Letter of Reply to Referee 1

Thank you for reading the manuscript and providing valuable suggestions to improve the paper. Our responses to your comments are shown below in blue, and changes to the manuscript are indicated in italics. Additionally, all modifications are marked in the revised manuscript.

Major comments

45-47 ". In the second step, FFMC, DMC and DC are used to model the rate of fire spread (ISI) and the potential fuel available for surface fuel consumption (BUI)." The wind speed is missing (ISI = FFMC + WS10), and it would correlate to the POTENTIAL rate of spread (I mean, it is a variable involving the combustion of surface fuel, like dry leaves and such + the wind). Also, the BUI = DMC + DC, so it is not surface fuel consumption, but it involves the potential for a surface fire to burn the deeper fuel (Build Up) and become a much more persistent fire.

Thank you for your comment. The suggestions have been included in the revised manuscript. The description of the FWI calculation reads now as follows:

"The FWI is calculated in two steps. First, the 2-meter temperature, 2-meter relative humidity, 10-meter wind speed, and 24h accumulated precipitation at local noon are used to calculate the moisture content of three separate fuel layers. These fuel layers are characterized by different depths and fuel consistencies, which result in varying water capacities and drying speeds. The Fine Fuel Moisture Code (FFMC) represents the moisture content of litter and other fine cured fuels at a nominal depth of 1.2 cm; the Duff Moisture Code (DMC) indicates the moisture content of loosely compacted layers (nominal depth ~7 cm); and the Drought Code (DC) denotes the moisture content of deep, compacted layers in a depth of around 18 cm (Van Wagner, 1987). DMC and DC respond slower to weather variations compared to the fast-drying fuel represented by the FFMC. Consequently, the effective day length, which determines the amount of drying that can occur during a given day, must be considered and monthly day length adjustment factors for DMC and DC are applied based on latitude (Lawson and Armitage, 2008). The fuel moisture codes are dependent on previous weather conditions; therefore, the preceding day's noon values for FFMC, DMC and DC are necessary for their calculations. In the second step, FFMC and the 10-meter wind speed are used to model the potential rate of fire spread (ISI). DMC and DC are used to calculate the Buildup Index (BUI), a numeric rating of the total amount of fuel available for combustion which comprises the potential of a surface fire to burn deeper fuel layers (build up) and thus evolve into more persistent fires. These fire behaviour indices are then used to calculate the FWI. The FWI values are always non-negative, with low numbers indicating low fire weather danger and high values indicating high fire weather danger. The FWI is often classified into danger classes and values above 50 are considered extreme. However, those levels can vary depending on local conditions (e.g., vegetation types). Consequently, what is considered a low or extreme FWI in one region may not be equivalent in another. A more comprehensive description of the FWI system can be found in Van

Wagner (1987) and Lawson and Armitage (2008). For this study, we implemented the calculation of the Canadian Forest Fire Weather Index (FWI) using Python programming language, following the source code provided by Wang et al. (2015), with modifications to utilize gridded input data."

76-78 "Figure 1b shows the scatter plot of analysis and observations for all stations and every time step. While the FWI derived from the forecasted weather parameters seems to generally underestimate the FWI values compared to the values derived from the observations (slope ∼ 0.63)" AND Figure 1b: Many doubts arise from this scatterplot:

- The plot itself shows too many points to use a scatter. A density plot NEEDS to be used in this case, or two if you want to show separately the data from Finland.

Thank you for your comment. We agree that the scatterplot may not have been the most suitable choice for representing our data and conveying our message. We decided to change the representation to histograms showing the distribution of the FWI calculated from high-resolution forecast (analysis) and from observation (Fig.2 in the revised manuscript). Additionally, we split the Figure into 2 parts to avoid overcrowding the Figure. The new Fig.1 contains the map of observation stations (a), with colour representing the linear correlation coefficient, and the time series of three example stations (b,c,d). The new Figure 2 contains the histograms of the frequency distribution of the FWIs (a,b) and the linear correlation coefficient.

- The regression which leads to the 0.63 slope seems to be off by looking at the scatterplot, which might be due to the lack of information about the point density. What causes the slope to be 0.63 (and not closer to 1, as the scatterplot would suggest)? Also, please specify the method used for the regression (I assume linear regression).

