Annals of Forest Science (2021) 78: 93
https://doi.org/10.1007/s13595-021-01109-5

RESEARCH PAPER

foreMast: an R package for predicting beech (*Fagus sylvatica* L.) masting events in European countries

Ugo Chiavetta[1] · **Sebastian Marzini**[1]

Received: 13 July 2021 / Accepted: 11 October 2021 / Published online: 27 October 2021

Abstract

Key message **Forecasting annual seed production will improve the management of forests across Europe. The foreMast R package we developed predicts current year masting probability in beech (*Fagus sylvatica* L.) using climate data easily accessible by any stakeholder.**

Context Modelling and predicting forest masting is one of the most challenging tasks in forest management, as it is a strategy shared by several species, very important for tree dispersion and forest regeneration, mainly related to climate and ecological processes.

Aims As many studies focus on European beech (*Fagus sylvatica* L.) masting without simple practical implementations, we developed a tool capable of predicting beech masting years.

Methods The tool is an R package (*foreMast*) made by three functions, which relies mainly on climate data. The algorithm performance is compared with the records of the MASTREE database, which gather several beech seed production series for various sites across European countries.

Results Overall, the results show a tight correlation with the compared sites ($\rho = 0.50$ to 0.61, p-value < 0.0001, respectively), especially when temperatures weigh three times more than precipitation. Nevertheless, in some sites, seed production seems to be more related to precipitation dynamics than to temperatures.

Conclusion foreMast can be used both for studying changes in mast events in relation to climate changes and in operative forest management and planning. It is flexible and thus amenable to future implementation of additional predicting variables or target species.

Keywords Climatic cues · *Fagus sylvatica* L. (European beech) · Forest management · Mast event · R package · Seed production

1 Introduction

Mast seeding is the highly variable and synchronous production of exceptionally large seed crops among years of many wind-pollinated plant species (Vacchiano et al. 2017b; Lucas-Borja and Vacchiano 2018; Nussbaumer et al. 2020). As an adaptive strategy for reproduction and seed dispersal (Drobyshev et al. 2010; Ascoli et al. 2017), it has serious direct and indirect implications on various ecosystem dynamics (Drobyshev et al. 2010, 2014; Cutini et al. 2013; Kelly et al. 2013; Pearse et al. 2016; Bajocco et al. 2021). Masting is a result of pressure selection, explained by the predator satiation (Janzen 1971; Kelly 1994) and the pollination efficiency (Kelly et al. 2001) hypotheses, concurrently with environmental constraints, such as microclimate and resource allocation (Kelly et al. 2001; Piovesan and Adams 2001; Hilton 2003; Pearse et al. 2016; Lucas-Borja and Vacchiano 2018). While there is strong evidence about the dependence of mast years to weather cues, difficulties arise when trying to analyse resource dynamics, which may be triggered and regulated by microclimate in the years prior the mast (Monks and Kelly 2006; Drobyshev et al. 2010, 2014; Pearse et al. 2016; Vacchiano et al. 2017b). Across

---

Handling Editor: Bruno Fady

✉ Sebastian Marzini
sebastian.marzini@crea.gov.it

Ugo Chiavetta
ugo.chiavetta@crea.gov.it

[1] CREA Research Centre for Forestry and Wood, Viale Santa Margherita 80, 52100 Arezzo, Italy

Europe, one of the most representative species with a significant importance under a forestry perspective is certainly beech (*Fagus sylvatica* L.), which is widely spread thanks to its strong competitiveness and its wide climatic and edaphic ranges. Many studies described its ecology and masting dynamics (Piovesan and Adams 2001; Hilton 2003; Nocentini 2009; Wagner et al. 2010; Vacchiano et al. 2017b; Bogdziewicz et al. 2020). Beech masting events are highly related to weather cues. In general, a large seed production occurs after a year with high summer temperatures and low precipitations, preceded by a year with low temperatures and high precipitations (Packham et al. 2008; Vacchiano et al. 2017b; Bogdziewicz et al. 2020). Different studies have highlighted a pattern of lagged-autocorrelated masting events with a returning interval varying between 2 and 10 years (Piovesan and Adams 2001; Hilton 2003; Drobyshev et al. 2010; Kasprzyk et al. 2014), although an increase in the frequency of masting events has been observed, in relation to climate change (Drobyshev et al. 2014; Kasprzyk et al. 2014; Nussbaumer et al. 2016). Overall, there is a negative correlation with temperatures in the summer 2 years prior to the mast event and a positive one with those of the previous year (Pearse et al. 2016; Vacchiano et al. 2017b), suggesting that temperature is a more consistent triggering factor than precipitation, although some studies found drought to play a key role (Piovesan and Adams 2001). Extreme events in the mast year, such as late frost in spring, wet conditions during the pollination period or very dry and hot summer, negatively affect masting, even preventing the production of seed (Hilton 2003; Packham and Hytteborn 2012; Kasprzyk et al. 2014; Nussbaumer et al. 2020).

Mast years are closely related to regeneration dynamics; thus, they are important in the management of forest areas and related activities (Hilton 2003; Cutini et al. 2010; Wagner et al. 2010; Vacchiano et al. 2018). In Europe, European beech is one of the most important forest species (Nocentini 2009; Leuschner and Ellenberg 2017), distributed in mountain areas at the southern limits of its spatial range (Ottaviani et al. 2019) and at lower elevations northwards where the climate shifts from oceanic to continental. At range margins, beech forests are more susceptible to climate changes, especially considering a future northward shift of the Mediterranean climate conditions (Vacchiano et al. 2017a). Over centuries, beech has been widely used as a major subsistence resource (Nocentini 2009; Vacchiano et al. 2017a; Ottaviani et al. 2019), and nowadays, it also represents one of the main species within the conservation network in southern Europe countries (Maesano et al. 2014). Therefore, the prediction of masting events might be very important under both perspectives of management for production or for conservation of beech forest. Usually, silvicultural activities require long times for planning, especially under a conservation regime. Thus, a mast prediction tool could inform and accelerate decision-making about regeneration cutting. Different models try to reproduce mast patterns and, in some cases, to predict them (Kelly et al. 2013; Pearse et al. 2016; Vacchiano et al. 2018), but generally they require much input information creating a gap between masting process understanding and operationality. Considering the constraints and the limits in the knowledge of mast dynamics, the aim of this study was to provide a simple tool that allows to predict the probability of seed production, i.e. a mast event in the current year, during the winter or early spring, with the lowest information level possible. To do that, we created an R package (R Core Team 2020) called *foreMast* that calculates the mast probability given the summer monthly mean temperature and precipitation values (June, July and August) of the 2 past years.

