The article describes a new index for improving the extreme fire behaviour predictability by considering the vertical atmospheric profile, namely, to evaluate if conditions are favourable to convective fire behaviour. In general, the article is well written, the methodology is interesting, and the topic is relevant for the fire research community and for practical applications. I have however some concerns and comments regarding the current version of the manuscript.

My main concern is that a discussion comparing the proposed index with existing works in the literature is lacking. In the introduction, the authors mention the Haines Index but quickly dismiss its usefulness citing Pinto et al. 2020 and the saturation problem of the Haines Index. The cited work by Pinto et al. 2020 does however address the same problem, proposing an enhanced Fire Weather Index that combines the FWI with the “Continuous Haines Index”. I was expecting to see some discussion or comparison to put the proposed EFBI in the context of those existing indices. For instance, since the EFBI uses the vertical profile of the atmosphere, I would assume that the authors have all the data necessary to compute the Continuous Haines Index (that is based on the temperature and dew point at two pressure levels) and the enhanced FWI. I believe such a comparison would be of great interest to the fire community and would set a new validation standard for future works on this topic.

My second concern is that, to my understanding, the EFBI was computed with one hour time steps and aggregated using the minimum, the maximum and the average and the three variables were used as predictors when applying the decision tree and the multilayer perceptron. Since the FWI is computed only daily it is possible that part of the improvement in the accuracy is due to the indirect incorporation of sub-daily conditions. It would be relevant to compare the change in accuracy when considering the minimum, maximum and average aggregates of EFBI individually before including the three aggregates all together.

Other comments

L20-21: I suggest rewording this sentence.

L116-117: Are the 222 small fires a subset of a larger initial selection matching the
described criteria? I would expect the number of small fires to be higher.

L136: If the initial day of the time window is increased by two days wouldn’t the correct day be missed for the cases where the MODIS MCD64A1 gives the correct or the following day? How is the 2-day value selected?

L141-142: Consider updating to: “using Scikit-learn (Pedregosa et al., 2011).”

L179-180: I suggest updating to: “than the percentile and value of drought code”. In fact, looking at Figure 4, the MI for dc_percentile is not significantly higher than zero.

Figure 5: The EFBI is named as “Index” in the figure. Consider renaming or updating the figure description to make the interpretation clear. The same comment applies to some of the other figures.

L197: What activation function was used in the multilayer perceptron?

L198-199: What is the standard error in these cases?

L199: I assume FWI is referring to the set of FWI components in percentile form, please clarify if this is the case.

L203-205: It is certainly expected that by removing the 50 most often misclassified cases, out of a total of 445 cases, the accuracy would rise substantially. Unless there is some manual checking of these 50 events, I don’t see the point of this exercise.

L231-232: The results of the case study are interesting; I would further comment that speeds greater than about 1 km/h are only present for EFBI values above ~220. This result is close to the threshold of about 200 in Figure 5 for “Index_max”.

L270: I suggest adding some comment regarding the need for future research towards constructing datasets of fire behaviour type and higher temporal resolution fire progression.

L274: The EFBI is misspelled as “EBI”.

L281: “ha” is missing after 10000.

L291-292: Does the 4-hour computation time considers the time to download the GFS data?