The regression line in the original Figure 1 appeared off due to the high concentration of values near zero. This becomes now evident considering the distribution shown in histogram Fig.2a. We also clarified in the text that we used linear regression to obtain the coefficient.

- I am aware that it is common practice to use the ECMWF analysis at minimal lead time in place of observations, but once you have found an important underestimate like you did, why did you dismiss it so fast? It seems like a very important matter that can have a huge impact on the paper's reliability. Please explain in depth why you can ignore this bias or what you did to correct it.

We agree that a correction of the analysis based on the observations would be beneficial. We found that the strongest discrepancies between observation and analysis occur in mountainous areas, e.g., Austria or Norway. This is likely caused by difficulties in capturing the small-scale weather phenomena occurring the complex terrain with relatively coarse resolution of the forecast model. However, in those regions the fire danger is generally rather small because of high precipitation and low temperatures. To further investigate this bias, additional station-specific characteristics such as elevation or land use would be necessary. In this article, we want to present an easy, straightforward method to calibrate the FWI and we believe that the correlation of analysis and observation is already quite good and justifies the use of the analysis as substitute for observations. However, the use of additional station-specific parameters could be topic of future research.

The section that previously related to the scatterplot has been changed to:

"To determine if the analysis is suitable to be used as observation substitute, we check their agreement with actual observation-based values, which is shown in Fig. observation and 2. We use observations available from the Finnish Meteorological Institute's observation database for the years 2021–2023. Figure 1a shows the stations, for which it is possible to calculate the FWI for more than 200 consecutive days in addition to the reference areas which are introduced in Chapter 3. In total 682 stations can be used. The time series of three example stations are shown in Fig.1a,b,c. The FWI analysis is shown in orange while the FWI calculated from observations is shown by the black dashed line. Overall, there is good correlation between the forecasted and observed FWI values. However, especially for high FWI values, the FWI derived from forecasted weather parameters tends to underestimate the values compared to those derived from observations. This is particularly evident during the summer months of 2022 and 2023 in Meiningen, Germany (Fig.1c) and Chrysopouli, Greece (Fig.1d).
Figure 2a overlapping histograms of the density distribution of the FWI analysis (orange) and FWI calculated from observations (grey) for all stations shown in Fig.1a and every time step. The distributions show a high frequency of low values (FWI< 1), and when plotted on a logarithmic x-scale, a bimodal structure becomes evident with a separation at an FWI value of 1. This bi-modality results from a necessary restriction imposed on the FWI calculation (Eq. 30a,b Wang et al. (2015) or Eq.40 in Van Wagner (1987)). Figure 2b again displays the density distribution of FWI values, but focuses on FWI values greater than 1, as these are generally more relevant for assessing wildfire danger. For higher FWI values, the forecast-derived FWI underestimates those calculated from observations, while for lower FWI values, it tends to overestimate. One contributing factor is the different spatial resolution of the data sources. The analysis uses gridded data with a resolution of 0.1° (approximately 9 km in Central Europe), whereas weather observations are local measurements. However, in general there is a fairly good positive correlation of FWI analysis and FWI calculated from observations. Figure 2c shows the histogram of the linear correlation coefficient of stations that observe FWI values greater than 1, the mean correlation coefficient is 0.72. Low correlation coefficients are mainly caused by differences between analysis and observation in mountainous areas, such as Austria, Switzerland, Romania, and Norway, as indicated by the colored markers in the station map (Fig. 1a). This is due to the difficulty in capturing the complex terrain and its small-scale weather phenomena with relatively coarse resolution (~9km) of the forecast model. Because there is a fairly good correlation of observation and analysis, we will use the forecasted FWI with short lead time (analysis) derived from ECMWF high-resolution forecasts as observation to compare with the FWI forecasts derived from ECMWF medium-range ensemble forecasts."