2 Material and methods

We based the development of our tool on three main aspects: (a) flexibility, (b) open data and (c) added value. The tool must be shareable among different users and adaptable to the local needs, thus open to further implementation and code editing. Data must be accessible and cost free, with a good quality in terms of resolution and accuracy. Consequently, the tool must provide reliable information, with a good level of reproducibility across different sites. In this way, we developed three functions on R, wrapped in a package (*foreMast*) released under Creative Commons Attribution 4.0 licence in GitHub (https://github.com/uchiavetta/foreMast). Assuming that synchronicity in mast events derives indirectly from shared microclimate, the aim was to compute the probability of large seed production in the current year for a given location using climatic data only, as they are tightly correlated with masting years (Piovesan and Adams 2001; Drobyshev et al. 2010, 2014; Vacchiano et al. 2017b). To do that, we considered the weather conditions of specific years (the 2 previous years) in a certain location as percent ranks in the whole observed climatic variability, instead of specific absolute thresholds. Indeed, we assumed that an arid or a wet year in two nearby sites with different precipitation ranges constitutes the same cue, despite a difference in the precipitation absolute values. The package includes three functions (Fig. 1). The first function (cdsDownload) allows users to download the climate data for a given site using the Copernicus Climate Data Store (CDS). To fill out the request, it is important to specify the period and the site of interest. The site is defined by a single point within the selected area, given the latitude and the longitude. After completing the request, the downloaded data are processed by the second function (mastFaSyl) which calculates the mast probability and returns the numeric data. A third function (probPlot) plots a graphic view of the same results.

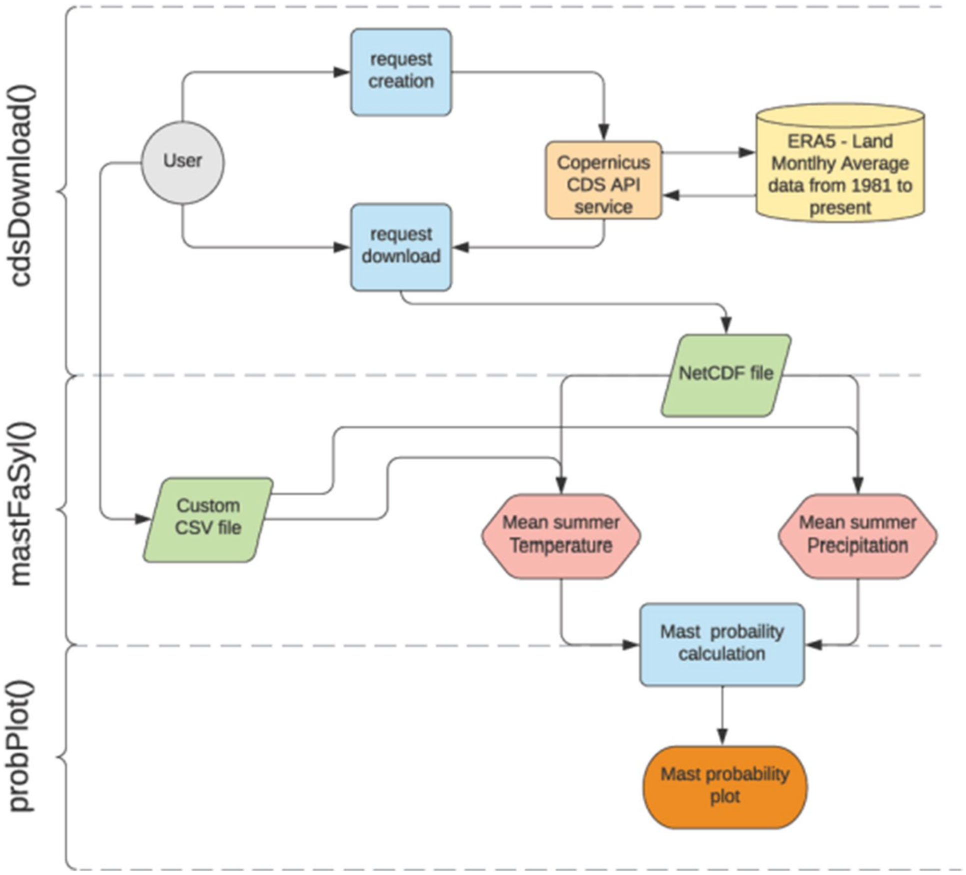

Fig. 1 Workflow of the foreMast package and its functions

2.1 Climatic data

The data used for calculating the mast probability in the current year originates from the Copernicus CDS service. Copernicus is the European Union's Earth observation programme and provides many useful tools for the study and analysis of the different aspects of the environment. Within the Copernicus Climate Change Service (C3S) and with the collaboration of ECMWF (European Centre for Medium-Range Weather Forecasts), several open climate datasets are available. Among all, we used "ERA5-Land monthly average data from 1981 to present" (Muñoz-Sabater et al. 2020) that provides data with a horizontal resolution of $0.1° \times 0.1°$ (around 9 km), starting from 1981, and with a monthly resolution. The update frequency of the dataset is monthly with a delay of about 3 months from the current date. We chose this dataset since it has the highest resolution, compared to others, with a long availability of data and a continuous update of the dataset. CDS data can be easily downloaded, selecting the variables of interest, from the web API or using R, through the package "ecmwfr" (Hufkens et al. 2019). The "ncdf4" package (Pierce 2019) is needed to manage such data. In addition, climatic data can also be uploaded using a csv file with records gathered from local weather stations or other sources.

