We very much thank the reviewers for their thorough and constructive reviews on the manuscript. Below we give a detailed response to each comment and we explain how we used the comments to improve the manuscript.

The reviewer comments are shown in italic fonts, our answers are in blue (normal fonts). **Proposed changes in bold.** When referring to figures in this response letter, the prefix 'R' is used before the figure number.

Reviewer #2 (Francesca Di Giuseppe):

We are very grateful to the reviewer for the very thorough review. The comments and proposed suggestions were very helpful to improve the paper.

The fire weather index improved for boreal Peatland Using Hydrological modelling and satellite based-L band microwave observations

By Mortelmans et al.

The paper investigates the use of hydrological model outputs specifically designed for peatland to replace part or all components of the FWI danger rating systems. It responds to a known limitation of the system that, while developed for above ground fires, has been, and still is, commonly used in global fire forecast systems.

It is well known that the FWI does not correlates very well with fire activities where the fuel is very dissimilar to forests. So I believe it is a valuable idea to address this problem. The paper is well written and explores in very much details the source of predictability that can arise by employing the water table and/or the soil moisture content. Certainly I recommend publication of the manuscript as it provides an useful framework to improve fire danger prediction.

However given the results that clearly highlight that the direct use of the water table is a better predictor than the FWI for peatland, I am asking the author if suggesting to rescale this variable to match the value of the FWI is really the right thing to do? Is it going in the right direction wanting to retain the infrastructure of the FWI at all costs? The water table is a physical measure that could be measured and even adjusted through the use of satellite observations while the FWI is an empirical transformations of the fire intensity that its calibrated on a specific ecosystems. To CDS-match the FWI seems a weird think to do if you have a metric (the FWI) that is not very correlated to the fire activities in peatland. Even more so as you indead then evaluate against fire activity and not fire danger.

The question for me would be why do not directly rescaling the water table by training it to detect actual fire activities ? Along those lines I developed the FOPI which was trained on observed fire activities and did not attempt to rescale the FWI while still using the FWI as a driver for the fire weather component.

Importantly, when you train for fire activity, your output is somehow a probability which is more intuitive to understand. Another benefit of using directly the water table would be that when these variables are improved by the model or the assimilation system, this improvement shoud benefit the fire danger indicators in cascade. With the empirical structure of the FWI if you improve the moisture content estimation of the duff layer do you really improve the DMC ?

Indeed the motivation to be using the FWI infrastructure is provided in the paper. The FWI is easily interpretable by fire agency. This is a good motivation. Still I am asking the author to elaborate a bit as I think they should discuss what would be the benefit of also shifting toward a more physical based indicators of fire danger.

We appreciate the thoughts about a possible fundamentally different algorithm for estimating fire danger and/or fire activity in peatlands. We agree shifting to another indicator may have many advantages. However, we also see some important practical disadvantages, especially with regard to the short-term impact of our study. Intentionally, we decided for the very conservative approach of CDF matching for two main reasons:

1) Proof of the positive impact of peat-specific hydrological information: We believe that it is a very transparent and valuable approach to prove the impact of peatland hydrological information by incorporating it in an existing fire danger prediction system. Having this very well known reference system, users may be more easily convinced to adopt the use of peatland hydrological information. In contrast, the performance of a totally different system is much more difficult to specifically analyze for the impact of the new peatland hydrological information. We see such an issue, for example, in the interpretation of the results in Mezbahuddin *et al.* (2023) in which a model with peat hydrological input was shown to perform better than one with weather data only. However, it remained unclear how well their reference system, i.e. a new machine learning setup, was able to extract the information from the weather data. Perhaps the weather-based machine learning model performed even worse than the original FWI? This was not tested in Mezabahuddin *et al*. (2023). In contrast, in our study, we show the value of the peatland hydrological input in an established system, i.e. the value of the peat hydrological information becomes unambiguously clear.
2) Interpretation based on established FWI framework: A completely new approach will complicate the interpretation of the results while the new PEAT-FWI can be interpreted in the same way as the results of the original FWI. With the different experiments presented in the paper, we show that for the early fires, keeping part of the original FWI structure in the system is better than solely relying on the hydrological information.

We have included the following text in Section 4.2.2: Opportunities for operational FWI products and advancing fire danger models

We employed CDF-matching to adjust select moisture codes within the FWI over peatlands. This conservative approach makes the inclusion of hydrological variables more accessible to operational centers and the user community accustomed to the existing FWI system. It provides a transparent and valuable means to demonstrate the impact of peatland hydrological data.

