

## RESEARCH ARTICLE

# Fawn birthdays: From opportunistically sampled fawn rescue data to true wildlife demographic parameters

Johanna Kauffert<sup>1</sup>  | Sophie Baur<sup>1,2</sup>  | Michael Matiu<sup>3</sup>  | Andreas König<sup>4</sup>  |  
Wibke Peters<sup>2,4</sup>  | Annette Menzel<sup>1,5</sup> 

<sup>1</sup>Ecoclimatology, TUM School of Life Sciences, Technical University of Munich, Hans-Carl-von-Carlowitz-Platz 2, D-85354, Freising, Germany

<sup>2</sup>Bavarian State Institute of Forestry (LWF), Hans-Carl-von-Carlowitz-Platz 1, D-85354, Freising, Germany

<sup>3</sup>Department of Civil, Environmental and Mechanical Engineering, University of Trento, I-38122, Trento, Italy

<sup>4</sup>Wildlife Biology and Management Unit, TUM School of Life Sciences, Technical University of Munich, Hans-Carl-von-Carlowitz-Platz 2, D-85354, Freising, Germany

<sup>5</sup>Institute for Advanced Study, Technical University of Munich, Lichtenbergstraße 2a, D-85748, Garching, Germany

**Correspondence**

Johanna Kauffert  
Email: [johanna.kauffert@tum.de](mailto:johanna.kauffert@tum.de)

**Funding information**

Bavarian State Ministry of Food, Agriculture, and Forestry, Grant/Award Number: A/19/17

**Handling Editor:** Jim Vafidis

**Abstract**

1. Spring mowing in May and June is one of the main causes of mortality of roe deer fawns in agricultural regions. Knowing the exact birth distribution of fawns is important to guide farmers in their pre-mowing precautions to avoid fawn deaths.
2. Wildlife volunteers searching fields prior to mowing can act as citizen scientists by producing data sets of rescued fawns and their approximate age at find. However, due to weather-dependent searches, the corresponding birth distributions can be highly skewed. We simulated virtual field data to examine the shortcomings of such data sources and introduced two algorithms for reconstructing reliable birth distribution parameters (mean and standard deviation) based on skewed samples.
3. We found that weather-dependent search data biased the calculated means and standard deviations by up to 14 and 5 days, respectively. However, the use of the proposed advanced algorithms (Grid Search and Machine Learning) resulted in better estimates of the sample means and standard deviations by reducing the root-mean-square error by 65% and 80% respectively. Furthermore, the Grid Search algorithm was able to capture birth distribution parameters based on real citizen science data in Bavaria, Germany, from 2021, which are close to the results of more systematic samples of the same year.
4. The simulation exercise highlighted the shortcomings and discrepancies of using non-probabilistic samples, for example on the occasion of mowing activities, to study demographic parameters compared to the true simulated distribution. Yet, the proposed algorithms can address these drawbacks and potentially turn citizen science data into an important data source for wildlife studies. This could ultimately help reduce wildlife losses during the mowing season by better knowing the distribution of births in a region.

**KEYWORDS**

*Capreolus capreolus*, mowing death, non-probabilistic sampling, parturition timing, simulation

This is an open access article under the terms of the [Creative Commons Attribution](https://creativecommons.org/licenses/by/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2023 The Authors. *Ecological Solutions and Evidence* published by John Wiley & Sons Ltd on behalf of British Ecological Society.

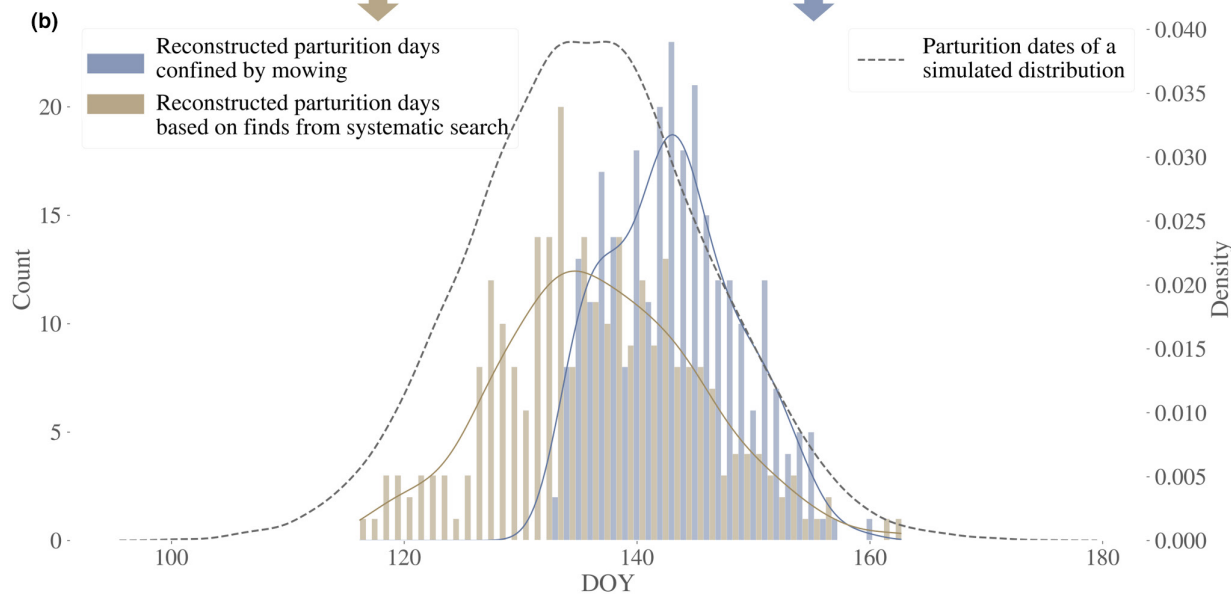
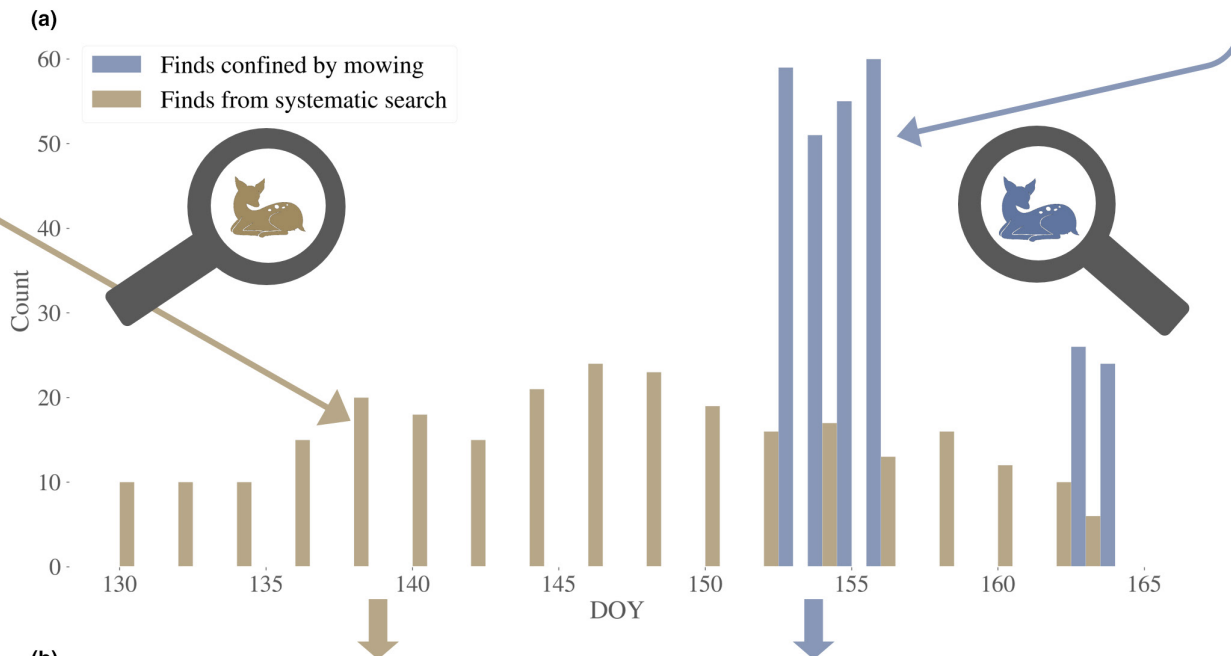
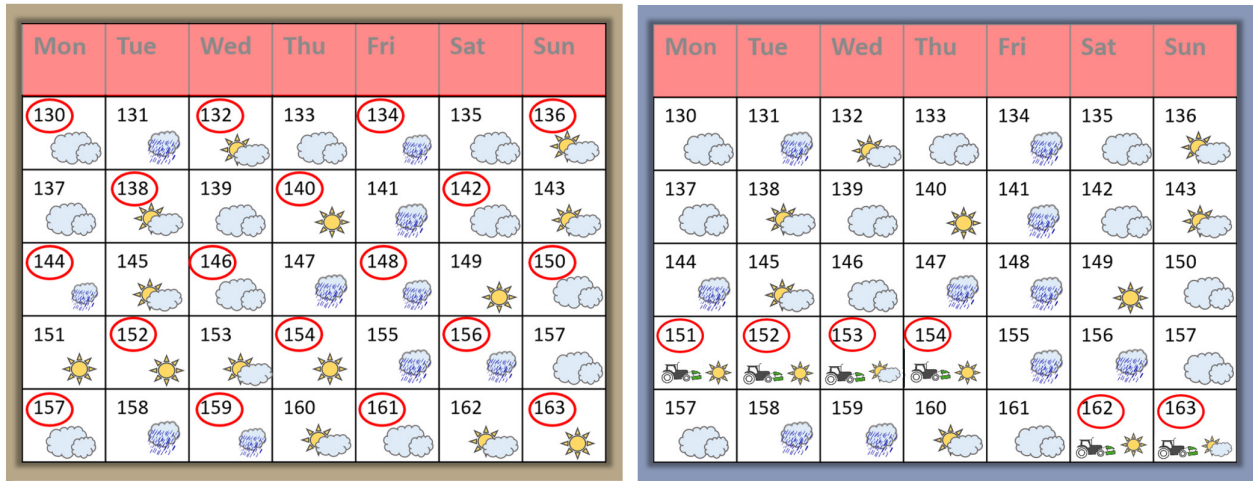
## 1 | INTRODUCTION

