

Authors' response to the first referee: M. Alban Farchi

We would like to acknowledge the meticulous review of our article done by M. Farchi, and we thank him for his relevant remarks and comments that will help us improve the article content further.

(1) On several occasions, the text alternated between two topics, which can be sometimes confusing:

- **In the introduction, between physical-and machine learning aspects of your problem.**
- **L91-92 between the description of the retrieval process and a post-processing that you apply (discarding negatives values + averaging)**
- **L212-213 between description of results and conclusion**
- **L362-367+ between NN results on the test (RMSE and correlation) and the application of the NN.**

I would advise to reorganise the text to avoid giving the reader the impression to go back-and-forth between topics, and also to take this opportunity to make the logical connections more visible. For example, it is only after reading the whole introduction that I understood the logical links between the various aspects of the problem.

RESPONSE 1.1: Thank you for these comments. We propose to simplify the Introduction by removing L. 40-46. This will ensure a better transition from SIF to assimilation.

About the specific lines mentioned:

- L92: we propose to gather all sentences related to the instrument together, as well as those related to the retrieval. Our own post-processing work will be summarised as: "In this study, since our model domain is defined on a regular grid, the SIF observations were interpolated onto the model grid. Only all positive instantaneous measurements for a given day covering a grid cell in the domain are averaged to provide a value for that grid cell."
- L 212-213: We agree. Presenting results in the Methods section is confusing. We propose moving the intermediate tests in a Supplement and focus on the final configuration that was used to produce the results. Namely, use of LAI observations as a predictor for training the NN, together with DOY, LAT, LON as metadata predictors.
- L362 -367+: We agree. We will move the following sentence at the beginning of the next paragraph: "The neural network was then successfully used ...".

(2) Eq 1-3: at first, I thought that there was a typo in 1a (that the argument if the model should be the analysis and not the background). It is only after reading Eq. 2 that I understood that it is not. Overall, the presentation method is highly confusing, first because there is a temporal aspect, which means that the method should be called "smoother" and not "filter" (I can understand that, for historical reason, you would prefer call it a filter, but you should at least mention this point), and second because the temporal aspect is never really explicitly explained: is the control vector the state at $t=i$ or $t=i-1$? From what I understand, the fact that M appears in the definition of the Kalman gain means that the control vector is the state at $time=i-1$, which means that there is a typo in 1b, it should be:

$$x^a_{i-1} = x^b_{i-1} + K. . .$$

or equivalently:

$$x^a_i = M x^a_{i-1} = x^b_i + M K. . .$$

in a case of a linear model.

Then, the sentence L 557-558 "The initial state is determined by the analysis performed over the previous 24-hour assimilation window." is confusing. The initial state for what? Or is this

***sentence only here to explain that the time step between $t=i-1$ and $t=i$ is 24h?
Finally, there is no mention about cycling. Are you using cycled data assimilation or is the background reset at each assimilation cycle?***

RESPONSE 1.2: The reviewer is right, there is a mistake in the SEKF equations. We propose to replace this set of equations by those published in the Supplement of Bonan et al. 2020.

(3) The LAI acronym introduced multiple times (L30 & L37)

RESPONSE 1.3: Thank you for pointing this out, the latest introduction will be removed.

(4) L37 "SIF observations" later in the introduction, these are called "TROPOMI SIF retrieval" (L46-47, 50). Are you referring to the same object? Please try to be consistent with your nomenclature.

RESPONSE 1.4: We agree that it is confusing. We propose to replace:

"To enable more real-time and operational monitoring of vegetation, the 8-day TROPOMI SIF retrieval has been assimilated into the IFS land surface model developed at ECMWF at a global scale (Garrigues et al., 2026). In this study, we propose to constrain simulations from the ISBA (Interactions between Soil, Biomass and Atmosphere) land surface model, which is developed at Météo-France and available through the SURFEX (SURFace EXTERNALisée) platform (Masson et al., 2013; Decharme et al., 2013), by assimilating the daily retrieval of TROPOMI SIF."

by

"To enable more real-time and operational monitoring of vegetation, Garrigues et al. (2026) have assimilated the 8-day SIF retrieval reprocessed from raw TROPOMI data has into the IFS land surface model developed at ECMWF at a global scale. In this study, we propose to constrain simulations from the ISBA (Interactions between Soil, Biomass and Atmosphere) land surface model, which is developed at Météo-France and available through the SURFEX (SURFace EXTERNALisée) platform (Masson et al., 2013; Decharme et al., 2013), by assimilating the daily retrieval of TROPOMI SIF instead of the 8-day TROPOMI SIF product used by Garrigues et al. (2026). Unlike Garrigues et al. (2026), our land surface model simulates LAI."

