Statistical calibration of probabilistic medium-range fire weather index forecasts in Europe

Stephanie Bohlmann[1] and Marko Laine[1]

[1]Finnish Meteorological Institute, P.O. Box 503, FI-00101 Helsinki, Finland

Correspondence: Stephanie Bohlmann (stephanie.bohlmann@fmi.fi)

Abstract. Wildfires are increasing in frequency and severity across Europe, which makes accurate wildfire risk estimation crucial for decision makers and emergency responders. Wildfire risk is usually estimated using ~~meteorological based~~ meteorological-based fire weather indices such as the Canadian Forest Fire Weather Index (FWI). By using weather forecasts, the FWI can be predicted for several days and even weeks ahead. Probabilistic ensemble forecasts require verification and calibration in order to provide reliable and accurate forecasts, which are crucial for informed decision making and an effective emergency response. In this study, we investigate the potential of non-homogeneous Gaussian regression (NGR) for statistically calibrating ensemble forecasts of the ~~Canadian Forest Fire Weather Index~~FWI. The FWI is calculated using medium range ensemble forecasts from the European Centre for Medium-Range Weather Forecasts (ECMWF) with lead times up to 15 days over Europe. The method is tested using a 30 day rolling training period and dividing the European region into three training areas (Northern, Central and Mediterranean Europe). The calibration improves FWI forecast particularly at shorter lead times up to 84 hours and in regions with elevated FWI values, i.e., areas with a higher wildfire risk such as Central and Mediterranean Europe.

The study demonstrates that NGR can be used to improve probabilistic FWI forecasts especially in the time range most critical for firefighting resource management and thereby supporting effective wildfire response strategies.

1 Introduction

Wildfires in Europe have become increasingly frequent and destructive in the recent decades. The wildfire season of 2023 alone burnt, according to the European Forest Fire Information System (EFFIS), an area of over 170 thousand hectares in Greece, killed at least 18 people and forced thousands to leave their homes (Faiola and Labropoulou, 2023). Similar devastating wildfires in 2017 and 2022 burned large areas in Portugal, Spain and Italy (Turco et al., 2019; Rodrigues et al., 2023). 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). Extended warm and dry periods raise the fire danger and may cause wildfires in regions that were previously not considered wildfire hotspots (San-Miguel-Ayanz et al., 2021; De Rigo et al., 2017). Examples are the 2018 heatwave, which caused wildfires in Sweden (San-Miguel-Ayanz et al., 2019) and across the United Kingdom (Sibley, 2019) or the wildfire outbreaks in Germany and the Czech Republic during the summer of 2022 (Skacel et al., 2022).

The rising frequency and severity of wildfires in Europe ~~emphasizes~~ emphasize the need for an effective management of forest fire emergencies. The SAFERS (Structured Approaches for Forest fire Emergencies in Resilient Societies, https://safers-project.eu/) project provides an integrated platform to assist first responders, firefighters, and decision-makers to become more resilient before, during, and after forest fire emergencies. An important component of SAFERS for identifying high wildfire risk areas are accurate and reliable weather forecasts, ranging from a few days ~~to several~~ up to six weeks. In this paper, we use the Canadian ~~forest fire weather index (FWI), which is~~ Forest Fire Weather Index (FWI, Van Wagner (1987); Di Giuseppe et al. (2016)), a widely recognized ~~numeric~~ numerical indicator for forest fire risk ~~(Van Wagner, 1987; Di Giuseppe et al., 2016)~~, to derive fire risk from weather forecasts with a lead time up to two weeks. The calculation of the FWI only requires four weather parameters ~~and can be calculated using~~: temperature, relative humidity, wind speed, and 24-hour accumulated precipitation, which are often available from deterministic or probabilistic weather forecasts. While deterministic forecasts provide a single forecast based on a given set of initial conditions, probabilistic ensemble forecasts offer a range of possible outcomes by using slightly perturbed initial conditions, giving a more comprehensive picture of potential weather conditions and providing an estimate of the forecast uncertainty. Probabilistic ensemble forecasts may require statistical calibration to ensure reliable and accurate forecasts, which are essential for making informed decisions and effectively allocating resources when responding to wildfires.

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~~ being efficient and easy to implement. It has been ~~proved~~ proven effective for various weather variables like temperature (Hagedorn et al., 2008), precipitation (Hamill et al., 2008) ~~or~~ and 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.

In this article, we investigate whether non-homogeneous Gaussian regression (NGR) can be used to calibrate the Fire Weather Index (FWI) derived from medium-range ensemble weather forecasts and assess the extent to which NGR improves the accuracy of the FWI predictions. The skill of the calibrated FWI forecast is shown and compared for three European regions.

2 Data and methods

2.1 Fire weather index calculation

A common method for indicating the danger of wildfires is the Canadian Forest Fire Weather Index (FWI) system (Van Wagner, 1987). Although originally developed for Canadian weather and vegetation, the FWI system is now used in various regions (de Groot et al., 2007; Di Giuseppe et al., 2016). For instance, the European Forest Fire Information System (EFFIS) employs the FWI to provide information on wildfires in the EU and neighbouring countries (Di Giuseppe et al., 2020). The ~~main~~

~~advantage of the FWI system is its relatively straightforward computation, requiring only~~ FWI system only requires four weather parameters and information about the season (time of year) and geographical location as input parameters ~~. The FWI~~ and 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~~.~~, i.e. Drought Code (DC), Drought Moisture Code (DMC), and Fine Fuel Moisture Code (FFMC). 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, i.e. Initial Spread Index (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~~, with values below 11 considered low and those above 50 ~~are considered extreme~~classified as extreme, according to the EFFIS fire danger classes (EFFIS, 2020). 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~~.

2.2 Forecast and observation data

In this study, we use ensemble forecasts from the European Centre for Medium-Range Weather Forecasts (ECMWF) to calculate ensemble forecasts of the FWI. ECMWF medium-range ensemble forecasts consists of 50-members, initialized twice a day at 0000 and 1200 UTC and provide forecasts up to 360 hours (15 days). The forecasts are derived from the TIGGE archive (Bougeault et al., 2010), which provides operational medium-range ensemble forecast data for non-commercial research purposes. The data is accessible through the ECMWF API[1]. The temporal resolution of the ensemble forecast data used in this paper is 6h for all lead times and the spatial resolution is 0.5° (∼50 km in Central Europe). We use only the forecasts initialized

---

[1]https://apps.ecmwf.int/datasets/data/tigge/, last access: 06/02/2024

at 0000 UTC. Although available forecasts cover the whole globe, our focus is on the European region from 25°N to 72°N and 25°W to 39.80°E and we specifically use forecasts from the years 2021 to 2023.

For the FWI calculation, we derive initial values for FFMC, DMC, and DC from ERA5 reanalysis data (Hersbach et al., 2020) to account for preceding conditions at forecast initialization. The initial values are determined using the climatological mean from 40 years of historical data (1980-2019) for each day of the year, using a centered 15-day rolling mean on each day. ERA5 reanalysis data can be retrieved from the C3S Climate Data Store (CDS) (Hersbach et al., 2017).