In temperate regions with intensive agriculture, spring mowing in May and June forms a great danger to many species inhabiting agricultural environments such as ground-nesting bird species like Corncrakes *Crex crex* (Tyler et al., 1998), pheasants (Deák et al., 2021) or grey partridge *Perdix perdix* (Kittler, 1979), leverets of brown hare *Lepus europaeus* (Kałuziński & Pielowski, 1976) as well as roe deer fawns *Capreolus capreolus* L. (Jarnemo, 2002; Kałuziński, 1982; Kittler, 1979). Particularly for roe deer fawns, mowing is one of the main causes of mortality. Roe deer fawns follow a hiding strategy for up to 2 months (Linnell, 1994) to protect themselves from predators and stay mostly hidden and secluded from their mothers (Jarnemo, 2004; Kurt, 1968; Lent, 1974). Therefore, particularly grassland and meadows are preferred bedding sites for roe deer due to their high seclusion and food abundance (Christen et al., 2018; Linnell et al., 1999, 2004; Panzacchi et al., 2010). Concurrent with parturition and lactation periods of roe deer, farmers begin with the first hay cut, usually in May, when grass protein content is at its highest. Due to the fawns' native instinct to stay hidden, the animals are at great risk from approaching machinery and many get killed (Jarnemo, 2002; Linnell, 1994). Jarnemo (2002) estimated for south central Sweden that over 3 years about 31% ( $n=54$ ) of new-born fawns are killed by mowing, while Kittler (1979) approximated 14.5% ( $n=3755$ ) of the game taken in North-Rhine Westphalia is lost due to mowing and Rehnus and Reimoser (2014) recorded that 19.5% of 930 reported deaths of marked fawns can be related to mowing activities in Switzerland. Not only in terms of wildlife welfare and conservation but also from an economic point of view, reducing mowing death is desirable as undetected animal carcasses can contaminate fodder silage for cattle (Driehuis et al., 2018; Israel, 2011; Jarnemo, 2002; Moeller Jr. & Puschner, 2007). In order to reduce the risk of mowing death of field-dwelling animals, conservational actions such as delaying mowing or adjusting mowing techniques have been found beneficial (Dicks et al., 2020) but are not always practicable. With respect to roe deer, several methods have been tested, such as antagonizing animals with scaring devices or searching fields before mowing with drones or volunteers (Cukor et al., 2019; Israel, 2011; Jarnemo, 2002). Despite high efforts by farmers, hunters, gamekeepers and volunteers, not all fields can be searched at the same time, as the time window for the first mowing of silage or hay is often restricted to a few days with favourable weather conditions. Furthermore, the limited knowledge about possible (temporal) plasticity of the underlying birth distribution of roe deer fawns and its corresponding (spatial) variations make effective and targeted searches difficult under climate warming. Particularly, the possible

plasticity of breeding phenology of roe deer to environmental cues or climate change is not yet fully understood. Although, information on demography, plasticity, and adaptability to climate change are of great interest (Gaillard et al., 2013; Marchand et al., 2021; Plard et al., 2013, 2014). For example, synchrony of reproduction and variability of peak parturition are thought to be responsive to environmental stimuli, but studies have produced conflicting results (Gaillard et al., 2013; Hagen et al., 2021; Plard et al., 2013, 2014; Rehnus et al., 2020). While Plard et al. (2014) were not able to detect a significant shift in birth timing in response to advancing plant phenology in Trois Fontaine, France, a fenced lowland area, Rehnus et al. (2020) reported a slight trend towards earlier births across different regions and most elevations in Switzerland and so did Hagen et al. (2021) in Baden-Württemberg, Germany, equally a very heterogeneous landscape. Hence, only regular (scientific) field survey data about parturition timing with high amounts of individuals in different areas and from free-ranging populations can lead to a higher validity of results (Rehnus et al., 2018, 2020; Weiser et al., 2020), however, these systematic surveys are expensive and thus rare.