(5) L40 a transition is missing before the second paragraph.

RESPONSE 1.5: Thank you for pointing this out. We propose to delete the first 6 sentences of this paragraph ("Advances in ... a large database ()"). This will ensure a better transition to the work of Garrigues et al. (2026).

(6) L42 "greater accuracy" → "higher accuracy"?

RESPONSE 1.6: Yes we agree, it should be "higher accuracy".

(7) "better suited to the requirements of an operational system" this statement is questionable. Taking the example of meteorology, machine learning weather prediction models sometimes show a lack of physical consistency in their forecasts, which is typically not what you would like in a operational setting.

RESPONSE 1.7: We would like to thank the referee for raising this issue. We propose to delete the first 6 sentences of this paragraph ("Advances in artificial intelligence ... a large database (...)).

(8) L44 "are beginning to emerge" the article cited has been published 6 years ago, I wouldn't say that it is the beginning.

RESPONSE 1.8: We would like to thank the referee for raising this issue. We propose to delete the first 6 sentences of this paragraph ("Advances in artificial intelligence ... a large database ()").

(9) L45 "Deep learning is based on the supervised training" there are many examples of applications of semi-supervised or even unsupervised learning in the geosciences.

RESPONSE 1.9: We would like to thank the referee for raising this issue. We propose to delete the first 6 sentences of this paragraph ("Advances in artificial intelligence ... a large database ()").

(10) L46 "To enable more real-time and operational monitoring. . ." Should'nt you start a new paragraph here?

RESPONSE 1.10: We would like to thank the referee for raising this issue. We propose to delete the first 6 sentences of this paragraph ("Advances in artificial intelligence ... a large database ()").

(11) L54 "seems more suitable" at this point, this is hypothetical, as it remains to be proven that such ML-based operator can be constructed and possesses sufficient accuracy for the task. Perhaps replace "seems" by "could" or "would"?

RESPONSE 1.11: We thank you for this relevant remark. We propose to replace "An alternative machine learning-based operator for SIF seems more suitable." by "An alternative machine learning-based operator for SIF could be more suitable."

(12) L63 "the methods including the dataset" I wouldn't say that a dataset can be considered as a method.

RESPONSE 1.12: Thank you for pointing this out. To clarify, we propose reorganising the paper by distinguishing a Data section from the Methods section.

(13) Figure 1: For panel(a), I would suggest to use a rectangular projection (like PlateCarree for example) to be consistent with the domain used. I assume that the colourmap is the same for all three panels (it is not explicitly mentioned). With this choice of colours, the map is very difficult to read. I would advise to choose a so-called categorical colourmap. Also, I am not sure whether the figure is intended to be single or double column, but I have the impression that a lot of space is lost with padding and margining.

RESPONSE 1.13: We agree. We have simplified Figure 1 as follows.