For calibration and verification purposes, we use ECMWF high-resolution deterministic forecasts initialized at 0000 UTC, which are available from ECMWF's Meteorological Archival and Retrieval System (MARS[2]). ECMWF high-resolution forecasts have a spatial resolution of 0.1° ($\sim$9 km) and a temporal resolution of 1 hour and can therefore give a more accurate picture of the actual weather conditions than medium-range ensemble forecasts with a coarser resolution. Ideally, the FWI forecasts would be verified using FWI values calculated from surface observations of the relevant weather parameters, since the FWI cannot be directly observed. However, measurement stations that provide continuous observations of all necessary weather parameters are sparse and only yield point-wise verification. Furthermore, for an operational calibration of the FWI, observation data would need to be rapidly available. We therefore use the FWI calculated using ECMWF high-resolution forecasts with the shortest lead time to the local noon with corresponding 24h precipitation as substitute for actual observations. These FWI values are hereafter called analysis.

To determine if the analysis is suitable to be used as observation substitute, we checked their agreement with actual observation based values, which is shown in Figs. 1 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. 1b-d. The FWI analysis is shown in orange while the FWI calculated from observations is shown by the black dashed line. Overall, there is good agreement 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 Chrysoupoli, Greece (Fig. 1d).

Figure 2a shows 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 very 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 in 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 very low values

---

[2]https://apps.ecmwf.int/mars-catalogue/, last access: 11/09/2024

(FWI < 1) can be disregarded when assessing wild fire danger. The analysis tends to overestimate FWI values below 6 and underestimate values above 6. Notably, these higher values (FWI > 10) are the most relevant for assessing wildfire danger. 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 yellow and red coloured markers in the station map (Fig. 1a). ~~This is due to the difficulty in~~ These differences result from the difficulty of capturing the complex terrain and its small-scale weather phenomena with the relatively coarse resolution (~~~9km~~approximately 9 km) of the forecast model. ~~Because there is a fairly good correlation of observation and analysis,~~ Notably, 53% of the available stations are situated in these aforementioned countries, while relatively few stations are located in southern Europe where high FWI values are typically expected. This uneven distribution of observation stations may influence the overall correlation assessment. Moreover, this demonstrates the challenges of using observations for verification and calibration over larger study domains. The spatial distribution of available data across the entire area and the representativeness of point-wise observations for larger regions, particularly in complex terrain, must be considered. Given the generally good correlation between observations and analysis, despite the underestimation of larger FWI values, 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~~ a proxy for observations in the subsequent evaluations, ensuring the complete spatial coverage across the entire domain.

2.3 Calibration and verification methods

2.3.1 Non-homogeneous Gaussian regression

For statistical post-processing the medium-range FWI forecasts, we apply the non-homogeneous Gaussian regression (NGR), which was originally proposed and employed for surface temperature and sea level pressure by Gneiting et al. (2005). This method was extended to non-negative weather variables (wind speed) by Thorarinsdottir and Gneiting (2010). The FWI is by definition non-negative and for the calibration of the FWI forecasts we assume that the FWI observations y follow a truncated normal distribution with cut-off at zero:

$$y \sim \mathcal{N}^0(\mu, \sigma^2). \tag{1}$$

The location and scale parameter are given by:

$$\mu_{kl} = a_l + b_l \overline{\mathrm{ens}}_{kl},$$
$$\log(\sigma_{kl}) = c_l + d_l \log(\mathrm{sd}_{kl}), \tag{2}$$
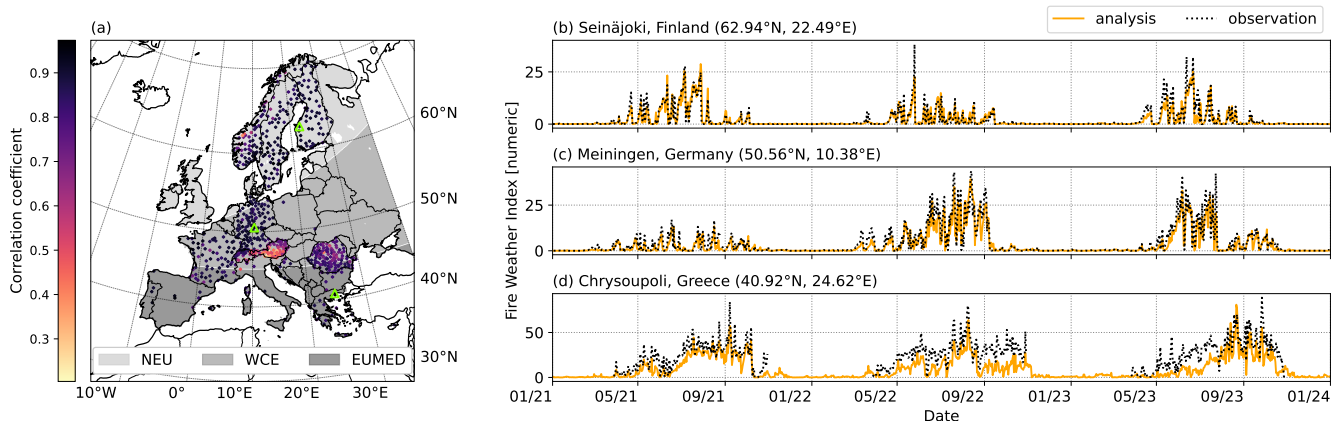
Figure 1. (a) Map of study area with modified AR6-WGI reference regions (Iturbide et al., 2020) shaded in grey (NEU=Northern Europe, WCE=West and Central Europe, EUMED=European Mediterranean). The location of observation stations for which the FWI can be calculated for at least 200 consecutive days are presented by dots with the colour representing the correlation coefficient (see. Fig. 2c) for each station. Example stations (b – d) are marked with green triangles. ~~(b)~~ Example of FWI time series calculated from observations (black dashed) and forecast analysis (orange) for a station (b) in NEU (Seinäjoki, Finland); (c) for WCE (Meiningen, Germany) and (d) EUMED (Chrysoupoli, Greece).

with ens_{kl} being the ensemble mean and sd_{kl} being the standard deviation of the 50 ensemble members for each location k and lead time l. a_l–d_l are regression coefficients. The logarithmic link $\log(\mathrm{sd}_{kl})$ is used to assure positive values for the scale parameter σ_{kl}.

