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. Wildfire risk is usually estimated using 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 ~~post-processing~~ 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 ~~post-processing~~ calibrating ensemble forecasts of the Canadian Forest Fire Weather Index. 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.

1 Introduction

Wildfires in Europe have become increasingly ~~prevalent~~ frequent and destructive in the ~~last~~ recent decades. The ~~recent wildfire in Greece~~ 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 ~~home (Faiola and Labropoulou, 2023). Also,~~ homes (Faiola and Labropoulou, 2023). Similar devastating wildfires in 2017 and 2022 ~~were years with devastating wildfires, burning~~ burned large areas in Portugal, Spain ~~, Greece~~ and Italy. ~~But not just~~ While the Mediterranean region ~~is affected by wildfires, also Middle~~ continues to face the highest occurrence of wildfires, Central and Northern Europe ~~is experiencing more and more often unusually dry and warm summers. Those extended~~ 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 ~~. One example is the heatwave~~ (San-Miguel-Ayanz et al., 2021; De Rigo et al., 2017). Examples are the 2018 heatwave, which caused wildfires ~~and over 20 thousand hectares burned land~~ 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 ~~seriousness~~ severity of wildfires in Europe emphasizes 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. ~~Accurate~~ An important component of SAFERS for identifying high wildfire risk areas are accurate and reliable weather forecasts, ranging from a ~~couple of days to multiple weeksto identify high wildfire risk areas is an important part of SAFERS. Here~~few days to several weeks. In this paper, we use the Canadian forest fire weather index ~~, short FWI~~(FWI), which is a widely recognized numeric indicator for forest fire risk ~~(Van Wagner, 1987)~~(Van Wagner, 1987; Di Giuseppe et al., 2016). The calculation of the FWI only requires four weather parameters and can be calculated using 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 ~~post-processing~~ calibration to ensure reliable and accurate forecasts, which are essential for making informed decisions and effectively allocating resources when responding to wildfires.~~One widely used method for calibrating ensemble forecast is~~ 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) ~~(Gneiting et al., 2005), which is commonly used for the calibration of~~ 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.

In this article~~we show that NGRcan also be used for the calibration of~~ , we investigate whether non-homogeneous Gaussian regression (NGR) can be used to calibrate the Fire Weather Index (FWI) derived from medium-range ~~fire weather index forecasts~~ 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 ~~Fire weather index calculation~~Data and methods

2.1 Fire weather index calculation

A common method ~~to indicate the danger for~~ 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, ~~it is used in many other regions, e. g., by~~ 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 counties (Di Giuseppe et al., 2020).~~

~~One advantage of using FWI is the relatively simple calculation only requiring~~ neighbouring countries (Di Giuseppe et al., 2020) . The main advantage of the FWI system is its relatively straightforward computation, requiring only four weather parameters ~~in addition to information of~~ and information about the season (time of year) and geographical location as input parameters. 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~~of different depth and diameter: the fine fuel moisture code~~ . 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) ~~, duff moisture code~~ represents the moisture content of litter and other fine cured fuels at a nominal depth of 1.2 cm; the Duff Moisture Code (DMC) ~~and drought moisture (DC). To consider the different~~ 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, ~~and therefore~~ 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 ~~used with regard to~~ 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 ~~, DMC and DC~~ and the 10-meter wind speed are used to model the potential rate of fire spread (ISI)~~and the potential~~ . DMC and DC are used to calculate the Buildup Index (BUI), a numeric rating of the total amount of fuel available for ~~surface fuel consumption (BUI)~~ 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. ~~Often the FWI is~~ 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 and~~ , vegetation types). Consequently, what is considered a low or extreme FWI in one region may not be ~~the same~~ 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.

~~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. We use the climatological mean values of FFMC, DMC and DC calculated using ERA5 reanalysis data as initial values. These climatological values are calculated from 40-years historical data (1980-2019) for each day of the year with 15 day rolling mean around each day.~~