2.2 Mast algorithm

The main function of the package returns the probability of seed production in the current year, i.e. the probability of a mast event. The first function downloads the climate data for the requested period, given as input the starting year (1981) and the coordinates of the site. To calculate the probability of seed production, the algorithm relies on three main parameters: (a) summer precipitation of the 2 past years; (b) summer temperature of the 2 past years; and (c) autocorrelation of mast probability of 2 consecutive years estimated only by parameters 1 and 2.

According to the literature reviewed (Hilton 2003; Drobyshev et al. 2010, 2014; Kelly et al. 2013; Kasprzyk et al. 2014; Vacchiano et al. 2017b), the summer 1 year before a mast event has to be warm and dry. Therefore, we calculated a score directly proportional to the summer temperature percentile rank and inversely proportional to the summer precipitation one. Moreover, the summer the second year before the masting event must be cooler and wetter than the next year. Therefore, we calculated a score inversely proportional to the summer temperature percentile rank and directly proportional to that of the summer precipitation. As in literature temperature showed to be a primary cue compared to precipitation, in both the scores, we considered the opportunity to give different weights to the two variables. Some authors (Kelly et al. 2013; Vacchiano et al 2017b) found ΔT (the difference between t-2 and t-1) as a good predictor of masting. Our algorithm implicitly considers this effect since a higher ΔT implies opposite temperatures percent ranks in 2 consecutive years. The third assumption regards autocorrelation, as mast years are generally followed by a year with scarce seed production and it is unlikely to have two consequent masting events (Piovesan and Adams 2001), since there is a very low probability to have a high ΔT the year soon after a very high seed production (likely a previous very high ΔT). To implement the lagged autocorrelation, we developed the algorithm to decrease the prediction probability in relation to the (potential) seed production in the year before. Therefore, monthly mean temperatures and precipitation are downloaded. Successively, the summer (June, July and August) temperatures ($\bar{T}_s$) and precipitation ($\bar{P}_s$) averages are calculated for each year, as they show to be strictly related with the occurrence of masting events (Drobyshev et al. 2010; Kelly et al. 2013; Vacchiano et al. 2017b), obtaining a time series with one value per year for each variable. Once the datasets are ready, the masting probability at the year t is calculated as described below. First, the five scores are calculated:

$$S_{t-1} = \frac{w_T \cdot R_{T_{t-1}} + w_P \cdot (1 - R_{P_{t-1}})}{w_T + w_P} \tag{1}$$

$$S_{t-2} = \frac{w_T \cdot (1 - R_{T_{t-1}}) + w_P \cdot R_{P_{t-1}}}{w_T + w_P} \tag{2}$$

$$S' = \frac{S_{t-1} + S_{t-2}}{2} \tag{3}$$

$$S_{ac} = 1 - R^2_{S'_{t-1}} \tag{4}$$

$$S_p = \frac{S_{t-1} + S_{t-2} + w_{Sac} \cdot S_{ac}}{2 + w_{Sac}} \tag{5}$$

where:

S_{t-1} is the score calculated for weather conditions at the year t-1.

S_{t-2} is the score calculated for weather conditions at the year t-2.

S' is the overall masting score at the year t calculated without considering autocorrelation.

S_{ac} is the autocorrelation score.

S_p is the overall final masting score at the year t considering also autocorrelation.

w_T and w_P are the weights given to summer temperature and precipitation, respectively, in the algorithm.

$R_{T_{t-1}}$ and $R_{T_{t-2}}$ are the percentile rank of summer temperatures registered in the summer at t-1 and t-2 respectively, during the available weather data observation series (1981–present), in the code they are used as decimal (range 0 to 1).

$R_{P_{t-1}}$ and $R_{P_{t-2}}$ are the percentile rank of summer precipitation registered in the summer at t-1 and t-2, respectively, during the available weather data observation series (1981–present), in the code they are used as decimal (range 0 to 1).

w_{Sac} is the weight given to the autocorrelation score.

The first two scores calculated by Eqs. 1 and 2 range from 0 to 1. The value is 0 when the worst combination of temperature and precipitation occurs in the observed period. Conversely, the value is 1 when the best combination of the two variables occurs. The score is the weighted average, and the weights can be tested against an observed masting dataset.

The average of the two scores calculated by Eq. 3 is a temporary calculation of the masting probability not considering masting autocorrelation.

To consider autocorrelation, we assumed that *S'* at the year *t-1* represents the past mast event. Then we calculated the current year mast score as the complement to the previous year squared percent rank of *S'* (Eq. 4). In this way, a potential mast event the year before will create a low score in the current one proportionally more than a low seeding event will do. The overall final score is calculated by the weighted mean of the scores obtained by Eqs. 1, 2 and 4. The first two scores have a weight of 1, while the weight for the fourth score can be customised (default is arbitrarily set to 0.5). This score ranges theoretically from 0 to 1, but since it is a very smoothed value, it needs a transformation to be compared with observed seed production percentiles. Then, the final probability of the masting event is given by the percentile rank of the overall final masting score S_p with the:

$$S_p = R_{Sp} \tag{6}$$

2.3 Testing data

To test our algorithm, we used the MASTREE database (Piovesan and Adams 2001; Ascoli et al. 2017), which is a set of data about the seed production of Norway spruce and European beech, collected throughout all Europe in the last two centuries. To analyse the functionality of the algorithm, we selected different sites among the MASTREE dataset, to compare the outcome with the observed data. Therefore, we filtered the dataset for the data pertaining to beech, which proxies where seed or fruit, with observation starting from 1981, a length of records of at least 9 years and reporting coordinates. Only 10 series, located in 6 European countries, satisfied our parameters (Table 1). After the selection, for each record, we downloaded the climate data, and we ran the algorithm to calculate the mast probability.

