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

Stephanie Bohlmann¹ and Marko Laine¹

¹Finnish Meteorological Institute, PO 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 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 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 and in regions with elevated FWI values i.e. areas with a higher wildfire risk.

1 Introduction

Wildfires in Europe have become increasingly prevalent and destructive in the last decades. The recent wildfire in Greece 2023 alone burnt, according to the European Forest Fire Information System (EFFIS), an area of over 170 thousand hectares, killed at least 18 people and forced thousands to leave their home (Faiola and Labropoulou, 2023). Also, 2017 and 2022 were years with devastating wildfires, burning large areas in Portugal, Spain, Greece and Italy. But not just the Mediterranean region is affected by wildfires, also Middle and Northern Europe is experiencing more and more often unusually dry and warm summers. Those extended warm and dry periods, raise the fire danger and cause wildfires in regions that were previously not considered wildfire hotspots. One example is the heatwave 2018, which caused wildfires and over 20 thousand hectares burned land in Sweden (San-Miguel-Ayanz et al., 2019).

The rising frequency and seriousness 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 and reliable weather forecasts ranging from a couple of days to multiple weeks to identify high wildfire risk areas is an important part of SAFERS. Here, we use the Canadian forest fire
weather index, short FWI, which is a widely recognized numeric indicator for forest fire risk (Wagner, 1987). The calculation of the FWI only requires four weather parameters and can be calculated using deterministic or probabilistic weather forecasts. Probabilistic ensemble forecasts may require statistical post-processing 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 non-homogeneous Gaussian regression (NGR) (Gneiting et al., 2005), which is commonly used for the calibration of various weather variables like temperature (Hagedorn et al., 2008), precipitation (Hamill et al., 2008) or wind-speed (Thorarinsdottir and Johnson, 2012). In this article we show that NGR can also be used for the calibration of medium-range fire weather index forecasts. The skill of the calibrated FWI forecast is shown and compared for three European regions.

2 Fire weather index calculation

A common method to indicate the danger for wildfires is the Canadian Forest Fire Weather Index (FWI) system (Wagner, 1987). Although originally developed for Canadian weather and vegetation, it is used in many other regions, e.g., by the European Forest Fire Information System (EFFIS) to provide information on wildfires in the EU and neighboring counties (Giuseppe et al., 2020). One advantage of using FWI is the relatively simple calculation only requiring four weather parameters in addition to information of the season (time of year) and geographical location.

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 (FFMC), duff moisture code (DMC) and drought moisture (DC). To consider the different effective day length, and therefore the amount of drying that can occur during a given day, monthly day length adjustment factors for DMC and DC are used with regard to latitude (Lawson and Armitage, 2008). In the second step, FFMC, DMC and DC are used to model the rate of fire spread (ISI) and the potential fuel available for surface fuel consumption (BUI). 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 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 what is considered a low or extreme FWI in one region may not be the same in another. A more comprehensive description of the FWI system can be found in Wagner (1987) and Lawson and Armitage (2008).

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

For FWI calculations we use ensemble forecasts of ECMWF’s operational ensemble forecast system (ENS). ECMWF medium-range ensemble forecasts consists of 51-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 TIGGE archive (Bougeault et al., 2010). TIGGE provides operational medium-range ensemble forecast data for non-commercial research purposes from 13 global NWP centres. The data is accessible through ECMWF API. The resolution of the used TIGGE data is 6h for all lead times and the spatial resolution is 0.5°. Although available forecasts cover the whole globe, we focus here on the European region from 25°N to 72°N and 25°W to 39.80°E. For this study, we use forecasts of 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 ECMWF high-resolution forecasts with the shortest lead time to the local noon with corresponding 24h precipitation forecast as substitute for FWI observations. ECMWF high-resolution forecasts have a spatial resolution of 0.1° and a temporal resolution of 1 hour and can therefore give a more accurate picture of the weather conditions than medium-range ensemble forecasts with a coarser resolution. To determine if those short term FWI forecasts, hereafter called analysis, are suitable to be used as observation substitute, we check their agreement with actual observation based values, which is shown in Fig. 1. 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. 1 shows the stations, for which it is possible to calculate the FWI for more than 200 consecutive days. In total 682 stations can be used. Stations from outside of Finland are not necessarily quality controlled and therefore shown separately. Figure 1b shows the scatter plot of analysis and observations for all stations and every time step. While the FWI derived from the forecasted weather parameters seems to generally underestimate the FWI values compared to the values derived from the observations (slope ~ 0.63), a correlation is apparent. This good correlation can also be seen in 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) as observation to compare with longer forecasts.

