# *Long-term annual seed production data of individual European beech (Fagus sylvatica) trees in the Netherlands*

Cherine C. Jantzen[1,2*], Marcel E. Visser[1,2]

[1]Department of Animal Ecology, Netherlands Institute of Ecology (NIOO-KNAW), The Netherlands

[2]Groningen Institute for Evolutionary Life Sciences (GELIFES), University of Groningen, The Netherlands

[*]Corresponding author: c.jantzen@nioo.knaw.nl; ORCID: 0009-0006-0723-2682

MEV: ORCID: 0000-0002-1456-1939

**Abstract**

Seed production of European beech (*Fagus sylvatica*) is highly variable between years and synchronised between individual trees (i.e., masting), creating years with high seed crops, separated by one or more years with little or no seed production. This pulsed reproduction has selective benefits for the trees, as it creates a cycle of predator satiation and starvation, and this way masting drives population dynamics of dependent species. Masting is sensitive to temperature, and thereby affected by climate change, making long-term data on annual seed production valuable for understanding changes in the masting patterns and their effects on seed consumers. This study presents a long-term dataset of 50 years containing annual, ground-plot beechnut counts for 81 individual trees in a study area of 111 ha in the Netherlands. Beechnut counts are categorized in sound, empty and predated nuts, complemented by gross and net weights of individual nuts, allowing for detailed analyses of masting patterns and annual variation in food abundance for seed predators.

## Background & Summary

European beech (*Fagus sylvatica*) is a forest forming tree species in Central Europe and its seeds (i.e., beechnuts) provide a valuable food resource for many forest species (e.g.,[1–3]). Since beeches reproduce by mast seeding (hereafter masting), meaning that their seed production is highly variable between years and strongly synchronized between individuals[4], this food resource is however variable, being superabundant in some years and almost absent to completely absent in other years, leading to demographic changes in seed consumer populations[1,5]. Since beechnuts are a crucial winter food resource for resident passerine birds, the dataset presented here was initially collected alongside a long-term nest box study on passerine birds[6] to analyse the effect of winter food availability on the bird's winter survival. It provides long-term, individual-level data on European beech seed production from the National Park De Hoge Veluwe in the Netherlands ranging from 1976 to 2025. The data consist of annual ground-plot based counts and weights of individual nuts from 81 individual trees resulting in 12.313 sampled plots.

Besides its original purpose, this dataset can contribute to our understanding of reproduction dynamics in beech and how they are affected by climate change[7]. Over the past decades, the field of masting research has grown rapidly with many comparative studies focussing on beech seed production across Europe (e.g.,[8,9]) and efforts have been put into collating reproductive time-series data from a diversity of plant species into one publicly available dataset (MASTREE+[10]). Our dataset can be a valuable addition to this data collection, since the Netherlands are so far mostly not included in comparative studies, due to a lack of available long-term data. Additionally, this dataset can be used for studies around pulse ecology, since high seed years serve as a strong resource pulse, and the resulting effects on interacting species (e.g.,[11,12]). Moreover, this long-term dataset can provide a baseline to capture climate change effects[7] and can be used in forecasting studies on changes in masting under different future climate change scenarios (as done, for example, in[13,14]) or changes in consumer populations with different seed production scenarios (e.g.,[15]).

Until now, the raw data, as described here, has not been used in any publications, but an index of beechnut production derived from the raw data has been used in several studies on bird phenology and demographics, utilizing the bird data from the nest box study in the same area. The earliest study by Perdeck et al.[2] showed that both juvenile and adult annual great tit (*Parus major*) survival is strongly influenced by the amount of beechnuts produced, which was confirmed in a later study using 16 years of additional data[16]. Additionally, beechnut availability has a positive effect on the probability of recruitment and the number of recruits in great tits[16,17]. Similarly, population dynamics and fluctuations of the same species are driven by the variation in available beechnuts[18,19] and, under simulated future scenarios of high beechnut availability, the vital rate of the birds and their population size increase[15]. In contrast, beechnut availability does not seem to have an effect on the selection for earlier egg laying in great tits[20,21], as well as in blue tits[20], but a positive effect on population growth[22]. Lastly, fluctuations in the population size of central European blue and great tit populations are synchronized by a common environmental force through large-scale synchrony in beech masting[19].