2.4 Statistical analyses

To analyse the algorithm outcomes in relation to the observed data, we performed the nonparametric Spearman's rank correlation, along with a significance test where α is set to 0.05. To compare the two variables, first we calculated the percentile rank of the MASTREE values related to the seed or fruit records in the different years. Secondly, for each site, we performed the Spearman's test among the two different types of data using the cor.test function on *R*, where for Spearman's test, *p*-values were computed using algorithm AS 89 (Best and Roberts 1975).

3 Results

The results can be observed from Tables 2 and 3. Spearman's correlation test has been performed on the whole dataset and for each site. Moreover, we ran the algorithm changing the temperature and precipitation weights (w_T and w_P) to observe possible variation in prediction accuracy and to determine the best variable combination to be set in the mast calculation function (mastFaSyl) of the *foreMast* package. In Table 2, the statistics related to the overall analyses are reported, while in Table 3, the values of ρ and *p*-value are shown for each site.

The best overall performance has been obtained when setting the w_T and w_P ratio to 3:1 ($\rho = 0.50$; *p*-value = 1.16e − 13). Single site trials showed different trends. For example, FASYCH0003A presented the highest correlation value among all, with a ρ of 0.88 and a *p* value of 0.0004, setting the temperature and precipitation weights ratio to 4:1 (Table 3). The worst case, represented by a low correlation value and a very low significance, pertained to FASYFR0017A (ρ: 0.13; *p* value: 0.56) with a w_T and w_P ratio to 1:2.

Table 2 Overall ρ and *p*-values for the trial ran with different temperature and precipitation weights

w_T	w_P	Overall rho (ρ)	Overall *p* value
1	1	0.4703691	8.53e − 12
1	2	0.4320298	5.38e − 10
1	3	0.4249535	1.09e − 09
2	1	0.4790876	3.08e − 12
3	1	0.505618	1.16e − 13
4	1	0.4904758	7.82e − 13
5	1	0.4888744	9.52e − 13
6	1	0.4729883	6.30e − 12

In Fig. 2, we plotted the results obtained with the best performing combination in the overall trial: $w_T = 3$ and $w_P = 1$. The calculated probability trend followed the observed data. Most sites (8 of 10) showed ρ values ranging between 0.51and 0.80, and with *p*-values generally lower than 0.05, except for FASYFR0017A (ρ, 0.26; *p* value, 0.24) and FASYSI0024A (ρ, 0.30; *p* value, 0.13).

Among all the trials, high values of *p* (> 0.05) were more frequent when the precipitation had a weight three times higher than the temperature ($w_T = 1$ and $w_P = 3$), where 5 out of 10 sites showed a significant correlation.

Despite the overall analyses, within the single cases, half of them presented an increase, even if slightly, in the prediction accuracy when the precipitation weighs more than the temperature (Table 3).

Table 1 Selected sites from the MASTREE dataset used to analyse the mast algorithm performance

ID	Proxy	Country	Lat (N)	Lon (E)	Start year	End year
FASYCH0003A	Seed	Switzerland	47.46	7.50	1999	2009
FASYCZ0014A	Seed	Czech Republic	50.73	15.55	1980	2006
FASYFR0007A	Fruit	France	48.08	7.66	1996	2005
FASYFR0017A	Seed	France	44.10	6.53	1994	2015
FASYFR0023A	Seed	France	43.15	-0.65	1994	2014
FASYIT0009A	Seed	Italy	43.80	11.81	1991	2010
FASYIT0115A	Seed	Italy	43.63	11.90	1992	2009
FASYIT0116A	Seed	Italy	43.63	11.90	1992	2009
FASYSI0024A	Seed	Slovenia	46.45	15.38	1987	2015
FASYSK0001B	Seed	Slovakia	49.56	19.53	1995	2004

Table 3 Values of ρ and p-value for each site using different weights of temperature (w_T) and precipitation (w_P). Values followed by * have a significance between 0.01 and 0.05, while ** highlights p-values lower than 0.01

Site ID	Metrics	$w_T=1$ $w_P=1$	$w_T=1$ $w_P=2$	$w_T=1$ $w_P=3$	$w_T=2$ $w_P=1$	$w_T=3$ $w_P=1$	$w_T=4$ $w_P=1$	$w_T=5$ $w_P=1$	$w_T=6$ $w_P=1$
FASYCH0003A	ρ	0.74	0.75	0.74	0.79	0.80	0.88	0.80	0.74
	p-value	0.01**	0.008**	0.009**	0.004**	0.003**	0.000**	0.003**	0.009**
FASYCZ0014A	ρ	0.39	0.37	0.3	0.48	0.54	0.44	0.43	0.5
	p-value	0.059	0.076	0.15	0.018*	0.007**	0.031*	0.038*	0.013*
FASYFR0007A	ρ	0.65	0.66	0.66	0.60	0.66	0.59	0.61	0.59
	p-value	0.049*	0.044*	0.044*	0.066	0.044*	0.08	0.062	0.08
FASYFR0017A	ρ	0.26	0.13	0.22	0.28	0.26	0.27	0.28	0.25
	p-value	0.24	0.56	0.33	0.21	0.24	0.23	0.20	0.27
FASYFR0023A	ρ	0.62	0.60	0.59	0.63	0.56	0.36	0.46	0.42
	p-value	0.003**	0.004**	0.005**	0.002**	0.009**	0.11	0.038**	0.05*
FASYIT0009A	ρ	0.66	0.67	0.67	0.66	0.61	0.64	0.67	0.63
	p-value	0.001**	0.001**	0.001	0.003**	0.005**	0.002**	0.001**	0.003**
FASYIT0115A	ρ	0.48	0.53	0.53	0.51	0.51	0.5	0.51	0.50
	p-value	0.043*	0.024*	0.024*	0.032*	0.032*	0.036*	0.032*	0.036*
FASYIT0116A	ρ	0.68	0.72	0.72	0.73	0.69	0.69	0.68	0.66
	p-value	0.002**	0.001**	0.000**	0.001**	0.001**	0.002**	0.002**	0.003**
FASYSI0024A	ρ	0.23	0.071	0.061	0.22	0.30	0.33	0.33	0.30
	p-value	0.23	0.72	0.76	0.26	0.13	0.082	0.089	0.12
FASYSK0001B	ρ	0.53	0.44	0.44	0.5	0.64	0.67	0.56	0.55
	p-value	0.028*	0.078	0.08	0.042*	0.005**	0.003**	0.019*	0.022*

