We would like to thank Anonymous Referee #3 for their helpful comments and suggestions. We address the referee's general and specific comments below:

General comments:

Comment: The authors have done a thorough job with the large task of overhauling an FWI system for an entire country, and I applaud them at that. Perhaps the issue I see not so much scientifically, but rather philosophically with the way in which the FWI was modified here. In principle, the FWI in Canada was designed to replicate fuel moisture conditions in a standard pine stand, and as a result is calculated using the same equations across the entire country (and indeed around the world where used), save the overwintering component of the DC. Your percentile-based approach appears very much scientifically sound, but there are complications to be found where in theory each grid cell may have differing thresholds or structure in the FWI, so that simpler operational guidelines (such as the commonly used FWI>19 threshold in Canada) would be more difficult to communicate. In practice in a large country like Canada, each jurisdiction has differing thresholds categorizing FWI into for "High", "Extreme", etc hazard levels, and these vary by climate. In practice, since an BUI of 40 is results from the same set of weather beforehand across the country, it is easier to train staff to the expectations of fire behaviour at those levels for a particular fuel complex. In any case, my concerns are not so much scientific as applied, but I think the application of such a variant of the FWI and its implications for firefighter training are worth considering.

***Response :** We agree that the benefits of an implementation of the native FWI system, with a comprehensive understanding of how it relates to fire behaviour UK fuels, would likely be superior to the approach proposed here for the purposes of operational understanding. However as highlighted in the manuscript, very limited data are currently available to perform such an adaptation in the UK. We see the approach taken here as somewhat of a 'stop gap' approach until these relationships are better understood.*

Comment: My other larger consideration is with the overlay of fuel type in this model. When you build Theil-Sen models by land cover type (as on page 7016), what happens when the land cover type changes?

***Response**: We envisage that an operational system built upon elements of this work would need periodic updates in all of these areas. In terms of land cover, the analysis presented here is only intended to give an insight into how effective percentiles of the different FWI components are for broad land cover types as a whole. However, we do envisage that if/when elements of this work are integrated into an operational forecasting system, each grid cell area of the UK would indeed have one or more associated fuel types derived from up-to-date land cover mapping. In terms of fire danger assessment, increased weighting could then be given to the FWI components that have the strongest relationships with fire activity for that particular land cover type. Ideally, once such a system is established, national land cover data would be updated as often as is reasonably possible, to minimise the issues arising from land cover changes. While updates to e.g. the UK Land Cover Map (http://www.ceh.ac.uk/services/land-cover-map-2007) are relatively rare, and suffer from a considerable lag between the year of observation and publication (LCM 2007 was only available in 2011, 4 years after the raw data were collected), more reliance could be placed upon products such as the Ordnance Survey Master Map (https://www.ordnancesurvey.co.uk/business-and-government/products/mastermap-products.html), which are updated more frequently. Additionally, if the system were used at a* *regional level (e.g. by county Fire & Rescue Services), local knowledge could be used to draw attention to major land use changes likely to affect fire activity, allowing the model to be updated relatively quickly.*

Comment: Would the percentiles here per land cover type be the only time that this FWI is calculated? Would future fire events (and more extreme weather) then also change the calculations behind the FWI? This could present problems for the deployment of such a model in the future.

Response: *In an operational system, as highlighted in section 5.2, we would hope that percentile data would be updated periodically (e.g. annually) as additional FWI data are made available, in order to better define 'extreme' values of the FWI components. Additionally, as more and improved fire data become available in the IRS database, we would anticipate that the relationships between fire activity and FWI component percentiles would also be updated when appropriate.*

Comment: My above concerns notwithstanding, the manuscript is skilfully clean and well laid out. I have very few editorial comments, listed below alongside more specific comments:

Specific comments:

Comment 1: The title on the right sidebar of the journal lists "FFWI System" when the proper name is "FWI System"

Response 1: *We will request the manuscript typesetters to change this.*

Comment 2: Page 7004: Given that the MOFSI has the same categories across the entire UK, how badly would a model with differing FWI thresholds per region perform across the UK?

Response 2: *We believe that a model with differing regional thresholds would indeed perform better than the current MOFSI approach; however there is often considerable variation in extremes within relatively small spatial regions, as e.g. as indicated by variation observed in the 99th percentiles of the FWI component shown in figure 4. These data could in theory be used to identify regions of similar extremes (e.g. diagonally across the UK, moving from NW to SE), and thresholds derived for each – perhaps this is a potentially useful application of this work in future. Our current approach of using percentiles however is well suited to investigating the relationships between FWI and fire activity nationally, as it avoids the introduction of artificial geographical boundaries in fire danger.*

Comment 3: Page 7019: Given a ~5 month fire season, would a 99th percentile not occur 1-2 times per year, rather than "one in several year". For indices that are less wind-sensitive and vary less per day such as BUI, I could see that the 99th percentile would indeed be only every few years.