In order to overcome this data deficit, observations of wildlife volunteers when searching for fawns in May and June, reported for example via online tools, may generate valuable records. However, data collected by these wildlife volunteers, like in our study in Bavaria, may have some inadequacies due to special circumstances of the fawn rescue search, such as non-random and opportunistic selection of locations for observations (Brown & Williams, 2019; Weiser et al., 2020) and biased sampling dates. While traditional study designs attempt to sample regularly, independently and randomly from the true population (structured survey with a probability sample design; Conn et al., 2016), wildlife volunteers involved in fawn rescue search only on mowing days and consequently, the sample dates are confined to those mowing dates. The time window for the first grassland harvest is again determined by the optimal quality and output of hay for forage along with periods of dry weather (Figure 1; Boob et al., 2019). Hence, fawns are searched by occasion and then sampled on a few, often consecutive days within a region, determined by the farmers' choices of mowing. This type of sampling falls under a non-probability sampling technique since not all fawns have the same probability of being found (Etikan et al., 2016; Fink, 2003). Only fawns being already born on the mowing days, young enough to be caught and laying in the particular fields are included which collates this particular case to convenience sampling. In contrast to probability sampling, the probabilities of errors and biases, such as temporal truncations (see Figure 1), are not known and the representatives of the sample cannot be quantified (Leiner, 2016). Therefore, inferring moments (population mean  $\mu$  or standard deviation  $\sigma$ ) of a

**FIGURE 1** Illustrative plots of the study's concept: Comparison of the effects of 275 simulated systematic searches (light brown) and 275 simulated by occasion searches (blue, confined to mowing dates) on (a) histograms of finds and (b) the reconstructed parturition timing and birth date distribution with the assumption that fawns older than 25 cannot be found anymore and age calculation of older fawns is error-prone. Day of the year (DOY) 130 in the calendar represents May 10th. Red circles indicate days of search. Reconstruction of birth dates (b) from finds (a) is based on age estimates of the fawns. The dashed line in (b) indicates the likely true distribution (virtual field data) of parturition dates.



distribution that are based on the sample's mean ( $\bar{x}$ ) and sample's standard deviation ( $s$ ) of a non-probability sample can lead to false conclusions. Similarly, due to the density of fawns across spring and year-by-year variations, a systematic search can also not be accounted for as a true probability sample design but is considered to be more reliable.

Some proposals have been made regarding the handling of truncated and non-probability sample data, like expanding the convenience sample by a smaller probability sample that potentially remedies the shortcomings and leads to higher validity (Brus & De Grijter, 2003). Speak et al. (2018) proposed to stratify a convenience sample by known categories like land cover types, while Elliott (2009) suggested pseudo-weights to expand a probability sample by a non-probability sample. Most research concentrates on examining the effects of non-probabilistic sampling within animal abundance and spatial distribution studies (Conn et al., 2017; Dambly et al., 2021; Weiser et al., 2020).

However, biases introduced by the time window of by-occasion observations are less in focus. Inferring precise information about breeding phenology poses unforeseen challenges (Moussus et al., 2010). Caughley and Caughley (1974) proposed probit-analysis to estimate the median birth date in impalas *Aepyceros melampus* by monitoring the ratio of young to females at different times. Johnson et al. (2004) dealt with issues of coarse scales in measurements and came up with a generalization of Sheppard's correction to predict the breeding variance of Dall's sheep *Ovis dalli*.

Tackling the knowledge gaps regarding phenological asynchrony in roe deer and spring green-up introduced by climate change, we introduce novel methods to infer reliable demographic birth parameters. This knowledge about the exact birth distribution of roe deer fawns in a region can have practical implications for the farmers and wildlife managers in terms of intensified searches during the peak period of breeding. In principle, the methods could be also of value to any other opportunistic environmental sampling, for example by citizen scientists.

In this paper, we study how well we can estimate the birth distribution of roe deer when the fawns are found based on a convenience sample driven by mowing dates and thus indirectly by weather conditions (see Figure 1). We develop and test algorithms to counteract systematic omissions resulting from convenience sampling. Further, we apply the proposed methods to field data and citizen science data from Bavaria, Germany in 2021. Finally, we evaluate the benefits and drawbacks of integrating citizen science data with their inherited biases and test if flowering dates of meadow foxtail *Alopecurus pratensis* could add value to birth distribution approximations.

## 2 | MATERIALS AND METHODS

The underlying data for our study are 2500 virtual fawn seasons with simulated distributions of fawn births with varying but known mean and standard deviation, upon which we sampled parturition events biased by various aspects of convenience sampling (see

Section 2.1). We compared two algorithms to infer better estimates of the first two moments (population's mean ( $\mu$ ) and population's standard deviation ( $\sigma$ )) of these birth distributions. First, an iterative approach was applied where we set up a Grid Search seeking the first two moments that fit the field data best (see Section 2.2). Second, we used machine learning algorithms (ML) to infer the mean with Support Vector Regression (SVR) and standard deviation with Gradient Boosting Regression (GBR; see Section 2.3). We compared our estimates to the true simulated moments as well as to the calculated sample's mean ( $\bar{x}$ ) and sample's standard deviation ( $s$ ). All computations were performed in Python 3.8.6 (see code Kauffert et al., 2023a). Algorithms were tested on multiple simulated test data (see Section 2.1) as well as on real data from a case study in Bavaria, Germany in 2021 (see Section 2.4 and fawn data: Kauffert et al., 2023b). We also checked whether plant phenological data may support fawn birth estimates (see Section 2.5).

### 2.1 | Simulating virtual field data

We generated virtual field data accounting for different mowing times as well as varying parameters of birth distributions. Simulations included in total 2500 fawn seasons varying in the mean and standard deviation of a Gaussian distribution of birth dates and 23 mowing scenarios. The resulting virtual field data was based on the following assumptions (see Table 1). Birth date distribution characteristics such as the mean or the standard deviation for simulations have been sampled from uniform distributions allowing only integer values and their ranges reflected estimated published literature values (Ellenberg, 1978; Hagen et al., 2021; Linnell & Andersen, 1998; Raganella-Pelliccioni et al., 2007). We used 23 different mowing scenarios which are characterized by different weather regimes beginning with the 25th of April and lasting to the 5th of July (range is within dates of first hay cut according to the German Meteorological Service [DWD: Deutscher Wetterdienst] Bock et al., 2013). Thus, mowing scenarios varied in the number of days and the number of consecutive day chunks. We evaluated their main effects by highlighting four representative mowing scenarios which describe (1) early mowing (Day of Year (DoY): 124–130), (2) midway mowing (DoY: 135–141), (3) late mowing (DoY: 155–156, 167–170) and (4) distributed mowing (DoY: 119–121, 138–141, 157–158).

