Reviewer comments for esd-2023-26

Title: Regionally optimized fire parameterizations using feed-forward neural networks

Authors: Yoo-Geun Ham, Seon-Ho Nam, and Jin-Soo Kim

In this manuscript, the authors have proposed a neural network based method to simulate Fire Radiative Power (FRP). The inputs to this network are 4 meteorological parameters. This new method to estimate FRP can be useful to scientists involved in understanding the fire intensity and weather/climate relationships, especially with less access to various fire-related datasets like ignition etc. However, there are some serious issues in the methods and interpretation of the results obtained. I recommend addressing the methodological issues and rethinking the interpretations of the results before proceeding with publication. Also, the composition of the paper needs to be changed significantly. I have segregated my comments into three sections, major, minor and language-related. Following are the points in the order of seriousness.

Major comments:

1. Methodological improvement with FFNN

I found several instances which do not strongly support the opinion of the authors, i.e., FFNN-based FRP estimates are indeed better than the FWI-based ones.

A. The whole argument is based on the comparison of FFNN-based FRP estimates with the FWI-based linear regression model. Why linear regression? The relationship between FWI and FRP is not at all linear. The correct approach would be to use various non-linear regression techniques, take the best of them and then compare them with the FFNNs. Only then can we compare the predictive abilities between FWI-based and FFNN-based methods. I expect significant changes in results and conclusion if instead of the linear method some other non-linear method is used.

: We fully understand the reviewer's concern. Our reference method to be compared to FFNN is FWI-based forecasts, and the linear regression is just to match the variability between FWI index and the FRP. In other words, we compared the parameterization quality seeking the nonlinearity between the meteorological variable and the FRP in the widely-used meteorology-based fire intensity estimation algorithm (i.e., FWI algorithm) to that seeking the nonlinearity using the neural network weights and the nonlinear activations.

To avoid the confusion, we modified the term 'FWI-based linear regression model' to 'FWI-based model' throughout the revised manuscript, and the brief description about the FWI-based model is also modified in the revised manuscript as follows.

Line 128-136 : "A FRP-estimation model based on the FWI was established as a baseline. The FWI is obtained from the daily averages of T2m, RH2m, WS10m, and PRCP, and .... To match the systematic amplitude differences between the FWI and FRP using the different units, a linear regression coefficient of the FRP with respect to the FWI, which was separately calculated for each grid point, is multiplied to produce the FWI-based model. Therefore, the nonlinearity between the meteorological variable and the FRP in the baseline model is purely originated from the procedure to derive the FWI."

B. Though FWI was originally developed for Canada, a large number of studies have used the system successfully to ascertain fire weather. It is true that the equations do not take into account some factors like species distribution etc. (lines 57-59). Even FFNN or in fact, any machine-learning-based model also does not take these into account.

: Thank you for point this out. Both FFNN and FWI cannot take into account the species distribution. The corresponding sentence is modified as follows "However, regional fire dynamics vary significantly depending on its unique climatological states (Flannigan et al., 2005, Kim et al., 2019)."

C. Inferences in lines 60-62 such as which variable contributes most to Arctic or Amazon fire activity can be determined by FWI-based studies too.

: We tried to argue that the regional difference between Amazon and Arctic region would not be fully considered as the algorithm for the FWI is purely based on the data in Canada. We modified the

sentence as follows.
Line 70-78 : “, however, its regional difference would not be fully considered as the FWI is purely derived by the data in Canada, while the relationship between the meteorological variables and the fire activity varies significant from regions to regions.”

D. Lines 156-158 and Figure 1.b: Even if this approach of linear regression is considered, from the figure it seems that the FWI-based model is statistically significant at all points as FFNNs. How is that an improvement?

: The significant test for the difference between FWI model and NN model is already performed in original Figure 1c. For the significant of the difference, we followed the method in Zou (2007) as follows. First, Fisher’s z transformation of correlation coefficient r for each forecasts is calculated $z = \frac{1}{2}\ln\left(\frac{1+r}{1-r}\right) = artanh(r)$. If (X, Y) has a bivariate normal distribution with correlation ρ and the pairs (Xi, Yi) are independent and identically distributed, then z is approximately normally distributed with mean of $\frac{1}{2}\ln\left(\frac{1+\rho}{1-\rho}\right)$, and standard deviation of $\frac{1}{\sqrt{N-3}}$, where N is the sample size, and ρ is the true correlation coefficient. Then, the difference in the standardized correlation, i.e., $\frac{1}{2}\ln\left(\frac{1+\rho 1}{1-\rho 1}\right) / \frac{1}{\sqrt{N1-3}} - \frac{1}{2}\ln\left(\frac{1+\rho 2}{1-\rho 2}\right) / \frac{1}{\sqrt{N2-3}}$, is over 0.05 (i.e., 95% confidence level) were marked.