The three-scale beech crop index used in these studies is however only based on a subset of the original dataset and is one possible interpretation of the raw data, which in most studies is summarised at the population level. The index categorises the data based on determined thresholds (see "*Calculation of beech crop index*"), which underrepresents the amount of variation in seed production and is restricted to the number of full beechnuts. Looking at the actual counts and weights of the beechnuts can therefore give a much more fine-tuned insight in the reproductive dynamics.

## Methods

### *Data collection*

Beechnuts are collected annually to assess the number of nuts produced by individual trees and the net weights of the whole nuts, because the resulting total annual net weight of the kernels of the beechnuts (i.e., their fleshy part) is relevant to seed consumers. Data collection takes place in autumn, around mid-October, in the Southern part of the National Park De Hoge Veluwe (52.04°N, 5.85°E) in the Netherlands (Figure 1, left map), a study area of 111 ha. The study area consists of a mixed forest with oaks, birch, pine, and larch, growing on poor sandy soils. Sampled trees are distributed across the whole study area, as seen in Figure 1 (right map).

Individual trees are marked and resampled every year, but deviations occur over time, as trees have died or were cut and replaced by proximate trees, or because sampling effort differed. Beechnuts are collected by placing metal squares (hereafter plots) of 25 cm x 25 cm underneath each tree (Figure 2b), collecting all full and partial nuts within the plots (if nuts are still attached to their cupula (Figure 2c), the cupula is removed), placing the nuts in separate bags per plot, drying them by room temperature for at least 24 hours to then sort them into different categories and count the number of nuts per category per plot. While the general collection method stayed the same over time, the involvement of many different people in conducting the fieldwork over a period of 50 years led to slight deviations in the methodology between years, which are described in more detail in the following.