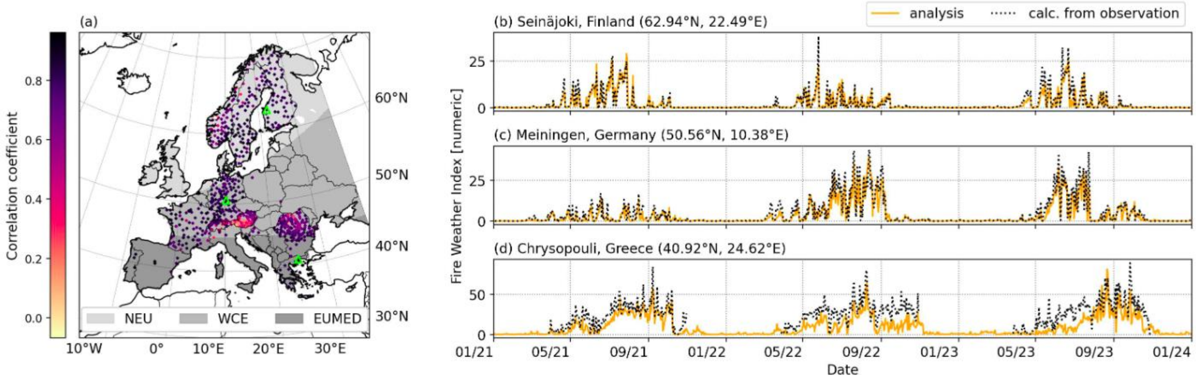

Fig.1

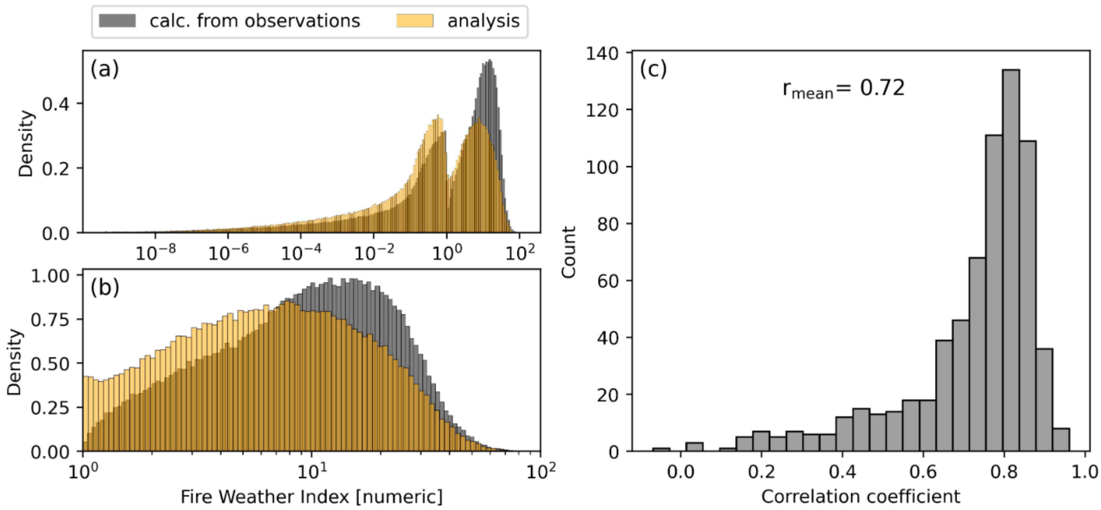

Fig.2

78-79 “a correlation is apparent. This good correlation can also be seen in the time series examples for a station in Finland and Greece”: Please provide us with the necessary quantitative information (e.g., correlation coefficients for all the stations) to support this claim, especially coming right after the previous comment. Two sample stations (Figure 1c) are not enough to validate a claim on over 600 others; by this I do not mean that Figure 1c must be removed.

We agree that only showing two examples is not enough to show that there is a good correlation. In the revised manuscript we included an additional figure (Fig.2c) which shows the distribution of the correlation coefficient for all stations. The mean correlation coefficient is 0.72, which indicates a fairly good positive relationship correlation. Additionally, we added a third station in Fig.1, providing an example for each subregion. Changes to the manuscript can be seen in the reply to the previous comment.

90-97 “Furthermore, data from all grid points in the training area is used to estimate a single set of coefficients for the given day (regional EMOS)”: some additional explanation is then needed, how do you go from the μ_{kl} to the estimate used (I guess μ_l ?)

We improved this section (now 2.3.1) to make the method better understandable and to clarify what training data is used. Specifically, we changed this sentence to:

"We adopt here a regional approach, pooling training data from all grid points within the training area to derive a single set of calibration coefficients (a_l - d_l) specific to each lead time for the given forecast."

We hope this clarifies that the calibration coefficients are estimated by the calibration method and then used to calibrate the ensemble mean of the given forecast.