We also considered the estimated fawn's age and its maximum possible age at find. A correct estimation of the fawn's age is challenging, especially if done by untrained volunteers. Even though the fawn's age was determined by attributes of behaviour, appearance (colour, coat pattern) and vocalization, age estimation can still be erroneous, in particular for older fawns (Rehnu et al., 2018; Stamm et al., 2017). However, age estimations can be expected to match actual births better if parameters such as weight, hindfoot, body length and umbilical cord appearance can be additionally included by trained staff (Galli et al., 2008; Sams et al., 1996). Most accurate age approximations can be yielded by monitoring habitat use and

**TABLE 1** Assumptions in simulating birth distribution and virtual field data with three scenarios (S1, S2 and S3) in age estimation and maximum age at the time of finding.

Parameter	V	Value	Reference
Distribution		Gaussian distribution	Gaillard et al. (1993), Linnell and Andersen (1998) and Ellenberg (1978)
Mean birth date		DoY: 130–144	Mean 143 in Bavaria 1970s (Ellenberg, 1978); mean 143 in Baden-Wuerttemberg (1973–2019), however, the advance of parturition dates between 0.16 and 0.33 d/year (Hagen et al., 2021); high variability in Europe (Linnell & Andersen, 1998)
Standard deviation		Days: 5–12	9.5 days (Hagen et al., 2021); high variability in Europe (Linnell & Andersen, 1998; Raganella-Pelliccioni et al., 2007)
Fawns max. age at time of finding	S1/S2 S3	25 days No age restriction	Jarnemo (2002)
Age estimation error	S1/S3 S2	None Age: <7 days: $\pm 2$ ; age: 8–14 days $\pm 6$ ; age: >15 days $\pm 5$	The age of recently born fawns can better be estimated than older fawns (Rehnus et al., 2018)
Mowing period		DoY: 115–185	Different combinations of mowing regimes, varying in length, timing and number of consecutive days. Based on phenological data from the German Meteorological Service (Deutscher Wetterdienst [DWD]; first cut for silage in Germany)

activity rates of GPS-collared mothers in conjunction with the aforementioned methods (Marchand et al., 2021).

At the age of 10 days, fawns start fleeing from an approaching human with a distance of less than 1 m (Jarnemo, 2002) but are still catchable. However, fawns aged 20 to 30 days are fleeing at a human distance of 1–2 m (Jarnemo, 2002), hence catching becomes more difficult. Thus, we used 25 days as an upper limit for age that a fawn can be found and aged appropriately.

To accommodate these restraints for our simulations, we tested algorithms based on three scenarios (S1, S2 and S3). Scenarios 1 (S1) and 2 (S2) assumed that only fawns younger than 25 days can be found and aged if born before the respective mowing day. This restriction may lead to left-truncated birth distributions. In scenario 2, we additionally assumed that fawns' age estimations are erroneous and depend on age (see Table 1). In contrast, in scenario 3 (S3), we did not apply an age estimation error and allowed fawns of any age to be found (no maximal age—no left truncation). Hence, S3 represents an ideal scenario, whereas S2 presumably represents reality best.

We sampled 500 parturition events from each of the 2500 simulated virtual field data seasons given the mowing days and the aforementioned restrictions and performed a sensitivity analysis by

increasing the sample size to test the robustness of our results (see Figure S1). Probability distributions of the samples (500 parturition events) and the virtual field data with the known birth distribution were estimated in a non-parametric way by using kernel density estimation with a Gaussian kernel.

## 2.2 | Grid Search algorithm

Inferring the first two parameters, the mean ( $\mu$ ) and standard deviation ( $\sigma$ ), based on a not necessarily normal, truncated sample forms a nonlinear inverse problem. Since the possible values of  $\mu$  and  $\sigma$  can be restricted to only a range (e.g.  $\mu$  must be between DoY: 130 and 144) due to prior knowledge and the mowing days are given, a Grid Search algorithm can be used to exhaustively iterate over every possible combination of the selected ranges of  $\mu$  and  $\sigma$  and compare each outcome to the sample (Menke, 2012). In this way, the combination of the first two parameters was reiteratively updated to match the sample best. This kind of procedure is commonly used during hyperparameter tuning in machine and deep learning algorithms to find the best parameters for training (Alibrahim &

Ludwig, 2021). The weakness of common Grid Search algorithms lies in the prior determination of parameter ranges and a limited number of parameters. Testing many different parameters and/or defining broad ranges results in high numbers of possible combinations and is thus computationally inefficient and time-consuming (Alibrahim & Ludwig, 2021).

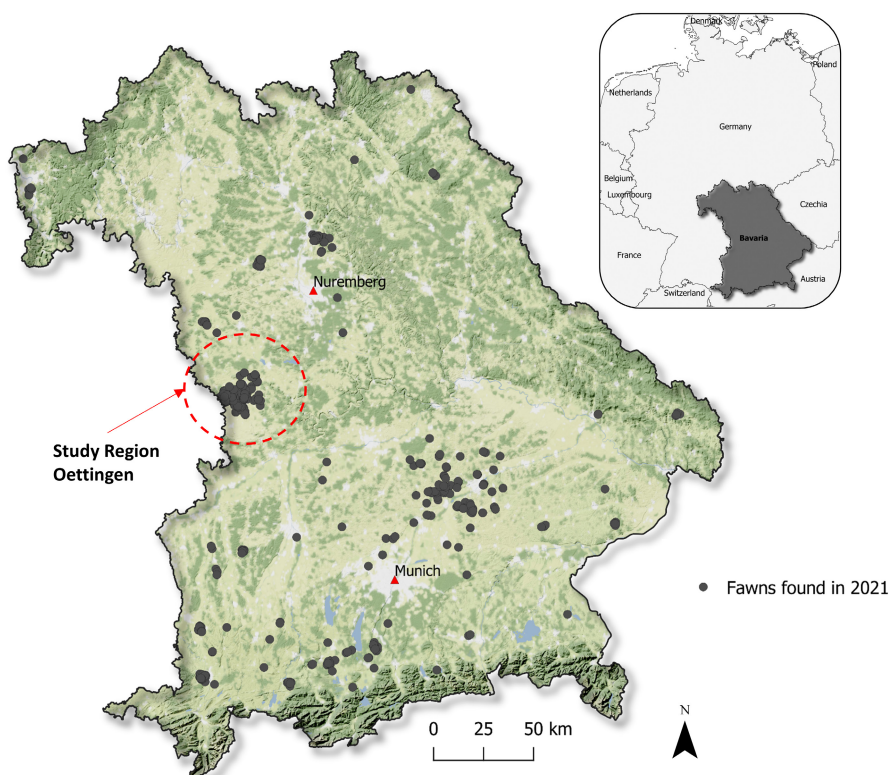
### 2.3 | Machine Learning algorithms

The second approach included supervised ML methods to infer the first two parameters of the underlying distribution. Estimations of ML approaches were not only based on the calculated birth date of fawns but also used the fawn's age at the day of find. The composition of the ages possibly added further information for the models to make better predictions by exploiting the relationship between birth date distribution and age distribution. The day of birth and the corresponding age were zipped together into arrays storing all observations of the iteration (Harris et al., 2020). The data set was then split into training (70%) and testing (30%). ML was performed with scikit-learn in Python 3.8.6 (Pedregosa et al., 2011). Used and presented supervised ML algorithms were selected based on the best accuracies of a broad comparison of models and algorithms. Estimations of  $\mu$  were predicted with a Support Vector Regression. SVR is a generalization of the popular Support Vector Machines (SVM), which are widely used in many fields (Awad & Khanna, 2015). In contrast to classification problems (like in SVM) where the output is restricted to a finite and determined set, the output of regression problems (here SVR) is allowed to return a continuous-value output.