E. Line 249: Since a linear regression method was used to map FWI into FRP, of course, the FRP estimates would not be able to capture the non-linear RH2m-FRP and PRCP-FRP relationships. The argument in section 4 needs to be rewritten after considering some non-linear regression of FWI and FRP.

: FWI can consider the nonlinear RH2m-FRP, or PRCP-FRP relationship as the procedure to derive the FWI from the meteorological variables are nonlinear (https://www.nwcg.gov/publications/pms437/cffdrs/fire-weather-index-system). For example, As shown in Figure 3, and 4 of the original manuscript (also shown in Figure A-1, and A-2 in this response letter), relationship between the RH2m and FWI, and PRCP and FWI is not linear, respectively; the differences between the bin is not a constant. As our main goal to compare the FWI model as a reference is that the simulation of the nonlinearity between the meteorological variables and the FRP in the FFNN is realistic compared to that in the FWI, we did not do any further corrections in FWI-based forecasts by applying non-linear regression of FWI and FRP.

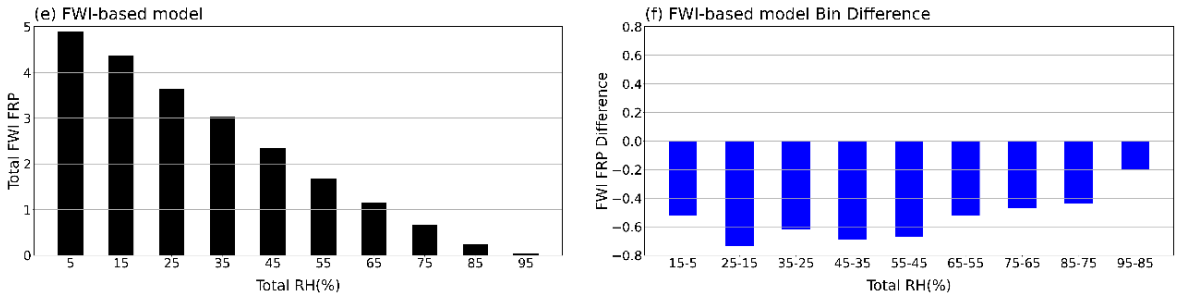

Figure A-1. Case-averaged FRP with respect to the RH2m with a 10% interval in the FWI-based model (left), and the differences in the case-averaged FRP at the upper bin from the lower bin in FWI-based model (right).

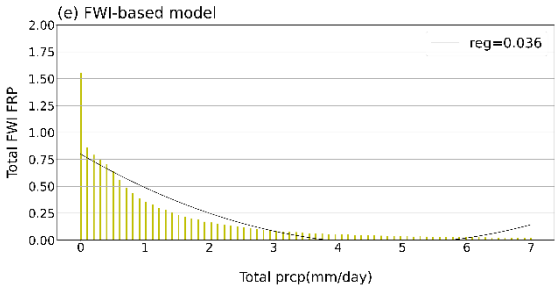

Figure A-2. Case-averaged FRP with respect to the PRCP with 0.1 mm/day interval in the FWI-based model

F. Methods like FFNNs need extensive validation. Though it is a common practice to train and validate ML models using the same datasets, if additional validation can be done with any station data or ground observations, then the FRP estimates will be more reliable.

: Thank you for the constructive suggestion. However, it is extremely hard to obtain the daily ground observations for the enough period to validate the fire activity. We note the limitation of our study by not validating our forecasts with ground observations in the revised manuscript as follows.

Line 179-186 : "We note that evaluating the skill of FFNN against FRP data may lead to an overestimation of its estimation abilities, given that the FFNN is trained using same type of data. Regrettably, the absence of ground-based observations on fire activity/intensity for the enough period deprives us of the opportunity to cross-reference FFNN-based FRP estimations with independent observations."