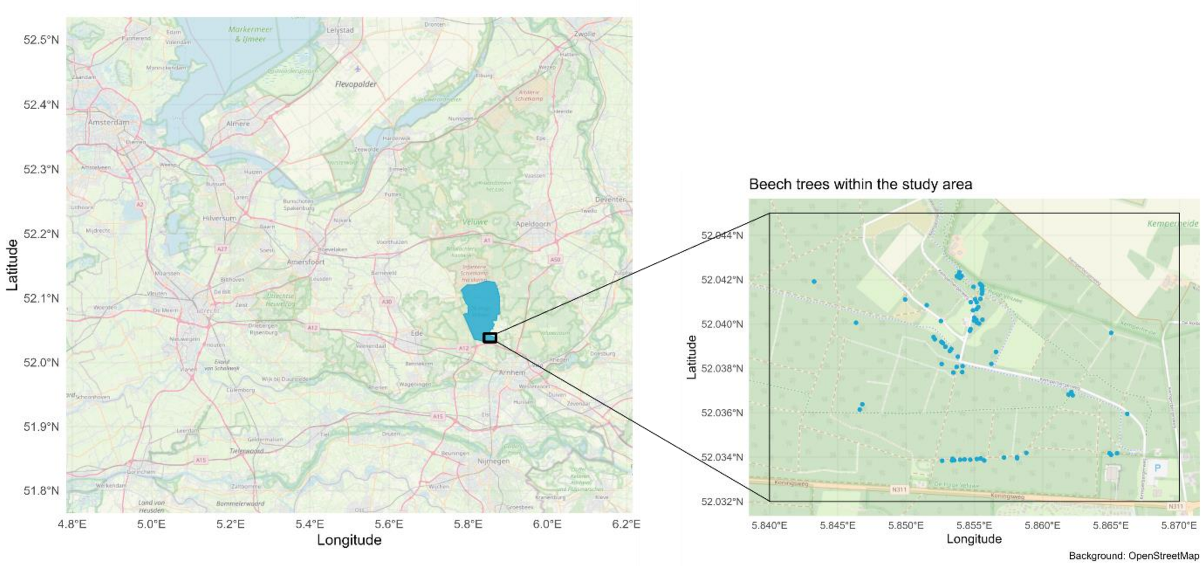


**Figure 1 Location of the study area and the studied beech trees.** The left map shows the location of the National Park De Hoge Veluwe (filled in blue) on the Open Street Map of the Netherlands. The Hoge Veluwe is located north of Arnhem and part of the Natura 2000 area, as well as the Long-Term Ecosystem Research (LTER) site *Veluwe*. The black box indicates the location of the study area within the Hoge Veluwe. The right map is zoomed into this black box, showing the distribution of the sampled beech trees (blue dots) within the study area.

*Field work*

Field work is done annually in mid-October, the standard sampling date (but sometimes more often, see below). However, in 1991 and 1998, samples were not taken in autumn but only in December and January (of the following year), respectively. While in most years, sampling was done on one day a year, there are 13 years in which sampling was spread over several sampling days. This is partly because not all trees could be sampled within one day (1977, 1995, 1979, 1980, 1986, 1992), samples were taken repeatedly within a month or across several months (1976, 1987, 1989, 1990, 2014), or both (i.e., several months have been sampled, but samples within a month were taken on two days; 1983, 2024). If the same trees have been sampled in different months, the plots were moved slightly to avoid resampling the exact same plot. Plots have always been placed approximately in the same direction from the tree, but deviations occurred when, e.g., branches fell off. In 2024, all plots were marked at the first sampling date and these plots were re-sampled on the following sampling occasions (in addition to the normal samples), leading to eight measurements per tree per date (instead of four measurements). While from 1976 to 1986 three plots per tree have been sampled, the number of plots increased to four from 1987 onwards. The plots were placed in a straight line with the first plot located half a meter from the trunk, the last plot perpendicular to the largest overhanging branch and the remaining two plots in equal distances in between.

Zero values are true zeros and there are 12 years (1978, 1981, 1988, 1991, 1996, 1997, 2001, 2003, 2005, 2008, 2017, 2019), in which no nuts could be found. Sampling effort within a season varied between years and there is one missing year (1982), in which no data was collected.