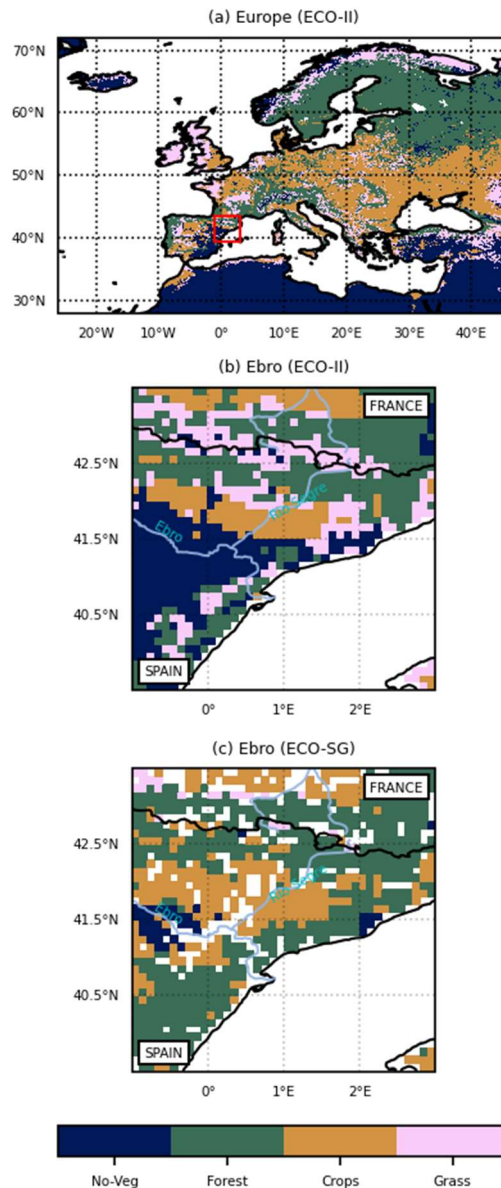


FIGURE R1.1 - Dominant Plant Functional Type according to land use maps. (a) Domain Europe with ECOCLIMAP-II [26°W to 46°E, 28 to 72°N], the delimited area in red represents the Ebro basin subdomain. (b) The Ebro basin subdomain [1°W to 3°E, 39.5 to 43.5°N] with ECOCLIMAP II. (c) The Ebro basin subdomain with ECOCLIMAP-SG. Four surface types are shown: grass (C3 grassland, C4 grassland), crops (C3 croplands, C4 croplands, flooded crops), forest (deciduous broadleaf forest, coniferous needleleaf forest, evergreen broadleaf forest), No-veg (bare soil, rocks, snow and ice, peat and wetlands).

(14) L86-91 I would reformulate these sentences. As a non-expert in land surface modelling, I got the impression that the same process is described three times and was totally unable to understand exactly what is done here.

RESPONSE 1.14: Yes. We propose to replace:

“The retrieval process involves the statistical modelling of TOA radiances over vegetation, while excluding the signal induced by the underlying bare surface. Surface vegetation is described using the MODIS2018 database (Friedl and Sulla-Menashe, 2015). Using prior knowledge of the solar-

induced fluorescence spectrum, the algorithm extracts the SIF-related part of the TOA radiance signal (Guanter et al., 2015). The estimated SIF is then computed by subtracting the expected radiance of the bare soil from the TOA signal. This process ensures that the distribution of TROPOMI SIF observations is centred on zero over bare soil.”

by

“The retrieval process involves statistically modelling TOA radiances over vegetation while excluding the signal induced by the underlying bare surface. The MODIS2018 database (Friedl and Sulla-Menashe, 2015) is used to identify areas of vegetation that are relevant for SIF retrieval. Using prior knowledge of the SIF spectrum, the algorithm extracts the SIF-related part of the TOA radiance signal (Guanter et al., 2015). This process ensures that the distribution of TROPOMI SIF observations is centred on zero over bare soil.”

(15) L91-92 *“This means that negative values for TROPOMI SIF can occur, which we have discarded”. I am a bit lost here (most probably because I don’t know what exactly SIF is): is SIF supposed to be positive?*

RESPONSE 1.15: Thank you for pointing this. We propose to replace:

“This means that negative values for TROPOMI SIF can occur (...)”

by

“This means that non-physical negative values for TROPOMI SIF can occur.”

(16) L98-100 *At this point, it is unclear for what these additional measurements will be used.*

RESPONSE 1.16: We agree. We will move the description of the LIAISE HyPlant dataset to the supplement.

(17) L107-110 *“if an observation is missing, the gap is not filled” and “using an arithmetic average if at least 50% of the grid points are observed (i.e. half the minimum amount)” this whole part is confusing. First, I think that it would be good to remind the reader why observations can be missing and if we are speaking of only parts of the observation missing or entire maps missing. Then, depending on the answer to that question, does the order of interpolation (space or time first) matter? Finally, the 50% threshold does not really make sense to me.*

RESPONSE 1.17: Thank you for your remark and for highlighting that this paragraph requires more detailed explanations.

LAI observations are available at a resolution of one kilometre, at best, every 10 days. In order to align with our daily TROPOMI SIF product during the NN training process, the LAI maps undergo linear time interpolation. This is only performed between two consecutive LAI observations separated by 10 days. If a LAI observation is missing, the gap is not filled. As in previous studies (Albergel et al., 2010; Barbu et al., 2011; Corchia et al., 2023, among others), LAI observations are aggregated to the spatial resolution of the model grid cells using an arithmetic average, provided that LAI observations cover at least 50% of the grid cell.

We checked that the order of interpolation (space or time first) does not impact the final maps. These explanations will be included in the revised version of the paper.

(18) L111 *“instead” is repeated*

RESPONSE 1.18: Thank you, it will be removed.

(19) L112 *"is spatially interpolated the same way" no time interpolation for these observations?*

RESPONSE 1.19: Yes. We propose to replace:

"This provides data every 10 days and is spatially interpolated in the same way."

by

"This provides data every 10 days and is spatially and temporally interpolated in the same way."

(20) Figure 2: *I would advise to use a grid, otherwise the plot is difficult to read. You can also set the green axis range to (0,5) so that the tick positions are exactly the same for all three axes. Furthermore, is there a real interest to present the data as points (with fancy markers that can be distracting) instead of lines? In the caption, you mention the domain Fig1, I would suggest to add the (a) label to remind the reader that it is the median over the European domain. Finally, are you showing the statistics for the raw or interpolated product?*

RESPONSE 1.20: We thank the referee for these suggestions. We will use lines and modify the axes to make the figure more readable. We will replace "Fig. 1" by "Fig. 1a". Regarding the displayed statistics, the LAI and SIF are interpolated products, whereas the GPP remains on its original grid (pre-interpolation).