4 Calibration and verification methods

4.1 Non-homogeneous Gaussian regression

For statistical post-processing the 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 we assume that FWI

1https://apps.ecmwf.int/datasets/data/tigge/, last access: 06/02/2024
Figure 1. Top left: Map of study area with modified AR5 regions shaded in grey. 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 the locations shown in the map. Data from Finnish stations is quality controlled and shown separately. The solid line illustrates a perfect correlation. Bottom: Examples of FWI time series calculated from observations (orange) and forecast analysis (dashed).

Observations $y$ follow a truncated normal distribution with cut-off at zero:

$$y \sim N^0(\mu, \sigma^2).$$  \hfill (1)

The location and scale parameter are given by:

$$\mu_{kl} = a_l + b_l \text{ENS}_{kl},$$

$$\log(\sigma) = c_l + d_l \log(\text{sd}_{kl}),$$  \hfill (2)

with $\text{ENS}_{kl}$ being the ensemble mean and $\text{sd}_{kl}$ being the standard deviation of the 51 ensemble members for each location $k$ and lead time $l$. $a_l$ and $d_l$ are regression coefficients. The logarithmic link $\log(\text{sd})$ is used to assure positive values for the scale parameter $\sigma$. 

https://doi.org/10.5194/nhess-2024-57
Preprint. Discussion started: 31 May 2024
© Author(s) 2024. CC BY 4.0 License.
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. 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 periods of different length, 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 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 each lead time separately, excluding the first time step of the forecast (T+12h) as these forecasts serve as observation. The obtained coefficients are then used to calibrate the forecast at the respective lead time in the selected training domain. Fitting of the regression model and prediction of location and scale parameters of the predicted 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.

4.2 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 which are shortly introduced hereafter.

A common method to evaluate the forecast reliability of probabilistic forecasts is the comparison of ensemble spread and root mean square error (RMSE) of the ensemble mean, calculated as

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

(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 and average observation, which is defined as the mean error (ME):

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

(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)) allows a probabilistic assessment (Hersbach, 2000). The CRPS compares the whole
distribution of ensemble members, represented as cumulative distribution function, with the observation:

\[
\text{CRPS}(P,x_a) = \int_{-\infty}^{\infty} \left[ F_{F_c}(x) - F_O(x) \right]^2 dx,
\]

where \( F_{F_c}(x) \) 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:

\[
\text{CRPSS} = 1 - \frac{\text{CRPS}_{cal}}{\text{CRPS}_{ref}}.
\]

where \( \text{CRPS}_{cal} \) and \( \text{CRPS}_{ref} \) 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.

5 Results

In this section, we present results of applying the introduced calibration method to FWI forecasts of the years 2021 to 2023. 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 main fire season in Europe is typically from May until October but varies strongly in length and intensity, e.g. the fire season starts later and is shorter in Northern 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 considerably high in all regions.

Figure 2 shows the spread-skill relationship for the raw (black) and calibrated (orange) FWI forecast averaged over the grid-points of 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 - S3. The climatology is shown by the solid blue line. 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) 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 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, respectively. The regional differences could be explained with the generally higher FWI values in the more southern, fire prone regions compared to Northern Europe where FWI values are often very small.

The mean error (ME) averaged over the respective area and the fire weather season is shown in Fig. 3. Uncalibrated forecasts have a negative bias for all lead times, which means the forecasted FWI is too low compared to observations. 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. This 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. 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.

Figure 4 shows the CRPSS with raw forecasts as reference. In all three regions the CRPSS is positive for the first 3 days of the forecast which suggests an improvement of the FWI forecast after calibration for short lead times. 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 of forecast, while in WCE and EUMED the calibrated forecast has no skill compared to the raw forecast after 8 and 10 days, respectively.

These regional differences are furthermore illustrated in the maps shown in Fig. 5, where the CRPSS averaged over the fire
Figure 3. 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: Northern Europe, Center: Western & Central Europe, Right: European Mediterranean.

Figure 4. 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: Northern Europe, Center: Western & Central Europe, Right: European Mediterranean.

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.
6 Conclusions

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 (ECMWF). To estimate the calibration coefficients 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 high-resolution weather forecasts with short lead time to calculate the FWI 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 compared to the analysis. However, calibration improves the forecasts especially at short lead times. 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. 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.

To further improve the calibration of 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.

190 **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.

195 **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