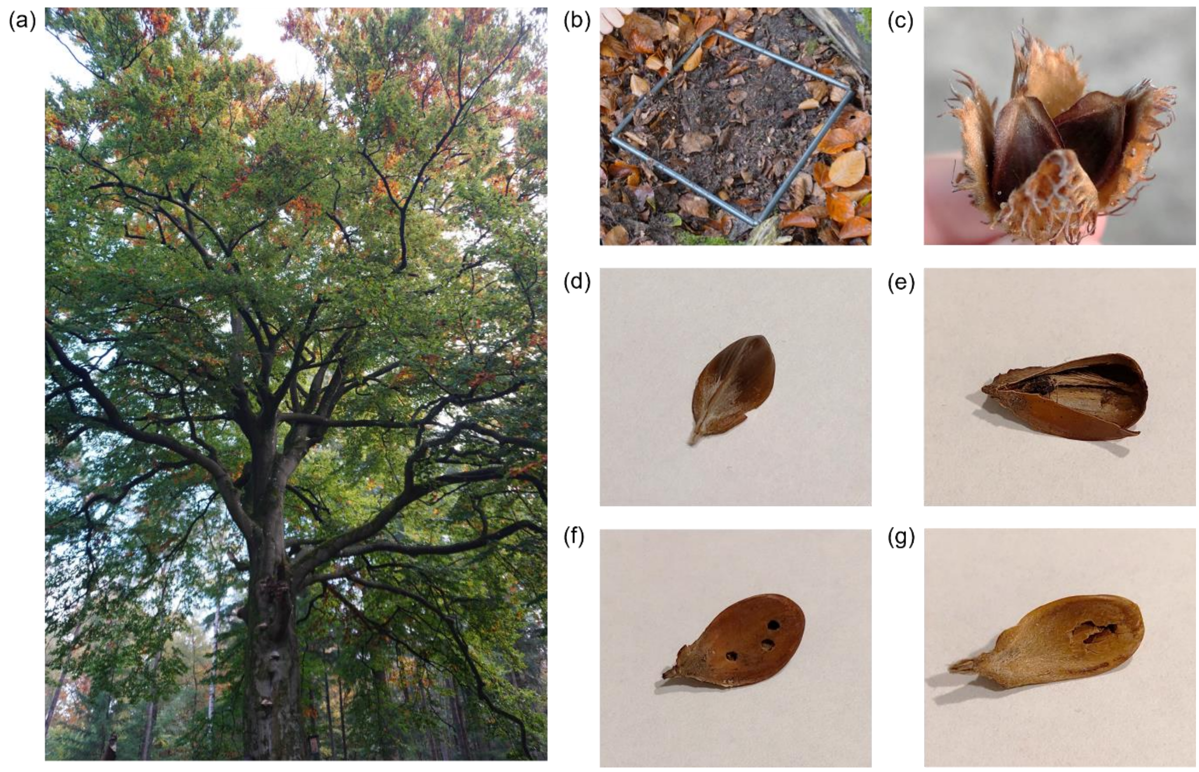


**Figure 2 Field- and lab work of beechnut counts.** (a) The dataset contains data of 81 beech trees (*Fagus sylvatica*). (b) Under each of them, three to four metal squares (25x25 cm) are placed, within which all beechnuts are collected. (c) One cupula (i.e., the spiky outer husk) contains two beechnuts. These cupulas are removed when collecting the nuts. (d-g) In the lab, the collected beechnuts are sorted into different categories: (d) full nuts that are firm and shiny on the outside, (e) empty nuts (here the inside of an empty nut) that are not firm but curved inwards from the outside and empty from the inside, (f) caterpillar nuts that show one or several round emergence holes of caterpillars, and (g) eaten nuts that show other signs of predation, like non-round holes, or signs of pecking by birds.

*Lab work*

All lab work is done per plot, i.e., tree-date-plot specific. Nuts were always sorted into whole nuts (i.e., freshly produced, firm nuts containing a developed, viable seed, Figure 2d), empty nuts (i.e., nuts that are completely empty inside, Figure 2e), predated nuts, which are split into nuts with caterpillars (i.e., showing small round entry holes caused by insects or containing caterpillar eggs when opened, Figure 2f) and nuts predated upon by birds or mammals (i.e., showing non-smooth larger holes, Figure 2g). Additionally, the categories of rotten nuts (i.e., decaying nuts, often from previous years, that contain small, dried-out black seeds) and other nuts (containing mainly pieces of nuts that cannot be put in any other category) have been used, but not in all years.

After drying and sorting, the whole nuts additionally are weighed. First, all whole nuts are weighed together to get the total gross weight and, since 1999, nuts were also weighed individually to get the weight per nut. When no individual weights were taken before 1999, the weight per nut is estimated as the mean, i.e., total weight divided by the number of nuts weighed. In high seed years, the number of nuts that have been weighed together is however sometimes lower than the total number of whole nuts that was found, so the total gross weight of all nuts is calculated. Besides the gross weight, net weights were collected by weighing the nuts after peeling of their outer shell. In more recent years, a maximum of five random nuts is weighed (gross and net weight) individually to reduce processing time. From these data the total net weight per plot can be calculated using the mean ratio between net and gross weight from these five nuts. The number of nuts that have been weighed is indicated in the data by either several entries for single nuts linked to the same plot or only one weight linked to a number of weighed nuts, when more than one nut was weighted together.

### ***Data interpretation and processing***

*Accounting for years with only three plots*

For the early years, in which only three plots have been sampled, the fourth position is calculated as the mean of the three measured plots and floored to get whole counts. These entries have a remark indicating that they were calculated (i.e., "4th position extrapolated from positions 1 to 3"). This is done to make the number of plots per tree consistent, facilitating analyses.

