understand the statement about the necessity to have good GPP performance to “successfully” (whatever this means) simulate SIF. One could imagine that, in order to correct for the large bias in SIF simulations, one could simply reduce LAI for instance, which would then degrades the GPP simulation performance.

45. L387-L388: The sentence is awkward.

46. L406: It is indicated here, in the Result section, that sustained NPQ is not considered for Sodankylä. This should rather be introduced in the “Methods” section. The best would however be to clearly show the impact of the accounting or not of sustained NPQ at the three sites in the first section of the “Results”.

47. L423: Could the lowest SIF performance at FI-Sod be explained by not accounting for sustained NPQ at this site?

48. Table 4: Why considering only the summer values, given one of the focus of the study is more on winter and summer times when the sustained component of NPQ can not be neglected. I would be very much interested in seeing the temporal variation of this variable (which is not common in the literature). Could the estimated values be compared against other works?

49. L443: Isn’t APAR expressed in µmol *of photons*?

50. L443: I do not understand what “ecosystem response” refers to here.

51. L449: “… there was no feedback” > “in the current framework, there is currently no feedback”? This part should be moved to “Discussion”.

52. L452: I do not understand the causality for the lower seasonal cycle of APAR at US-RN1. Is FAPAR constant for all sites (as it could be interpreted from L329)?

53. L457: Is it ΦF or ΦFt?

54. §3.3. Why focusing on summer time for this analysis, when the sustained component of NPQ is less important?

55. L467: Have the authors tried to fit two separate hyperbolic functions for morning vs afternoon values to support that statement?

56. L470: "…lower level". For SIF, GPP or both?

57. L473 about the stomatal conductance. Difficult to understand as how it is implemented in QUINCY has not been presented. This part fits better in the Discussion section.

58. L476: What does "smaller uncertainties" for the parameter *a* means? I am puzzled that the calibration of the parameters of these hyperbolic functions (used for a rather "qualitative" evaluation) is performed "rigorously" with a dedicated optimization routine, while this is not the case for some of the main parameters of the model.

59. L478: I have not understood the sentence.

60. L480: Which is the Figure / Table that show the lower performance at US-NR1? Why not providing a Table in which the scores are calculated over the whole of the temporal window for the three sites? Again, why focusing on a limited window in summer?

61. L485: Isn't it Table 5 instead?

62. L487: About understorey vegetation, this should be moved to the "Discussion" section. Because no information has been provided in the "Methods" section on how the vegetation is described in QUINCY for the different sites (100% boreal evergreen needleleaf forest?), the implications for potential model improvements are difficult to grasp.

63. L495-L497: This is more for Discussion.

64. §3.4: The title of the section is very misleading and does not reflect what is presented here (a comparison between model simulation and satellite estimates).

65. L500: "better agreement"… This should be supported by quantitative metrics.

66. L502-L503: A bit awkward given that the averaging characteristics are determined by the authors. I have understood that the authors use daily SIF data; do those "rapid **phenological** changes" occur at a higher temporal resolution? What do "phenological" refer to here? Do the authors mean "physiological"? Note also that the temporal monitoring of SIF by satellite observation is limited by acquisition around noon.

67. L503: The authors could have also adopted a less stringent criterion on cloud fraction.

68. L510-L514: Why not showing the corresponding time series?

Discussion:

69. Several pieces of information about modelling assumptions and protocol are provided in the Discussion section. Presenting them earlier in the Methods section would make it easier to interpret the results while reading the manuscript.

70. Table S3: The caption is not clear. Couldn't the "overestimation" be calculated as a ratio or as a bias?

71. L523: Is this statement supported by other studies?

72. L529: Which are those "other ecosystems" and tests, in relation with the model developments presented in this study (sustained NPQ and RT models)?

73. L533: Has the assessment of satellite versus *in situ* data not been addressed in previous studies?

74. L535: About the relevance of "simple approaches" and the assessment of the modelling SIF performance based only on R^2: The conclusion is really dependent on what ones want to do with SIF! If the objective is to assimilate SIF observations to improve TBMs' simulation of GPP, isn't it preferable to use a process-based representation of SIF that is consistent with that of GPP?
Given the substantial mismatch between the modelling performance of GPP and SIF in QUINCY, it seems likely that there is an issue with the SIF modelling framework. Under the current model, would it be possible to assimilate SIF data without degrading GPP?