The regression coefficients a_l–d_l are estimated by minimizing the average continuous ranked probability score (CRPS, Hersbach (2000)) from training data. The training data is defined using a specific training area and a 30-day rolling window preceding the forecast as training period. As training areas, we use here climatic reference regions, defined by the sixth IPCC Assessment Report (AR6-WGI (Iturbide et al., 2020)), which divide the European domain into Northern Europe (NEU), West and Central Europe (WCE), and the Mediterranean (MED). In this study, however, we use only the European part north of the Mediterranean Sea, ~~referred to hereafter~~ hereafter referred to as the European Mediterranean (EUMED). The reference regions are marked grey in Fig. ~~1(a)~~ 1a. The calibration can also be performed over smaller geographical areas, e.g., individual countries. Training ~~periods of different length~~ windows from 15 to 40 days were tested, as shorter training periods allow a faster adaptation to seasonal differences. On the other hand, longer periods provide more data, thereby reducing statistical variability. ~~Training windows from 15 to 40 days were examined and only~~ Only minor differences in the calibration performance were found when using big training areas, as in this example. When using smaller geographical training areas, however, a training period of 30 days seemed to most suitable (see Supplementary ~~Figure~~Fig. S1). Therefore, we chose a 30-day window for our analysis. 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. The initial forecast time step (T+12h) is excluded from calibration, as these forecasts are used as observation. The obtained coefficients are then used to calibrate
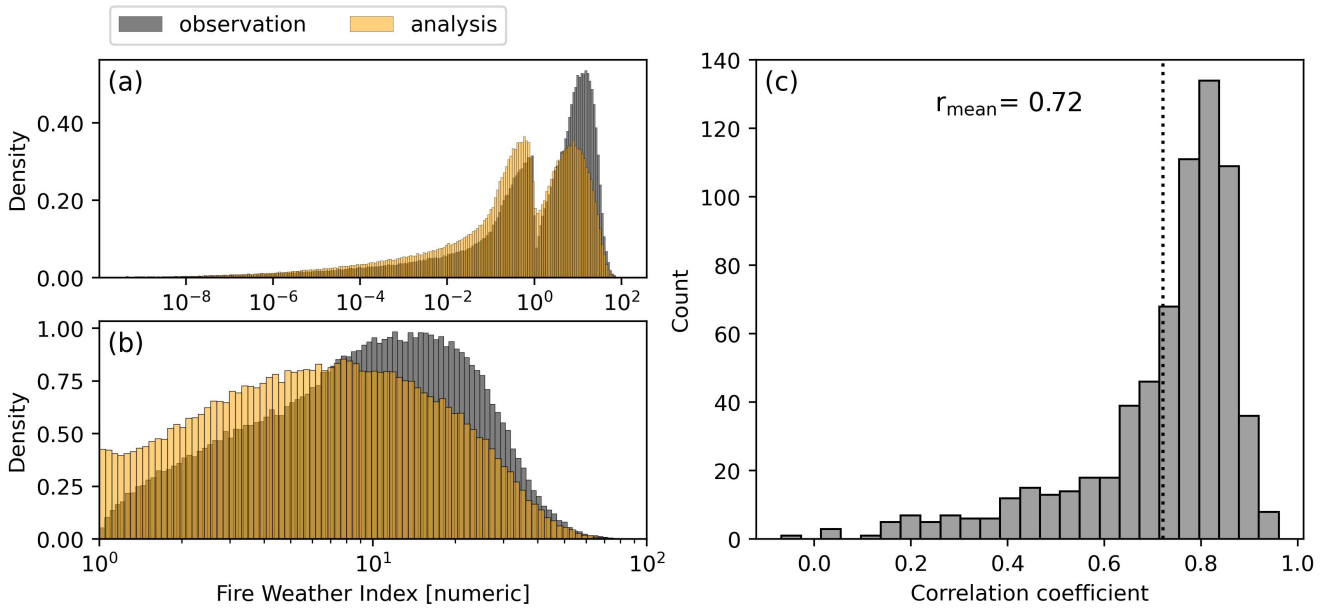
Figure 2. (a) Density plot of the FWI calculated from observation data (grey) and from FWI analysis (orange) for all the stations shown in Fig. 1a. (b) Same as in (a) but only showing FWI values greater than 1. (c) Frequency (count) of the correlation coefficient of FWI high-resolution analysis and FWI calculated using observation data for FWI values greater than 1. The vertical line marks the mean correlation coefficient.

the forecast at the respective lead time in the selected training area. Fitting of the regression model and prediction of location and scale parameters of the predicated distribution is done using the R-package *crch*, which provides censored regression with conditional heteroscedasticity (Messner et al., 2016) and uses the Broyden–Fletcher–Goldfarb–Shanno algorithm (Nocedal and Wright, 2006) to minimize the CRPS.

2.3.2 Verification metrics

The aim of forecast calibration is to correct forecast errors deriving from both structural deficiencies in the dynamical models and forecast sensitivity to uncertain initial conditions (Wilks and Vannitsem, 2018). To evaluate the predictive performance of calibrated forecasts compared to the raw forecasts, we are using ~~several verification metrics~~ the spread-skill relationship, mean error, and the Continuous Ranked Probability Skill Score (CRPSS), which are shortly introduced hereafter.

A common method to evaluate the forecast reliability of probabilistic forecasts is the ~~comparison of ensemble spread and skill.~~ ~~The~~ assessment of the spread-skill relationship (Weigel, 2011). A frequently used measure for the skill of an ensemble forecast

is ~~often defined in terms of~~ the root mean square error (RMSE) of the ensemble mean, calculated as

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(F_i - O_i)^2}, \tag{3}$$

where F_i and O_i are the predicted and observed value, respectively, at time step i of n forecasts. The ensemble spread is calculated ~~using~~ as the square root of the average ensemble variance, where the average ensemble variance is the mean of the variances of all ensemble members (Fortin et al., 2014). In a well calibrated forecast model, the ensemble spread should be on average equal to the RMSE. The ensemble forecast is considered underdispersive if the spread is smaller than the ~~skill~~ RMSE and overdispersive otherwise.

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):

$$\text{ME} = -\frac{1}{N}\frac{1}{n}\sum_{i=1}^{n}(F_i - O_i). \tag{4}$$

While the spread-skill relationship and ME are deterministic scores and applied to the ensemble forecast mean, the Continuous Ranked Probability Score (CRPS, Eq. ~~(5)~~ 5) allows a probabilistic assessment (Hersbach, 2000). The CRPS compares the whole distribution of ensemble members, represented as cumulative distribution function, with the observation:

$$\text{CRPS}(P, x_a) = \int_{-\infty}^{\infty}[F_{Fc}(x) - F_O(x)]^2 dx, \tag{5}$$

where $F_{Fc}(x)$ and $F_O(x)$ are the cumulative distribution functions of the forecast and the observation, respectively. The CRPS is negatively oriented, which means smaller values indicate a better performance of the ensemble forecast. In this study, we assume a truncated Gaussian distribution of the FWI forecasts and apply the truncated Gaussian form of the CRPS for raw and calibrated ensemble forecasts. The skill of the calibrated forecast with respect to a reference forecast can be accessed by calculating the Continuous Ranked Probability Skill Score (CRPSS), defined as:

$$\text{CRPSS} = 1 - \frac{\text{CRPS}_{cal}}{\text{CRPS}_{ref}}. \tag{6}$$

where CRPS_{cal} and CRPS_{ref} denote the CRPS of the calibrated and reference forecast, respectively. Positive values indicate a higher skill of the calibrated forecast, while negative values indicate a lower skill of the calibrated forecast with respect to the raw forecast.

3 Results

In this section, we present results of post-processing FWI forecasts of the years 2021 to 2023 by applying the ~~introduced calibration~~ NGR method. We use here the AR6-WGI reference regions, NEU, WCE and EUMED, introduced in Sec. 2.3.1 and shown in Fig.~~1(a)~~ 1a.