*Calculating net weights*

To calculate the total net weight of all full nuts per year, the total gross weight of all full nuts can be multiplied with the mean ratio between gross and net weight (net/gross) of the individual nuts. This calculation should be done per plot, if all three values (i.e., total gross weight, gross weight per nut, net weight per nut) are available on the plot-level. As this is however not always the case, the net-gross-weight ratio can then be calculated as follows: If either the gross or net weight of individual nuts is missing, another net/gross-ratio has to be used. If there is a ratio for the same tree (calculated from other plots at that date), the average ratio over these other plots should be used. Otherwise, the average ratio of the plots of all other trees in the same year should be used, and if that is also not available, the weighted mean of the ratio over the whole study period (being 0.61). For some plots, the total gross weight is missing. These are however only the plots from additional sampling days outside of the normal collection in mid-October, as nuts have not been weighed for these dates. If net weights for these dates should be calculated, the average net weight per nut of the same tree on other dates in the same year can be multiplied with the number of full nuts.

*Including the correct dates*

For the years where data has been collected on several days (see above), it is important to choose one representative date to be used for analyses that focus on annual beechnut production. Given that the standard sampling date is around mid-October, the dates closest to that should be selected, and it should be a date at which most trees have been sampled. For years, where sampling was split over several days due to time constraints, all these days have to be included. If there were several sampling events in October, the dates on which the largest number of unique trees has been sampled should be included. To facilitate selection of dates, the dataset contains an "Include_dates" table, listing all sampling dates and marking the dates to include for each year (see section "data record").

*Calculation of beech crop index*

The beech crop index (BCI) that has previously been used in studies on this population was introduced by Perdeck et al.[2] to make this dataset integrable with an older time series that consists of visual estimates on a nine-point scale by the Dutch State Forestry Service (SBB). Based on the cyclical behaviour of the beechnut production, with high seeding years followed by low years and some intermediate values, the measured counts of whole nuts per square metre were translated into a three-level index (1 = low, 2 = intermediate and 3 = high). The cut-off points were defined as 0 to 50 nuts/m² for index 1, 50 to 400 nuts/m² for index 2 and > 400 nuts/m² for index 3 (Figure 1). This index is only based on the number of whole nuts, because it was always used in the context of food availability for seed consuming species, for which only filled nuts are important.

If the data are not used together with other seed production data, we recommend to use the original counts instead of the index, as masting patterns are changing with climate change[7,23], so that the cut-off points might no longer be adequate and less suitable to capture fine-scale variation in beechnut production between years.