75. L544: A more detailed description of the clumping and description of the canopy architecture would be helpful in the section presenting the model.

76. L550: I agree with the authors that a consistent representation of the radiation regime within the canopy is required for SIF and GPP. However, as said before, the current way the model is presented does not allow understanding the improvements to be made in that direction. How would the authors "include" NIRv in their framework? Is QUINCY equipped with a RT scheme allowing to simulate the corresponding TOC reflectances as it is the case for SIF? What datasets would they consider? *In situ* or space-borne data? The values of the calculated escape factors could have been compared to literature.

77. L564: Could the authors provide references on the variation of SIF emission spectrum across species? Given that all RT approaches yield approximately the same degree of overestimation in SIF simulations compared to observations, how important is this assumption relative to other potential sources of error? It would be helpful if a ranking of these possible sources of error could be provided. What is the relative contribution of incorrect parameter values (which could potentially be corrected through data assimilation) compared to incorrect process representation?

78. L567: "… in vivo spectrum". Could a reference be provided? I do not understand the implication of this, compared to the other choices made. Could the authors elaborate?
Does this part related to the spectral dependency of SIF really fits within a "radiative transfer" oriented section?

79. L569-L585: Same remark as before, this part related to data processing do not fit within this RT section (which is already very long).

80. L605: Although leaf chlorophyll content is the main factor of leaf absorptance in the visible, this is not the case in the near-infrared.

81. L606-L607: So, this is not the case in QUINCY? What to the authors mean by "include"? Does it relate to the representation of the temporal variations and/or vertical gradients?

Conclusions:

82. L615: While the study focuses only on QUINCY TBM, this general statement seem to apply to other similar models. Could the authors elaborate on this? What is expected from a better accounting of the canopy structure by proximal sensing measurements?

83. L619: I thought that sustained NPQ was not applied at Sodankylä… A clearer presentation of the simulation configurations is needed.

84. L620: Do this "process" refer to sustained NPQ? Why "likely"? Isn't air temperature explicitly used already in the formulation of NPQs?

85. L620-L622: TROPOSIF is not a satellite, it is a product derived from the observations by the TROPOMI instrument. What does "seems feasible" mean given that this is exactly what is done in this study (Figure 8)?

86. L625: On the link between SIF and latent heat, I am really unsure that this fits in the conclusion of this manuscript given this has never been addressed earlier. This could be expended in the discussion.

87. L627: Given the concerns I have expressed, I have some reservations about the concluding remark. Does it concern QUINCY or does it apply to other TBMs as well?

Tables:

88. Table 1: (°) for the coordinates should be added. LAI and Air temperature vary with time; Are those mean values? The years considered in the study could be indicated.

89. Table 5: Couldn’t the presentation of Table 3 be adopted also for this table (R^2 and RMSE on the same line)?

Figures

90. Figure 1: TROPOSIF should be indicated in place of TROPOMI. The wavelength of the SIF product is 740 nm, not 743.

91. Figure 2: The values on y-axes on the left should also be in purple, not only the label, to facilitate interpretation. The same unit is repeated multiple times; this could be simplified. The years are missing on the left figures. For improved clarity, it would be more consistent to present the panel labels (a, b, c) prior to their corresponding descriptions.
Metrics quantifying the model-data discrepancy (RMSE, R^2) could be provided.
These remarks apply to all similar figures.

92. Figure 3 - Figure 7: Please use “LZ” or “Liu/Zeng” or “Zeng” consistently throughout the manuscript and figures.

93. Figure 5. Couldn’t the different components of K_{NPQ} be also presented?
“photosynthetically active” is missing.
Is it phi_Ft that is shown here (eq 10)?

94. Figure 6: Dots and crosses are not distinguishable.

95. Figure 7: “Far-red” should appear more clearly in the corresponding figure insets.

96. Figure 8: “TROPOSIF” instead of “TropoSIF”. 740 nm should be considered instead of 743 nm.
What is the “uncertainty” of those TROPOSIF data averaged at 0.5°? Is this a kind of averaged retrieval error or the standard deviation of all individual retrievals ? How does it relates to the “standard deviation” cited in L509?
What does “TROPOSIF values are daily” mean? It is not clear whether the authors have considered the daily corrected variable in the TROPOSIF product or not.