1. Introduction

Forest fires play a significant role in ecosystem dynamics in many parts of the world. Ongoing climate change may alter existing relationships between abiotic variables and fire patterns, and fire regimes may deviate from their historical patterns since climate and weather are closely linked with forest fires (Müller et al., 2015; Pausas and Keeley, 2021). Alterations caused by climate change are already visible in some parts of the world, such as extensions of the fire season (Jolly et al., 2015), changes in total burned area (Amatulli et al., 2013) and the more frequent occurrence of extreme fire events (Moreira et al., 2020). Moreover, even strongly increasing fire suppression expenditures will not be able to prevent large fires since current policies in many regions cause the ‘firefighting trap’, i.e. that fuel build-up due to strict fire suppression policies leads to larger and more severe fires (Collins et al., 2013; Moreira et al., 2020). Such deviations from historical fire regimes are likely to result in significant changes in the ecological impact of fire, including vegetation shifts and losses of ecosystem services, despite extraordinary fire suppression efforts (Keeley and Pausas, 2019; Pausas et al., 2008; Pausas and Keeley, 2019).

Fire indices were initially developed for assessing the relationship between fire behavior, weather elements and other aspects such as fuel
AGRICULTURAL AND FOREST METEOROLOGY 342 (2023) 109746

I. Bekar et al.

Therefore, different methodological approaches were used in the literature to understand large fires. For instance, logistic regression models (Arpaci et al., 2013; Freeborn et al., 2015). Moreover, fire agencies to manage available resources more efficiently. In particular, weather conditions at the starting day and during the first days of the fire to spread (Stocks et al., 1989). These indices combine multiple variables that characterize fire weather and allow for a better representation of the determinants of fire occurrence than single meteorological variables, since they integrate multiple aspects of fire weather (Bedía et al., 2015). Consequently, they were used for multiple purposes such as understanding the past and future effects of climate change on fire danger (Abatzoglou et al., 2019, Dupire et al., 2017), studying changes in fire regimes (Pezzatti et al., 2013) or the impacts of fire-suppression policies (Ruffault and Mouillot, 2015). Understanding the relationship between the properties of forest fires and weather as a fundamental causing factor is important not only for short-term fire management activities, but also for assessing the long-term ecological consequences. In this context, fire weather-based assessments of forest fire indices are pivotal.

Developing a fire index requires extensive research and testing (Hardy and Hardy, 2007). Over the years, a large number of fire indices have been developed around the world for a variety of purposes and regions. Some of them were re-calibrated or re-scaled to be used in other contexts (San-Miguel-Ayanz et al., 2012; de Jong et al., 2016). For example, the European Forest Fire Information System uses the Canadian Fire Weather Index System to provide a standardized assessment of fire danger in all the countries of Europe. However, many of the indices are used in areas for which they had not originally been intended and calibrated. Despite the understandable reasons of saving time and money, using existing indices outside the area of calibration may cause an under- or overestimation of fire danger and eventually questions the index’s suitability (Arpaci et al., 2013). This is especially crucial for regions where fires are not common, i.e. where overestimation of the fire danger can cause false alarms and reduce the credibility of the system (García et al., 1995). Such problems may arise from local differences in fire incidence, short-term weather variability, fire-vegetation feedbacks or anthropogenic factors that are strongly co-shaping the fire regime in most areas (Tian et al., 2011; Pausas and Paula, 2012; Sturm et al., 2012).