3 ~~Forecast and observation data~~

2.1 Forecast and observation data

~~For FWI calculations~~ In this study, we use ensemble forecasts ~~of ECMWF's operational ensemble forecast system (ENS)~~ from the European Centre for Medium-Range Weather Forecasts (ECMWF) to calculate ensemble forecasts of the FWI. ECMWF medium-range ensemble forecasts consists of ~~51-members~~50-members, initialized twice a day at 0000 and 1200 UTC and provide forecasts up to 360 hours (15 days). The ~~spatial resolution of ECMWF medium-range ensemble forecasts was 0.2° during the period considered here. However, to have a larger set of available data for verification, we used forecast data derived form~~ forecasts are derived from the TIGGE archive (Bougeault et al., 2010). ~~TIGGE~~ , which provides operational medium-range ensemble forecast data for non-commercial research purposes~~from 13 global NWP centres~~. The data is accessible through the ECMWF API[1]. The temporal resolution of the ~~used TIGGE data~~ ensemble forecast data used in this paper is 6h for all lead times and the spatial resolution is 0.5° ~~.~~ (∼50 km). We use only the forecasts initialized at 0000 UTC. Although available forecasts cover the whole globe, ~~we focus here~~ our focus is on the European region from 25°N to 72°N and 25°W to 39.80°E ~~. For this study, we use forecasts of~~ and we specifically use forecasts from the years 2021 to 2023.

~~The FWI can not be observed directly and needs to be calculated using surface observations of the relevant weather parameters. 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 needs to be available rapidly.We therefore use FWI calculated using~~

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 from 40 years of historical data (1980-2019) for each day of the year, using a 15-day rolling mean centered 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 forecasts ~~with the shortest lead time to the local noon with corresponding 24h precipitation forecast as substitute for FWI observations~~initialized at 0000 UTC. ECMWF high-resolution forecasts have a spatial resolution of 0.1° (∼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. ~~To determine if those short term FWI forecasts,~~ The FWI values calculated using ECMWF high-resolution forecasts with the shortest lead time to the local noon with corresponding 24h precipitation act as substitute for actual observations and are hereafter called analysis~~, are~~ . Ideally, FWI forecasts would be verified using FWI values calculated using surface observations of the relevant weather parameters as the FWI cannot be observed directly. 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 needs to be available rapidly.

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. **??**1 and 2. We use observations available from the Finnish Meteorological Institute's observation database for the years 2021–2023~~for Finland and other European countries. The map in Fig. **??** .~~ 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. ~~Stations from outside of Finland are not necessarily quality~~

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

~~controlled and therefore shown separately.Figure 1b shows the scatter plot of analysis~~ 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 ~~observations~~ 2023 in Meiningen, Germany (Fig.1c) and Chrysoupoli, 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. ~~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~~ 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 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 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 coloured 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 (~~0.63), a correlation is apparent. This good correlation can also be seen in~~ 9km) of the forecast model.

Because there is a fairly good correlation of observation and analysis, we will use the ~~time series examples for a station in Finland and Greece. In the following, we are using the~~ forecasted FWI with short lead time (analysis) derived from ECMWF high-resolution forecasts as observation to compare with ~~longer forecasts~~ the FWI forecasts derived from ECMWF medium-range ensemble forecasts.