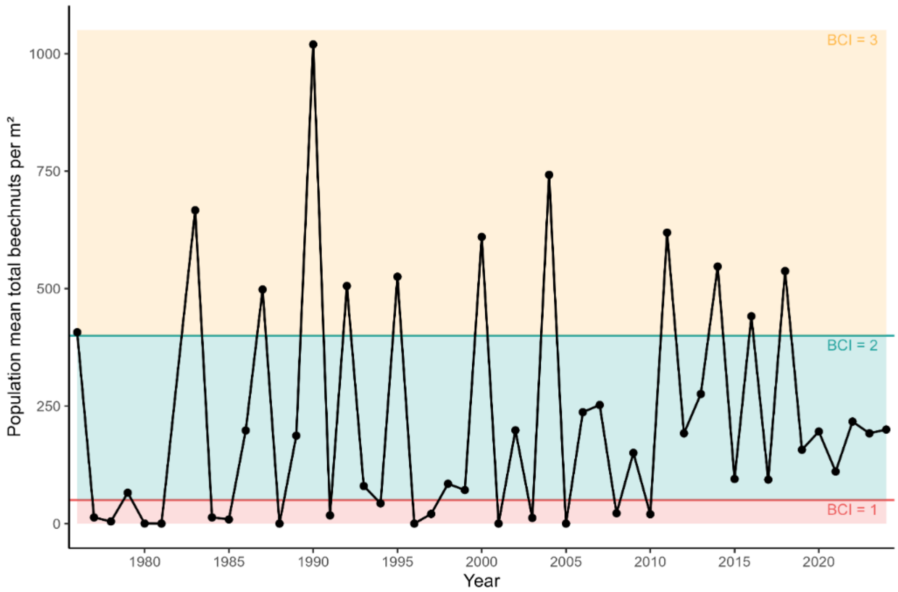


**Figure 3 Cut-off point to calculate the beech crop index (BCI) as defined by Perdeck et al.**[2] The black dots show the population mean of the total number of beechnuts per m² per year. The coloured bands indicate which category of the BCI would be assigned to the years. The cut-off points for each level of the index are 50 nuts/m² for BCI = 1, 400 nuts/m² for BCI = 2 and everything above 400 nuts/m² is BCI = 3.

### *Making data FAIR*

Before sharing the data, we enhanced their reusability by making them compliant with the FAIR (findable, accessible, interoperable, reusable) principles[24], following the workflow described in[25]. This included applying a machine-readable standard to both data and metadata, using ontological terms, where possible, to describe data contents and restructuring the data. For the data, we used Darwin Core[26], a data standard tailored to biodiversity data, while for the metadata we used the Ecological Metadata language (EML)[27]. We additionally, restructured the data into a Darwin Core Archive[28], where data contents are organised in a core file and several extension files, which are linked to the core file (see below for more details). Together, these steps allow for easier integration of the dataset with others, easier understandability and less ambiguity, ultimately enhancing the reusability. As we do not use a knowledge representation language, the data is however not fully FAIR, but FAIR enough to be reused in the ecological community (as discussed in more detail in[25]).

## Data records

The dataset[29] is available under a CC BY-SA 4.0 licence on DataverseNL, a Dutch installation of Dataverse, and contains a total of 81 individual trees and 12.313 plot samples over 50 years. It consists of three tabular data files stored as CSV files using Darwin Core terms as column headers and a separate CSV file that is not in a standard format as a helper file (Include-dates.csv, see below). Following the terminology and structure of a Darwin Core Archive for sampling-event data, the data contents are organised in an event core file (beechcrop_event.csv) and two extension files, occurrence (beechcrop_occurrence.csv) and extended measurement or fact (beechcrop_extendedmeasurementorfact.csv), that are both linked to the core file. The event core stores information about each sampling event and describes a hierarchical event structure. The first and highest event level describes the sampling of an individual tree on a certain date, the second level describes the sampling of one plot of a tree on a certain date and the third and lowest level describes the weighing of individual nuts (or groups of nuts) in the lab. Within the event file, each event is assigned two dates, an eventDate and a verbatimEventDate. While the eventDate describes the calendar day on which the data has been collected, verbatimEventDate contains the year of the season to which this date belongs (if data is collected throughout winter and some collection events are in the following calendar year, they should be assigned to the same year as measurements in October). Additionally, the event core describes details on space, sampling protocol, and sampling effort of each event and contains a unique eventID for each event by which it is linked to the records in both extension files. The occurrence file contains taxonomic and quantitative information of each sampling event and the extended measurement or fact file contains the actual counts and weights of the beechnuts per sampling event. To use this data in analysis, the three tables have to be merged based on their IDs.

Metadata is stored in two files: an XML file (beechcrop_meta.xml) that contains the structural metadata (i.e., how files are linked) and refers to the URI of each Darwin Core term and a second XML file (beechcrop_EML.xml) using EML to describe the remaining metadata, e.g., data collection, provenance, author and contact details.

Additionally, there is a separate CSV file (Include-dates.csv) that is not part of the collected data but facilitates its usage by indicating which dates to include in analyses of annual beechnut production for years with several sampling dates (see "*Including the correct dates"*). It contains all sampling dates, the number of plots that were sampled at that day and a column indicating which dates to include (i.e., records that have a year in the include column should be included in that year).