2. **Incomplete information in the methods**

A. Methods should include all the calculations and steps taken for complete analysis leading to figures and also stating the reasons for conducting that particular analysis. The sensitivity experiments are not explained elaborately. A flowchart or graphical representation of these steps or at least a detailed explanation would be appropriate. Otherwise, it is very cumbersome for readers to understand the inferences discussed in the subsequent sections.

: Sorry for the inconvenience. As a reviewer mentioned, we added a detailed procedure about the sensitivity experiments for Figure 2, and layer-wise relevance propagation (LRP) for main Fig. 5 as follows.

For sensitivity experiments in Figure 2, Line 231-237 : "… in the RH2m Clim experiment, the prescribed values of RH2m as an input of the FFNN is the daily climatology during the whole period (i.e., 2001-2020), therefore, its year-to-year variations in the RH2m is removed. Then, the correlation skill difference between the control simulation, that prescribes all input values at the corresponding date, and the RH2m Clim experiment is calculated to assess the importance of the RH2m in FRP parameterization."

For LRP in Supp. Fig. 7, Line 261-274 : "It provides a so-called relevance score, which linearly decompose the importance of each input variables as follows by propagating the output value backward toward the input variables using a chain rule.

$$f(RH2m, PRCP, T2m, WS) = R_{RH2m} + R_{PRCP} + R_{T2m} + R_{WS}$$

where f is a nonlinear model (i.e., FFNNs) to derive the FRP, and R_{RH2m}, R_{PRCP}, R_{T2m}, R_{WS} is a relevance score of RH2m, PRCP, T2m, and WS10m, respectively. The relative importance of any particular variable to the estimated FRP can be quantified by calculating the degree of the similarity between the relevance scores. For this purpose, we obtained the relevance score of each variable for each day during the whole testing period (i.e., 2001-2020), and calculated the correlation with the estimated FRP in the FFNNs."

B. Input variables for the FFNN model as well as for computing FWI are discussed in lines 95-96. Are these taken at 12 Noon local time? Or daily averages? For precipitation, was it the 24-hour sum or

average? Ideally, 12 Noon UTC values or daily maximum temperature, minimum relative humidity, average wind speed and 24-hour accumulated precipitation should be used. Please specify in the manuscript.
: Sorry for missing an important information. It is 24-hour averaged value. This temporal treatment in the input variable is identical to the input for the FWI. We noted this information in the revised manuscript.

C. In lines 118-121, some techniques related to the neural networks are just mentioned. Please explain the terms or provide citations which explain the terms like dropout rate, batch normalization, and ReLU function for readers not familiar with these. Also, what does a dropout rate of 0.2 physically signify? Please explain these in the text.
: Sorry for the possible inconvenience. We added a short description of each term with references in the revised manuscript.

3. **Miscellaneous**
A. An anomalous behaviour of increasing FRP with increasing RH when humidity is less than 30% has been reported in section 4 (lines 237-241). This is supported by an argument reported by Abatzoglu and Kolden (2013). However, I could not find any such observation in the cited article. Can the authors clarify exactly where in the paper this is mentioned? Also, fire activity usually translates to fire frequency. FRP gives us more information about fire intensity. It is unusual that fire intensity at 20% RH will be lesser than at 30% RH.
: Sorry for causing the inconvenience. In Abatzoglou and Kolden (2013), even though they showed that the positive correlation between the soil moisture and the burned area in non-forested regions (compared their figure 4 to figure 3), they did not discuss its mechanism. About this positive relationship, Xystrakis et al. (2014) discussed that the increased wetness is associated with the build-up of the fuel during the drying season, which eventually contribute to increase the burned area. This is consistent with our results to some extent that the increased FRP with the increased relative humidity occurs in the low relative humidity regime, which might imply that the increased relative humidity contributes to increase the fire activity (i.e., FRP) by increasing the fuels to burn. We added the aforementioned discussions in Line 308-314 of the revised manuscript.

B. In supplementary figure S5, the result of precipitation and temperature values are almost similar. Line 182 in the main text contradicts this.
: Sorry for causing the confusion. This sentence is modified as follows, “It clearly indicates that the RH2m are the main factors influencing the accuracy of the FRP estimations in the FFNNs.”.