(21) L 118 *"which is present only in the TROPOMI SIF and LAI V1 datasets". I only saw that spike for the LAI V1 in the figure after zooming (the green cross is almost hidden by a blue triangle). I think that this is a good example why lines would be perhaps more appropriate for this figure.*

RESPONSE 1.21: Yes, we will use lines instead of points.

(22) L 136-137 *"This process involves identifying the optimal control vector according to a model that minimises discrepancies between observations and a prior background estimates" This needs to be reformulated. As is, one understands that the choice of the control vector is part of the DA system, that the model does the minimisation, that discrepancies are computed between the background and observations, and that the background is to be optimised! Furthermore "prior background" seems redundant.*

RESPONSE 1.22: Yes. We propose to replace:

"This process involves identifying the optimal control vector according to a model that minimises discrepancies between observations and a prior background estimates."

by

"This process involves updating the control variables in the ISBA model by minimising discrepancies between the variables and observations, and between the variables and a prior estimate."

(23) L 139 *"(SEKF)(Mahfouf and Bilodeau, 2009)" → "(SEKF, Mahfouf and Bilodeau, 2009)"*

RESPONSE 1.23: We will correct this, thanks for pointing this out.

(24) L140 *"x, where the subscript denotes the temporal step" Please reformulate. As is, you are referring to a subscript in "x", which has no subscript.*

RESPONSE 1.24: Yes. We propose to replace:

“Thereafter, the control vector will be referred to as x , where the subscript denotes the temporal step.”

by

“Thereafter, the control vector will be referred to as x_i , where the subscript denotes the temporal step.”

(25) L144 *“where he superscripts “a”, “b”, and “o” stand for analysis, background and observations, respectively.” I think that you can remove the “o” superscript. Also, I would use an italic font for “a” and “b” here, to match what is used in the equations.*

RESPONSE 1.25: We will change the notations and fontstyle use in the equation to be consistent.

(26) Eqs 3a & 3b, K should be in `mathbf`

RESPONSE 1.26: Thank you for pointing this out, it will be corrected.

(27) L160 is “assessed” the appropriate word here?

RESPONSE 1.27: Yes. We propose to replace:

“Assimilation is assessed when the difference in observation space between the observation and the background simulation exceeds the final departure in the observation space between the observation and the analysis simulation.”

by

“Since the purpose of assimilation is to bring the model closer to the observations, the difference between the analysis and the observations in the observation space should be smaller than the difference between the prior simulation and the observations.”

(28) Before section 2.3 (but also before section 2.2) I think that it would be good to provide some transitions.

RESPONSE 1.28: Thank you for this suggestion, we will consider it in our corrections.

(29) L172-173 *“we will train a single neural network to cover all pixels.” This gives the impression that you design a NN that process the map at once, which, as far as I understand later on, is not the case: each grid point is processed independently.*

RESPONSE 1.29: This sentence actually needs to be rephrased. Although an NN is applied to each cell independently, only one NN is trained on all the model grid cells. We are not training a dedicated NN for each grid cell in the model.

(30) L174 *“probability of belonging to a given category” Now I am confused, I thought it was a regression problem, why do you mention classification?*

RESPONSE 1.30: Yes. We propose to replace:

“A feed-forward neural network’s basic architecture comprises an input layer, one or more hidden layers of neurons, and an output layer, which provides an estimate or probability of belonging to a given category.”

by

“A feed-forward neural network’s basic architecture comprises an input layer, one or more hidden layers of neurons, and an output layer, which provides an estimate.”

(31) L183 “with a maximum close to zero” do you mean here the maximum of the SIF values or the mode of the distribution?

RESPONSE 1.31: Yes. We propose to replace:

“The distribution of TROPOMI SIF values is similar to a log-normal distribution with a maximum close to zero (see in the appendix).”

by

“The distribution of TROPOMI SIF values resembles a log-normal distribution where the maximum of the mode of the distribution is close to zero (see in the appendix).”

(32) L184 “below zero predictions” → “negatives predictions”

RESPONSE 1.32: Yes. “below-zero” will be replaced by “negative”.

(33) L 184-185: “we applied the following strictly increasing function to the TROPOMI SIF values to predict. The sentence should end with ”:”. Also, to prove your point, it would be perhaps easier to give the expression of the inverse function and show that it has positive values, no?

RESPONSE 1.33: Yes. The inverse function writes :

$SIF = \text{Exp}(SIF_I)$ if $SIF_I \leq 0$, and $SIF = \ln(SIF_I+1)+1$ is $SIF_I > 0$.

