General text

The study “statistical calibration of probabilistic medium-range fire weather index forecasts in Europe” of Bohlmann and Leine shows how the Fire Weather Index (FWI) can be calibrated in medium-range weather forecasts to improve FWI in weather forecast and enhance preparedness of fire-fighting resources during high FWI periods. This topic is scientifically important and suits well into the scope of NHESS.

The authors show that their chosen method, i. e., non-homogenous Gaussian regression (NGR) improves the FWI derived from medium-range weather forecasts at shorter lead times by presenting results of different skill metrics, i.e., RSME, ME and CPRSS. I appreciated that the manuscript is well-written and in general easy to follow. Unfortunately, a clear research question is missing, which makes it hard for the reader to know what to expect from the paper. Further, it is not clear in methods and data section for what post-processing steps which datasets are used. This can be improved by revising the manuscript carefully as outlined in the comments below. The results are presented in a clear structure and the figures are easy to interpret, because of the good metric description in the method section. However, the visualization can be improved by minor adjustments. The discussion section is missing, which is unfortunate as a reflection of the authors on the strengths and weaknesses of their method and achieved results, in comparison to other studies would be very valuable for other researchers in this field.

Before publication, the manuscript needs major revisions. I suggest the authors to revise the manuscript carefully, correct and clarify the methods, data and results section and add a discussion section. Please find my specific comments below:

Major comments

1. Study workflow is not very clear, e. g. which datasets are used in which step. This is due to missing research questions and the structure of the manuscript. This should be addressed by:
    a. adding research questions, e. g. can NGR be used to calibrate FWI derived from mid-range weather forecasts, how well does NGR improve the FWI derived from mid-range weather forecast. The research questions should be placed at the end of the introduction, as they help the reader to know what to expect from the following chapters.
    b. Restructuring the manuscript by summarizing chapters 2 to 4 to a data and methods section with subchapters (the following is a suggestion):
        i. 1. Introduction, providing research questions at the end of the chapter
        ii. 2. Data and Methods,
            1. 2.1 FWI and FWI calculation
            2. 2.2 Forecast and observation data
            3. 2.3. Validation and Calibration methods
                a. 2.3.1 NGR
                b. 2.3.2 Verification methods
        iii. 3. Results
        iv. 4. Discussion (missing, see other comment.)
        v. 5. Conclusion

2. Introduction: Please add a paragraph why you chose the NGR method over other methods, e. g. other variations of EMOS calibrations or bias-correction methods (see Whan et al. 2021, https://doi.org/10.1016/j.wace.2021.100310).
3. Fire weather index calculation: It is not clear how the FWI is derived. Please clarify in section 2 how you derive the FWI, i. e. which R-package of python package you are using, as there are differences between cffrds and the ECMWF fire products derived from the ECMWF GEFF-Model (see Vitolo et al. 2019, https://doi.org/10.1038/sdata.2019.32)

4. Forecast and observation data:
    a. Lines 57 – 65: You state that you derive the FWI from the ECMWFs operational forecast system (ENS). Later you state that you use the TIGGE dataset. Can you clarify if you used the ENS dataset, the TIGGE dataset or both datasets later in your analysis? I understand that the TIGGE dataset has a higher temporal resolution than the ENS dataset, however, the spatial resolution is coarser (0.5° vs 0.2°). Later in your results section you show the earliest results for 36h lead time. Can you add a sentence why you are more interested in the increased temporal resolution of the TIGGE dataset over the spatial resolution of the ENS dataset in your manuscript?
    b. Lines 65 – 68: These sentences should be moved to the paragraph where you discuss how you verify the FWI from ECMWF data to observation data.
    c. Line 68: "We therefore use the FWI calculated ECMWF high-resolution forecasts …". You did not introduce what the ECMWF high-resolution forecasts are yet. You can optimize this by merging this sentence with the next sentence (i. e. "ECMWF high-resolution forecasts have …"), but it remains now unclear why you introduced the ENS and the TIGGE dataset before. Please clarify in this section on which datasets you derive the FWI from and for which later steps you use which datasets (NGR regression and verification).

5. Results / Figures:
    a. All figures with subplots: add labels for subplots, i. e., (a), (b), (c) as suggested in the NHESS publication guidelines https://www.natural-hazards-and-earth-system-sciences.net/submission.html#figurestables)
    b. In your written text you relate to lead times as days while in the figures you show lead time in hours on the x-axis. Can you synchronize this information? In the current state the reader has to transform between written "7 days" to 7*24h in the x-axis of the respective plot. You could write the hour also in brackets next to the days in the text.

6. Terminology "short lead times": You state multiple times that your results work best for short lead times. Can you clarify in your manuscript how you define short lead times (e. g. 72h or 132h) or be more specific which lead time you still find good performing (e. g. rephrasing to something like: "for short lead times up to 84h").