Within SVM, an optimal hyperplane is fitted to the training points with support vectors (Vapnik et al., 1996). To make this algorithm applicable to regression problems, an alternative loss function that includes a distance measure is introduced to the model (Awad & Khanna, 2015). Gradient Boosting Regression was used to infer the standard deviation of the underlying distribution. GBR, and boosted regression trees in general, are tree-based methods yielding high prediction accuracy by strategically combining and assembling simple tree models and thus forming an additive regression model (Elith et al., 2008; Friedman, 2001). Further, we made use of bagging regressors to perform bootstrapping on ML algorithms to improve accuracy (Breiman, 1996).

### 2.4 | Case study

In contrast to our simulated virtual field data, true moments of the population are not known for real case studies but are always an approximation of the truth (Moussus et al., 2010). Nonetheless, algorithms are meant to be tested and examined on field data. Thus, as a case study, we used data from our own research teams and data provided by volunteers and citizen science projects in Bavaria in 2021 (see Figure 2). For this, we set up an online form (LimeSurvey Project Team, 2022) for volunteers to record observations of roe deer fawns in Bavaria. Additionally, we used data from an established tool by the Bavarian State Ministry of Food, Agriculture and Forestry (Wildtierportal—<https://www.wildtierportal.bayern.de/>), which already had a good reputation and traffic among gamekeepers and hunters. Both forms contained questions about the date of



**FIGURE 2** Locations of fawns reported in Bavaria, Germany, in 2021 and localisation of study region Oettingen.

the find, whereabouts as well as features of the fawn. Mowing days were approximated by periods of dry weather in Bavaria and an increased number of reported fawns. The occurrence of false positive observations was neglected since fawns are handled in person by volunteers, and there are low chances of confusion between other species (Ruiz-Gutierrez et al., 2016).

Altogether, in 2021, 813 fawns were reported from which 458 fawns have been found in association with mowing (from fawn rescues), whereas 355 have been found irrespective of mowing (from own research teams). Only fawns with sufficient attributes regarding behaviour, appearance (colour, coat pattern) and vocalization have been used in analysis to validate age estimations. For better validation, the study region of Oettingen was searched intensively from the end of April to late June 2021 (see Figure 2), producing a data set with a high number of observations irrespective of mowing (115/309 on mowing days). This rather randomly selected data set provided a good baseline for comparisons of inferred moments with respect and irrespective of mowing. Age estimations of the data set in Oettingen can be expected to match actual births better since attributes such as weight, hindfoot, body length and umbilical cord appearance extended estimations of a fawn's age (Galli et al., 2008; Jullien et al., 1992; Sams et al., 1996). Additionally, 19 roe deer females were captured with box and up-net traps, handled and fitted with a Lotek LiteTrack 360 collar (Lotek Wireless, Newmarket, Ontario, Canada) in Oettingen between January and March 2021. The collars were scheduled to take GPS positions every 15 min during the fawning season. Visual inspection of the transmitted data was carried out daily. We searched for the fawns on a frequent basis and as soon as a change in the females' space use was visible. The age estimate of the fawns was calculated following a weighted combination of Jullien et al. (1992) and Rehnus et al. (2018). Visual appearance was connected to the change of female space-use to determine the day of birth. Age estimation of fawns matching GPS-collared mother does was expected to be the most accurate. Therefore, we assumed that the actual birth peak in Oettingen should match the mean of the aforementioned data sets closely. Ethical standards for capturing and handling females and fawns were in agreement with German law (capture permit RUF-55.2.2-2532-2-1160-25 by the government of lower Franconia, Germany [18.11.2020]).

**TABLE 2** Root mean square error of estimations of mean and standard deviation of fawn birth based on simulated distributions with 23 mowing scenarios ( $n = 2500$ ).

RMSE of	S1		S2		S3	
	Mean	Standard deviation	Mean	Standard deviation	Mean	Standard deviation
Virtual field data	4.845	2.393	4.845	2.127	4.389	1.588
Grid search	2.731	1.135	2.597	1.155	2.315	0.858
ML	2.158	1.596	2.166	1.647	1.785	1.270

Note: S1, S2 and S3 represent different scenarios of restrictions: S1 with max. age of 25 days and no age estimation error, S2 with max. age of 25 days and age estimation error, S3 with no age restriction and no age estimation error (see Table 1).

## 2.5 | Link to plant phenology

Lastly, we tested if parturition dates can be associated with flowering dates of meadow foxtail across space. It is assumed that birth phenology follows Hopkins' Bioclimatic Law (correspondence of birth distribution to latitude and elevation), but Peláez et al. (2020) also remarked, that parturition seems to be more likely driven by phenology than by elevation exclusively. Similar observations were made in Norway, where fawns were born on the island of Storfosna on average 2 weeks earlier than in the more southern county of Hedmark (Linnell & Andersen, 1998). According to these findings, we related flowering dates of meadow foxtail from the German Meteorological Service (DWD) in 2021 to our parturition dates and sites as well as elevation (mean elevation of fawn locations: 501.3 m, standard deviation: 128.4 m, SRTM 30 m resolution). Flowering dates were provided as an interpolated raster with a spatial resolution of 1000 m based on phenological observations in Germany (compare Yuan et al., 2021). We sampled raster values of interpolated flowering dates of 2021 and elevation to the locations (points) of the found fawns. We tested for linear relationships and examined the environmental conditions of the study region in Oettingen to the remaining parts in Bavaria.

## 3 | RESULTS

### 3.1 | Simulation study

The analysis of the 2500 samples of simulated virtual field data seasons (see Section 2.1) clearly indicated that incomplete field data would lead to considerable errors when estimating the population's mean ( $\mu$ ) and its standard deviation ( $\sigma$ ) solely based on the averages of the sample. The calculated sample's mean ( $\bar{x}$ ) and sample's standard deviation ( $s$ ) diverged substantially from the simulated true  $\mu$  and  $\sigma$  with a root-mean-square error (RMSE) of more than 4.845 days for the mean (S1 and S2) and RMSE of 2.393 days for the standard deviation (S1). Only scenario 3 (no age restriction, no age estimation error, thus optimal survey conditions) yielded a slightly smaller RMSE for the mean with 4.389 days and for the standard deviation with 1.588 days (Table 2).

For all three simulation scenarios, S1–S3, Grid Search and ML algorithms were able to yield better results in estimating  $\mu$  and  $\sigma$  than the respective averaged virtual field samples (see Table 2).

ML was superior concerning estimations of  $\mu$  of the virtual field data, while Grid Search improved the estimations of  $\sigma$ . Further, results of S2 suggested that an error in age estimation did not necessarily impact the estimation's quality. If fawns of any age were allowed to be found (S3), meaning no left truncation, estimations of algorithms as well as of averaged field data improved.