Many studies measuring the performance of fire indices focused on fire occurrence, i.e. the probability of a fire to start in a specific reference area (Wotton, 2009). While it is critical to know the probability of fire occurrence, it is equally important to know the likelihood of those fires to spread and become large. The definition of large fires may change depending on the region and ecosystem of interest. However, in relative terms, fires can be classified as “large” if the area burned is in the upper range of fire sizes (e.g., fires with a burned area corresponding to the upper 95 or 99 percentile, see Gill and Allan (2008) for a review). These fires, although small in number, account for a significant proportion of the total area burned by all fires and thus have a disproportionately large impact on ecosystems and society (Pezzatti et al., 2020; San-Miguel-Ayanz et al., 2013). Understanding the dynamics of large fires is however difficult, as they are relatively rare and irregular (Gill and Allan, 2008). Among others, a better understanding of the meteorological conditions that can lead to large fires would help firefighting organizations to manage available resources more efficiently. In particular, weather conditions at the starting day and during the first response by firefighters have a strong influence on the final burned area (Wotton, 2009). This calls for a daily resolution of weather data and fire danger assessment. However, most of the studies on the relationship between burned area and fire indices have investigated this relationship on weekly or monthly time scales, mostly in the case of climate change projections (Amatulli et al., 2013; Carvalho et al., 2008; Flannigan et al., 2005), and only a few have attempted to study it at a daily resolution (Arpaci et al., 2013; Freeborn et al., 2015). Moreover, fire databases do not always include data on fire spread through time. Therefore, different methodological approaches were used in the literature to understand large fires. For instance, logistic regression models were used to estimate the probability of the occurrence of a large fire on a given day (Bradstock et al., 2016; Preisler et al., 2004). However, such studies require a precise definition of “large”, which obviously is region-specific. In some regions, this may be > 1000 hectares, whereas in others it can be 50 hectares (Linley et al., 2022). This makes it hard to apply and compare the same approach to different regions.

Pezzatti et al. (2020) proposed a novel metric, the Cumulative Area Ranking Efficiency (CARE), to evaluate the ability of a fire index to correctly rank fire events according to burned area. CARE evaluates the total burned area of all fires started on a given day and can be used as a complementary metric to performance measures of fire danger ratings. Furthermore, the metric is easily applicable to most fire databases since it is based on daily total burned area instead of the burned area of a single fire or the sum of the daily newly burned area by ongoing fires. To date, however, CARE has been developed and tested only for a small number of regions and for single fire indices (Pezzatti et al., 2020), but not with models based on multiple variables. De Angelis et al. (2015) trained regional multi-variable models of fire occurrence and demonstrated that models based on raw meteorological variables can in some cases outperform complex fire danger indices, such as the Canadian Fire Weather Index. Finding a way to include the CARE metric in the process for selecting the best multi-variable models (model selection) has the potential of indicating which models best predicting the fire occurrence are also informative regarding the final burned area (Pezzatti et al., 2020).

Various statistical modeling and machine learning approaches, such as Generalized Linear Models (GLM; Adámek et al., 2018), zero-inflated regression models (Bekar and Tavangragh, 2017), or Maxent (Parisien and Moritz, 2009) have been used in the fire ecology literature. Although a considerable impact of methodological choices on the modeling results has been shown in a variety of fields (Elith et al., 2006; Elith and Graham, 2009; Syphard and Franklin, 2009), only few have studied the effect of the modeling method in fire ecology (Massada et al., 2013; De Angelis et al., 2015). For instance, Massada et al. (2013) showed that while the performance of machine learning methods was only marginally better than that of GLMs, the ignition probability maps generated by these models showed notable differences (Massada et al., 2013). In addition, model performance can be influenced by various internal factors, including the complexity of the model and the types of variables considered (Fernandez et al., 2017; Brun et al., 2020).

The aims of this study are (1) to evaluate the suitability of integrating the CARE metric as a model selection criterion for identifying the best combinations of fire indices or meteorological variables in predicting fire danger. The primary purpose of testing CARE across a wide range of environmental conditions was to complement the AUC metric by taking burned area into account when selecting the best model for assessing daily fire danger (Pezzatti et al., 2020). In this perspective, the CARE approach was tested in eleven case study regions along a large environmental gradient from European Alpine and pre-Alpine (selected regions in Austria, Switzerland, and France) to Mediterranean conditions (Spain and Italy). Further, we aim at evaluating the differences in the performance depending on (2) the type of the explanatory variables being used, i.e., the elaborated fire indices vs. raw meteorological variables or simple derivations of them; (3) the complexity of the model, i.e., single-variable models vs. multi-variable models; and (4) the modeling approach, i.e., Generalized Linear Models vs. Maxent.