The main fire season in Europe is typically from May until October but varies strongly in length and intensity across regions, e.g. the fire season starts later and is shorter in Northern Europe compared to Southern Europe (San-Miguel-Ayanz et al., 2012). For the calibration verification, we therefore only focus on forecasts during the months May to October, when the FWI in all regions is substantially higher compared to off-season months (see Fig. 1b-d).

Figure 3 shows the spread-skill relationship for the raw (black) and calibrated (orange) FWI forecast averaged over the grid-points within the three study areas NEU (left), WCE (middle) and EUMED (right) and the wildfire season (May to October). Monthly averages can be found in the Supplementary material, Figs. S4 - S6. The RMSE of the climatology, shown by the solid blue line, is calculated similarly to the forecast RMSE. However, it uses the climatology derived from 40 years of ERA5 reanalysis data (see Sec.2.2) for each day of the year instead of using the forecast. Both calibrated as well as raw forecasts have a smaller RMSE compared to the climatology which indicates a general skill of the FWI forecasts compared to the climatology even at longer lead times. For raw and calibrated forecasts the spread (dashed line) is constantly smaller than the RMSE (solid line). This implies that the forecast is underdispersive and lacks spread. However, after calibration the ensemble spread is closer to the RMSE, which indicates that the reliability of the forecast is improved. The calibration also improves forecast accuracy, as indicated by the decreased RMSE, especially during the first forecast days. In Northern Europe, the RMSE of the calibrated forecast is slightly above the RMSE of the raw forecast after 180 hours (7 days) of forecast, whereas the RMSE of raw and calibrated forecast in Central and Mediterranean Europe is similar for forecasts longer than 228 hours (9 days) and 300 hours (12 days), respectively. The FWI forecast in Northern Europe lacks skill for lead times longer than 180 hours (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 in NEU and the lower variability of values in this region compared to WCE and EUMED (see Figs. S2 and S3). The applied calibration appears to be more effective for higher FWI values and shows limitations for smaller FWI values, i.e., values < 10. 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.

The mean error (ME) averaged over the three subregions and the fire weather season is shown in Fig. 4. Uncalibrated forecasts have a negative bias for all lead times, indicating that the forecasted FWI values are consistently lower than the observed values. In Northern Europe the mean error before calibration is around -0.6 for all lead times. In Middle and Southern Europe the mean error is more negative but increasing with lead time. The improvement of the mean error is especially contributed to forecasts in the months with high FWI, July and August for WCE (Fig. S8) and June to September in EUMED (Fig. S9). After calibration the ME is considerably improved. Best results seem to be achieved in the Mediterranean region, where the mean error after calibration is around zero. In Northern and Central Europe, the bias is slightly positive after calibration especially for longer lead times, ranging from less than 0.1 to 0.4.
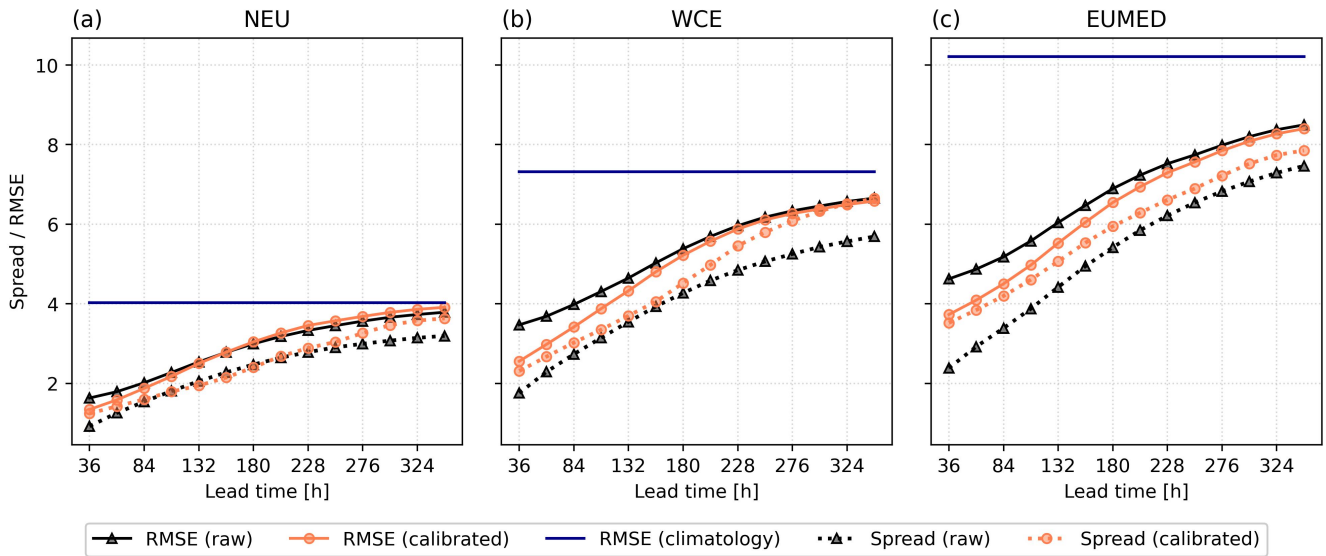
Figure 3. Spread and RMSE for raw and calibrated FWI forecasts averaged over the region of interest: (a) Northern Europe (NEU), (b) West & Central Europe (WCE), (c) European Mediterranean (EUMED) and fire season (May-October). The RMSE of the climatology for the respective region is additionally provided.
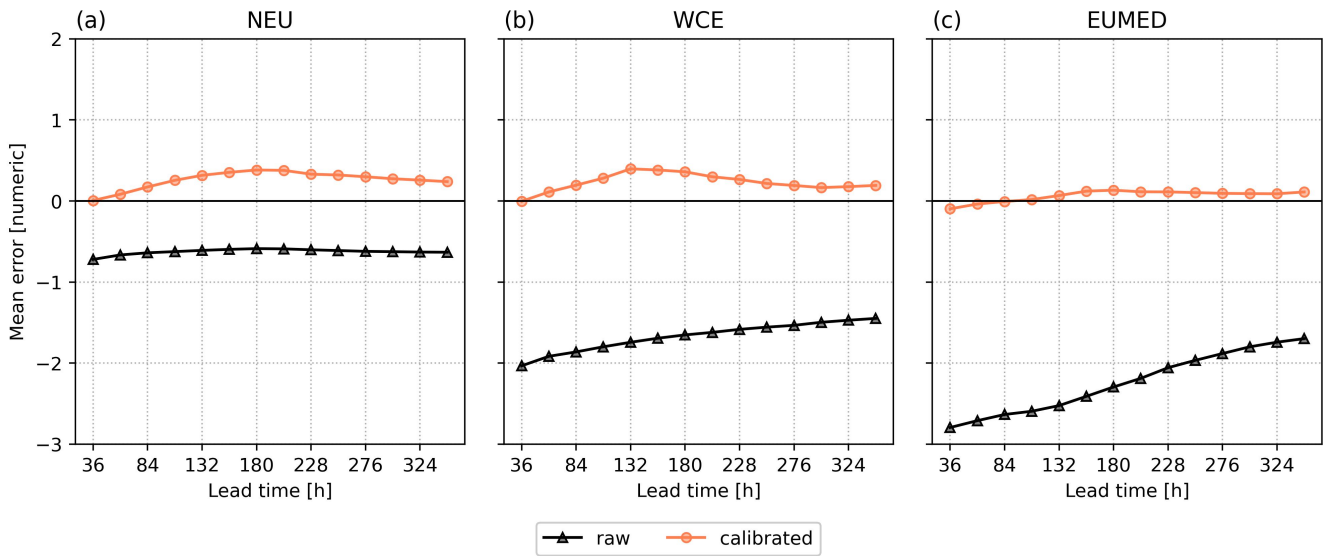
Figure 4. Mean Error for the raw and calibrated FWI forecasts averaged over the fire season (May-October) and the respective region: (a) Northern Europe (NEU), (b) Western & Central Europe (WCE), (c) European Mediterranean (EUMED).
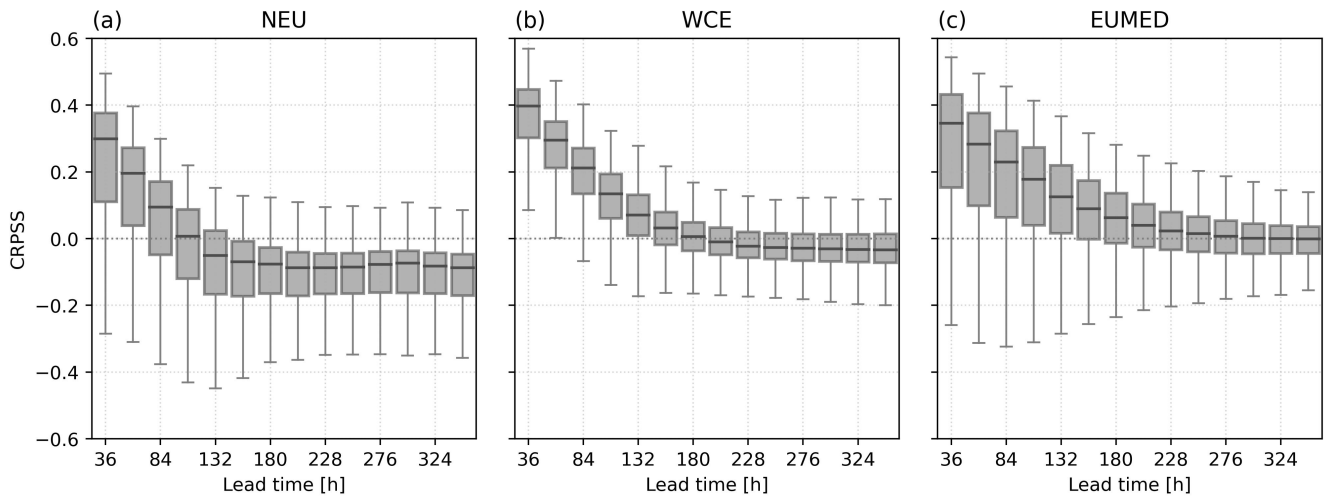
Figure 5. Continuous ranked probability skill score (CRPSS) for the calibrated forecasts against raw forecasts averaged over the fire season (May-October) and the respective region: (a) Northern Europe (NEU), (b) Western & Central Europe (WCE), (c) European Mediterranean (EUMED).