Selected examples of our simulation study (Figure 3) underlined for one identical mowing scenario (DoY: 128–133, 161–162) that both assumptions on age estimation error and maximum find age (S1–S3, columns), as well as fitting algorithms (Grid Search, ML, rows) mattered. Grid Search performed slightly better in S1 than

ML algorithms and estimated  $\mu$  to be DoY: 135. The sample's  $\bar{x}$  was almost 5 days earlier. Estimations for  $\sigma$  matched the underlying distribution closely for Grid Search ( $\sigma$ : 9), ML algorithms ( $\sigma$ : 10) and based on the sample ( $s$ : 10.5). With imperfect age estimation (S2) Grid Search algorithms might deliver too early birth dates, whereas ML algorithms can cope with imperfect age estimates and provide better estimates for V2. With respect to V3 (no restrictions), Grid Search was out by 1 day for the mean, while ML algorithms yielded very close results for  $\mu$  as well. The sample's  $\bar{x}$  was off by 5.5 days. Again, results for  $\sigma$  of Grid Search ( $\sigma$ : 10), ML ( $\sigma$ : 10.5) and the reconstructed sample ( $s$ : 9.6) matched the true value closely.

For a better understanding of the impact of specific mowing scenarios, RMSEs of representative scenarios are displayed in Table 3. Representative scenarios that mirror early, midway, late

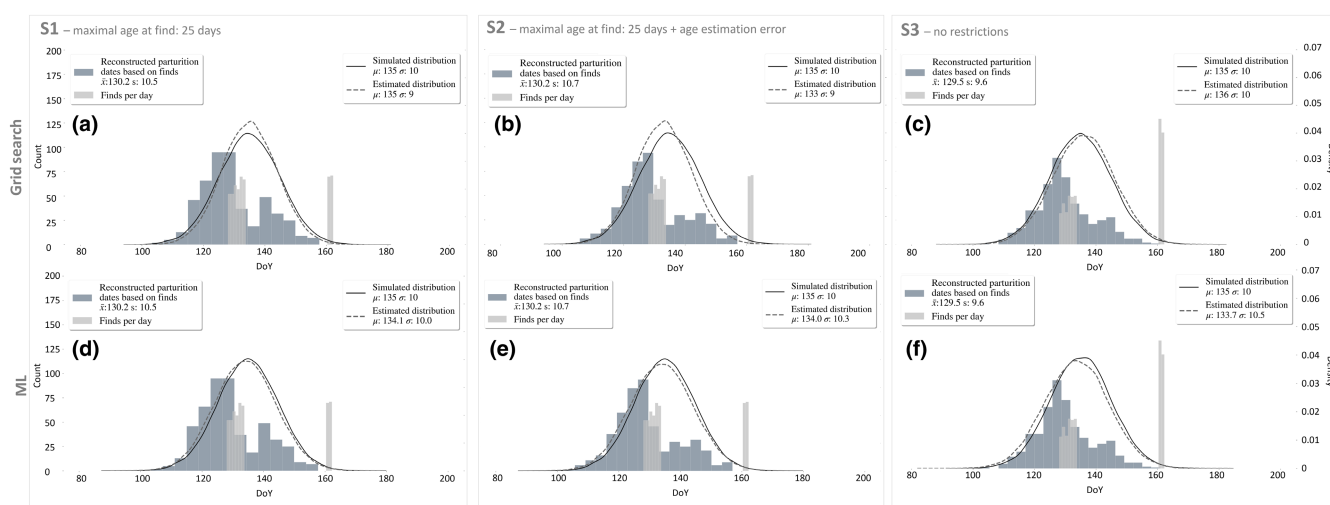


FIGURE 3 Example for one simulated find scenario and its reconstructed parturition dates of found fawns and distribution moments based on three scenarios of age estimation (S1–S3) and two algorithms (Grid Search, ML). All results are based on a specific simulated birth distribution ( $\mu$ : 135,  $\sigma$ : 10, solid line) and finds are based on an identical mowing scenario (light grey bars). The reconstructed parturition dates (dark grey bars and  $\bar{x}$  and  $s$  given in the upper left corner) depend on the scenario of restrictions: S1 (a, d) with max. age of 25 days and no age estimation error, S2 (b, e) with max. age of 25 days and age estimation error, S3 (c, f) with no age restriction and no age estimation error (see Table 1). The estimated birth distributions are given as dashed lines (estimates of  $\mu$  and  $\sigma$  in the upper right corner) and refer to the Grid Search method (a–c) and the ML method (d–f).

TABLE 3 Root mean square error of estimated distribution moments in comparison to simulated distributions of test data set for four representative mowing scenarios.

RMSE of	S1			S2				
	Virtual field data		Grid Search	ML	Virtual field data			
	Mean	Standard deviation	Mean	Standard deviation	Mean	Standard deviation		
Early mowing	13.846	4.814	5.628	1.168	4.785	0.944	13.845	4.514
Midway mowing	6.603	3.603	3.645	1.405	2.793	1.198	6.603	3.295
Late mowing	6.355	2.507	3.982	0.729	2.694	1.058	6.358	2.058
Distributed mowing	2.610	1.196	1.653	0.668	1.751	1.036	2.611	0.952

Note: DoYs of 'Early': 124–130, DoYs of 'Midway': 135–141, DoYs of 'Late': 155, 156, 167, 168, 169, 170; DoYs of 'Distributed': 119, 120, 121, 138, 139, 140, 141, 157, 158.