2. Materials and methods

2.1. Study areas and data sources

We selected eleven regions from five countries across Europe (Fig. 1), i.e. two regions each from Austria, France, Spain, and Switzerland, and three regions from Italy. They represent a wide range of environmental conditions from the more moist and cold weather in the Alps to the warmer and drier Mediterranean regions, with distinctly different fire regimes (see Table 1). We used the official fire event records (i.e., start date and total burned area) from the corresponding national agencies.
already retrieved and used by Bekar et al. (2020) and Pezzatti et al. (2020), and merged them with the additional data hosted in the Promethee database (Forest fires database for Mediterranean area in France), which is freely available at https://www.promethee.com/. In order to allow for a more straightforward comparison among all considered regions, only the fires occurring in the vegetation period (defined here as May to November) were retained for the analysis. Meteorological data were obtained from a meteorological station in each region (Table S1). In Italy, we removed 2002 for Cilento and Chilivani, and 2008 for Chilivani from the analyses, due to a lack of meteorological data.

2.2. Meteorological variables and fire indices

We used nine meteorological variables (three of them related to precipitation) and 14 widely used fire indices of different complexity (Table 2). The selected indices are primarily used as fire danger indicators rather than a direct measure of fire weather conditions. Concerning the fire indices, the Canadian Forest Fire Weather Index system includes six indices (Van Wagner, 1987). Three of these are the Fine Fuel Moisture Code (FFMC), the Duff Moisture Code (DMC), and the Drought Code (DC). They keep track of moisture content in different types of fuels. A combination of these indices with wind speed results in the Initial Spread Index (ISI) and the Buildup Index (BUI). Finally, a combination of ISI and BUI results in the overall Fire Weather Index. The Angström Index and the Fuel Moisture Index (FMI) are simple indices that require only relative air humidity and temperature. Furthermore,
Table 2
Meteorological variables and fire indices used in the study. PET is the potential evapotranspiration [mm/day]. Meteorological variable names with the suffix "12" refer to values measured at noon. Otherwise, they refer to the daily mean value or daily sum in the case of precipitation.

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Name</th>
<th>Variables used in calculation</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Meteorological variables</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>T [°C]</td>
<td>Air temperature</td>
<td>–</td>
</tr>
<tr>
<td>Tdew [°C]</td>
<td>Dewpoint temperature</td>
<td>–</td>
</tr>
<tr>
<td>H [%]</td>
<td>Relative air humidity</td>
<td>–</td>
</tr>
<tr>
<td>VPD [kPa]</td>
<td>Vapor pressure deficit</td>
<td>–</td>
</tr>
<tr>
<td>U [m/s]</td>
<td>Wind speed</td>
<td>–</td>
</tr>
<tr>
<td>P (mm/day)</td>
<td>Precipitation</td>
<td>–</td>
</tr>
<tr>
<td>WeekRain [mm/week]</td>
<td>Total rainfall over the last seven days</td>
<td>–</td>
</tr>
<tr>
<td>DaysSinceRain [day]</td>
<td>The number of days since the last rainfall event</td>
<td>–</td>
</tr>
<tr>
<td>LastRainSum [mm]</td>
<td>The sum of the last rainfall event (consecutive days with rain)</td>
<td>–</td>
</tr>
<tr>
<td><strong>Fire indices</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FFM</td>
<td>Fine Fuel Moisture Code</td>
<td>T12, H12, U12, P12</td>
</tr>
<tr>
<td>DMC</td>
<td>Duff Moisture Code</td>
<td>T12, H12, P12, day length</td>
</tr>
<tr>
<td>DC</td>
<td>Drought Code</td>
<td>T12, P12, day length</td>
</tr>
<tr>
<td>ISI</td>
<td>Initial Spread Index</td>
<td>FFM, U12</td>
</tr>
<tr>
<td>BUI</td>
<td>Buildup Index</td>
<td>DMC, DC</td>
</tr>
<tr>
<td>FWI</td>
<td>Fire Weather Index</td>
<td>BUI, ISI</td>
</tr>
<tr>
<td>Angstrom Index</td>
<td></td>
<td>–</td>
</tr>
<tr>
<td>Baumgartner Index</td>
<td></td>
<td>–</td>
</tr>
<tr>
<td>FMI</td>
<td>Sharples Fuel Moisture Index</td>
<td>T, H</td>
</tr>
<tr>
<td>Sharples</td>
<td>Sharples Forest Fire Danger Rating Index</td>
<td>T, H, U</td>
</tr>
<tr>
<td>FFVI</td>
<td>Fosberg Fire Weather Index</td>
<td>T, H, U</td>
</tr>
<tr>
<td>Nesterov Index</td>
<td></td>
<td>T, Tdew, P</td>
</tr>
<tr>
<td>FFDI</td>
<td>McArthur Forest Fire Danger Index</td>
<td>T, H, U, P</td>
</tr>
<tr>
<td>KBDI</td>
<td>Keetch-Byram Drought Index</td>
<td>T, P, mean annual precipitation</td>
</tr>
</tbody>
</table>