Figure 5 shows the CRPSS with raw forecasts as reference. In all three regions, the CRPSS is positive for the first ~~3 days (~~84 hours (3 days) of the forecast~~which suggests~~, suggesting an improvement of the FWI forecast after calibration for short lead times up to 84 hours. ~~The~~ However, the lead time up to which the calibration is improving the forecasts varies with region. In NEU, the calibration actually worsens the ~~calibration after 5 days (~~skill of the forecast after 132 hours (5 days) of forecast ~~, while in WCE and EUMEDthe~~ indicated by a negative median CRPSS. In WCE, the median CRPSS is negative after 204 hours (8 days), also indicating a decline in skill for longer lead times. In EUMED, however, the median CRPSS remains positive for all lead times. After 276 hours (11 days), the median CRPSS approaches zero, indicating that beyond this point, the calibrated forecast has no skill compared to the raw forecast~~after 7 (180 hours) and 11 days (276 hours), respectively. This is probably caused by the small~~. The declining skill in NEU and WCE is likely due to lower FWI values in ~~NEU and for longer lead times~~these regions (see Figs. S2 and S3). For low FWI values~~close to 0, the calibration is not effective. These regional differences are furthermore illustrated in the maps shown in Fig. 6, where the CRPSS averaged over the fire season (May-October) is given.~~, particularly those below 10, the calibration becomes less effective. 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~~and in Atlantic-influenced regions, e.g., northwest Spain and the United Kingdom (see Fig. 6). In these regions the FWI is generally low throughout the fire season ~~.~~(see Fig. S2).
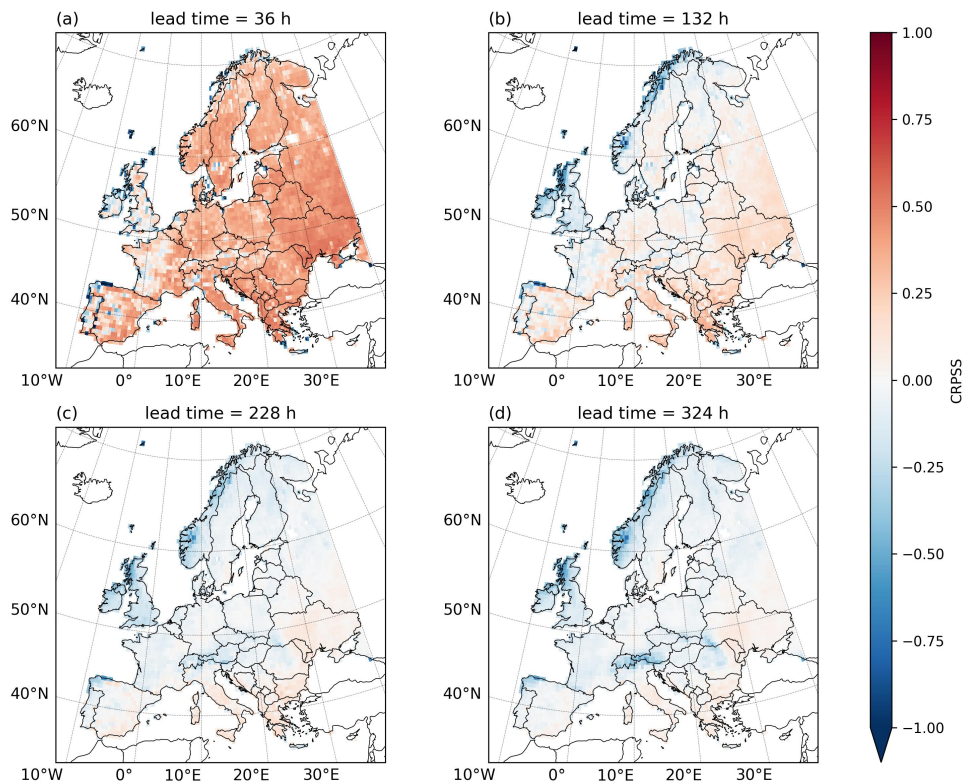
Figure 6. Continuous ranked probability skill score (CRPSS) with the raw forecast as reference averaged over the fire season (May-October) for different lead times.