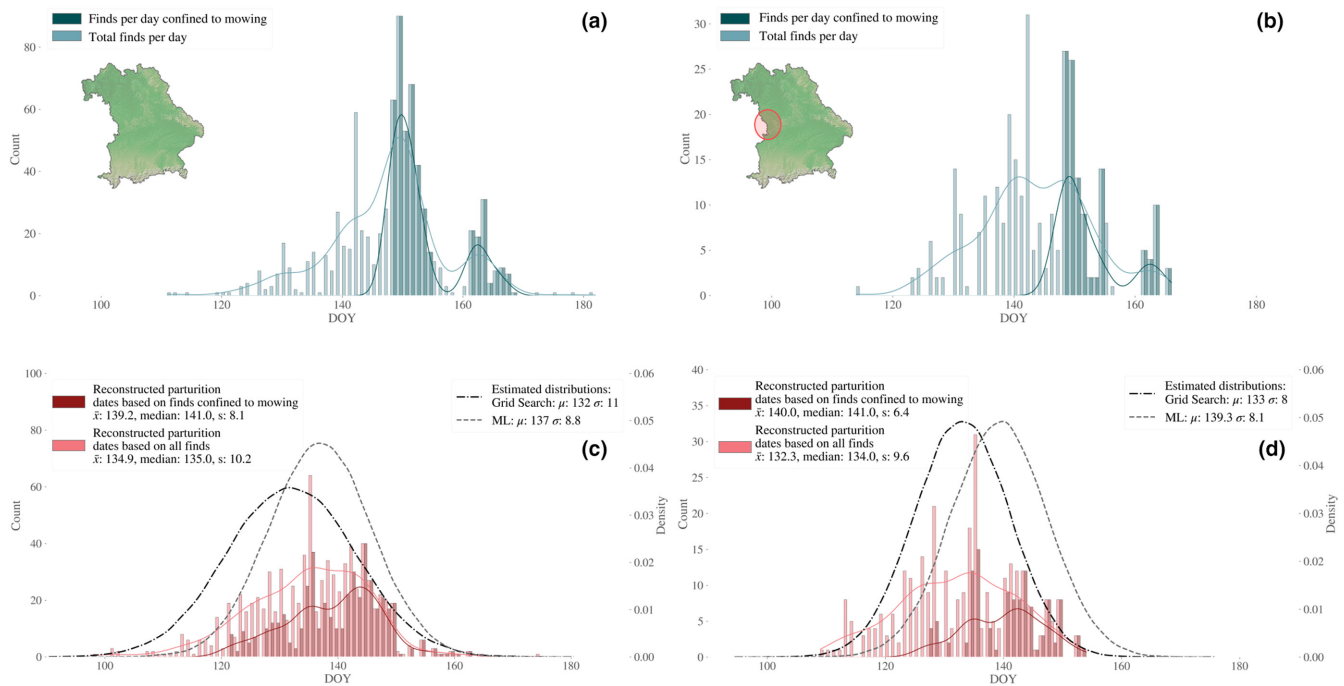


or distributed mowing days were selected. Results indicated that with distributed mowing across the spring season best results were expected (S1 and S2). Early mowing impeded the accuracy of fawn birth distribution estimation most (up to almost 2 weeks for the mean and 5 days for the standard deviation). Midway and late mowing yielded almost similar results, however, if all fawns can be found and aged (no age restriction—S3) the late mowing scenario performed by far best. Within this particular case, estimating  $\mu$  and  $\sigma$  solely based on the sample's  $\bar{x}$  and  $s$  gave the best results. In most other cases (except for distributed mowing) ML algorithms

yielded the best results in estimating the moments of the underlying distribution.

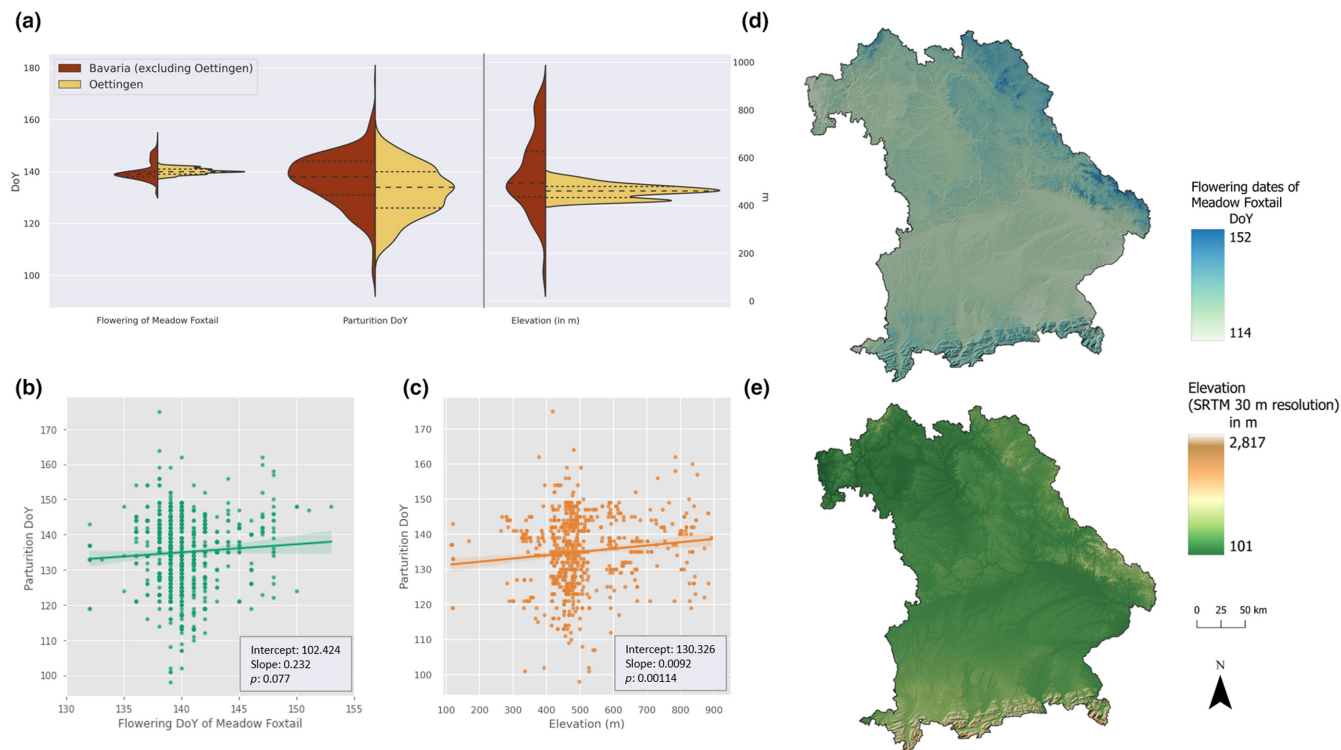
### 3.2 | Case study 2021

For the recorded fawn birth data in 2021 in Bavaria (see Figure 2) and the subset for our intensive study region Oettingen in the West, Figure 4b,d show the obvious discrepancies between  $\bar{x}$  and  $s$  of the different sampling designs (sampling confined to mowing



**FIGURE 4** Actual finds in 2021 (a, b) and estimated parturition dates of roe deer fawns (c, d) in Bavaria (left column) and Oettingen (right column). (a and b): Finds per day confined to mowing (dark cyan) and total finds (light cyan). (c and d): Reconstructed parturition dates based on finds confined to mowing (dark red) and based on all finds (light red) with respective  $\bar{x}$  and  $s$ . The lines show the estimated distribution of birth dates of the Grid Search algorithms (black dot-dashed) and of the ML algorithms (grey dashed) with respective estimates of  $\mu$  and  $\sigma$  based on parturition dates confined to mowing.