the Sharples Forest Fire Danger Rating Index combines FMI with wind speed (Sharples et al., 2009). The Fosberg Fire Weather Index uses temperature, relative air humidity, and wind speed to provide information on the impact of small-scale/short-term weather variations on fire potential since it is calculated based on hourly data (Goodrick, 2002). The Nesterov Index requires temperature, dew point temperature, and precipitation and is suitable for assessing the ignition potential (Nesterov, 1949). The Keetch-Byram Drought Index requires daily temperature, daily and annual precipitation as input and aims to represent the flammability of the organic material in the ground (Keetch and Byram, 1968). The McArthur Forest Fire Danger Index was originally developed for assessing fire danger in Eucalyptus forests and is mostly used in Eastern Australia. It is based on temperature, relative air humidity, wind speed, and fuel availability (McArthur, 1967; Noble et al., 1980). Lastly, the Baumgartner index was developed based on the assumption that fire danger is primarily driven by fuel dryness. It requires daily precipitation and potential evapotranspiration (Baumgartner et al., 1967).

All fire indices were calculated on a daily basis with the Fire Danger Indices Calculator software (https://github.com/Insubric/fire-calculator) developed by the Swiss Federal Institute for Forest, Snow, and Landscape Research (WSL).

2.3. Modeling approach and performance evaluation

We used two statistical modeling approaches, i.e., Maxent and Generalized Linear Models (GLM) to model the occurrence of fire days (i.e., days in which there is at least one ignition). In most fire danger studies, Generalized Linear Models (GLMs) have been used (Costafreda-Aumedes et al., 2017). However, ecological niche modeling that predicts the presence/absence of species in space represents an alternative. Niche modeling approaches such as Maxent are becoming increasingly popular for forest fires studies (Parisien and Moritz, 2009; Arpaci et al., 2014; Bekar et al., 2020). Maxent is particularly getting attention due to its presence-only nature which allows Maxent to estimate the relationship between presence and background data instead of absences. This fits well with fire occurrence data, where fire statistics may be incomplete (missing registration of fires) or anthropogenic ignitions may not occur in days with very high fuel flammability since people are aware of the danger (Massada et al., 2013). Moreover, Maxent shows high performance even with small sample sizes (Guisan et al., 2007). De Angelis et al. (2015) used Maxent to implement the principles of ecological niche modeling to the temporal scale of fire where fire weather conditions were considered as the environmental space.

Furthermore, we used two groups of variable pools: meteorological variables from which we derived what we hereafter call ‘meteorological models’, and fire indices (‘index models’). We also considered single-variable and multi-variable models separately and compared their performance when based on different variable groups (all models referred to in the paper are multi-variable unless they are specifically denoted as single-variable models). Using multi-variable models consisting of several fire indices instead of a single index allows us to better characterize fire weather conditions (cf. De Angelis et al., 2015). The correlation among the predictor variables was checked and combinations containing highly correlated ones (i.e., Pearson’s $r > 0.9$) were removed during the model building procedure. For each region, we used k-fold cross-validation to obtain more robust results from the models. For this purpose, we selected a combination of two consecutive years for the test and the remaining folds as training. We repeated this for all combinations of two years. Finally, we calculated the performances on each test fold and averaged them.