Response 3: *We first discuss the idea of "one in several year" extreme conditions on p7011. While it would be entirely defensible to work on the basis of a fire season of ~5 months, our calculations and discussion here are based upon the 3-month periods that broadly equate to* *the UK's distinct meteorological seasons (Dec-Feb = Winter, Mar-May = Spring, Jun-Aug = Summer, Sep-Nov = Autumn), and are in-line with the temporal divisions we have used throughout the manuscript. These seasons also broadly coincide with periods of distinct fire activity in the UK (e.g. see figure 2c – also, burned area data reinforces these temporal trends, but have yet to be added to the manuscript). Our analysis makes use of 287, 276, 321 and 334 days of winter, spring, summer and autumn FWI data, respectively, giving a total of 2.9, 2.8, 3.2 and 3.3 days in each season exceeding the 99th percentile value. Given that the dataset is ~3.5 years in length, this equates to ~ 1 day > 99th percentile per year, on average. However as highlighted on p7011, extreme FWI component values are likely to be auto-correlated, suggesting that the days on which FWI > 99th percentile may be grouped together temporally in one of the 3.5 years. As you propose however, it maybe that less clustering occurs in the more variable components e.g. FFMC, ISI. We will investigate the FWI dataset to determine whether this is the case, and amend the '1 in X year' statement as required.*

Comment 4: Page 7019: the detailed narrative for the Berkshire fire is far too detailed for a paper of this broad scope. I'd suggest cutting this back a fair bit.

Response 4: *This description will be shortened in the final manuscript.*

Comment 5: Page 7021: Is there a statistically-sound way of convincing the reader that the skill of your model on these extreme weather days is better than random chance? What is holding you back from just painting everything exceptional? Can you link this back to AUC or some other statistic that would account for false positives and better reflect the true skill of the model?

Response 5: *In theory, there is nothing stopping everything being classified as exceptional here, though clearly in practice the model would be unusable in this case due to the number of false alarms that would be generated – we will update the analysis here to include statistical testing to better demonstrate the model skill.*

Comment 6: Page 7024: The paragraph on lines 7-20 is excellent, and more discussion of this nature should be fleshed out in the paper in my opinion.

Response 6: *This section will be revisited in the final manuscript, and added to where appropriate.*

Comment 7: Page 7038, figure 2: in panels b and c, is the trend similar when examining area burned rather than just the number of fires? Given that fire size varies by a few orders of magnitude, a few small fires ~1 ha may be quite different that one fire of 100-200 ha.

Response 7: *We propose adding a summary of burned area during the study period by landcover type to Table 1. Figure 2(c) will also be amended to indicate the monthly distribution of burned area relative to fire occurrence. The spatial distribution of burned area has been investigated, however this was found to be extremely similar to the spatial distribution of wildfire counts. We will include a statement to this effect, and discuss more generally these findings in Section 5.1.*

Comment 8: Figure 7: in the rank-percentile curves, is there any consideration of fire size? Is there a good justification for not weighting this by fire size somehow?

Response 8: *The rank percentile curves used here do not modify the approach of Eastaugh et al. (2012), and as a result do not take account of burned area. Some initial exploration of the relationships between burned area and FWI has been carried out; however these relationships appear to be quite weak. While this lack of correlation may be a true reflection of reality, it may also be related to the quality of burned area data in the IRS. The IRS dataset has undergone a considerable amount of QA/QC, however it is a recent development, and we believe that there may be relatively large uncertainties associated with the accuracy of the burned area estimates and the thoroughness with which these data are recorded, as their collection is not a high operational priority. Additionally, use of burned area data may also introduce a degree of spatial bias. Regional UK FRS' have quite different response strategies to wildfire, in terms of both response time and number of appliances deployed. In areas where wildfire response is considered to be of lower priority than in others, area burned may be much larger than in other areas, even if weather conditions are very similar (because the fire is able to spread for longer) – this was the principle reason for the Scottish Wildfire definition being based upon multiple metrics, in addition to area burnt. For these reasons we felt that until the burned area data have been evaluated in much more detail, it was better to use the occurrence data that the IRS database provides as the basis for the fire activity analyses, and only use burned area to highlight general fire activity over the study period, and to indicate whether a fire maybe considered a 'wild' fire or not, rather than using it in a more quantitative fashion.*