7. The discussion is missing. In my opinion the discussion is an integral part of a research paper and as the of the study presents a novel way of calibrating mid-range weather forecasts, it would be good to critically reflect on the results:
    a. Is the FWI a suitable predictor for fire events in all three regions? What are the challenges regarding wildfire hazard in the three regions? For example, you could address why the postprocessing works particularly well in the MED and summer

months of WEU and why not for NEU? Further, you could reflect on how low FWI values, as present in NEU, affect your method.

b. How does your method (NGR) compare to other post-processing methods? Select a two to three different studies, with a similar research question and set your results in a broader context. For example, you could discuss why you are correcting the FWI instead of the input variables of the FWI, why you chose the NGR method and not a bias correction method or machine learning based method (see Whan et al. 2021 https://www.sciencedirect.com/science/article/pii/S2212094721000086) and Worsnop et al. 2021 (https://journals.ametsoc.org/view/journals/wefo/36/6/WAF-D-21-0075.1.xml?alreadyAuthRedirecting)

c. What can stakeholders take away from your study. You illustrated quite nicely in the introduction that post-processing helps to make accurate forecasts helping first responders. What do you wish this target group takes away from your results, e. g. will more firefighting resources be placed at locations with elevated FWI?

Minor comments

1. Line 6 & Line9: you use the terms post-processing and calibration interchangeable, please choose one term.
2. Lines 9 – 11: Be more specific about what you mean by short lead times (e. g. 84h?) and regions with elevated FWI (e. g. MEU)
3. Lines 13: I would drop the word “recent” and replace “wildfire in Greece 2023” by “wildfire season of 2023”.
4. Line 15: Drop “Also” at beginning of sentence.
5. Line 16: Drop “But” or make this sentence sound more formal.
6. Line 17: Provide references for your statement.
7. Line 19: Provide references for your statement.
8. Line 20: Here it would be great if more than one example could be provided, e. g. one for each subregion.
9. Lines 22 – 25: Switch the order of the sentence to stress more clearly that weather forecast is part of SAFERS or drop mentioning the project.
10. Line 26: Add Di Giuseppe et al. 2016 (https://doi.org/10.1175/JAMC-D-15-0297.1) as a reference.
11. Line 26: Watch out that your citation tool takes the names correctly. It is van Wagner and Di Giuseppe not Wagner and Giuseppe.
12. End of line 27: Here I am missing a short explanation of what is the difference between deterministic and probabilistic weather forecast. A short explanation would be helpful to emphasize that the topic of the manuscript is postprocessing probabilistic forecasts.
13. Line 28: drop the word “may” or provide a clear statement whether post-processing is needed. Consider also my first comment on your interchangeable usage of post-processing and calibration. This confuses the reader.
14. Line 39: Chose a more scientific formulation than “is a relative simple calculation” for the FWI.
15. Line 43: Add the depth of the moisture levels.
16. Line 48: Rephrase the sentence starting with often, e. g. “The FWI can be classified”
17. Lines 52 – 54: I would change the order of the sentences to make the statement at the beginning of the paragraph clearer, e. g. “we use climatological mean values …, to account for preceding conditions at the initialization”.

18. Lines 59, 63 and 70: Please provide an approximation of the grid resolution in km in brackets?
19. Line 60: it should be “derived from **the** TIGGE archive”.
20. Line 61 and following: Please add dataset after TIGGE, i. e., “The TIGGE dataset …”.
21. Line 62: Please add “the” to ECMWF API.
22. Line 62: Please rephrase sentence to “the temporal resolution of the TIGGE dataset …”.
23. Lines 65 – 78: Please rephrase this paragraph. Keep the statement that the FWI has multiple input variables. Explain for which later steps you use which dataset to calculate the FWI. Mentioning the station data here is confusing.
24. Line 71: I am not sure which dataset you are meaning by “those”, please clarify.
25. Line 75: How many of the 682 stations are in Finland and how many are outside of Finland. Can you provide values.
26. Fig 1 / Line 74 (first mentioned):
    - Add letters for subfigures.
    - Fig 1a
        - Use a different projection, e. g. Lambert Conformal Conic, as the northern latitudes are strongly distorted.
        - It would be very nice to have the stations shown in Fig 1c on the map of Fig 1a as well.
        - Place the region legend (i.e., NEU, WCE, EUMED) inside Fig 1a.
        - Place the legend of the countries (i.e., Finland ant others) at a position, where it is clear the legend belongs to both Figures (i. e. Fig 1a and Fig1b)
    - Fig 1b:
        - Provide a legend for the regression line. Is this the line for all stations, or for only “other” stations (outside of Finland) or for only Finland stations?
        - Can you provide a line for Finland as well?
    - Fig 1c:
        - Please show a 3rd station for WCE.
        - Add the location of the stations in Fig 1a.
        - Why are there missing values in the Greece station in the winter of 2022 and 2023? You previously stated that all your selected stations have a sufficient data coverage. Please clarify that your consecutive 200 days refer to the summer half (?) in the Figure caption.
27. Line 77: Provide a reference to Fig 1b as your statements originate from the figure.
28. Line 77 and Fig 1b: Is your, I assume linearly derived correlation, mainly driven by the large number of low FWI values? Fig 1c shows quite apparent that for high FWI values the underestimation is much stronger pronounced than for low FWI values. This would be also a good point to be discussed in the discussion.
29. Line 79: Provide a reference to Fig 1c.
30. Line 80: Please clarify which datasets you use for longer forecasts.
31. Line 83: Please clarify which dataset(s) you mean by the FWI forecasts.
32. Line 86: Please specify what you are calibrating, e. g. the parameters of the NGR or the whole post-processing pipeline.
33. Line 86: It should be: “**the** FWI”.
34. Line 91: Please clarify from which dataset the 51 ensemble members come, e. g. by adding (ENS) in brackets. You describe the statistical part very clearly, but it is not clear to which datasets you are applying the formulas.
35. Line 96: Please specify what the training area is.