The Area Under the ROC (Receiver Operating Characteristic) Curve, also known as AUC, is commonly used to evaluate the performance of fire occurrence models (Fielding and Bell, 1997; Parisien and Moritz, 2009). It refers to a model’s accuracy in discriminating between days with and without fire (i.e., fire occurrence). We visually combined AUC with the newly developed metric CARE (Cumulative Area Ranking Efficiency) in order to get an assessment of the model’s ability to provide information on both fire occurrence and burned area (Pezzatti et al., 2020) and to explore the patterns between the two metrics. CARE relies on the cumulative curve of the total burned area of the fires started on a particular day by ranking the fire days according to decreasing predicted fire occurrence probability. Here, the higher the area under the curve is, the better the performance since this means that large fire days are associated with days with high fire danger levels (for the full details of the methodology, see Pezzatti et al., 2020). Therefore, CARE can be used as a complementary performance metric to AUC, as it evaluates the ability of a specific fire model to assess the risk of large fire days. It ranges from 0 to 1, where 0.5 represents a random prediction. We used the logarithmically transformed version of CARE, i.e., CLARE (Cumulative Logarithmic Area Ranking Efficiency) to handle the generally strongly skewed distribution of burned area (see Pezzatti et al. 2020 for details).

All analyses were run using the R statistical software version 3.6.1 (R Core Team, 2021). We used the ‘dismo’ package with its default parameter settings (Hijmans et al., 2021).

3. Results

3.1. Performance of AUC vs CLARE

The proposed approach highlighted different performance patterns according to CLARE vs. AUC plots, of which the extreme cases are shown in Fig. 2. In Campidano, CARE performances were aligned and in
Fig. 2. CLARE vs. AUC performances of all index and meteorological models for Western Hautes Alpes and Campidano. Numbered boxes highlight the highest performance in AUC (box 1), CLARE (box 3), and both metrics (box 2).

Fig. 3. CLARE vs. AUC performances of the Maxent models for all regions. To enhance the readability, in the plots multi-variable model performances are represented by lines encompassing the entire range of models (models with 2 or more explanatory variables). The symbols represent all single variable models.
I. Bekar et al.

In accordance with those based on AUC, while in Western Hautes-Alpes, the performances were scattered, as highlighted by the much broader point clouds (Fig. 2). In both cases, the number of models that achieved an AUC above 0.8 was very high; however, in Western Hautes-Alpes, models that showed high AUC performance featured a wide range of CLARE values, from 0.3 to 0.8, and many of them were characterized by CLARE values <0.5 (Fig. 2). Here, models with the highest AUC performance (Fig. 2, box 1) had a very low CLARE performance, whereas models with the highest CLARE performance (Fig. 2, box 3) had much lower AUC performance than the highest AUC. However, some models (Fig. 2 box 2) had a good performance in both metrics. Thus, using the additional CLARE metric allows for a fine-tuning in the selection of the best models, such as in Campidano, or it may be given a greater importance according to a user-defined weighting that considers a trade-off between AUC and CLARE.

Fig. 3 summarizes the patterns found in all regions, differentiating between single and multi-variable models and between index and meteorological models. The observed patterns can be more similar to the monotonic relationship found in Campidano, e.g. in the other Italian regions; more scattered like in the Eastern Hautes Alpes, Campo Arcis, Carinthia, and the Swiss regions; or an intermediate situation, like in Tyrol and Liliria.

Absolute values of model performance varied strongly for both CLARE and AUC depending on the region. AUC values were generally above the random guess value of 0.5, which is understandable given that the models were built on fire occurrence. Instead, five regions had a considerable number of models with CLARE values below 0.5. Maximum AUC values among the models were found in Campidano, Tyrol and Western Hautes-Alpes, while the highest values in CLARE were in the Western Hautes Alpes, Tyrol, Valais and Liliria.