4 Discussion

Beech is an important species of the European forests, characterising the landscape and providing different ecosystem services. Understanding its ecology and developing prediction models related to its dynamic would be of great benefit for the improvement of forest management practices under different perspectives (Vacchiano et al. 2018). By trying to contribute to this purpose, we developed an algorithm, implemented in an R package, which aims at predicting the masting probability of the current year by considering as main mast event drivers the temperature and the precipitation of the preceding 2 years (Drobyshev et al. 2010). In our study, we focused on the prediction of large seed production of a population in a certain location. We did not directly considered synchronicity among the different areas, which we assumed deriving indirectly from common weather conditions on a large scale, both spatial and temporal, in the framework of the mesoclimate. For example, an arid or a wet year in two nearby sites with different precipitation ranges constitutes the same cue, despite a difference in the precipitation absolute values. We tested our algorithm on 10 sites in 6 European countries, selected from the MASTREE dataset (Ascoli et al. 2017), with available coordinates that are needed to run the model. In most of the cases, the results showed a significant strong or mild positive correlation between the observed data and the prediction outcome (Fig. 1). The prediction trends consistently followed the observed ones, with some irregularities and few peak shifts. In two sites, we observed a weak correlation. The first is FASYFR0017A, which represented the worst prediction case. In this site, the Klaus cyclone occurred in January 2009 with gusts of wind up to 190 km/h in that area. Severe windstorms can strongly damage the vegetation (crown, buds, branches or even stems) compromising the current or next year seed production. Accordingly, in Fig. 2, the missing seed production in 2009 and 2010 is very evident. Conversely, in the same period, the algorithm forecasted a strong predisposition to a mast event since the algorithm does not consider windstorm events. In this site, excluding these 2 years from observations, the Spearman test improved giving a ρ of 0.46 (p-value = 0.043). Another weak case was FASYSI0024A. In this case, the weak correlation can be due to the observation method which, differently from the other sites, is based on ordinal data deriving from field visual classification. Besides the subjectivity of the method, the class 1 (percentile rank = 0) is assigned in both the case of small and absent seed productions. This leads to repeating zero production years, as it is possible to observe in Fig. 2, where many years of no seed production are present, although the real seed quantity would have probably been small but higher than 0. In Table 4, we showed the overall correlation and significance values for the different tests, omitting the above-mentioned sites. It is worth noting that the algorithm performance increases for all the tests, following the general trend observed in Table 1.

Despite a general good correlation of the method, several aspects influence the algorithm performances and are important to be underlined to correctly understand its results. As already mentioned, our algorithm relies on the