36. Line 97: Why do you switch terminology from NGR to EMOS, I understand that this is the approach, but it would be good to decide for one name.
37. Lines 98 -100: Can you provide results, e. g. a table or a plot, for these findings. This could be part of your supplementary material.
38. Line 100: Here it becomes apparent, that it is not clear what the training area and hence smaller geographical training areas should be. Please clarify and provide results in the supplementary material.
39. Line 114: Please add a note that you call RMSE later spread and skill metric (i. e. line 122 and line 144).
40. Line 142: please clarify that you compare fire season length of Northern Europe to Southern Europe.
41. Line 145: the grid points "within" rather than "of" the three study areas.
42. Line 146: I can't follow how you derived the RSME of the climatology. Can you describe this briefly.
43. Line 151, Line 163: clarify that your calibration is done as a post-processing.
44. Line 152: Specify what you mean by short lead times, e. g. 132h ?
45. Line 154: Provide lead time in hours in brackets after "7 days", i. e. 7 days (168 h).
46. Figure 2:
    - Add labels (letters a, b, c) to subplots.
    - Place legend outside of NEU to make it clear it belongs to all three subplots.
    - Adjust the y-axis label (Spread/ RMSE) to your figure caption, which is currently "spread and skill". I would expect them to be the same, e. g. spread and RSME in the figure caption or spread and skill in the y-axis label.
47. Line 157: Rather the three subregions than the respective area.
48. Line 158: Rephrase "too low" to something like "lower than observations".
49. Line 160: "The improvement" instead of "This improvement".
50. Line 164: Provide numbers for what you think is slightly positive.
51. Line 167: specify short lead times.
52. Line 166 – 169: This finding would be a good point to discuss in the discussion. For example, I would be interested what these findings imply for the application of your suggested post-processing technique.
53. Figure 3: please add letters to subregions.
54. Figure 4: Please add letters to subregions.
55. Figure 4 caption: drop "the grid of".
56. Lines 171 – 172: Can you discuss this in your discussion section? Does your approach perform well for higher FWI values, suggested by the better performance in WEU in July and August and MED? Does your approach need to perform well on or low no-fire danger days?
57. Figure 5:
    - Add labels to the subplots (i. e., a, b, c, d)
    - Plot the land-sea boundary to improve the visualization.
    - Drop the large white space in the south and west of the plot in such a manner that the plot is filled with results.
58. Line 176: Here you mention the first time that you calibrate the coefficients of the NGR, this is not clear in your previous description of the post-processing method. Please clarify this in the method section. In Line 176 add "of the NGR" after "to estimate the calibration coefficients".
59. Line 177 - 179: This sentence belongs to the discussion section and not the conclusion session. Also, I suggest adding more meaning to this sentence, e. g. what are the implications of sparsely available data?

60. Line 180: Drop “Thus”.
61. Line 180: which dataset do you mean by high-resolution weather forecast with short lead time? Is this the third dataset you introduced in the data section?
62. Line 181: Make clear that you mean the dataset “analysis” and not the analysis of the FWI.
63. Line 189: At the end of your conclusion, I would expect a last sentence coming back to your initial statement that this is important for fire resource management and the SAFERS project. Please add such a sentence.
64. Line 209: Add “Di” to “Di Giuseppe”.
65. Line 249: Add “Van” to “Van Wagner”.