3.2. Performance of index vs. meteorological models

The role of the different explanatory variables on model performance varied strongly for both CLARE and AUC, depending on the region (Fig. 3). The index models showed a slightly better AUC performance in most regions than the meteorological models (Fig. 3). In some regions, such as Campo Arcis, the best performing meteorological models (represented by the 100 models with the highest performance) achieved a mean AUC of 0.66, showing a performance close to that of the best performing index models, which had a mean AUC of 0.68. However, the same index models showed a higher mean CLARE performance (0.70) than the meteorological models (0.61) even though their AUC performance was similar to that of the meteorological models. The CLARE performance of the index models was also slightly higher than that of the meteorological models in most regions. The strongest exception was found in Valais, where the meteorological models outperformed the index models in terms of CLARE performance, achieving as high as 0.75, compared to the index models’ maximum performance of 0.69. However, it should be noted that the meteorological models that had the highest CLARE values in Valais did not feature an acceptable AUC performance. The magnitude of the difference between the index and meteorological models varied strongly depending on the region. For instance, while the differences in Western Hautes Alpes were very clear, they were hardly noticeable in Carinthia.

3.3. Performance of single-variable vs. multi-variable models

The multi-variable index models typically showed an AUC performance that was at least similar to or higher than that of all single-variable models (cf. dots in Fig. 3). A similar pattern was found for CLARE, with a few exceptions. In Ticino and Eastern Hautes Alpes, the single-variable index models showed a slightly higher CLARE performance than the multi-variable models. However, these single-variable index models were not among those with the highest AUC performance in the respective region. In the Western Hautes Alpes, a single-variable meteorological model with a CLARE of 0.65 outperformed all multi-variable meteorological models, which were all below 0.58. However, these single-variable models were not among those with the highest AUC performance in the respective region. It is also important to highlight that multi-variable meteorological models showed a performance at least similar to or higher than single-variable index models for both AUC and CLARE in most instances. For example, in Valais, the best multi-variable meteorological models achieved a CLARE as high as 0.75, outperforming all single-variable index models (Fig. 3). Also, the performance of the best meteorological models was nearly identical or even higher to that of the index models in some regions, such as Carinthia, Ticino, Valais and Campo Arcis.

3.4. Performance of the GLMs vs. Maxent modeling approach

No single modeling approach consistently outperformed the others across regions or variable group (Figs. 4 and 5). For instance, CLARE and AUC for both modeling approaches was nearly identical in Carinthia and Chilivani (Figs. 4 and 5). However, for fire indices, GLMs had higher CLARE and AUC than the respective Maxent models in Valais (Fig. 5), while the index Maxent models had clearly higher performance than the index GLMs in Campidano (Fig. 4). The impact of the modeling method also changed depending on the input variable groups. Maxent performed at least similar to or had a higher performance than the GLMs for the index models (Fig. 4). However, contrasting patterns were observed where the GLMs had at least similar to or higher performance than Maxent models for the meteorological models (Fig. 5). In most cases, the variation in CLARE performance between the modeling methods was higher than in AUC performance (Figs. 4 and 5). Lastly, the GLM performances were generally distributed over a wider range of values compared to the Maxent model performance.

5. Discussion

5.1. Suitability of CLARE in a model selection framework

In this study, we successfully implemented CLARE in the model selection process for widely different environmental conditions and fire regimes. Using CLARE makes it possible to select models with a better prediction capability for burned area among the ones displaying the best AUC performance. However, it should also be highlighted that the primary selection criterion must remain the AUC performance. Therefore, one should not overlook the AUC performance of the models and choose models that perform best in CLARE only, but rather identify and select models that have a sound trade-off between predicting fire occurrence (AUC) and burned area (CLARE). Our results further highlight that not all models with high AUC performance also perform well in terms of predicted burned area (Pezzatti et al., 2020). This discrepancy may be understandable as the short- and long-term drivers of fire ignition and burned area may differ within a region (Bedia et al., 2014). In such cases, models that can accurately predict fire ignition may fail to identify the risk of fires to get large (Tanskanen and Venäläinen, 2008). As a result, identifying models with good AUC performance that perform well also in terms of burned area should be done carefully for each region. Overall, including CLARE as a complementary metric in addition to AUC for evaluating model performance allows for a more thorough investigation of the role of meteorological factors, and ultimately results in a more robust model selection process.