While the Darwin Core Archive provides a standardised format that facilitates integration with other datasets, it can be difficult to use in analyses. To facilitate data use, the dataset therefore comes with a short R pipeline (pipeline_DarwinCore-Archive-to-analysis-format.R) that converts the Darwin Core Archive structure into a format more easily usable for data analyses. This pipeline excludes re-sampling plots (SampleTypeID = 3) and those sampling dates that should not be included according to the Include-dates table. If the user wishes to include this data, the script needs to be adapted.

## Technical validation

The data collection in the field and the processing of samples follows a standard operating procedure (SOP) and the ground-plot counting method used for this dataset is considered a highly reliable measure of seed production[30]. Intensive data cleaning was performed before publication of the dataset based on the original field books by manually comparing entries in the books and database.

## Usage Notes

To use the data as done in previous studies, the beech crop index has to be calculated based on the mean number of whole nuts produced by the population in a year, using the cut-off points provided. As in some, but not in all years, beechnuts were collected on several days, it is important to choose the correct dates for analyses that use annual counts of beechnuts. Additionally, when modelling the data, it is important to consider that individual-tree beechnut count data is by nature highly zero-inflated, which should be properly accounted for by using, for example, zero-inflated negative binomial models (as done in[7]). To facilitate the use of the data, the dataset contains an R pipeline that converts the Darwin Core Archive format of the data into two data frames directly usable for analysis.

## Funding

This publication is part of the project *The heartbeat of the forest* with file number OCENW.M.22.426 of the research programme NWO Open Competition Domain Science which is financed by the Dutch Research Council (NWO).

**Data availability**

The dataset is available on Dataverse under the DOI: https://doi.org/10.34894/TQY74M.

**Code availability**

The R script converting the data into a different format more easily usable for analysis is stored along the data on DataverseNL[29].

**References**

1. Ostfeld, R. S. & Keesing, F. Pulsed resources and community dynamics of consumers in terrestrial ecosystems. *Trends in Ecology & Evolution* **15**, 232–237 (2000).
2. Perdeck, A. C., Visser, M. E. & van Balen, J. H. Great Tit Parus major survival, and the beech-crop cycle. *Ardea* **88**, 99–108 (2000).
3. Zwolak, R., Celebias, P., Zduniak, M., Bogdziewicz, M. & Wróbel, A. Scatterhoarder abundance and advantages of seed burial drive dynamics of a tree–rodent interaction. *Journal of Ecology* 1365-2745.14356 (2024) doi:10.1111/1365-2745.14356.
4. Kelly, D. The evolutionary ecology of masting. *Trends in Ecology & Evolution* **9**, (1994).
5. Bogdziewicz, M., Zwolak, R. & Crone, E. E. How do vertebrates respond to mast seeding? *Oikos* **125**, 300–307 (2016).
6. van Balen, J. H. A Comparative Study of the Breeding Ecology of the Great Tit Parus major in Different Habitats. *Ardea* **61**, 1–93 (1973).
7. Jantzen, C. C., Burant, J. B., Gamelon, M., Bakker, E. S. & Visser, M. E. Masting breakdown in European beech reduces fitness benefits of masting, partly explained by climate change [Manuscript under review].
8. Szymkowiak, J. *et al.* Tail-dependence of masting synchrony results in continent-wide seed scarcity. *Ecology Letters* **27**, e14474 (2024).
9. Vacchiano, G. *et al.* Spatial patterns and broad-scale weather cues of beech mast seeding in Europe. *New Phytologist* **215**, 595–608 (2017).