(34) L190-200: I am confused here. First, you should mention explicitly what is coming from observations and what is coming from simulations. I guess that GPP, LAI V1, TROPOMI SIF are observations, but for the rest it is not explicitly said. Then, you mention a simulation database: where does it come from? Did it already include assimilation? Does it use the same model as the one you are using for the assimilation experiments? This last point is especially important: if the NN is trained to predict observations based on simulated variables (i.e. more or less something close to a background), then in a data assimilation experiment I would expect the innovation (y-Hxb) to be essentially noise, right? In the end, I don’t think it is an issue, because you only selected LAI-V1 (which, as far as I understood, is an observation product) plus some extra predictors, but this is definitely confusing.

RESPONSE 1.34: Yes. In this sentence, only LAI V1 and TROPOMI SIF correspond to observations. This section will be clarified by moving the description of preliminary tests (including Appendix A2) in a Supplement. We will focus on the actual configuration used in the paper.

(35) L204-207: I did not understand this paragraph, in particular connections.

RESPONSE 1.35: Thank you for pointing out this missing information. This section lacks a reference to Annex Figure A2, which displays the average Pearson correlation between each input and output. This step is intended to identify possible linear relationships between certain variables. This section will be clarified by moving the description of preliminary tests (including Appendix A2) in a Supplement.

(36) Fig 4: perhaps it would be better to present these results as a table instead of a figure?

RESPONSE 1.36: We thank you for your suggestion, we will show the Table associated to this Figure. This section will be clarified by moving the description of preliminary tests (including Appendix A2) in a Supplement.

(37) L214: ZSE and LON seem to have similar impact. Why not dropping LON as well?

RESPONSE 1.37: ZSE seems to have a little less impact, and keeping both coordinates as metadata predictors felt more logical.

(38) L215 "even if the observation operator has learnt the model biases": how could the observation learn the model biases, since it only relies on observation products in the training?

RESPONSE 1.38: There is a misunderstanding due to the lack of clarity in this section. We meant that, since the NN is trained using LAI observations and LAI is the only biophysical predictor, the NN can be used with any land surface model. The following sentence will be deleted: "This configuration ensures that, even if the observation operator has learnt the model biases, the assimilation of SIF will not simply adapt to the model and that it will remain robust to any improvements to the model."

(39) L217 "to estimates" → "to estimate"

RESPONSE 1.39: Thank you for pointing out this typo, it will be corrected.

(40) L218-225: I would reformulate this paragraph a bit to make it a bit smoother. Also, you mention "from 1 June 2018 to 31 May 2021" for the full data, but it is mentioned in 2.1.2 that Tropisif provides data from May 2018 to December 2021. Did you discard part of this data? Also the final year is not used at all?

RESPONSE 1.40: Yes, this was a mistake. The full data range used is from June 2018 to December 2021, including these dates. This will be corrected. LAI-V1 was used for training until June 2020. Therefore, the training/testing dataset could not extend beyond June 2020. However, the availability of LAI-V1 did not constrain the assimilation experiment, which is why it was conducted up to December 2021.

(41) L228-229: "As mention earlier" → "As mentionned earlier"

RESPONSE 1.41: Thank you again, we will correct this typo.

(42) L227: our focus is on a small area" it is unclear whether the small domain is used only for the assimilation or for the selection of the training dataset as well? Also, it is unclear whether the forecast model used in the assimilation experiments is able to run on the small domain only or needs to be run on the whole European domain?

RESPONSE 1.42: Training was carried out across the entire European domain. However, the assimilation experiments were only conducted over the Ebro basin. This will be made clearer in the revised paper.

(43) L231 "The assimilation period runs from 1 January 2018 to 31 December 2021" I thought that the Tropomi data only started in May 2018?

RESPONSE 1.43: You are right that the TROPOMI data were not available in early 2018, but we also run an assimilation experiment involving the LAI product, which was available. We have chosen to run the experiment for a full year.

(44) L233: why did you change the land cover map? If the second is more accurate and more recent, then why not using it for the training database?

RESPONSE 1.44: At the time of the study, ECOCLIMAP-II-driven ISBA simulations were the only ones available for the entire European domain. As ECOCLIMAP-II tended to prescribe excessive bare soil in the Ebro basin region, the ECOCLIMAP-SG updated framework was included for the data assimilation experiments.

(45) L241-243: It is not entirely clear in the text that you are talking about the observation error for the SIF observations (and not for the LAI observations).

RESPONSE 1.45: Yes. We propose to replace:

"To determine the amount of observation error to consider,..."

by

"To determine the amount of SIF observation error to consider,..."