5.2. Importance of the nature of the predictor variables

Fire indices have been specifically developed to better represent environmental and fire weather conditions than meteorological variables. Therefore, we expected that index models perform better than meteorological models. Indeed, the index models showed a higher CLARE or AUC performance than the meteorological models in most
regions (De Angelis et al. 2015; Fig. 3). However, it should be noted that the meteorological models showed nearly identical performance as the index models in regions such as Carinthia and Ticino, which are low-to-medium fire activity regions in comparison to the high fire activity regions such as Spain and Italy, where index models performed better although to varying degrees. Moreover, meteorological models clearly outperformed many of the single-variable index models in most regions. Similar findings that show a high performance of the raw meteorological variables compared to fire indices have been previously reported by several studies including one from Austria (Padilla and Vega-García, 2011; Arpaci et al., 2013). Such high performance by the meteorological models could be promising for regions where the calculation of fire indices is not possible. However, it has to be considered, that the findings also reflect the fact that the tested fire indices have not been adjusted to the Alpine / European conditions. So, fuel moisture might be biased by the different vegetation types and fuel beds (Arpaci et al., 2013).

5.3. Niche modeling approach: multi-variable vs. single variable models

Model complexity is an important factor that strongly affects performance (Brun et al., 2020). The number of variables considered is an important part of the model complexity where too few variables may cause a model to fail to capture the relationship between response and environmental factors. At the same time, too many variables can create unnecessary noise in models and may lead to overfitting (Moreno-Amat et al., 2015; Brun et al., 2020; Low et al., 2021). In our study, the largest difference in overall performance between multi-variable and single-variable models was found in the Western Hautes Alpes and Tyrol, which had the lowest number of fire days and the lowest amount of burned area, respectively. In contrast, the differences were small in regions such as Campidano and Chilivani, where the number of fire days was high, which shows that number of variables in the model has a smaller effect on model performance when the sample size is high. It is known that sample size has a strong effect on model performance from multiple aspects (Wisz et al., 2008). Models may not be able to describe the ecological niche of the modeled organism to its full extent and can only describe relatively large patterns when the sample size is low (Foody, 2011). In our case, multi-variable models increased the model performance in regions with small sample sizes, whereas the effect was low in high fire activity regions. This is understandable since model performance shows more variability at small sample sizes and strong response to number of variables (Wisz et al., 2008; Brun et al., 2020). The high performance of some of the single-variable index models can be explained by the relationship between the fire indices and fire weather since those indices take multiple meteorological variables into account simultaneously with the specific aim of representing fire weather conditions. Moreover, it is important to emphasize that the use of multi-variable models that take into account multiple fire indices allows for leveraging their best features by weighting them in a model that performs an indirect regional calibration by taking local fire history into account (De Angelis et al., 2015). Overall, our findings provide further evidence to De Angelis et al.’s (2015) conclusion that the multi-variable modeling approach allows for a better characterization of fire danger conditions than single-variable models. This approach is not only beneficial in most of the studied regions, but it is particularly effective in areas with low to moderate fire activity and/or with a strong anthropogenic component in the fire regime, which make the selection and the
calibration of a suite of fire indices is usually quite problematic (De Angelis et al., 2015). A multi-variable approach based on weather data alone may be particularly valuable in regions where the calculation of fire indices is not feasible or where the application of fire indices poses challenges (e.g., calibration issues).