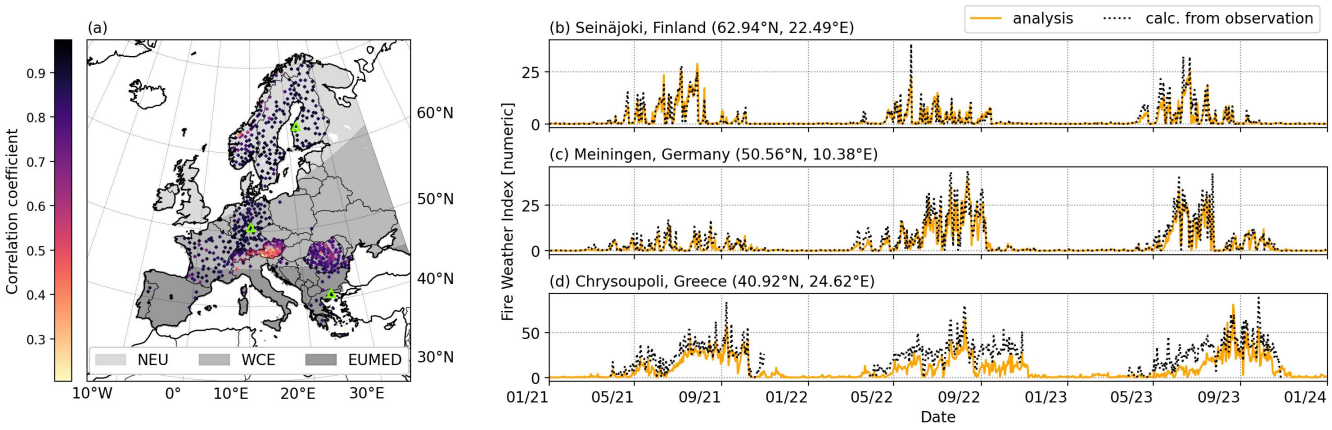

Figure 1. ~~Top left:~~ (a) Map of study area with modified ~~AR5~~ AR6-WGI reference regions (Iturbide et al., 2020) shaded in grey (NEU=Northern Europe, WCE=West and Central Europe, EUMED=European Mediterranean). ~~Location~~ The location of observation stations ~~(dots)~~ for which the FWI can be calculated for at least 200 consecutive days ~~. Top right: Correlation of FWI high-resolution analysis and FWI calculated using observation data at~~ are presented by dots with the ~~locations shown in~~ colour representing the ~~map~~correlation coefficient (see. ~~Data from Finnish stations is quality controlled and shown separately~~Fig.~~The solid line illustrates a perfect correlation~~2) for each station. ~~Bottom: Examples~~ Example stations are marked with green triangles. (b) Example of FWI time series calculated from observations (~~orange~~black dashed) and forecast analysis (~~dashed~~orange) for a station in NEU (Seinäjoki, Finland); (c) for WCE (Meiningen, Germany) and (d) EUMED (Chrysoupoli, Greece).

3 ~~Calibration and verification methods~~

2.1 Calibration and verification methods

2.2 ~~Non-homogeneous Gaussian regression~~

2.1.1 Non-homogeneous Gaussian regression

For statistical post-processing the medium-range FWI forecasts, we apply the non-homogeneous Gaussian regression (NGR), ~~also called ensemble model output statistics (EMOS) approach,~~ 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:

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

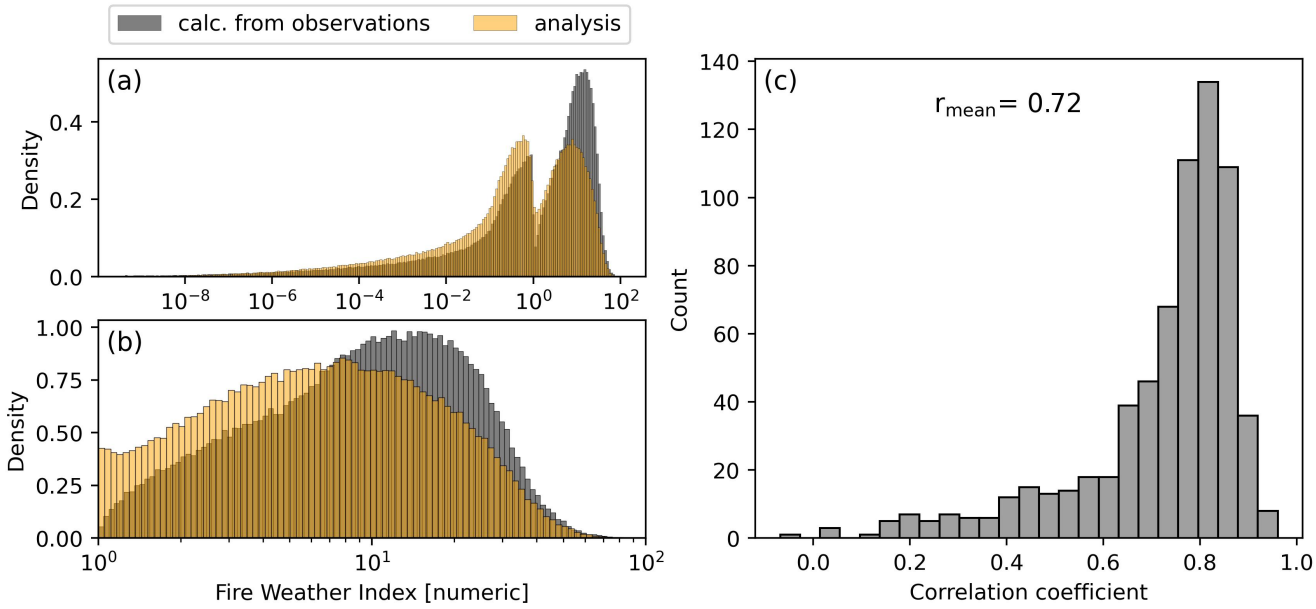

Figure 2. (a) Density plot of the FWI calculated from observation data (grey) and from FWI analysis (orange) for all the stations shown in Fig.1a. (b) Same as in (a) but only showing FWI values greater than 1. (c) Frequency (count) of the correlation coefficient of FWI high-resolution analysis and FWI calculated using observation data for FWI values greater than 1.

