

The authors improved the manuscript and addressed the reviewers' remarks in most parts of the manuscript. However, the manuscript still lacks consistency in terminology and appropriate order of arguments/ statements to have a good reading flow. A critical discussion of the study's results is still missing. The authors point out flaws in their results quite clearly, but do not set them in context with other studies using similar methods, datasets, or parameters.

These issues are of minor magnitude but should be addressed. Therefore, I am happy to read the revised manuscript once again after the following points have been addressed:

General minor comments

1. Terminology
 - a. Be consistent in how you refer to figures in your text (see comment 22): You use Figure 2a (line 107), while in line 99 you use Fig. 1 and 2. Also make sure the blank space is consistent between figure and figure number as well.
 - b. Name your metrics consistently, e. g. RSME and skill are used interchangeably. I pointed it out in comments 35,39 and 41 but there might be locations I oversaw. Please check this carefully before resubmission.
2. Definitions:
 - a. You use relative statements throughout your paper. Make clear what are low / high FWI values or what is a good correlation. Add numbers to these relative terms and state why you use these numbers as thresholds to define a FWI to be high or low, or a correlation to be good (see comments 20 and 24)
 - b. Same for short- and long-term forecasts. Define these terms in your text, e. g. short term (up to 3 days) and long term (more than 3 days).
3. General comment on verification and Fig. 2b:

You state several times that your approach does not work well for small/ low FWI values. What are the thresholds for you to consider FWI to be low? From your study I read that this is below 1, but from the EFFIS danger classes (<https://forest-fire.emergency.copernicus.eu/about-effis/technical-background/fire-danger-forecast>) low FWI values are considered to be below 11.2. Your underestimation of the FWI in the analysis mainly affects this relevant scale (10 -100) (see Fig 2b).

 - Can you mention this in the description of the figure (i.e. line 110)
 - Discuss and state stronger that your calibration does not address this discrepancy since you calibrate on the (orange) analysis and not on the (grey) observations density?
 - What are the impacts of your decision to use the analysis as observations in the calibration process. Don't you systematically underestimate large/ relevant FWI (FWI > 10) values by calibrating to the analysis over observations? Please add your elaboration on this to the discussion and conclusion.

Detailed minor comments

1. Line 2: "makes accurate wildfire risk estimation crucial" add a target group, e. g. decision makers, emergency responders, etc.

2. Line 6: You already introduced the abbreviation FWI for the Canadian Fire Weather Index. I suggest to not spell it out here and just use FWI.
3. Add a strong last sentence to your abstract, in which you point out what the target group and the impacts of your study should be.
4. Lines 13 - 21: You added good references for your description of the more northern parts in the second part of your paragraph (line 16ff.). Could you underline your statements in the first part as well?
5. Line 27: In your abstract you say you look into forecasts up to 15 days which is 2 weeks, while here you state you are interested into "forecasts ranging from a few days to several weeks". Do you refer to your study or the outline of the SAFERS project here and what are several weeks?
6. Line 28: Rephrase the sentence and add what you use the FWI for, e. g. "in this paper, we use the FWI ... for deriving fire risk from weather forecasts".
7. Line 28ff: Here you could improve the reading flow by two things:
 - You should shortly mention the parameters of the FWI, i.e. naming the four weather parameters in it (temperature, precipitation, wind speed and rel. humidity).
 - Add a transition to weather forecasts by adding a statement that these parameters are available in deterministic and probabilistic forecasts.
 - Then you can continue with explaining the pros and cons of deterministic vs. probabilistic forecasts.
8. Line 35ff: Thank you for adding this paragraph about other methods and adding your purpose of the study. This section is much clearer now.
9. Line 49: Add a reference after various regions, e. g. Di Giuseppe et al. (2016)
10. Line 51: Please rephrase "relatively straightforward computation" to a more scientific language. I disagree that the many empirical formulations of the FWI and its subindices are straightforward to develop, though they are straightforward to apply once a source code is available.
11. Line 55: Add ", i.e. Drought Code (DC), Drought Moisture Code (DMC), and Fine Fuel Moisture Code (FFMC)".
12. Line 65: Add ", i.e. Initial Spread Index (ISI)."
13. Line 70: Add reference to EFFIS fire danger classes.
14. Line 70-71: I like that you reflect on the fire danger level thresholds here, but you do not show results for FWI values later in your results section and you do not group your FWI values into these classes. Therefore, you can remove these two sentences about the fire danger levels.
15. Line 74: Drop "with modifications to utilize gridded input data" or provide the code in a GitHub repository, because I am curious to see/ use it now.
16. Line 86: You missed a word ("mean"?) after climate.
17. Line 87: Rephrase to "using a centered 15-day rolling mean on each day".
18. Line 89: Here, I do not understand where the ECMWF high-resolution forecast comes from. Is this part of the TIGGE archive or from a different archive? Please clarify.
19. Line 93ff: This sentence is confusing. Aren't you showing weather station data in Fig. 1a)? Please restructure this in accordance with the beginning of your new paragraph and make this a uniform/ clear statement.
20. Line 103: Please put your correlation coefficient ($r=0.72$) in brackets here. The definition of a good correlation is varying in different science domains, and I would rather read that you found a correlation coefficient of 0.72 and why you think this is sufficient for your analysis.
21. Line 107: Missing word after Figure 2a.