4   Discussion and conclusion

We investigated whether non-homogeneous Gaussian regression (NGR) can be used to calibrate fire weather index (FWI) forecasts based on ~~medium range~~ medium-range ensemble weather forecasts by the European Centre for Medium-Range Weather Forecasts ensemble forecasts (ECMWF). To estimate the calibration coefficients of the NGR, we employ a truncated Gaussian distribution with a cut-off at zero and use forecast and observation pairs of the last 30 days preceding the forecast. ~~We~~ NGR is a well established and effective method for calibrating probabilistic forecasts and provides good results compared to other calibration approaches, such as bias correction (Cannon, 2018) or correcting the individual input parameters, as described by Worsnop et al. (2021). While correcting individual input parameters can be beneficial, it requires different models for each variable, demanding careful verification and access to quality controlled observational data over the whole study region. Instead of observed weather parameters, we used ECMWF high-resolution weather forecasts with the shortest possible lead time to calculate the FWI~~analysis and used these as~~. The FWI values from these forecasts are referred to as the FWI analysis and are used as a substitute for observations. Although the FWI analysis seems to slightly underestimate the FWI ~~slightly~~in the

range associated with elevated fire danger (FWI > 10), a good correlation is observed. Utilizing observations would have been possible, but would have required careful consideration of the quality, availability and representativity of the observation data.

Specifically, wind and precipitation observations might not be of sufficient quality at all stations, and point observations might not accurately represent larger areas. To ensure a consistent coverage of the entire study area and time period we chose to use the FWI analysis for calibration and verification instead of observations.

FWI forecasts using medium range ensemble weather forecasts perform generally quite well when compared to the analysis. However, calibration improves the forecasts especially at short lead times up to 84 hours (3 days). In the Mediterranean region and Central Europe an improvement of FWI forecast with respect to the FWI analysis is also apparent for longer lead times up to 8 to 10 days, respectively. This is likely caused by the generally higher values in those regions ~~and is~~ (shown in Figs. S2 and S3), and further supported by the monthly averaged metrics in the appendix (Figs. S7-S9), which show a stronger improvement caused by the calibration in the months with high FWI values. Hence, it can be concluded that the calibration performs better for higher FWI values (FWI > 10), as indicated by the better performance during the summer months in West and Central Europe and in the European Mediterranean. However, the calibration method shows limitations for low FWI values (FWI < 10), which could be observed in the NEU reference region and especially for longer lead times. Although ~~it would be ideal for the method to~~ ideally the method would perform effectively across the entire range of FWI values, ~~it is generally more critical that it demonstrates a good~~ its performance at higher FWI values ~~, where the potential fire danger is more significant. While long-range forecasts~~ is generally more critical, as higher values are associated with greater fire danger. Forecasts of potential fire danger ~~are valuable, short-range forecasts , especially~~ for extended periods (1-2 weeks) are valuable; however, short-term forecasts for the first 1-3 days ~~,~~ are usually more critical for firefighting resource management.

To further improve the presented calibration method for fire weather index forecasts, it could be tested if calibration of individual components of the FWI system (e.g. FFMC, DMC and DC) would improve overall skill of the forecast. Furthermore, more advanced models using additional predictors, e.g. elevation or land-use, could improve the calibration but were not tested here.

Reliable and accurate probabilistic forecasts, particularly when the fire risk is high and over the short-term time frame, are crucial for decision-makers and emergency responders to effectively coordinate resources. 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~~, e.g., bias correction (Cannon, 2018) or the correction of the input parameters instead of the FWI as in Worsnop et al. (2021). When correcting the individual input parameters, different models need to be applied for the individual variables, which require careful verification and access to quality controlled observation over the whole study region. To further improve the presented calibration method for fire weather index forecasts, it could be tested if calibration of individual components of the FWI system e.g. FFMC, DMC and DC would improve overall skill of the forecast. Furthermore, more advanced models using additional predictors, e.g. elevation or land-use, could improve the calibration but were not tested here.~~.

Code and data availability. The TIGGE data that was used for demonstrating the method is freely available in the TIGGE archive (https://apps.ecmwf.int/datasets/data/tigge/ Bougeault et al. (2010)). ERA5 reanalysis data which was used to calculate climatologies is freely available on the Climate Data Store (https://doi.org/10.24381/cds.143582cf, Hersbach et al. (2017)). Other data and code can be made available from the authors upon request.

310

Author contributions. SB wrote the manuscript with the help of ML. SB and ML developed the FWI calculation and calibration methods.

Competing interests. The authors declare that there are no competing interests.

Acknowledgements. This work has been supported by the Horizon 2020-funded project SAFERS "Structured Approaches for Forest Fire Emergencies in Resilient Societies" (H2020/Innovation Action), grant agreement No. 869353. The verification is based on TIGGE data.

315 TIGGE (The International Grand Global Ensemble) is an initiative of the World Weather Research Programme (WWRP).

References

Barriopedro, D., Fischer, E. M., Luterbacher, J., Trigo, R. M., and García-Herrera, R.: The Hot Summer of 2010: Redrawing the Temperature Record Map of Europe, Science, 332, 220–224, https://doi.org/10.1126/science.1201224, 2011.

Bougeault, P., Toth, Z., Bishop, C., Brown, B., Burridge, D., Chen, D. H., Ebert, B., Fuentes, M., Hamill, T. M., Mylne, K., Nicolau, J., Paccagnella, T., Park, Y.-Y., Parsons, D., Raoult, B., Schuster, D., Dias, P. S., Swinbank, R., Takeuchi, Y., Tennant, W., Wilson, L., and Worley, S.: The THORPEX Interactive Grand Global Ensemble, Bulletin of the American Meteorological Society, 91, 1059 – 1072, https://doi.org/10.1175/2010BAMS2853.1, 2010.

Bremnes, J. B.: Probabilistic Forecasts of Precipitation in Terms of Quantiles Using NWP Model Output, Monthly Weather Review, 132, 338 – 347, https://doi.org/10.1175/1520-0493(2004)132<0338:PFOPIT>2.0.CO;2, 2004.

Cannon, A. J.: Multivariate quantile mapping bias correction: an N-dimensional probability density function transform for climate model simulations of multiple variables, Climate Dynamics, 50, 31–49, https://doi.org/10.1007/s00382-017-3580-6, 2018.

de Groot, W. J., Field, R. D., Brady, M. A., Roswintiarti, O., and Mohamad, M.: Development of the Indonesian and Malaysian Fire Danger Rating Systems, Mitigation and Adaptation Strategies for Global Change, 12, 165–180, https://doi.org/10.1007/s11027-006-9043-8, 2007.