C. In lines 167-170. Supplementary figures S2 and S3 have been discussed. These figures refer to Fig.2 in the main article which has not been discussed yet. Such issues in the chronology of the paper disrupt the course of reading. The whole of the result sections need to be reorganised.
: We looked through the paragraph describing Supp Fig. S2 and S3, but cannot find any statements mentioning Figure 2. Supp. Fig. S2 and S3 is the correlation skill after managing the data, which is the follow-up analysis of Figure 1, therefore, we think the results section is organized in orders.

D. Lines 186-194: Inferences related to FFNNs are discussed here, however, the inferences from the FWI-based method (Figure 2d-f) are in lines 201-208. These inferences related to the same figure should be kept together. Also, between Fig 2c and Fig 2f, how do we conclude which is more correct?
: Sorry for the inconvenience. We modified the orders of the paragraph as a reviewer suggested.

E. Line 195: Explain the physical interpretation of the LRP method for unfamiliar readers. How is the result obtained from this analysis different from the sensitivity study with FFNNs? It is still unclear which method gives a more accurate estimate of the major factors influencing FRP. Also, explain

supplementary figure S7 in the supplementary or main text.

: We added a description of the LRP method with a physical explanation in Line 261-274. It is hard to say which method between the LRP and the sensitivity experiment by replacing an input variable is more accurate to estimate the relative importance, and that is why we introduced both methods in the article. The LRP method provides an output so-called 'relevance score' for each variable, and, main Fig. 5 is the correlation skill between the estimated FRP and the relevance score. This information is also added in the revised manuscript.

F. Line 221: Why was this particular 0.05 number chosen? The sentence in lines 223-225 is unclear. Please rephrase. Also, figure S8 can be moved to the main article.

: We tested a threshold of 0.05 and 0.1 of correlation improvement, and found that the general conclusion is still obvious with different threshold (Figure A-3). The final threshold is chosen as 0.05 as the large number of selected grid points would provide a rigorous result. We noted this point in Line 294-296 of the revised manuscript.

Sentences in lines 287-289 is modified as follows "… (2) RH2m is the most sensitive variable for FRP estimation in FFNNs (green color in Fig. 2c), and (3) PRCP is the most sensitive variable in the FWI-based model (blue color in Fig. 2f)."

Also, we included Figure S8 in the main figure as a reviewer suggested.

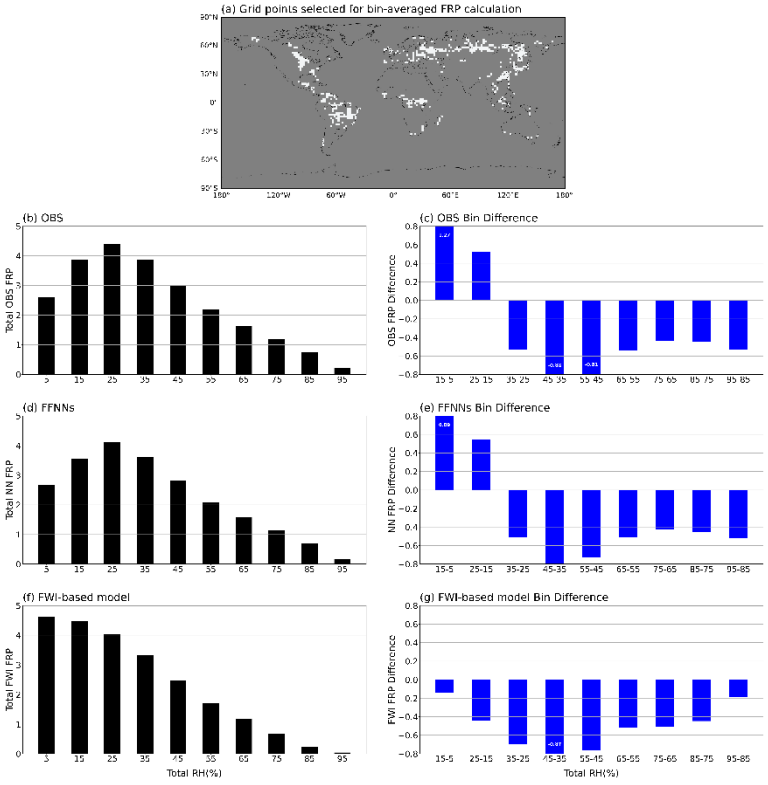

Figure A-3. Same is main Fig. 6, but for the threshold of 0.1 for the correlation skill improvement.