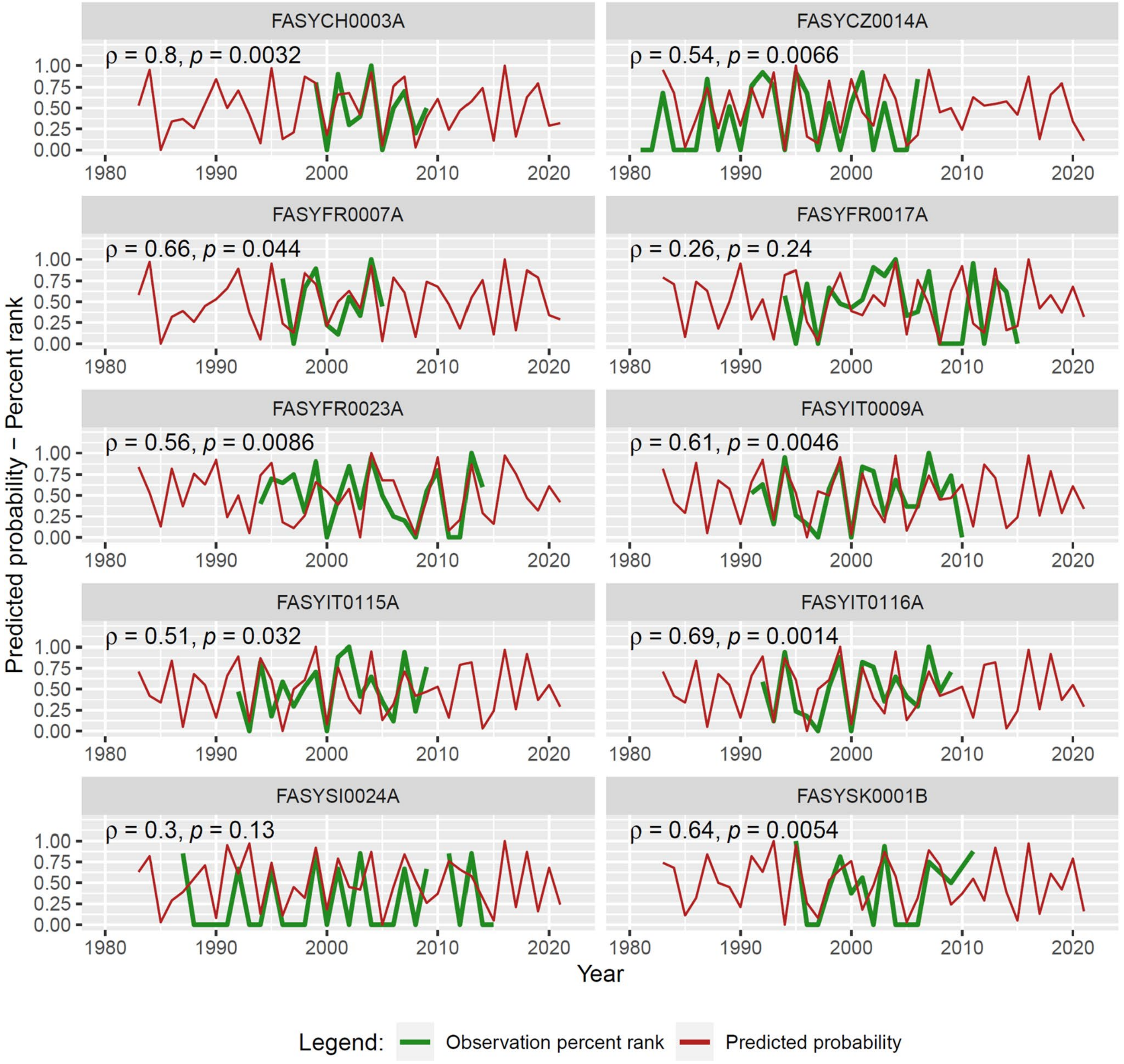

Fig. 2 Plots representing the predicted mast probability of the best overall case scenario ($w_T=3$ and $w_P=1$) and the observed seed production events for the related sites selected from the MASTREE database. For each plot, there are two different data representations, one for the predicted mast probability, i.e. the algorithm outcome (red line) and one for the data collected within the MASTREE dataset (green line). High and low peaks correspond respectively to high and low seed production probability (expressed in percentile rank). For each site, the Spearman's rank correlation coefficient (ρ) and the p-value are reported

weather cues that are strongly related to a mast event. We used CDS data, related to the temperature and the precipitation of the different sites. Although this type of data can be easily downloaded with no costs, the low resolution may hamper the algorithm performance. Drobyshev et al. (2010) observed that temperatures, as a regional component, are less affected by site factors than precipitation, whose dynamics are mainly related to the soil and the stand profile. Nonetheless, using the percentile rank transformation, anomalies in the series are avoided, and data are easy to be compared also with a low spatial resolution. Nevertheless, further studies should investigate which scale is better related to the climate cues, since within a forest, there might be non-uniform weather patterns in relation to the morphology. Looking at the seed production, precipitation is often reported to have a weaker consistency as a predictor with low correlation values than temperatures (Vacchiano et al. 2017b), although other authors (Piovesan and Adams 2001;

Table 4 Overall statistics for the various experiments after removing the sites FASYFR0017A and FASYSI0024A, which present probable issues related to the data collection

w_T	wP	Overall rho (ρ)	Overall p value
1	1	0.5783661	8.91e – 14
1	2	0.5792045	8.05e – 14
1	3	0.5607539	7.02e – 13
2	1	0.5937003	1.33e – 14
3	1	0.6147845	8.23e – 16
4	1	0.579739	7.54e – 14
5	1	0.5779112	9.41e – 14
6	1	0.5658254	3.92e – 13

Vacchiano et al. 2017b) suggest that the water balance represents the major constraint in triggering mast events. We did not use other parameters related to the water balance, such as the air humidity or the evapotranspiration, but we found that, from an overall perspective, precipitation plays a minor role in predicting mast events. In fact, looking at the results (Tables 2 and 3), it is possible to observe that when the precipitation has a weight higher than the temperature, the overall ρ and p-value decrease, which is the contrary to what happens after increasing the weight of the temperature, relatively to the precipitation one, in accordance with the studies that suggest temperature being the main trigger of masting years (Drobyshev et al. 2010; Kelly et al. 2013; Vacchiano et al. 2017b). In any case, the overall significance increases until the ratio between the temperature and precipitation weights reaches the threshold of 3:1 (Tables 2 and 3), decreasing soon after. Anyway, when increasing the temperature weight, despite a general improvement, some sites showed a slight decrease for ρ and p-value. This is evident also when setting the algorithm with precipitation weighing more than temperature, where, instead, some sites showed an improvement in the prediction accuracy (FASYFR0007A, FASYFR0023A, FASYIT009A, FASYIT0115A, FASYIT0116A). This might probably be linked to (a) the local constraints of the sites such as the exposition, the altitude and the soil characteristics, which affect the water availability and the temperature variation during the year, and (b) the climate region where the sites are located and, consequently, the limiting factors. Therefore, we decided to release our R package (*foreMast*) by setting the weights of the temperature and the precipitation of the mastFaSyl function in three ways that can be decided by the user: (1) $w_t = 3$ and $w_p = 1$, as it showed an overall good performance (Table 2); (2) automatically set with the most performing weights of the nearest MASTREE site (Table 3), (3) manually by the user. Mast events also depend strictly on the current year dynamics; precipitation plays a major role during the spring and early summer months of the current year (Kasprzyk et al. 2014), as they negatively affect the pollination efficiency of beech (Bogdziewicz et al. 2020), decreasing the probability of large seed productions. Also, late frost and summer drought concur to limit mast events. Late frosts usually happen in spring, burning the leaf buds and the flowers, therefore hampering, or importantly limiting, the pollen production (Augspurger 2009; Pearse et al. 2016; D'Andrea et al. 2019b). Reproduction failures have also been observed after extreme hot and arid summer, which lead to the fruit abortion (Nussbaumer et al., 2020) or severe windstorms. Also, an auto-nullifying option has been implemented in the mastFaSyl function. This option requests user to report a severe adverse event able to compromise seed production in the current growing season. Our algorithm relies on data that are updated monthly with a lag of 3 months. Since late frosts usually happen between April and May, it might be possible to implement them as parameters in the mast function. Nonetheless, it would be counterproductive when aiming at supporting the preparatory procedures required in planning forestry activities, as the data would be available too late, when other masting cues can be directly observed (e.g. flowering or fruit abundance). Same issue regards summer drought, mainly occurring in August, where the relative data would probably be available for November at least. These additional parameters may increase the performance of the algorithm in predicting a mast event in the current year but would be of non-practical use. Nonetheless, other climate datasets, which provide continuous updated data with possible no lag in their availability, could be used, but we were not able to find a climate service with such a large coverage and with a higher resolution. Another solution to obtain updated climate variables might be to use custom weather stations, which provide continuously detailed data with a high accuracy. Nevertheless, they require several costs for their installation and maintenance, especially for long-term use.