The location and scale parameter are given by:

$$\begin{aligned} \mu_{kl} &= a_l + b_l \overline{\mathrm{ens}}_{kl}, \\ \log(\sigma_{kl}) &= c_l + d_l \log(\mathrm{sd}_{kl}), \end{aligned} \tag{2}$$

with ens_{kl} being the ensemble mean and sd_{kl} being the standard deviation of the ~~51~~ 50 ensemble members for each location k and lead time l. a_l–d_l are regression coefficients. The logarithmic link ~~log(sd)~~ $\log(\mathrm{sd_{kl}})$ is used to assure positive values for the scale parameter ~~σ~~σ_{kl}.

The regression coefficients a_l–d_l are estimated by minimizing the average continuous ranked probability score (CRPS, Hersbach (2000)) ~~over a selected training period~~from training data. The training ~~period is here defined as a rolling window of 30 days prior to the forecast . 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). We tested 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 as the European Mediterranean (EUMED). The reference regions are marked grey in Fig.1(a). The calibration can also be performed over smaller geographical areas, e.g., individual countries.

Training periods of different length 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 ~~tested~~ examined and 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 ~~.The coefficients are estimated for~~ (see Supplementary Figure S1). 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 ~~separately, excluding the first time step of the forecast~~ for the given forecast. The initial forecast time step (T+12h) is excluded from calibration, as these forecasts ~~serve~~ are used as observation. The obtained coefficients are then used to calibrate the forecast at the respective lead time in the selected training ~~domain~~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.

2.2 ~~Verification metrics~~

2.1.1 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 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 skill of an ensemble forecast is often defined in terms of the root mean square error (RMSE) of the ensemble mean, calculated as

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

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 the square root of the average ensemble variance (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 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):

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

While the ~~spread/skill~~ spread-skill relationship and ME are deterministic scores and applied to the ensemble forecast mean, the Continuous Ranked Probability Score (CRPS, Eq. (5)) allows a probabilistic assessment (Hersbach, 2000). The CRPS

compares the whole distribution of ensemble members, represented as cumulative distribution function, with the observation:

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

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:

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

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 ~~applying the introduced calibration method to~~ post-processing FWI forecasts of the years 2021 to ~~2023.~~ 2023 by applying the introduced calibration method. We use here ~~climatic reference regions , defined by the 6th IPCC Assessment Report (AR6 (Iturbide et al., 2020)), to divide the European domain into Northern Europe (NEU), West and Central Europe (WCE) and the Mediterranean (MED).However, here only the European part, north of the Mediterranean sea, was used and called European Mediterranean (EUMED)hereafter.Other regions can be selected as well, e.g. the calibration can also be done country-wise or at even smaller level.~~the AR6-WGI reference regions introduced in Sec.2.1.1 and shown in Fig.1(a).

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 is considerable high in all regions.

Figure 3 shows the spread-skill relationship for the raw (black) and calibrated (orange) FWI forecast averaged over the grid-points ~~of~~ 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, Fig. ~~S1~~ S2 - ~~S3. The climatology is~~ S4. 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.2.1) for each day of the year instead of using the forecast. Both calibrated as well as raw forecast 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)