(46) L236-246: you don't mention the dates for the experiments assimilating SIF. EDIT: From the caption of figure 7, I get the impression that the SIF assimilation experiments cover June 2018 - May 2020, which includes the training set of the NN (June 2018 - June 2019). I would heavily advise to remove this period from the statistics.

RESPONSE 1.46: To clarify, we will detail the assimilation periods by adding a dedicated column in Table 1. We will produce another version of Figure 7, where the different sub-periods will be distinguished (June 2018-May 2019, and June 2019-May 2020).

(47) Table 2: is there an interest in including the line without data points? Also, you haven't defined ubRMSE (I haven't checked, but there might be other undefined acronyms)

RESPONSE 1.47: We will check throughout the paper that all the acronyms were defined.

"ubRMSE" stands for "unbiased root-mean square error" and is equivalent to standard deviation of differences.

(48) Fig. 5: I would advise reducing the padding of the figure overall (I don't think that the ticklabels are needed) and using continuous (instead of discrete) colourmaps. Perhaps you could also change the projection to PlateCarree. Furthermore, I don't understand the choice of a diverging colourmap for correlation, unless there is something specific happening when the correlation is 0.5? Finally, even if this figure is intended for one column, I think you can use a bit more horizontal space.

RESPONSE 1.48: Thank you for your suggestions on how to improve this figure. We will adjust the settings for the plots. Regarding the colour maps, we ran several tests and found that the current

settings, which are based on discrete colour maps, produce the most readable figures.

(49) L279: "The saved weights of the neural network are implemented in the assimilation scheme" please reformulate.

RESPONSE 1.49: Yes. We propose to replace:

"The saved weights of the neural network are implemented in the assimilation scheme as an observation operator for the SIF."

by

"In a first phase, the NN was trained over the whole European domain. In this section, the neural network's saved weights are implemented in the assimilation scheme as an observation operator for the SIF."

(50) L281: "LAI representation will be our main focus on the control vector variables" I don't understand what you mean here?

RESPONSE 1.50: Yes. We propose to replace:

"As the observation operator only takes LAI as inputs, LAI representation will be our main focus on the control vector variables."

by

"As the observation operator only considers LAI among the control variables, the accuracy of the LAI representation is the primary criterion for evaluating the assimilation results."

(51) L282: "the monthly-averaged maps over the C3 crops" It is unclear to me whether the figure will show the increments only over this land category or over the entire domain. Also, is there a reason why you want to focus on this category?

RESPONSE 1.51: Figure 6 only shows the increments for this specific C3 crops vegetation type. The reason for focusing on this category is twofold: firstly, the NN tends to perform better where this category is dominant; and secondly, the Ebro basin is an agricultural area which is typically characterised by this specific vegetation type. This enables us to observe that LAI increases in known irrigated areas in summer and that this increase is applied to the correct type of cover. To clarify Figure 6, we removed the grid cells displaying less than 10% C3 crops (see FIGURE R1.2 below).

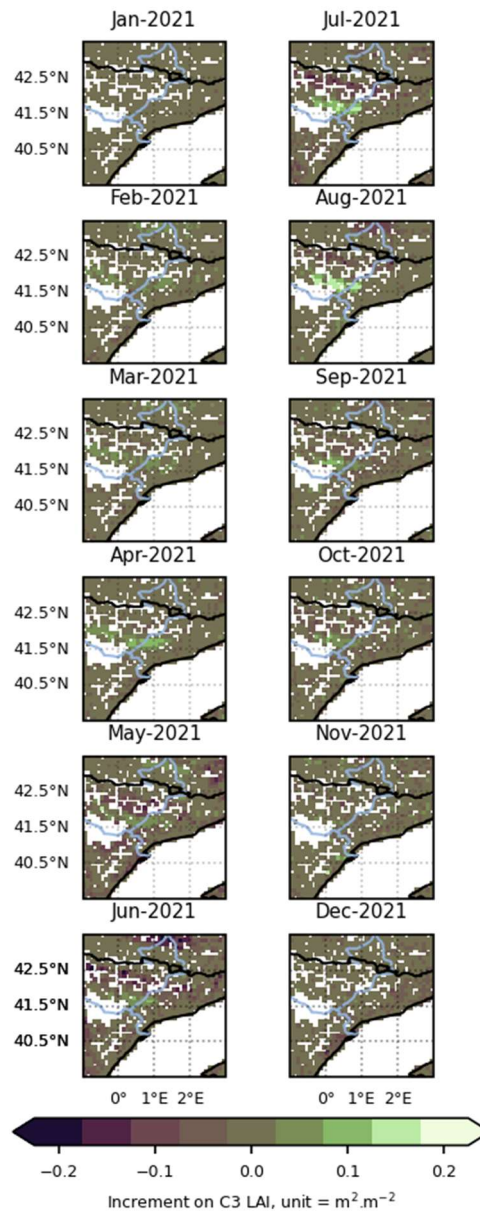


FIGURE R1.2 - Mean monthly analysis increments of the SIF20 experiment on the LAI of C3 crops due to SIF assimilation during the year 2021. Grid cells displaying less than 10 % C3 crops are masked.