5 Conclusions

Masting years are very important events for forest dynamics, linked to different ecological constraints and climate cues. When managing a forest, the prediction of such events might be of great importance for planning silvicultural activities, aiming at both a productive and a conservative approach (Ottaviani et al. 2019). Understanding masting dynamics and predicting them may concur in the adaptation of forest management planning to climate change (Wagner et al. 2010). European beech is the most representative European forest species, linked to different ecosystem services. Aiming at an improvement of forestry planning, we developed a tool allowing the prediction of mast events of beech forest for the next growing season. The tool is an R package (foreMast), which uses the temperature and the precipitation from an open access global database or from local observations, as main cues triggering mast events, as well as considering a negative autocorrelation factor. The results showed that overall, the algorithm performed well (Tables 2 and 3), especially increasing the temperature weight over the precipitation one, suggesting that temperature is a more significant trigger of masting events than precipitation. At the same time, the correlation of the predicted mast probability with the record of seed production for each site changed slightly among the different trials without a constant improvement (Table 3). Same happened when increasing the precipitation weight over the temperature one, showing the lowest overall correlation and significance but an improvement in the prediction ability for some sites. Nevertheless, when using the CDS dataset, the algorithm relies on data that are updated with a lag of several months, being therefore constrained by the exclusion of current year events that play a significant role in masting prediction, such as late frosts or summer droughts (D’Andrea et al. 2019a, b; Nussbaumer et al. 2020). Anyway, the tool offers the advantage of having a preliminary probability of the mast event for the next growing season; the lack of information about late frosts and summer heat waves and droughts might be compensated using weather stations or direct observations, until a more updated climate service server is available. Moreover, since different sites are related to slight changes in the significance of climate cues, users have the possibility to modify the algorithm code to make it more correspondent to the local dynamics based on the past observations. Further analyses should be made, implementing late frosts and summer drought as parameters and observing if they increase the algorithm performance over the mast time series. Other climate-related parameters could be implemented, especially looking at the water balance and water exchanges through evapotranspiration mechanisms and the air humidity along with dendrochronological data, since masting years are linked to a reduction in the diametral increment (Drobyshev et al. 2010). Moreover, high-resolution data should be tested to understand how microclimate is linked locally to mast events, despite regular trends on regional scale. In conclusion, we believe that this tool might be useful in planning forestry activities, which usually relies on long time procedures, allowing also to analyse European beech forests mast trends and fostering their adaptation to climate change scenarios. Moreover, beech mast events also affect wildlife species depending directly or indirectly on beech fruit abundance such as bears, rodents and ungulates. Thus, this tool can also support wildlife conservation and management. Finally, the package is an open, flexible and dynamic project. New variables can be considered to predict European beech forests, and new algorithms and functions can be added for other species.

Acknowledgements This work was realised with the contribution of the LIFE Programme of the European Union-Project AForClimate “Adaptation of forest management to climate variability: an ecological approach” (LIFE15 CCA/IT/000089).

Author contribution All authors have contributed equally to the development of the research and to the writing of the article.

Availability of data and material Data sharing is not applicable to this article as no original datasets were generated or analysed during the current study.

Code availability The foreMast package and its codes generated during the current study are available in the foreMast repository, following this link https://github.com/uchiavetta/foreMast.

Declarations

Ethics approval The authors declare that the study was not conducted on endangered, vulnerable or threatened species.

Consent to participate The authors declare that no human participants have been involved in this study.

Consent for publication All authors gave their informed consent to this publication and its content.

Conflict of interest The authors declare no competing interests.

References

Ascoli D, Maringer J, Hacket-Pain A et al (2017) Two centuries of masting data for European beech and Norway spruce across the European continent. Ecology 98:1473. https://doi.org/10.1002/ecy.1785

Augspurger CK (2009) Spring 2007 warmth and frost: phenology, damage and refoliation in a temperate deciduous forest. Funct Ecol 23:1031–1039

Bajocco S, Ferrara C, Bascietto M et al (2021) Characterizing the climatic niche of mast seeding in beech: evidences of trade-offs between vegetation growth and seed production. Ecol Indic 121:107139

Best DJ, Roberts DE (1975) Algorithm AS 89: the upper tail probabilities of Spearman’s rho. Appl Stat 24:377

Bogdziewicz M, Kelly D, Tanentzap AJ et al (2020) Climate change strengthens selection for mast seeding in European beech. Curr Biol 30:3477-3483.e2. https://doi.org/10.1016/j.cub.2020.06.056