G. Line 233: What is the relevance of this reference here?

: Sorry for the confusion. We changed the references to clearly state the relationship between the relative humidity and the wildfire as follows.

Papagiannaki, K., Giannaros, T. M., Lykoudis, S., Kotroni, V., & Lagouvardos, K. (2020). Weather-related thresholds for wildfire danger in a Mediterranean region: The case of Greece. Agricultural and Forest Meteorology, 291, 108076.

Ying, L., Cheng, H., Shen, Z., Guan, P., Luo, C., & Peng, X. (2021). Relative humidity and agricultural activities dominate wildfire ignitions in Yunnan, Southwest China: Patterns, thresholds, and

implications. Agricultural and Forest Meteorology, 307, 108540.

H. Line 257 onwards: Is the precipitation considered here the daily average or the daily sum? The ideal approach would be to use daily sum. Clarify.
: We utilized the daily averages for the precipitation. We noted this point in Section 2.1.2. As an input of the statistical model, a difference between daily average and daily sum should not affect to the final results, as a simple multiplication of the constant value converts daily averages into daily summation.

I. Line 285-289: Is quadratic the best fit? Also, please explain why the heightened sensitivity of FRP to precip is incorrect. That very well might be the case.
: In our view, the quadratic fit works relatively well for the bin ranging from 0.1 mm/day to 5 mm/day. It is to roughly quantify the changes in the observed or estimated FRP to the changes in the precipitation. The heightened sensitivity in the FWI-based model is caused by the abrupt drop of the estimated FRP to the increase of the precipitation amount from 0 to 3 mm/day as shown in main Fig. 7e. On the other hand, degree of the changes in the FRP in the observation or FFNNs according to the changes in the precipitation is systematical weaker than that in the FWI-based model. The discussion about the heightened sensitivity of FRP to precipitation in the FWI-based model is elaborated in Line 356-365 as follows
Line 356-365 : "As a result, the regression coefficient between the FRP estimation and the PRCP is systematically greater in the FWI-based model. For observations, the quadratic coefficient is 0.022 MW/(mm/day)^2 (black in Figure 4a), and that for the FFNNs 0.023 MW/(mm/day)^2 (black in Figure 4c), denoting similar amplitude. On the other hand, the FWI-based model is 0.036 MW/(mm/day)^2, which is almost twice to that of the others (black in Figure 4e). This suggests that the FWI-based model is more responsive to changes in PRCP, resulting in a more pronounced FRP decrease with increasing PRCP. This excessive sensitivity in the estimated FRP to PRCP changes can contribute to the excessive influence of PRCP on the FRP estimations in the FWI-based model, as shown in Figure 4f."

J. ERA5 has been used in this study. What will be the scenario if we use some other weather data/ observation? How sensitive is the FFNN model to the kind of dataset used?
: It might be worthwhile to check as a reviewer asked, however, we did not test with different dataset (e.g., MERRA2, NCEP), as 1) ERA5 is one of widely used reanalysis product, and 2) it is important to compare FFNNs and FWI by utilizing same dataset.

K. In line 136, it is said that the entire period is divided into three-year periods. But if the test period is 1st Jan 2001- 31st Dec 2004, it is actually a four-year period. Such confusion should be removed.
: Sorry for the mistake. It is corrected to 'four-year periods' as a reviewer pointed out.

L. Supplementary figure S4: Why these particular stations and years? This is nowhere mentioned in the methods/ results/ supplementary text. Please elaborate.
: We picked the stations and years to check the simulation quality of the major wildfire events.
(a,b) 2019 Amazon wildfire : https://en.wikipedia.org/wiki/2019_Amazon_rainforest_wildfires
(c,d) 2016 Congo wildfire : https://www.mdpi.com/2072-4292/8/12/986
(e,f) 2003 Siberian wildfire : https://www.tandfonline.com/doi/pdf/10.1080/01431160802541549
(g,h) 2007 Southern China wildfire : https://link.springer.com/article/10.1007/s13753-017-0129-6
We added this information in Line 216-228 of the revised manuscript.