10. Hacket-Pain, A. *et al.* MASTREE+: Time-series of plant reproductive effort from six continents. *Global Change Biology* **28**, 3066–3082 (2022).
11. Kleef, H. L. & Wijsman, H. J. W. Mast, mice and pine marten (Martes martes): the pine marten's reproductive response to wood mouse (Apodemus sylvaticus) fluctuations in the Netherlands. *Lutra* **58**, 23–33 (2015).
12. Schmidt, K. A. & Ostfeld, R. S. Numerical And Behavioral Effects Within A Pulse-Driven System: Consequences For Shared Prey. *Ecology* **89**, 635–646 (2008).
13. Foest, J. J. *et al.* No Refuge at the Edge for European Beech as Climate Warming Disproportionately Reduces Masting at Colder Margins. *Ecology Letters* **28**, e70284 (2025).
14. Journé, V., Hacket-Pain, A., Oberklammer, I., Pesendorfer, M. B. & Bogdziewicz, M. Forecasting seed production in perennial plants: identifying challenges and charting a path forward. *New Phytologist* **239**, 466–476 (2023).
15. Gamelon, M., Jenouvrier, S., Lindner, M., Sæther, B.-E. & Visser, M. E. Detecting climate signals cascading through levels of biological organization. *Nat. Clim. Chang.* **13**, 985–989 (2023).
16. Reed, T. E., Jenouvrier, S. & Visser, M. E. Phenological mismatch strongly affects individual fitness but not population demography in a woodland passerine. *Journal of Animal Ecology* **82**, 131–144 (2013).
17. Ramakers, J. J. C., Gienapp, P. & Visser, M. E. Comparing two measures of phenological synchrony in a predator–prey interaction: Simpler works better. *Journal of Animal Ecology* **89**, 745–756 (2020).
18. Grøtan, V. *et al.* Spatial and temporal variation in the relative contribution of density dependence, climate variation and migration to fluctuations in the size of great tit populations. *Journal of Animal Ecology* **78**, 447–459 (2009).
19. Sæther, B. *et al.* The extended Moran effect and large-scale synchronous fluctuations in the size of great tit and blue tit populations. *Journal of Animal Ecology* **76**, 315–325 (2007).
20. Jantzen, C. C. & Visser, M. E. Climate change does not equally affect temporal patterns of natural selection on reproductive timing across populations in two songbird species. *Proc. R. Soc. B.* **290**, (2023).
21. Visser, M. E., Lindner, M., Gienapp, P., Long, M. C. & Jenouvrier, S. Recent natural variability in global warming weakened phenological mismatch and selection on seasonal timing in great tits (Parus major). *Proceedings of the Royal Society B: Biological Sciences* **288**, (2021).

22. Reed, T. E., Grøtan, V., Jenouvrier, S., Sæther, B.-E. & Visser, M. E. Population Growth in a Wild Birds Is Buffered Against Phenological Mismatch. *Science* **340**, 488–491 (2013).

23. Bogdziewicz, M., Kelly, D., Thomas, P. A., Lageard, J. G. A. & Hacket-Pain, A. Climate warming disrupts mast seeding and its fitness benefits in European beech. *Nature Plants* **6**, 88–94 (2020).

24. Wilkinson, M. D. *et al.* The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data* **3**, (2016).

25. Jantzen, C. C. & Vriend, S. J. G. Putting FAIR into practice for ecologists: how to make ecological data more reusable [Manuscript under review].

26. Wieczorek, J. *et al.* Darwin Core: An Evolving Community-Developed Biodiversity Data Standard. *PLoS ONE* **7**, e29715 (2012).

27. Fegraus, E. H., Andelman, S., Jones, M. B. & Schildhauer, M. Maximizing the Value of Ecological Data with Structured Metadata: An Introduction to Ecological Metadata Language (EML) and Principles for Metadata Creation. *Bulletin of the Ecological Society of America* **86**, 158–168 (2005).

28. GBIF. Darwin Core Archives – How-to Guide, version 2.2. Copenhagen: GBIF Secretariat. https://ipt.gbif.org/manual/en/ipt/3.0/dwca-guide (2021).

29. Netherlands Institute of Ecology (NIOO-KNAW) - Department of Animal Ecology. Long-term seed production data of European beeches in the National Park De Hoge Veluwe, Netherlands. DataverseNL https://doi.org/10.34894/TQY74M (2026).

30. Touzot, L. *et al.* The ground plot counting method: A valid and reliable assessment tool for quantifying seed production in temperate oak forests? *Forest Ecology and Management* **430**, 143–149 (2018).