5.4. The importance of the modeling approach (GLM vs. Maxent)

The role of the modeling approach in forest fire studies is becoming an increasingly important question since an increasing number of approaches have been used over the past decade (Oliveira et al., 2012; Bekar and Tavşanoğlu, 2017; Massada et al., 2013). Our multi-regional set-up with a wide range of environmental conditions and fire regime characteristics provides evidence that the role of the modeling approach depends on factors such as the region that is being modeled or the predictor variables that are used. For instance, while Maxent performed slightly better for index models (Fig. 4) in high fire activity regions (Spain and Italy), GLMs performed better for meteorological models (Fig. 5) in low to intermediate fire activity regions (France and Switzerland). The underlying mechanisms behind the modeling approaches may explain this, since the more complex the model is in terms of response shapes and interactions, the more data are needed for a reliable model (Barry and Elith, 2006). Maxent has the ability to model more complex relationships between the response and environmental variables than the relatively simple GLMs (Low et al., 2021). Massada et al. (2013) showed that Maxent featured a slightly higher prediction accuracy than GLMs when modeling the distribution of fire ignition. However, their study was based on a single area and did not investigate the effects of variable groups on model performance. Based on the results from our study and the literature, we prefer to avoid making generalizations since evidence strongly suggests that the model performance varies with several factors such as environmental variables, the traits of the modeled ‘organism’ (i.e., specific fire characteristics in our case), or the environmental conditions of the region that is being modeled (Elith et al., 2006; Guisan et al., 2007).

Detailed analyses of the performance and predictive capability of species distribution models (SDM) suggest that it is crucial to go beyond a single performance measure such as AUC, since such a metric does not reveal a comprehensive picture (Syphard and Franklin, 2009). For instance, different modeling approaches with similar predictive performance may produce different prediction maps (Massada et al., 2013). Using multi-modeling approaches on the same dataset – like in this study – might therefore be a promising way to overcome such shortcomings (Arpaci et al., 2014). Our finding that models with similar AUC performance may not show a similar performance when predicting burned area was valid for both modeling approaches. However, even when the maximum predictive performance of Maxent and GLM was similar, the general performance of the models had a larger spread in the GLMs than when using Maxent. This is in agreement with the findings from the literature that suggest that Maxent has a robust performance across a wide range of sample sizes (Wisz et al., 2008; Moudrý and Simová, 2012). Overall, the use of CLARE as a complementary performance metric offered a more comprehensive assessment of models with similar AUCs, both within and between modeling approaches. This highlights that particular attention should be paid to selecting the modeling approach and the evaluation criteria. Moreover, we suggest that this topic needs further research since various models have been used and compared to each other in the fire literature, albeit not always in a

Fig. 5. Comparison of the meteorological GLM and Maxent models for all regions. To enhance the readability, in the plots multi-variable model performances are represented by lines encompassing the entire range of models (models with 2 or more explanatory variables). Larger bubbles indicate that the performance of the multi-variable models is distributed over a larger area, while smaller bubbles indicate a narrower distribution.
systematic fashion.

6. Conclusion

Understanding the probability of the occurrence of large fires has critical importance since they are responsible for large-scale impacts on ecosystems as well as human livelihoods (Liang et al., 2008). Therefore, it is critically important to be able to estimate both the probability of fire occurrence along with its likelihood to develop into a large event. Our results provide evidence that using CLARE in the model selection process does provide better assessments when evaluating models by showing that not every model that is successful in predicting fire danger is able to identify days when large fires occur (De Angelis et al., 2015). We furthermore showed that multi-variable models yield better results than single-variable models, especially in regions with low fire activity. Moreover, multi-variable meteorological models had a similar performance as single-variable index models, which may allow for fire danger assessments in regions where the calculation of specific fire indices is not possible. Finally, the differences between the modeling approaches (GLMs vs. Maxent) were mainly related to the region being modeled or the input variable groups used for model building. Overall, we recommend using CLARE in model assessments as a complementary metric to AUC since the use of models that perform well in predicting both fire occurrence (AUC) and burned area (CLARE) has substantial theoretical and practical benefits. Our results also highlight the importance of testing different approaches in fire modeling studies since the selected algorithms may influence the predictive performance as well as the model outcomes.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The authors do not have permission to share data.

Acknowledgments

This project has been funded partly by a Swiss Government Excellence Scholarship to IB.

Supplementary materials


References