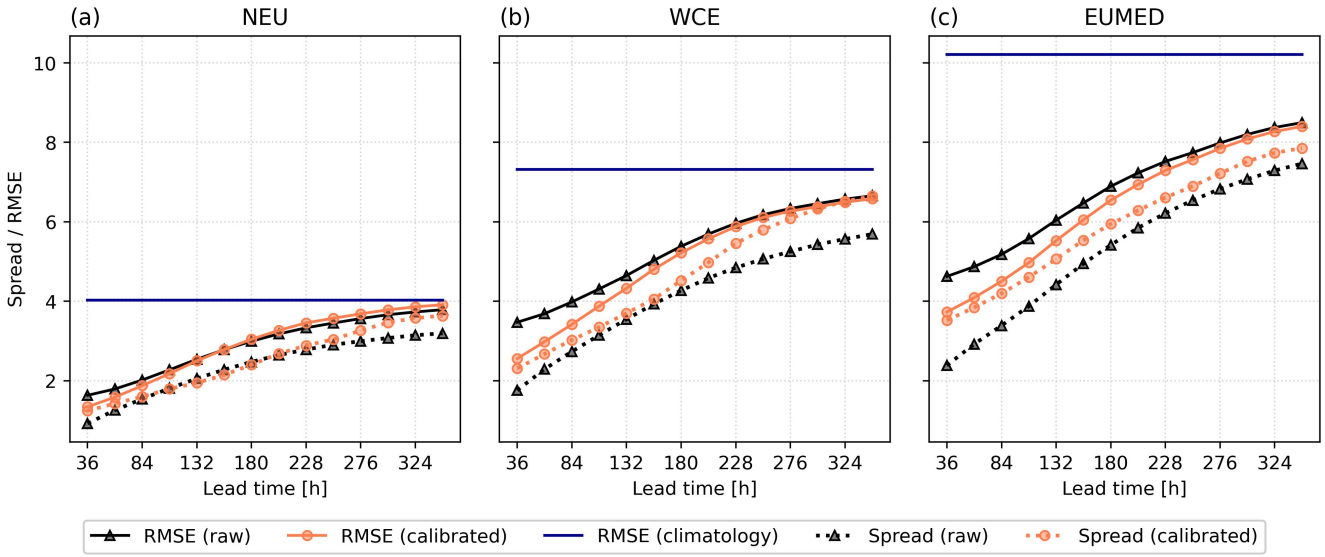

Figure 3. Spread and ~~skill (~~RMSE ~~)~~ for raw and calibrated FWI forecasts averaged over the region of interest~~(left~~: (a) Northern Europe (NEU), ~~middle:~~ (b) West & Central Europe (WCE), ~~right:~~ (c) European Mediterranean (EUMED) and fire season (May-October). The RMSE of the climatology for the respective region is additionally provided.

is constantly smaller than the skill of the ensemble mean (as measured by 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 decreases the RMSE especially during the first forecast days~~slightly~~, which means the accuracy of the forecast is improved~~especially for short lead times~~. 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 ~~8 and 10 days~~ 9 days (228 hours) and 12 days (300 hours), respectively. The ~~regional differences~~ 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 ~~with the generally~~ 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.

The mean error (ME) averaged over the ~~respective area~~ three subregions and the fire weather season is shown in Fig. 4. Uncalibrated forecasts have a negative bias for all lead times, ~~which means~~ indicating that the forecasted FWI ~~is too low compared to observations~~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