De Rigo, D., Libertà, G., Houston Durrant, T., Vivancos, T. A., and San-Miguel-Ayanz, J.: Forest fire danger extremes in Europe under climate change: variability and uncertainty, Research report, Publications Office of the European Union, https://doi.org/10.2760/13180, 2017.

Di Giuseppe, F., Pappenberger, F., Wetterhall, F., Krzeminski, B., Camia, A., Libertá, G., and San-Miguel, J.: The Potential Predictability of Fire Danger Provided by Numerical Weather Prediction, Journal of Applied Meteorology and Climatology, 55, 2469 – 2491, https://doi.org/10.1175/JAMC-D-15-0297.1, 2016.

Di Giuseppe, F., Vitolo, C., Krzeminski, B., Barnard, C., Maciel, P., and San-Miguel, J.: Fire Weather Index: the skill provided by the European Centre for Medium-Range Weather Forecasts ensemble prediction system, Natural Hazards and Earth System Sciences, 20, 2365–2378, https://doi.org/10.5194/nhess-20-2365-2020, 2020.

EFFIS: EFFIS Annual Statistics for Greece, https://effis.jrc.ec.europa.eu/apps/effis.statistics/estimates/GRC, last accessed: 27/02/2024.

EFFIS: The European Forest Fire Information System- User Guide to EFFIS applications, available at: https://effis-gwis-cms.s3-eu-west-1.amazonaws.com/apps/effis.viewer/userguide.pdf, last access: 11/09/2024, 2020.

Faiola, A. and Labropoulou, E.: How wildfires are threatening the Mediterranean way of life, The Washington Post, https://www.washingtonpost.com/world/2023/09/02/greece-fires-2023-rhodes/, 2023.

Fortin, V., Abaza, M., Anctil, F., and Turcotte, R.: Why Should Ensemble Spread Match the RMSE of the Ensemble Mean?, Journal of Hydrometeorology, 15, 1708 – 1713, https://doi.org/10.1175/JHM-D-14-0008.1, 2014.

Gneiting, T., Raftery, A. E., Westveld, A. H., and Goldman, T.: Calibrated probabilistic forecasting using ensemble model output statistics and minimum CRPS estimation, Mon. Wea. Rev., 133, 1098–1118, https://doi.org/https://doi.org/10.1175/MWR2904.1, 2005.

Hagedorn, R., Hamill, T. M., and Whitaker, J. S.: Probabilistic Forecast Calibration Using ECMWF and GFS Ensemble Reforecasts. Part I: Two-Meter Temperatures, Monthly Weather Review, 136, 2608 – 2619, https://doi.org/10.1175/2007MWR2410.1, 2008.

Hamill, T. M. and Colucci, S. J.: Verification of Eta–RSM Short-Range Ensemble Forecasts, Monthly Weather Review, 125, 1312 – 1327, https://doi.org/10.1175/1520-0493(1997)125<1312:VOERSR>2.0.CO;2, 1997.

Hamill, T. M., Whitaker, J. S., and Wei, X.: Ensemble Reforecasting: Improving Medium-Range Forecast Skill Using Retrospective Forecasts, Monthly Weather Review, 132, 1434 – 1447, https://doi.org/10.1175/1520-0493(2004)132<1434:ERIMFS>2.0.CO;2, 2004.

Hamill, T. M., Hagedorn, R., and Whitaker, J. S.: Probabilistic Forecast Calibration Using ECMWF and GFS Ensemble Reforecasts. Part II: Precipitation, Monthly Weather Review, 136, 2620 – 2632, https://doi.org/10.1175/2007MWR2411.1, 2008.

355 Hersbach, H.: Decomposition of the Continuous Ranked Probability Score for Ensemble Prediction Systems, Weather and Forecasting, 15, 559 – 570, https://doi.org/10.1175/1520-0434(2000)015<0559:DOTCRP>2.0.CO;2, 2000.

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., Nicolas, J., Peubey, C., Radu, R., Schepers, D., Simmons, A., Soci, C., Abdalla, S., Abellan, X., Balsamo, G., Bechtold, P., Biavati, G., Bidlot, J., Bonavita, M., De Chiara, G., Dahlgren, P., Dee, D., Diamantakis, M., Dragani, R., Flemming, J., Forbes, R., Fuentes, M., Geer, A., Haimberger, L., Healy, S., Hogan, R., Hólm, E., Janisková,

360 M., Keeley, S., Laloyaux, P., Lopez, P., Lupu, C., Radnoti, G., de Rosnay, P., Rozum, I., Vamborg, F., Villaume, S., and Thépaut, J.-N.: Complete ERA5 from 1940: Fifth generation of ECMWF atmospheric reanalyses of the global climate, Copernicus Climate Change Service (C3S) Data Store (CDS), https://doi.org/10.24381/cds.143582cf, last accessed: 22/03/2024, 2017.

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., Nicolas, J., Peubey, C., Radu, R., Schepers, D., Simmons, A., Soci, C., Abdalla, S., Abellan, X., Balsamo, G., Bechtold, P., Biavati, G., Bidlot, J., Bonavita, M., De Chiara, G., Dahlgren,

365 P., Dee, D., Diamantakis, M., Dragani, R., Flemming, J., Forbes, R., Fuentes, M., Geer, A., Haimberger, L., Healy, S., Hogan, R. J., Hólm, E., Janisková, M., Keeley, S., Laloyaux, P., Lopez, P., Lupu, C., Radnoti, G., de Rosnay, P., Rozum, I., Vamborg, F., Villaume, S., and Thépaut, J.-N.: The ERA5 global reanalysis, Quarterly Journal of the Royal Meteorological Society, 146, 1999–2049, https://doi.org/https://doi.org/10.1002/qj.3803, 2020.

Ibebuchi, C. C. and Abu, I.-O.: Characterization of temperature regimes in Western Europe, as regards the summer 2022 Western European

370 heat wave, Climate Dynamics, 61, 3707 – 3720, https://doi.org/10.1007/s00382-023-06760-4, 2023.

Ionita, M., Tallaksen, L. M., Kingston, D. G., Stagge, J. H., Laaha, G., Van Lanen, H. A. J., Scholz, P., Chelcea, S. M., and Haslinger, K.: The European 2015 drought from a climatological perspective, Hydrology and Earth System Sciences, 21, 1397–1419, https://doi.org/10.5194/hess-21-1397-2017, 2017.

Iturbide, M., Gutiérrez, J. M., Alves, L. M., Bedia, J., Cerezo-Mota, R., Cimadevilla, E., Cofiño, A. S., Di Luca, A., Faria, S. H., Gorodet-

375 skaya, I. V., Hauser, M., Herrera, S., Hennessy, K., Hewitt, H. T., Jones, R. G., Krakovska, S., Manzanas, R., Martínez-Castro, D., Narisma, G. T., Nurhati, I. S., Pinto, I., Seneviratne, S. I., van den Hurk, B., and Vera, C. S.: An update of IPCC climate reference regions for subcontinental analysis of climate model data: definition and aggregated datasets, Earth System Science Data, 12, 2959–2970, https://doi.org/10.5194/essd-12-2959-2020, 2020.