By contrast, using a totally different system would make it difficult to evaluate any performance gain specifically introduced by new peatland hydrological information. For example, Mezbahuddin *et al.* (2023) showed that a machine learning model with peat hydrological input performed better than one purely relying on weather data. However, this study did not offer a comparison with the original FWI system, leaving uncertainty about the true effectiveness of the peatland hydrological data. A study comparing our proposed PEAT-FWI against such a machine learning algorithm could offer new perspectives for future peatland-specific fire danger rating system frameworks.

Few more comments:

1. *I would put figure 4 and all the detailed explanation of the ROC curve derivations in an appendix. I think is quite a standard metric and while is nice to have a refresh in the paper it is distractive to have it in the methids section.*

We agree and moved it to Appendix 1.

2. *I would not mention satellite L-band Microwave Observation in the title. The focus here is not much how the water table is calculated but the use of the Hydrological model. Probably the same results would hold with another model.*

We agree and based on this comment and on a comment from of Reviewer 2, we changed the title into:

Improving the Fire Weather Index System for peatlands using peat-specific hydrological input data

3. *How easily available would this prediction of water table be for the forest agency to be able to calculate the new FWI-PEAT Index? Maybe a short discussion of the complexity of creating this index could be provided.*

The hydrological data is freely available to download from https://nsidc.org/data/spl4smgp/versions/7 and has a latency of 2.5 days. We decided on the CDF-matching of the hydrological data to keep the complexity of this index rather low and to allow for an easy interpretation of the results. We will add a short paragraph to the discussion about the operational use of the proposed index on the availability of data and complexity of creating the PEAT-FWI.

4. *You use fire ignition as a fire activity indicator but FWI expresses a measure of fire intensity. I am not sure if this has an impact on your results.*

This definitely has an impact on the results. Based on your comments and those of Reviewer 2, and to be more in line with your previous work (Di Giuseppe *et al.*, 2020; https://doi.org/10.5194/nhess-20-2365-2020), we decided to change from fire ignition to fire presence, i.e. every day that a fire was present is taken into account instead of only the first day. This changed our results for the better. We will replace Figures 7, 8, and 9 with the following figures:

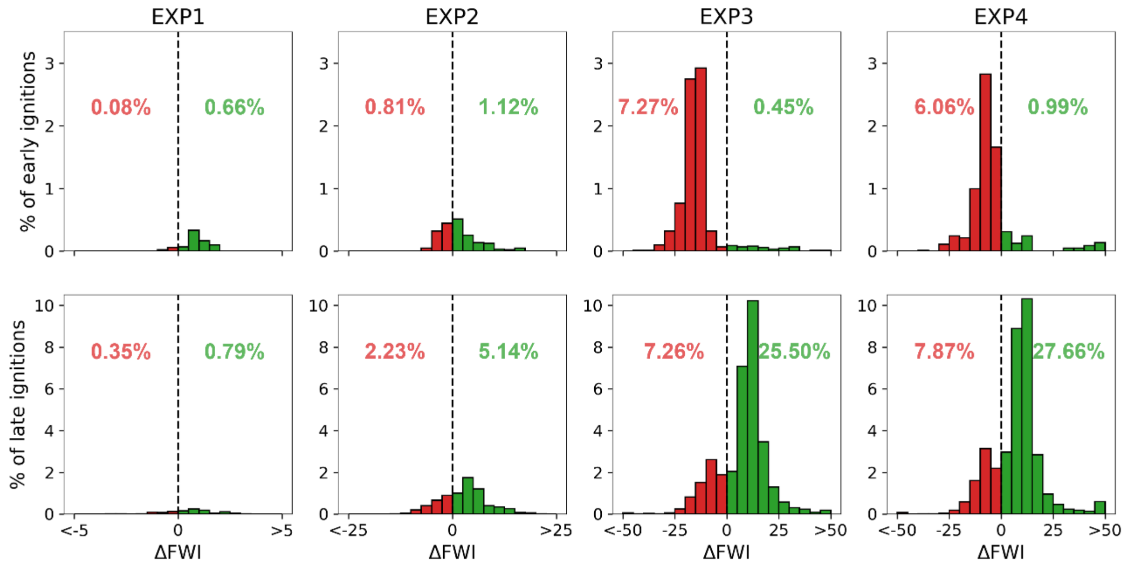

Figure R1: Replacement of Figure 7. This figure is now based on the evaluation against fire presence, instead of fire ignition alone.