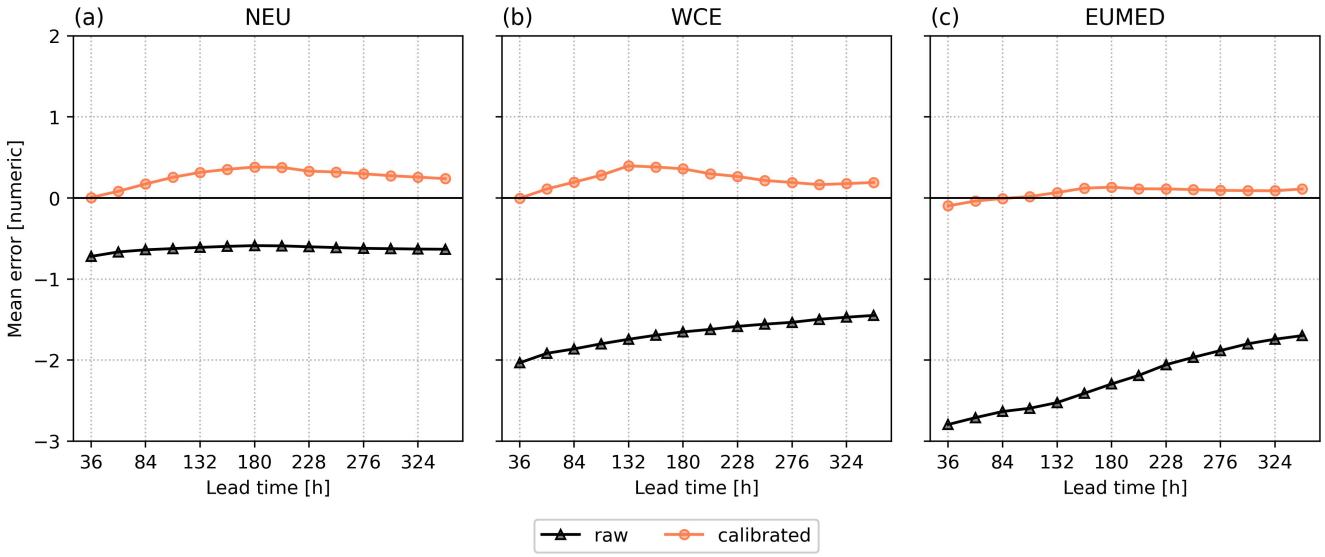

Figure 4. Mean Error for the raw ~~(black)~~ and calibrated ~~(orange)~~ FWI forecasts averaged over the ~~grid of the respective region and~~ fire season (May-October) ~~. Left~~and the respective region: (a) Northern Europe (NEU), ~~Center:~~ (b) Western & Central Europe (WCE), ~~Right:~~ (c) European Mediterranean (EUMED).

with lead time. ~~This~~ The improvement of the mean error is especially contributed to forecasts in the months with high FWI, July and August for WCE and June to September in EUMED, which can be seen in the monthly averaged mean error in Supplementary Figures ~~S4 to S6~~S5 to S7. 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 5 shows the CRPSS with raw forecasts as reference. In all three regions the CRPSS is positive for the first 3 days (84 hours) of the forecast which suggests an improvement of the FWI forecast after calibration for short lead times up to 84 hours. 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 (132 hours) of forecast, while in WCE and EUMED the calibrated forecast has no skill compared to the raw forecast after ~~8 and 10 days~~ 7 (180 hours) and 11 days (276 hours), respectively. This is probably caused by the small FWI values in NEU and for longer lead times. 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. 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.

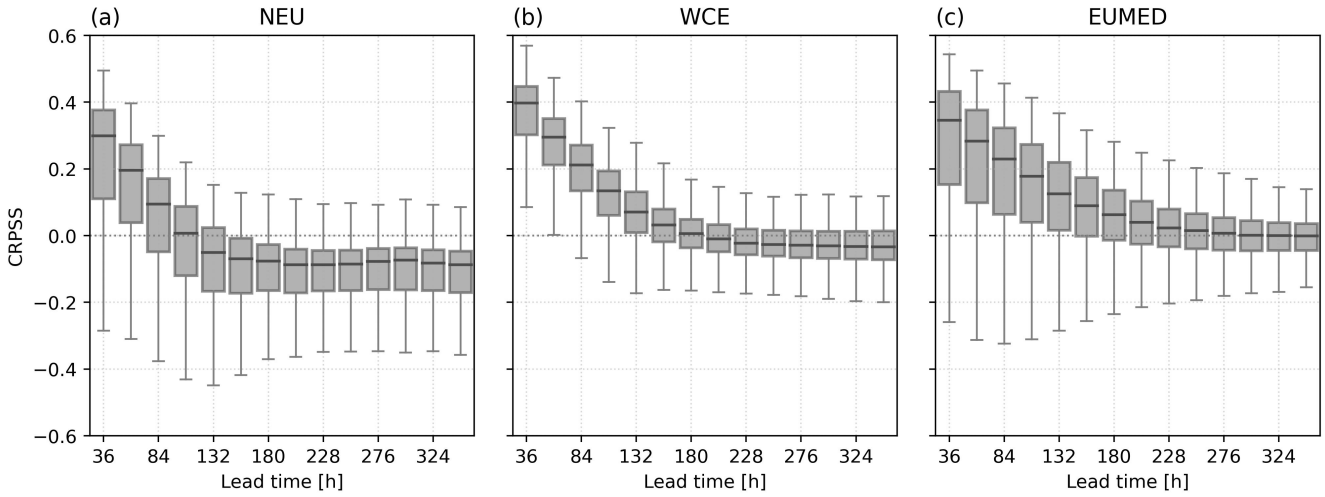

Figure 5. Continuous ranked probability skill score (CRPSS) for the calibrated forecasts against raw forecasts averaged over the fire season (May-October) and the ~~grid of the~~ respective region~~, left~~: (a) Northern Europe (NEU), ~~Center:~~ (b) Western & Central Europe (WCE), ~~Right:~~ (c) European Mediterranean (EUMED).