Figure 2:

- It can be made clearer if the legend was more explicit (dotted line with triangles: spread (raw) / solid line with dots: RMSE (calibrated) / etc.)

We made the legend more explicit as suggested.

- The legend, which is relative to all the three graphs must be outside the first graph. Consider also putting everything in a single column.

We placed the legend outside as suggested.

The previous Figure 2 is now Figure 3 in the revised manuscript:

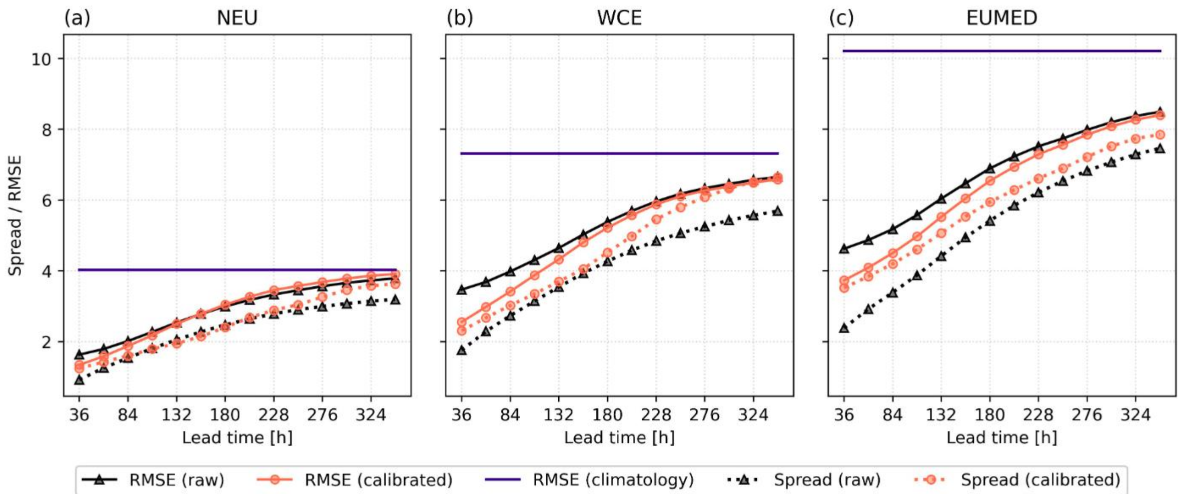

Fig.3

152-153 "In Northern Europe, the RMSE of the calibrated forecast is slightly above the RMSE of the raw forecast after 7 days of forecast," : please provide at least a hypothesis as to why this happens. The sentence on the subsequent lines "The regional differences could be explained with the generally higher FWI values in the more southern, fire prone regions compared to Northern Europe where FWI values are often very small" addresses the regional differences, but the difference from uncalibrated and calibrated NEU RMSE is not addressed.

We added a hypothesis trying to explain the slightly decreasing RMSE of the calibrated forecasts in NEU as follows:

"In Northern Europe, the RMSE of the calibrated forecast is slightly above the RMSE of the raw forecast after 7 days (180 hours) of forecast, whereas the skill of raw and calibrated forecast in Central and Mediterranean Europe is similar for forecasts longer than 9 days (228 hours) and 12 days (300 hours), respectively. The FWI forecast in Northern Europe lacks skill for lead times longer than 7 days and calibration based on 30 day rolling window fails to improve and even worsens the forecast. This finding could be explained by the rather small FWI values and large relative uncertainty in NEU, unlike FWI values in WCE and EUMED (not shown here). The applied calibration appears to be more effective for higher FWI values and shows limitations for smaller FWI values close to 0. This also explains the regional differences and the better calibration results in the more southern, fire prone regions with generally higher FWI values compared to Northern Europe, where FWI values are often very small."

Figure 3:

- The legend, which is relative to all the three graphs must be outside the first graph. Consider also putting everything in a single column.