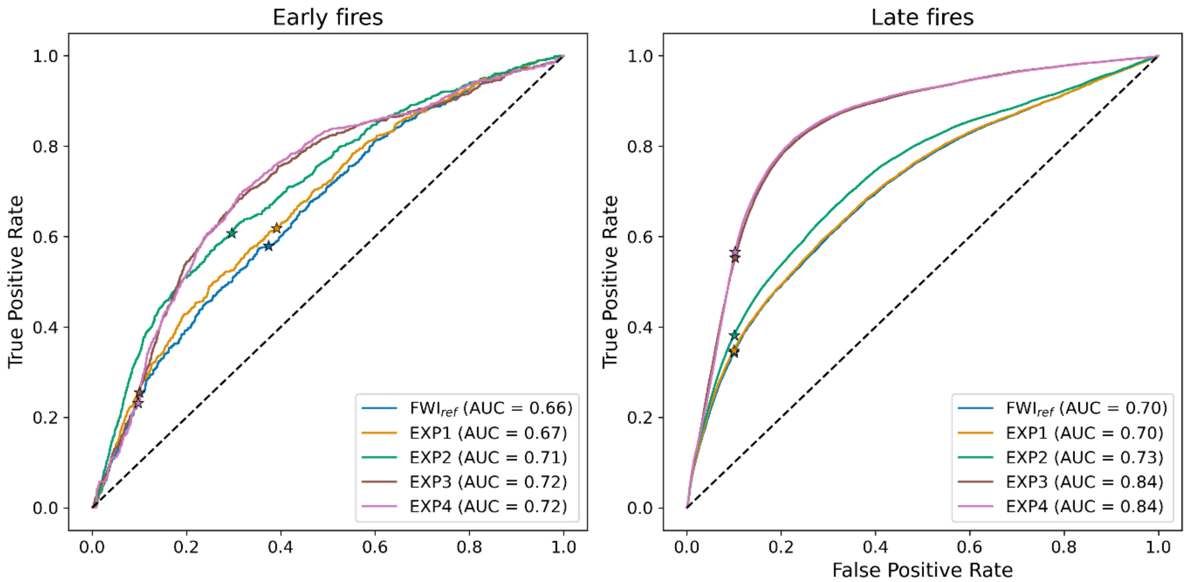

Figure R2: Same as Figure R1, but as replacement for Figure 8.