Cutini A, Chianucci F, Chirichella R et al (2013) Mast seeding in deciduous forests of the northern Apennines (Italy) and its influence on wild boar population dynamics. Ann for Sci 70:493–502

Cutini A, Chianucci F, Giannini T (2010) Effect of the silvicultural treatment on canopy properties, litter and seed production in beech coppices under conversion to high forest. Ann Silvicultural Res 36:109–124. https://doi.org/10.12899/asr-825

D’Andrea E, Rezaie N, Battistelli A et al (2019a) Winter’s bite: beech trees survive complete defoliation due to spring late-frost damage by mobilizing old C reserves. New Phytol 224:625–631. https://doi.org/10.1111/nph.16047

D'Andrea E, Rezaie N, Prislan P et al (2019b) Frost and drought: effects of extreme weather events on stem carbon dynamics in a Mediterranean beech forest. bioRxiv

Drobyshev I, Niklasson M, Mazerolle MJ, Bergeron Y (2014) Reconstruction of a 253-year long mast record of European beech reveals its association with large scale temperature variability and no long-term trend in mast frequencies. Agric Meteorol 192–193:9–17

Drobyshev I, Övergaard R, Saygin I et al (2010) Masting behaviour and dendrochronology of European beech (Fagus sylvatica L.) in southern Sweden. For Ecol Manage 259:2160–2171

Hilton GM (2003) Variation in the masting of common beech (Fagus sylvatica L.) in northern Europe over two centuries (1800–2001). Forestry 76:319–328

Hufkens K, Stauffer R, Campitelli E (2019) khufkens/ecmwfr: ecmwfr. Zenodo

Janzen DH (1971) Seed Predation by Animals. Annu Rev Ecol Syst 2:465–492

Kasprzyk I, Ortyl B, Dulska-Jeż A (2014) Relationships among weather parameters, airborne pollen and seed crops of Fagus and Quercus in Poland. Agric for Meteorol 197:111–122

Kelly D (1994) The evolutionary ecology of mast seeding. Trends Ecol Evol 9:465–470. https://doi.org/10.1016/0169-5347(94)90310-7

Kelly D, Geldenhuis A, James A et al (2013) Of mast and mean: differential-temperature cue makes mast seeding insensitive to climate change. Ecol Lett 16:90–98. https://doi.org/10.1111/ele.12020

Kelly D, Hart DE, Allen RB (2001) Evaluating the wind pollination benefits of mast seeding. Ecology 82:117–126

Leuschner C, Ellenberg H (2017) Beech and mixed beech forests. Ecol Central Eur For 351–441

Lucas-Borja ME, Vacchiano G (2018) Interactions between climate, growth and seed production in Spanish black pine (Pinus nigra Arn. ssp. salzmannii) forests in Cuenca Mountains (Spain). New for 49:399–414

Maesano M, Lasserre B, Marchetti M (2014) Increasing wood mobilization through sustainable forest management in protected areas of Italy. Annals of Silvicultural Research 38:13–21. https://doi.org/10.12899/asr-754

Monks A, Kelly D (2006) Testing the resource-matching hypothesis in the mast seeding tree Nothofagus truncata (Fagaceae). Austral Ecol 31:366–375

Muñoz-Sabater J, Dutra E, Agustí-Panareda A, et al (2020) ERA5-Land: a state-of-the-art global reanalysis dataset for land applications. 10.5194/essd-2021-82

Nocentini S (2009) Structure and management of beech (Fagus sylvatica L.) forests in Italy. iForest – Biogeosciences For 2:105–113

Nussbaumer A, Meusburger K, Schmitt M et al (2020) Extreme summer heat and drought lead to early fruit abortion in European beech. Sci Rep 10:5334. https://doi.org/10.1038/s41598-020-62073-0

Nussbaumer A, Waldner P, Etzold S et al (2016) Patterns of mast fruiting of common beech, sessile and common oak, Norway spruce and Scots pine in Central and Northern Europe. For Ecol Manage 363:237–251. https://doi.org/10.1016/j.foreco.2015.12.033

Ottaviani G, Götzenberger L, Bacaro G et al (2019) A multifaceted approach for beech forest conservation: environmental drivers of understory plant diversity. Flora 256:85–91

Packham JR, Hytteborn H (2012) Swedish beech forests and the storm gap theory. Arboricult J 34:151–159

Packham JR, Thomas PA, Lageard JGA, Hilton GM (2008) THE English beech masting survey 1980–2007: variation in the fruiting of the common beech (Fagus sylvatica L.) and its effects on woodland ecosystems. Arboricult J 31:189–214

Pearse IS, Koenig WD, Kelly D (2016) Mechanisms of mast seeding: resources, weather, cues, and selection. New Phytol 212:546–562. https://doi.org/10.1111/nph.14114

Pierce D (2019) ncdf4: Interface to Unidata netCDF (Version 4 or Earlier) Format Data Files

Piovesan G, Adams JM (2001) Masting behaviour in beech: linking reproduction and climatic variation. Can J Bot 79:1039–1047

R Core Team (2020) R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing

Vacchiano G, Ascoli D, Berzaghi F et al (2018) Reproducing reproduction: how to simulate mast seeding in forest models. Ecol Modell 376:40–53. https://doi.org/10.1016/j.ecolmodel.2018.03.004

Vacchiano G, Garbarino M, Lingua E, Motta R (2017a) Forest dynamics and disturbance regimes in the Italian Apennines. For Ecol Manage 388:57–66

Vacchiano G, Hacket-Pain A, Turco M et al (2017b) Spatial patterns and broad-scale weather cues of beech mast seeding in Europe. New Phytol 215:595–608. https://doi.org/10.1111/nph.14600

Wagner S, Collet C, Madsen P et al (2010) Beech regeneration research: from ecological to silvicultural aspects. For Ecol Manage 259:2172–2182