M. In Line 300, FRP behaviour or rather correlation of FFNN estimated FRP and observations over certain regions are discussed. However, why certain regions show high/low correlation has not been discussed anywhere.
: The mechanism of the improvement is given in the subsequent paragraph; the sensitivity

experiments with the examination of the relationship between FRP and RH2m, or PRCP in Figure 3 and 4 is to understand the improvement of the FFNNs compared to the FWI-based model. That is why we included grid points whose correlation skill improvement in the FFNNs is greater than a threshold value of 0.05 for plotting Figure 3 and 4. As this would be good to clearly refer that the sensitivity experiments is to understand the improvement of the FFNNs, we modified the sentence in Line 373 as follows.
Line 379 : "To identify the mechanism of the skill improvement in the FFNNs, a series of sensitivity experiments were performed …"

4. **Regarding figures:**
A. What does except 0 mean in supplementary figure S2? Mention in the text/ supplementary the methodology of how these figures were obtained, their necessity and inferences.
: Sorry for the lack of explanations. It is to evaluate the forecast skill only with the fire events. In some case of many non-wildfire days (i.e., FRP = 0), the skill is quite high even though the model always predicts a constant value (i.e., 0). We added the brief meaning of FRP > 0 in Line 210-211.

B. A reader has to constantly toggle between main article figures and supplementary figures. The order in which figures are discussed is random.
: Sorry for the confusion. We double-checked that the figures are discussed in order throughout the revised manuscript.

Minor comments:
1. Lines 26-27: This sentence requires rephrasing. The FFNNs captured the 'relationship' accurately. Correlation is a method by which we can ascertain this relationship. Also, what do the authors mean by 'as well as precipitation'? Is precipitation well correlated or not? Please clarify.
: Thank you for pointing this out. The corresponding sentence is corrected as "The FFNNs accurately captured the observed nonlinear RH2m-FRP and precipitation-FRP relationship."

2. Line 28: Ideally, we expect an inverse relationship between FRP and precipitation. How is this 'excessive' relationship a concern in this context?
: Sorry for the possible misunderstanding. We deleted the corresponding sentence in the revised manuscript.

3. Line 34: What kind of fires? Wildfires or agricultural fires? Not all fires cause ecological and socio-economic impacts. Please specify.
: The corresponding sentence is modified with references as follows "Wildfires are inflicting substantial terrestrial and economical impacts in numerous regions globally (NOAA, 2005; Bowmman et al., 2009).".

4. Lines 36-37: I see no relevance of this statement here, as the authors have not discussed the ignition factor anywhere in the manuscript.
: The corresponding sentence is deleted in the revised manuscript.

5. Line 49: Moisture codes provide no exclusive information about any deceased organic matter.
: We tried to convey the fact that the initial spread index (ISI) and buildup index (BUI), which combines to obtain the FWI index requires the moisture code as an input. To avoid the confusion, we rephrase the corresponding sentence as follows "…, the moisture codes are provided as an input of the fire behavior indices, such as initial spread index and buildup index to finally calculate the FWI, providing an estimation of fire intensity."

6. Line 63: It is unclear what the authors are trying to convey here. Also, how is it relevant to the rest

of the paragraph?
: Sorry for the possible misunderstanding. We modified the corresponding sentence as follows “while the relationship between the meteorological variables and the fire activity varies significant from regions to regions.”.

7. Line 65: How can we ‘calibrate’ sensitivity? Consider changing it to ‘estimate/ calculate’.
: The corresponding phrase is modified as follows “To understand the varying sensitivities of wildfire activity to the meteorological variables from different regions,”

8. Line 70: I guess there is a typo. It should be ‘fuel’ moisture code instead of fire.
: Corrected as a reviewer suggested.

9. Line 116: consider changing ‘responsible for’ with ‘representing’
: It is corrected as a reviewer suggested.

10. Line 127: there is a typo in the equation. It should be yi instead of y1.
: It is i, but looks like 1 or l in the manuscript, once we put the upper brackets notation. To avoid the confusion, it is now modified n.

11. Line 132: Is the title ‘Experimental design’ ’for section 2.3 suitable? It is rather an explanation of the cross-validation strategy only.
: It is corrected as a reviewer suggested.

12. Line 143-145: Rephrase the sentence. Anomalies were compared and assessed for accuracy.
: Thank you for pointing this out. The corresponding sentence is rephrased as follows “The FRP anomalies, which were calculated by subtracting the estimated daily climatology during 2001–2020 period, were compared and assessed for the FRP estimation accuracy.”.

13. Line 161: Please provide a map of FRP climatology in the main or supplementary article. What does the citation here convey?
: We deleted the citation, and provide a map of FRP climatology as Supplementary Fig. S2.