(52) Figs. 6 & 7: Same as for the previous figure, I think that you can remove the ticklabels and reduce padding overall. For this data, you absolutely need a diverging colourmap centered on zero.

RESPONSE 1.52: A linear colour map was used to meet the colour blindness requirement. If we use a diverging colour map, the difference between the two extreme values will not be significant enough. We will consider using a diverging colour map if possible.

(53) L290: You mention an RMSE here. Can you be explicit and say between what and what this RMSE is computed? From the caption in the figure, I would assume that this is the RMSE between the LAI observations and the LAI analysis. I think it would be good to remind the reader that LAI observations are not assimilated in this setup and hence that they can be used to

evaluate the analysis, provided that their errors are independent of the errors of the SIF observations.

RESPONSE 1.53: The referee is right, the RMSE is computed between the LAI from the analysis SIF20 and LAI-V1 observed, in units of $m^2 m^{-2}$. In this experiment, the LAI is not assimilated and thus we can use the LAI-V1 as reference. We will improve this section to clarify it. We will add "SIF20 experiment" in the caption of Fig. 6, as already done for Fig. 7.

(54) L295: "all cells with a p-value higher than 0.01 were removed" can you please specify which statistical test is used to compute this significance test? Is this mask also applied in the delta(RMSE) panel?

RESPONSE 1.54: In our case, we use the F-test. The p-value is the result of a test of the null hypothesis that two normal populations have the same variance. This enables us to eliminate grid cells with insufficient data. This mask is used for both the RMSE and correlation panels.

(55) L300: "are closer to the observed values" according to which metric?

RESPONSE 1.55: In this case both metrics are used (i.e. RMSE and correlation). We will clarify it.

(56) L306: "the reference LAI-V1 experiment" I think that you should be careful with the results of that experiment since here you are evaluating against LAI observations, i.e. the observations that are assimilated.

RESPONSE 1.56: In Fig. 9, the LAI-V1 experiments is used as a benchmark for other experiments: SIF01, SIF20, and SIFFR.

(57) Section 3.2.2: same remark for the experiments here.

RESPONSE 1.57: In Fig. 10, the LAI-V1 experiments is used as a benchmark for other experiments: SIF20, SIFFR, SL20, and SLFR.

(58) Fig. 11: The caption is not very clear. It doesn't give a description for (a). Which analysis is plotted? I assume the whisker show the in-situ measurements but it is not mentioned.

RESPONSE 1.58: Sorry for that. We will describe the LIAISE campaign in the supplement and we propose to replace the current caption of Fig. 11:

"Comparison to the in-situ measurements of SIF during the LIAISE campaign (OL: Open-loop, AN: analysis). Box represents the 25th and the 75th percentiles, the whiskers the 5th and 95th percentiles. The red line represents the 50th percentile and the green triangle the mean."

by

"Comparison of the OL and SIF20 experiments with the SIF airborne measurements of the HyPlant instrument within the LIAISE campaign. Boxes represents the 25th and the 75th percentiles, the whiskers the 5th and 95th percentiles of the airborne SIF observations. The central horizontal red lines represent the 50th percentile and the green triangles the mean of the airborne SIF observations."

(59) L337-338: "we must compare all in-situ SIF measurements on a given day with the simulated SIF value resulting from the analysis for the two situations shown in Fig. 11(b)" I don't

understand this.

RESPONSE 1.59: It is a typo, it should read “for the two simulations” (i.e. SIF20 and OL). We will correct it.

(60) L340-341: *“the values are closer overall to the average airborne measurement value”: but I thought that absolute values should not be compared because of different wavebands?*

RESPONSE 1.60: We have produced a clearer version of Figure 11 (FIGURE R1.3 below), in which the TROPOMI SIF data and model experiments are more visible (FIGURE R1.3c). The evolution of SIF20 appears to be more consistent with the airborne SIF time series than that of TROPOMI SIF and OL. However, this could be an artefact due to the lack of TROPOMI SIF data on 17 and 22 July 2021. As these results are inconclusive, we will move this section to the supplement.