22. Line 107 and wherever you refer to a figure, here you use Figure 2a, while in line 99 you use Fig. 1 and 2. Please make this uniform across your manuscript.
23. Line 112: Do you mean analysis by “forecast-derived”? Please stick to one term across your manuscript.
24. Line 115: Why is it here only a “fairly good” correlation while it was a “good correlation” in Line103?
25. Line 119: Make the statements about the markers more specific, e. g. “the bright red colored markers”.
26. Fig. 1b-c legend + Fig 2 legend: shorten “calc. from observation” to just “observations”.
27. Fig. 1 caption:
 - “See Fig. 2” to “see Fig. 2c”.
 - Add “(b – d)” after example stations.
 - Move “(b)” in front of “in NEU”.
28. Fig 2c: Can you add the r_{mean} as a vertical line to the plot?
29. Line 143 – 146. These two sentences need to be restructured. Mention the length of your training periods in the beginning, e. g. “training periods between 15 to 40 days were tested”, then add your benefits and disadvantages of different training lengths.
30. Line 147 – 148: What training period are you using in your analysis? You say you found 30 days appropriate for small geographical areas and then that you adopt a regional approach, which is in my understanding a large geographical area. Please clarify which training period you stick with here explicitly.
31. Line 158: Please replace several with the verification metrics you are using.
32. Line 160: Drop introductory sentence.
33. Line 164/ 165: I don’t understand how you derive the ensemble spread. Is the ensemble spread the standard deviation (square root of the ensemble variance) of the ensemble? What is the average ensemble variance? The mean of the variance of all ensemble members? Please rewrite this.
34. Lines 166/ 167: Please be clear in your terminology. Is skill in line 167 the same as RMSE in line 166? Please be consistent with these terms throughout your manuscript and figures, i.e. Fig. 3.
35. Line 185: Replace “introduced calibration method” with “the NGR method”.
36. Line 186: Name the regions again, i.e. NEU, WCE and EUMED.
37. Line 190: What is a considerably high FWI? Please find a reference using this fire seasons to refer to your Fig 1b-d.
38. Line 197-199: Here it becomes apparent that you should be consistent with naming it either skill or RMSE.
39. Line 200: Move the statement that your method improves the forecast to the beginning of the sentence to make your statement stronger.
40. Line 202/ 203: Consistent naming of RMSE and skill becomes apparent again. From the reading flow it might make more sense to stick to RMSE when you talk about the measured RMSE (line 202: “skill of raw and calibrated”) and use skill when you talk about the performance of your calibration method (line 203: “the forecast lacks skill”).
41. Line 205: What do you mean by large uncertainty? If you mean that the signal-to-noise ratio in NEU is higher than in other regions, because the FWI values are lower and more unlikely to exceed the variability range, then bring this statement more to the point. You could underline this statement by whatever you mean with “(not shown here)”.
42. Line 206: What are you not showing here and where do you show the relative uncertainty in different regions?! Drop this statement, refer to supplementary material or refer to a Figure.
43. Line 214: Just add in brackets the reference to the figures in your supplementary material.

44. Line 222: What is your reference in Fig. 5 for stating that the skill worsens in NEU but not in WCE? In Fig 5. the median in WCE goes below 0 at lead time between 180 and 220 and therefore, I think that the skill worsens here as well. In EUMED the skill is not worsening if you refer to the median of the boxplots.
45. Line 223/ 224: I don't understand the context of these two statements: Is this worsening of skill caused by small FWI values or by longer lead times? Didn't you say earlier you excluded FWI values below 1? Can you refer to a figure why you think low FWI values are the cause of the worsening? Throughout your manuscript you bring up this argument frequently, but you never show how high/ low the FWI values in the different regions are, and why you come to this conclusion.
46. Line 225: Merge this with the paragraph above. Shorten this by dropping the first sentence and adding the second sentence to the paragraph above by referring to the figure in brackets: "With increasing lead time the skill worsens especially in mountainous areas in Scandinavia and the Alps, which are also the regions with generally low values throughout the fire season (see. Figure 6)."
47. Figure 6 comment: You claim that your method does not perform well over mountain regions. Which is true for lead times of 228h and 324h. But at lead time 132h, I see that the atlantic influenced regions all do not perform very good. Has this something to do with the ocean/land interface? Could you investigate this and add a statement about this to your manuscript?
48. Line 234: please shortly reintroduce the term FWI analysis, e. g. FWI derived from the ensemble forecast (analysis). For people skimming your paper it might appear as you analyze the FWI, which you don't.
49. Lines 236 -239: Please be consistent when using lead time units. It is hard to transfer from 84 hours to 8 to 10 days. You can state the lead time in hours and add the days in brackets after the hours as you stated in your response to the reviewers.
50. Line 240: this statement is weak and a reference to appendix (e.g. "see Fig S5-7") is missing. From these figures you can only read that the mean error is smaller in the summer months, but not how high or low the FWI is in the specific subregions. Therefore, you cannot state here that your findings are likely caused by low FWI values.
51. Lines 240: It is not clear what you consider high and low FWI values, please add values or define low or high FWI somewhere, e. g. "high FWI values (FWI > 1)".
52. Line 244: Rephrase "Although it would be ideal" to more scientific language.
53. Line 246: Define what you mean by "long-range" and "short-range" forecasts.
54. Line 251: Move everything after "compared to more complex approaches" to the beginning of the discussion and conclusion chapter, e. g. to line 233 before you start with the sentence "We used..."; to end your paper with this very good and strong statement: *The improvement of FWI forecasts using the presented calibration method improves the ability to anticipate fire danger, ultimately supporting better response management and shows that a relatively simple method can provide good results compared to more complex approaches.* In line 233 you can discuss that NGR is a good method in comparison to other methods (e.g. bias correction) for your research purpose.