The legend has been placed below the figure to make it clear that it relates to all three subplots. Figure 3 is now numbered as Fig.4:

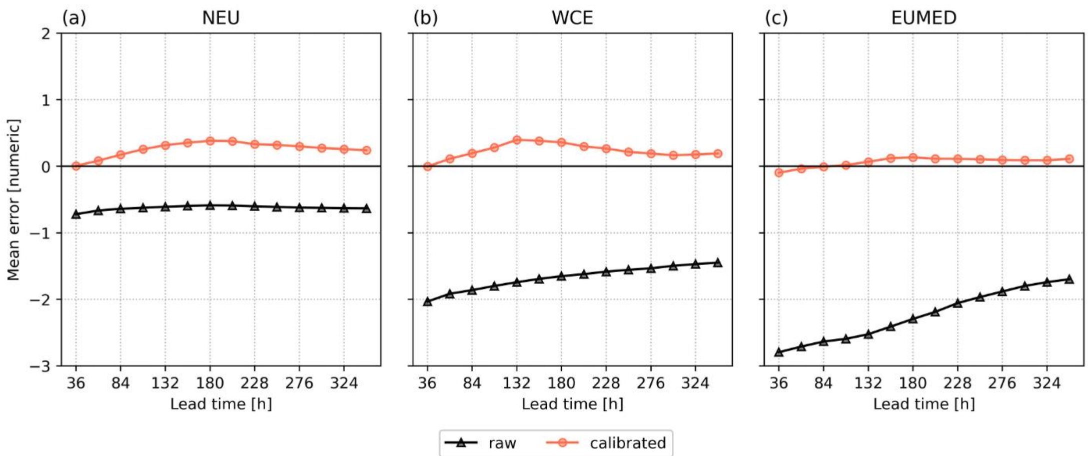

Fig.4

Minor comments (typos and formalities)

13: "prevalent": word choice

We changed the wording to "*frequent*".

16: “But not”: cannot start a sentence with “but not”

We rephrased this sentence as follows:

“While the Mediterranean region continues to face the highest occurrence of wildfires, Central and Northern Europe have experienced an increase in extreme temperature events and heatwaves in recent years (Ibebuchi and Abu, 2023; Rousi et al., 2023; Ionita et al., 2017; Barriopedro et al., 2011).”

Furthermore, we added references to support the statement.

18: unnecessary comma after “periods”

The unnecessary comma has been removed.

19 “heatwave 2018”: either heatwave of 2018 or 2018 heatwave

We changed it to *“the 2018 heatwave”*.

24 missing (Oxford) comma after “during”

The comma has been added.

24-25: “Accurate and reliable weather forecasts ranging from a couple of days to multiple weeks to identify high wildfire risk areas is an important part of SAFERS”: Accurate and reliable weather forecasts, ranging from a couple of days to multiple weeks, are an important part of SAFERS for identifying high wildfire risk areas. (Or equivalent paraphrasis)

We rephrased this sentence and changed the order of the sentence to make it more clear that the weather forecasts are part of SAFERS.

“An important component of SAFERS for identifying high wildfire risk areas are accurate and reliable weather forecasts, ranging from a few days to several weeks.”

25 “Here,”: “in this paper,”

Wording has been changed to *“in this paper”*

26 “short FWI”: (FWI)

This has been changed.

26: “Wagner, 1987”: Author’s last name is Van Wagner, throughout the paper

Thank you for pointing this out. This was a mistake in the used Bibtex reference and has been fixed.

29: “One widely used”: A widely used

A paragraph mentioning other calibration methods has been added to the introduction following the suggestion of referee 2 (comment 2). The wording has therefore changed:

"Various methods have been developed for calibrating probabilistic ensemble forecasts. Commonly used calibration methods include Bayesian model averaging (Raftery et al., 2005), non-homogeneous Gaussian regression (Gneiting et al., 2005), logistic regression (Hamill et al., 2004) and non-parametric ensemble post-processing methods such as rank histogram techniques (Hamill and Colucci, 1997), quantile regression (Bremnes, 2004) and ensemble dressing approaches (Roulston and Smith, 2003). Non-homogeneous Gaussian regression (NGR) is one of the most commonly used calibration methods and adjusts both ensemble mean and spread, while still be efficient and easy to implement. It has been proved effective for various weather variables like temperature (Hagedorn et al., 2008), precipitation (Hamill et al., 2008) or wind-speed (Thorarinsdottir and Johnson, 2012) and can be applied using a truncated or censored distribution to account for a constraint to non-negative values."

36-39 "Although originally developed for Canadian weather and vegetation, it is used in many other regions, e.g., by the European Forest Fire Information System (EFFIS) to provide information on wildfires in the EU and neighboring counties (Giuseppe et al., 2020)": sentence needs to be more orderly and written better; also, author's last name is "Di Giuseppe"

We improved this sentence in the revised manuscript and corrected the author's name.