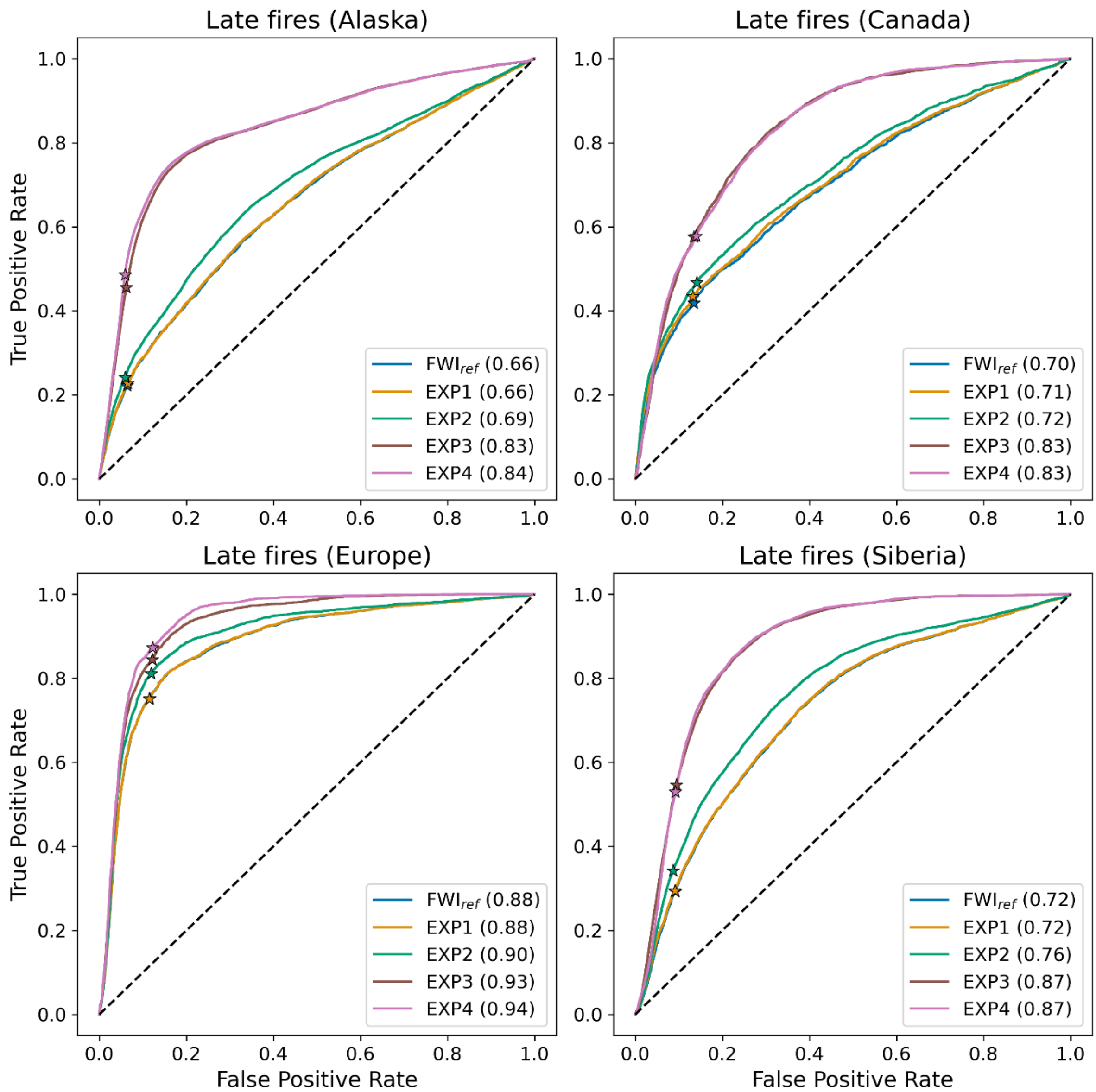

Figure R3: Same as FigureR 1, but as replacement for Figure 9.

5. *I haven't quite worked out why a CDS matching would remove a bias ? Could you give me more datails ?*

The bias we were talking about, is a spatial bias in the PEATCLSM product. PEATCLSM has likely a dry bias e.g. over the Boreal Plains as indicated by field measurements (Bechtold *et al.*, 2019). By applying the CDF-matching on the timeseries of each grid cell individually, we remove any possible spatial bias in PEATCLSM data. In other words, the spatial pattern of long-term statistics of PEATCLSM is matched to that of the FWI. This however, assumes that the FWI data are unbiased. If there is any spatial bias in the FWI, this will also be introduced in the new PEAT-FWI. However, as stated earlier, we are targeting in this paper the demonstration of the value of the temporal information in peatland hydrological data in an established fire danger system, ignoring any possible potential of the hydrological information to reduce spatial bias in the original FWI.

To further clarify this in the paper, we changed lines 139-145 of the old manuscript to:

… While this ensures that the CDFs of the PEATCLSM output variables and those of the corresponding FWI moisture codes match, the approach preserves the dynamical features (short-term and long-term anomalies as well as seasonal dynamics) of the PEATCLSM output. By performing this CDF-matching on a per-grid-cell basis, any spatial biases present in the PEATCLSM output, such as a possible dry bias over the Boreal Plains in Canada (Bechtold *et al.*, 2019), were removed. It is worth noting that any spatial bias in (the moisture codes of) the FWI will thus be maintained in the new PEAT-FWI with this per-grid-cell approach. However, our primary objective is to underscore the value of temporal peatland hydrological data within an established fire danger rating system.

6. *Finally I think the discussion is a bit fragmented. For exemple, the limitations of the GFA for peatland are nicely discussed but are not put in the contest of how (or if) they could affect the results presented. Similarly for the discussion about the new index. A clear statement of in how many more cases you are likely to get a good prediction compared to the use of the standard FWI (which you can read from the ROC curves) would certainly benefit the readibility of the conclusions.*

Based on your suggestion and those of Reviewer 2, we revised the discussion, focusing mainly on the net change of the different experiments compared to the reference FWI.

For the net change, i.e. in how many cases we would get a better prediction with PEAT-FWI compared to the reference FWI, we focus on the difference in AUC. Based on the new ROC curves, an improved fire danger prediction is achieved by EXP4 in 6% of the cases for the early fires and in 14% of the cases for the late fires. We added the following part to the discussion:

When subtracting the deterioration from the improvement, one can evaluate the net effect of the PEAT-FWI for the different experiments. This shows that EXP1 shows a small net improvement compared to the FWI_{ref} for both the early and late fires of 0.58% and 0.44%, respectively, i.e. 0.58% of the fires in the early season are better predicted with EXP1 than with FWI_{ref} using the 70th percentile threshold. EXP2 also shows a net improvement for both seasons (0.31% and 2.91% for the early and late season, respectively). For the early season, EXP3 and EXP4 show a net deterioration of 6.82% and 5.07%, respectively. For the late season, both experiments show a large improvement of 18.24% and 19.79% for EXP3 and EXP4, respectively.