Lawson, B. D. and Armitage, O. B.: Weather guide for the Canadian Forest Fire Danger Rating System, Natural Resources Canada Canadian

380 Forest Service Northern Forestry Centre, Edmonton, Alberta, 2008.

Messner, J. W., Mayr, G. J., and Zeileis, A.: Heteroscedastic Censored and Truncated Regression with crch, The R Journal, 8, 173–181, https://doi.org/10.32614/RJ-2016-012, 2016.

Nocedal, J. and Wright, S. J.: Quasi-Newton Methods, pp. 135–163, Springer New York, New York, NY, https://doi.org/10.1007/978-0-387-40065-5_6, 2006.

385 Raftery, A. E., Gneiting, T., Balabdaoui, F., and Polakowski, M.: Using Bayesian Model Averaging to Calibrate Forecast Ensembles, Monthly Weather Review, 133, 1155 – 1174, https://doi.org/10.1175/MWR2906.1, 2005.

Rodrigues, M., Cunill Camprubí, A., Balaguer-Romano, R., Coco Megía, C. J., Castañares, F., Ruffault, J., Fernandes, P. M., and Resco de Dios, V.: Drivers and implications of the extreme 2022 wildfire season in Southwest Europe, Science of The Total Environment, 859, 160 320, https://doi.org/https://doi.org/10.1016/j.scitotenv.2022.160320, 2023.

390   Roulston, M. S. and Smith, L. A.: Combining dynamical and statistical ensembles, Tellus A: Dynamic Meteorology and Oceanography, https://doi.org/10.3402/tellusa.v55i1.12082, 2003.

Rousi, E., Fink, A. H., Andersen, L. S., Becker, F. N., Beobide-Arsuaga, G., Breil, M., Cozzi, G., Heinke, J., Jach, L., Niermann, D., Petrovic, D., Richling, A., Riebold, J., Steidl, S., Suarez-Gutierrez, L., Tradowsky, J. S., Coumou, D., Düsterhus, A., Ellsäßer, F., Fragkoulidis, G., Gliksman, D., Handorf, D., Haustein, K., Kornhuber, K., Kunstmann, H., Pinto, J. G., Warrach-Sagi, K., and Xoplaki, E.: The extremely
395   hot and dry 2018 summer in central and northern Europe from a multi-faceted weather and climate perspective, Natural Hazards and Earth System Sciences, 23, 1699–1718, https://doi.org/10.5194/nhess-23-1699-2023, 2023.

San-Miguel-Ayanz, J., Schulte, E., Schmuck, G., Camia, A., Strobl, P., Libertà, G., Giovando, C., Boca, R., Sedano, F., Kempeneers, P., McInerney, D., Withmore, C., de Oliveira, S. S., Rodrigues, M., Durrant, T., Corti, P., Oehler, F., Vilar, L., and Amatulli, G.: Comprehensive Monitoring of Wildfires in Europe: The European Forest Fire Information System (EFFIS), in: Approaches to Managing Disaster,
400   edited by Tiefenbacher, J., chap. 5, IntechOpen, Rijeka, https://doi.org/10.5772/28441, 2012.

San-Miguel-Ayanz, J., Durrant, T., Boca, R., Libertà, G., Branco, A., de Rigo, D., Ferrari, D., Maianti, P., Vivancos, T. A., Pfeiffer, H., Loffler, P., Nuijten, D., Leray, T., and Jacome Felix Oom, D.: Forest Fires in Europe, Middle East and North Africa 2018, Tech. Rep. EUR 29856 EN, Publications Office of the European Union, Luxembourg, https://doi.org/10.2760/1128, JRC117883, 2019.

San-Miguel-Ayanz, J., Durrant, T., Boca, R., Maianti, P., Libertà, G., Artes Vivancos, T., Jacome Felix Oom, D., Branco, A., De Rigo, D.,
405   Ferrari, D., Pfeiffer, H., Grecchi, R., Nuijten, D., Onida, M., and Löffler, P.: Forest Fires in Europe, Middle East and North Africa 2020, Tech. Rep. EUR 30862 EN, Publications Office of the European Union, Luxembourg, https://doi.org/10.2760/216446, JRC126766, 2021.

Sibley, A. M.: Wildfire outbreaks across the United Kingdom during summer 2018, Weather, 74, 397–402, https://doi.org/https://doi.org/10.1002/wea.3614, 2019.

Skacel, J., Kahn, M., and Hovet, J.: Czech and German firefighters battle blaze in national park, Reuters, https://www.reuters.com/business/
410   environment/czech-german-firefighters-battle-blaze-national-park-2022-07-26/, 2022.

Thorarinsdottir, T. L. and Gneiting, T.: Probabilistic forecasts of wind speed: ensemble model output statistics by using heteroscedastic censored regression, Journal of the Royal Statistical Society: Series A (Statistics in Society), 173, 371–388, https://doi.org/10.1111/j.1467-985X.2009.00616.x, 2010.

Thorarinsdottir, T. L. and Johnson, M. S.: Probabilistic Wind Gust Forecasting Using Nonhomogeneous Gaussian Regression, Monthly
415   Weather Review, 140, 889 – 897, https://doi.org/10.1175/MWR-D-11-00075.1, 2012.

Turco, M., Jerez, S., Augusto, S., Tarín-Carrasco, P., Ratola, N., Jiménez-Guerrero, P., and Trigo, R. M.: Climate drivers of the 2017 devastating fires in Portugal, Scientific Reports, 9, 13 886, https://doi.org/10.1038/s41598-019-50281-2, 2019.

Van Wagner, C. E.: Development and structure of the Canadian forest fire weather index system, Canadian Forestry Service, Headquarters, Ottawa,, 35, 1987.

420   Wang, Y., Anderson, K., and Suddaby, R.: Updated source code for calculating fire danger indices in the Canadian Forest Fire Weather Index System, Natural Resources Canada Canadian Forest Service Northern Forestry Centre, Edmonton, Alberta, 2015.

Weigel, A. P.: Ensemble Forecasts, in: Forecast Verification, chap. 8, pp. 141–166, John Wiley & Sons, Ltd, https://doi.org/https://doi.org/10.1002/9781119960003.ch8, 2011.

Wilks, D. S. and Vannitsem, S.: Chapter 1 - Uncertain Forecasts From Deterministic Dynamics, pp. 1–13, Elsevier,
425   https://doi.org/https://doi.org/10.1016/B978-0-12-812372-0.00001-7, 2018.

Worsnop, R. P., Scheuerer, M., Giuseppe, F. D., Barnard, C., Hamill, T. M., and Vitolo, C.: Probabilistic Fire Danger Forecasting: A Framework for Week-Two Forecasts Using Statistical Postprocessing Techniques and the Global ECMWF Fire Forecast System (GEFF), Weather and Forecasting, 36, 2113 – 2125, https://doi.org/10.1175/WAF-D-21-0075.1, 2021.