"Although originally developed for Canadian weather and vegetation, the FWI system is now used in various regions. For instance, the European Forest Fire Information System (EFFIS) employs the FWI to provide information on wildfires in the EU and neighboring countries (Di Giuseppe et al., 2020)."

39-40 "One advantage of using FWI is the relatively simple calculation only requiring four weather parameters in addition to information of the season (time of year) and geographical location": Rephrase, e.g. "The main advantage of using FWI is its relatively simple computation, only requiring four weather parameters and information about the season (time of year) and geographical location"

We rephrased this sentence to:

"The main advantage of the FWI system is its relatively straightforward computation, requiring only four weather parameters and information about the season (time of year) and geographical location as input parameters."

42-43 "the moisture content of three separate fuel layers of different depth and diameter": this is one interpretation of the three parameters (of course, the main one), meaning that -more or less- they contribute to the fire danger with the same time scale of a certain fuel layer. For example, the DC can also be an index of the lack of precipitation for a long time. I tend to be more cautious when interpreting these indices, but it is a relatively small issue.

We rephrased this section to add the depth of the moisture levels (see comment 15, referee 2). We also added a comment about the response of the moisture codes to weather variations, please see reply to the first comment.

48 “Often the FWI is classified”: the FWI is often classified

We changed the order of this sentence accordingly.

50 “e.g. vegetation types”: since this is the second time it appears, I have to point out that you cannot put “e.g.” in the middle of a sentence without it being in parentheses or in a parenthetical expression (between commas). This is not formal enough for a paper, in my opinion.

We corrected this throughout the paper. Thank you for this comment!

52-53 “Fuel moisture codes (FFMC, DMC, DC) and consequently FWI values are dependent on preceding conditions. Thus, the preceding days noon values are used for FWI calculations and the calculations need to be initialized”: To be put above, together with the input variables, and to be written more clearly.

We moved the sentence about the initial values to the paragraph about the moisture code calculation (see reply to first comment). The sentence about the dataset used for initial values was moved to the Forecast and observation data section (2.2).

62 “The resolution of the used TIGGE data”: The resolution of the TIGGE data used in this paper

We rephrased this sentence accordingly.

65 “can not”: cannot

This typo has been fixed.

Figure 1 (caption): please provide a reference (IPCC) for the AR5 regions.

We wrongfully called the regions here AR5 when it was supposed to be the AR6-WGI regions. We corrected that and added the reference for the AR6 regions.

90 Formula (2): shouldn’t it be $\log(\sigma_{kl})$?

It should indeed be $\log(\sigma_{kl})$. We apologize for this mistake and correct the formula.

91-92 “The logarithmic link log(sd)”: define sd (ensemble standard deviation?). Besides, isn’t it sd_{kl} ?

sd is defined as the standard deviation of the 50 ensemble members in the previous sentence. We agree with the reviewer that it should be sd_{kl} and corrected this mistake.

119 “The bias of the forecast can be accessed by simply evaluating the difference between the average forecast and average observation, which is defined as the mean error (ME)”:

The bias of the forecast can be accessed by simply evaluating the difference between the average forecast $F\ i$ and average observation $O\ i$, which is defined as the mean error (ME)

We added the symbols F_i and O_i to the sentence.

137 “defined by the 6th IPCC Assessment Report (AR6 (Iturbide et al., 2020))”: did you not show the AR5 regions before (Figure 1a)? It needs to be coherent, at least in the name in the caption of Figure 1a, if the regions did not change.

Thank you for making us aware of this mistake. The reference regions did not change and it is supposed to be AR6. We changed it to be coherent throughout the manuscript.

140 “Other regions can be selected as well, e.g. the calibration can also be done country-wise or at even smaller level.”: colloquial, rephrase like “the calibration can also be performed over smaller areas (e.g., single countries)”

We rephrased this sentence to:

“The calibration can also be performed over smaller geographical areas, e.g., individual countries.”

158 “the forecasted FWI is too low compared to observations”: rephrase in a more formal “the forecasted FWI underestimates the observations[…]” or similar.

We rephrased the sentence to:

“Uncalibrated forecasts have a negative bias for all lead times, indicating that the forecasted FWI values are consistently lower than the observed values.”