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 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 cut-off at zero and use forecast and observation pairs of the last 30 days preceding the forecast. ~~Because direct FWI observations are not possible, the most accurate estimation of the true value would be obtained using observed weather parameters to calculate the FWI. However, observations of all necessary weather variables over a longer time period are only sparsely available. Thus, we used~~ We used ECMWF high-resolution weather forecasts with ~~short~~ the shortest possible lead time to calculate the FWI analysis and used these as substitute for observations. Although the FWI analysis seems to underestimate the FWI slightly, a good correlation is observed.

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. 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 supported by the monthly averaged metrics in the appendix, 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, 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, which could be observed in the NEU reference region and especially

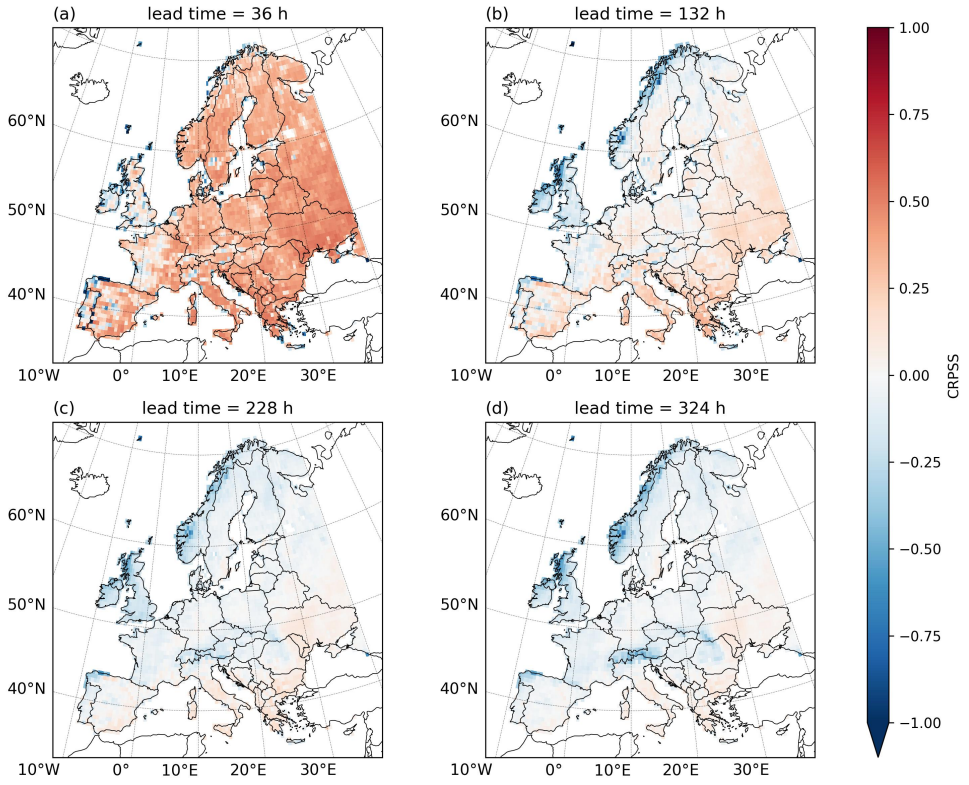

Figure 6. Continuous ranked probability skill score (CRPSS) with the raw forecast as reference averaged over the fire season (May-October) for different lead times.

for longer lead times. Although it would be ideal for the method to perform effectively across the entire range of FWI values, it is generally more critical that it demonstrates a good performance at higher FWI values, where the potential fire danger is more significant. While long-range forecasts of potential fire danger are valuable, short-range forecasts, especially for the first 1-3 days, are usually more critical for firefighting resource management. Reliable and accurate 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 ~~calibration of~~ 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.

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. 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 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.

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.

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, 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, 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 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., Gorodetskaya, 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 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.

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.

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 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, 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., 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/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 Weather Review, 140, 889 – 897, https://doi.org/10.1175/MWR-D-11-00075.1, 2012.

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

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.

Wilks, D. S. and Vannitsem, S.: Chapter 1 - Uncertain Forecasts From Deterministic Dynamics, pp. 1–13, Elsevier, 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.