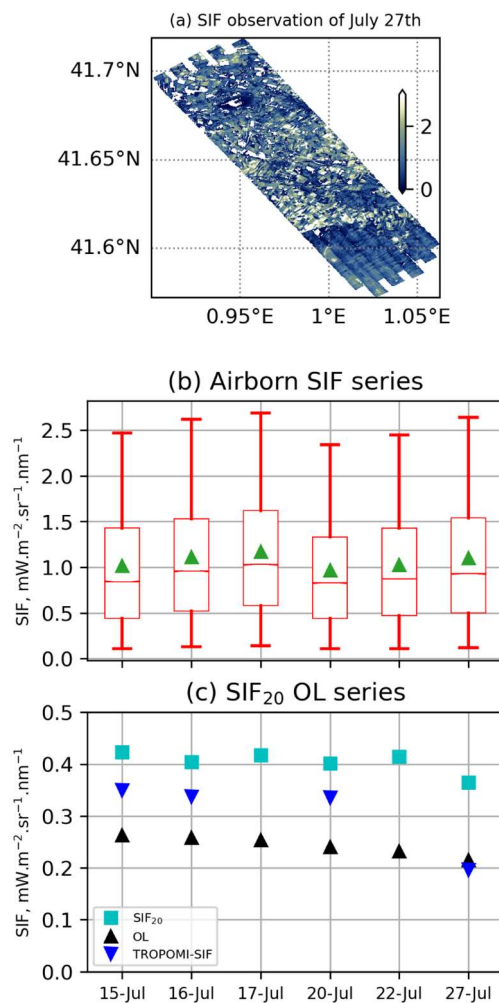


FIGURE R1.3 - Comparison of the OL and SIF20 experiments with the SIF airborne measurements of the HyPlant instrument within the LIAISE campaign. Boxes represents the 25th and the 75th percentiles, the whiskers the 5th and 95th percentiles of the airborne SIF observations. The central horizontal red lines represent the 50th percentile and the green triangles the mean of the airborne SIF observations.

(61) Section 4.1 *After reading the entire section, I am confused. I don't understand what you*

want to demonstrate here. The figure seems to show that the output of the NN more or less matches the values of TROPOMI SIF (which is what it has been trained for) and does not match another instrument: this is absolutely to be expected, or am I missing something here?

RESPONSE 1.61: This section is confusing because we are comparing the emulated SIF with the airborne SIF, which does not follow the same retrieval process. The key point is that the observation operator is not a surrogate model for airborne SIF measurements. Since these results are inconclusive, we will move this section to the supplementary material.

(62) L361: "Using this simple architecture" you are presenting the setup here, not really the architecture.

RESPONSE 1.62: We agree. "architecture" will be replaced by "setup".

(63) L362: "no overfitting evident" -> "no evident overfitting"?

RESPONSE 1.63: Thank you for this remark, we will correct it.

(64) L370: "update rate yielded by TROPOMI SIF assimilation is five times higher than that of 10-day synthesis assimilation" I don't understand this point.

RESPONSE 1.64: Yes. We propose to replace:

"However, it should be noted that the update rate yielded by TROPOMI SIF assimilation is five times higher than that of 10-day synthesis assimilation."

by

"However, it should be noted that daily TROPOMI SIF assimilation yields a higher update rate than assimilation of 10-day LAI observations."

(65) L374: "Using a neural network operator instead of a physical model makes propagating uncertainties difficult" What do you mean here? The uncertainties are all accounted for in the B and R matrices, no?

RESPONSE 1.65: Yes. We propose to replace:

"Using a neural network operator instead of a physical model makes propagating uncertainties difficult, so quantifying observation errors relies more on empirical choices."

by

"This study has shown that, when assimilating SIF alone, an observation error of 20% leads to the best assimilation performance. In a co-assimilation framework incorporating LAI, however, the best performance is obtained using a more complex error description (FR)."

(66) L375-376: "It is also important to understand that the neural network is trained to emulate observations with instrument uncertainties rather than the true physical variable, making quantifying the observation error of the true value challenging". I don't fully agree with that statement. The instrumental errors are (in principle) unpredictable and hence the NN cannot learn them. It is true that the use of a NN will introduce modelling error (because the process from state to observations is not perfectly represented) but this is also the case for any observation operator.

RESPONSE 1.66: Yes. We propose to replace:

“It is also important to understand that the neural network is trained to emulate observations with instrument uncertainties rather than the true physical variable, making quantifying the observation error of the true value challenging.”

by

“More research is needed to properly quantify the observation error.”

(67) L416-417: “information that cannot be changed” this is unclear: why couldn’t you change the values of the support parameters?

RESPONSE 1.67: Designating 'metadata predictors' as 'support parameters' was not relevant. We propose to replace:

“These support parameters provide information on the expected seasonal cycles. Yet, their use is questionable since they will provide information that cannot be changed, which could prevent the model from generalising well.”

by

“These metadata predictors provide information on the expected seasonal cycles.”