				S3					
Grid Search		ML		Virtual field data		Grid Search		ML	
Mean	Standard deviation	Mean	Standard deviation	Mean	Standard deviation	Mean	Standard deviation	Mean	Standard deviation
5.449	1.448	4.781	1.131	14.709	4.879	4.877	1.267	4.757	0.783
3.736	1.560	2.797	1.468	6.733	3.088	4.357	1.349	2.890	1.39
4.074	0.865	2.730	1.473	0.425	0.330	0.790	0.525	0.610	0.736
1.617	0.645	1.758	1.317	4.072	0.992	2.108	0.616	1.466	0.791



**FIGURE 5** (a) Flowering dates (DoY) of meadow foxtail (2021), estimated parturition dates and elevation (m) in Bavaria excluding Oettingen (red) and Oettingen (yellow). Flowering dates of meadow foxtail (b) and elevation (c) against all estimated Bavarian roe deer parturition dates (including Oettingen). Flowering dates of (d) meadow foxtail in Bavaria are based on interpolated phenological maps of Deutscher Wetterdienst, DWD and elevation (e) based on a DEM of SRTM.

versus total finds). When calculating  $\bar{x}$  of all fawns found in Bavaria in 2021, the  $\bar{x}$  and median parturition date can be approximated to the 135th day of the year (15th May), while 68% of births took place in ca. 20.4 days (standard deviation = 10.2 days). Limiting the observations to mowing days, the mean parturition date of the sample was later on the 139th day, the median on the 141st and the standard deviation was smaller ( $s = 8.1$ ; Figure 4a,c). A similar but even more distinctive divergence was observed when only looking at data from the study region of Oettingen (Figure 4b,d). While the mean date of parturition timing was the 132nd day of the year (median 134,  $s = 9.6$ ) when taking all finds into account, the mean of parturition timing based on finds confined to mowing was over 1 week later ( $\bar{x} = 140$ , median: 141) and births were less spread out ( $s = 6.4$ ). The fawns of 19 GPS-collared roe deer females were born on average on DoY 131 (median: 132,  $s = 10.9$ ), which clearly underlines that a mowing-only survey in 2021 might have led to late birth distribution estimates in comparison to additionally surveying systematically.

We applied our previously developed Grid Search and ML algorithms to the data, which was confined to mowing of Bavaria and Oettingen.

The Grid Search algorithm approximated the mean parturition date as DoY 132 for the whole of Bavaria, whereas the ML algorithm estimated the mean parturition day to be the 138th. Both estimations were again closer to the mean of all finds ( $\bar{x} = 134.9$ ; Figure 4c). With respect to Oettingen, results showed similar tendencies. Grid

Search estimates ( $\mu$ : 133 and  $\sigma$ : 9) were close to the reconstructed dates based on all finds in Oettingen ( $\bar{x}$ : 132.3 and  $s$ : 9.6) and to parturition dates of GPS-collared roe deer females ( $\bar{x}$ : 131 and  $s$ : 10.9), whereas ML algorithms estimated mean parturition time considerably later ( $\mu$  139.7 and  $\sigma$  7.1; Figure 4c,d).

The assessment of roe deer breeding phenology and site-specific related plant developmental conditions showed that flowering dates of meadow foxtail are more synchronized than parturition dates. For the smaller and environmentally more uniform subset region Oettingen, flowering and parturition dates were naturally less scattered (Figure 5a). Flowering dates of meadow foxtail and elevation at the site showed a significant relationship with parturition dates in Bavaria across space, with  $p = 0.077$  and  $p = 0.00114$  respectively, with  $\alpha = 0.1$  to reduce the chance of Type 2 error due to a smaller sample size compared to Peláez et al. (2020; Figure 5b,c).

## 4 | DISCUSSION

The study results demonstrate the adverse effects of estimating parameters of roe deer birth distribution when non-probabilistic sampling techniques are used. However, the proposed algorithms were found to be suitable for obtaining reliable results. Estimations of ML algorithms worked well when applied to similar data compared to its training data set. There were no big differences in estimation quality regarding different mowing times, which indicates the

models' satisfactory ability to generalize the data. The Grid Search algorithm performed less sharply on the test data set, especially if mowing was early, but yielded reasonable results with late mowing and better results with distributed mowing scenarios. Our simulations have also examined the effects of erroneous age estimations. Even though results in Tables 2 and 3 do not show that age estimation error considerably affected the estimations, this could be due to a loose assumption of the age assessment error. We expected the RMSE to be higher with erroneous age estimations and additionally higher when mowing was late as fawns would be on average older when found. The precision of age approximation decreases with advanced age and is more challenging. Especially if fawns are fleeing, estimations are often flawed due to only a glimpse or a very short examination of appearance. However, a gradual increase of higher errors with later mowing events could not be seen in the data or was negligibly low. Even though age estimation error varied with the age at which a fawn was found (Table 1), the errors were treated as a normal random phenomenon. It must be acknowledged that age estimation error could be skewed by guessing fawns' ages either systematically older or younger. Again, this error is hardly graspable and needs further attention in future work.

As expected, when mowing days were spread over a longer time frame, higher precision of estimates for S1 and S2 were achieved (Table 3). Early mowing combined with early searches resulted in estimates with the highest errors due to a heavily right-sided truncation. While left truncation due to age restriction was lower (since fawns at this time of the season are rarely older than 25 days), right truncation (fawns have not been born by the day of search) impeded the estimations even more. Particularly predominant is this seen in S3 where the bias of parameter approximation of the sample is intensified due to more fawns on the left side of the distribution (compare Figure 3). Nonetheless, the ML algorithm was able to reduce this RMSE for the mean (early mowing scenario) by ~65% (S1 and S2) and for the standard deviation by 80% (S1) and 75% (S2).

For late mowing scenarios, however, the left truncation of the data considerably impaired estimates (compare late mowing S1–S3 for which S3 yielded the best results), since fawns that flee or already follow the mother are not found or correctly aged. However, only when mowing was late, S3 provided considerably better results. Only then, the algorithms can rely on a sample from the full distribution. In this particular case, the parameter estimations solely based on virtual field data yielded the best results. This illustrates the unrecoverable error by left truncation in late mowing scenarios of S1 and S2 in contrast to S3.

Another bias within estimations of real-life studies could be introduced by neglecting the survival rate of roe deer fawns. Next to mowing, roe deer fawns encounter a handful of other threats in their first weeks of life. Hypothermia, starvation or disease are less likely than predation by red fox *Vulpes vulpes* in areas with high predator abundance (Aanes & Andersen, 1996; Jarnemo & Liberg, 2005). A long-term study in Sweden with high predator abundance has shown that predation risk is lowest shortly after the birth peak and significantly highest at both ends of the birth distribution. The majority of

fawns that are killed by a red fox are younger than 30 days and the risk is highest within the first week of life and thereafter declines almost linearly (Jarnemo et al., 2004). Thus, our algorithms could be affected by a reduced number of very young fawns (<7 days) and accordingly misinterprets the age distribution of the sample (Gilbert et al., 2014). Yet, Aanes and Andersen (1996) and Panzacchi et al. (2009) found contrary results regarding habitat fragmentation, movement, mobility, sex and, age. Thus, the results of other studies cannot be readily transferred to other regions without considering predator abundance or landscape composition (Aanes & Andersen, 1996; Jarnemo et al., 2004). It is known that red fox abundance and density have increased in many parts of Europe in recent decades due to successful rabies vaccination and adaption to urban areas and human-modified environments (Goszczyński et al., 2008; König et al., 2018; Soe et al., 2017). However, due to a knowledge gap on the exact red fox densities in Bavaria, predation was not included in our simulations. If predation risk within areas is known, simulations should be adjusted.

When applying the algorithms in the case study in Bavaria in 2021, it must be acknowledged that a perfect validation of the methods cannot be accomplished, since the true distributions are hardly measurable (Moussus et al., 2010). Birth distribution in Oettingen can be inferred with higher confidence due to sound data from GPS-collared animals and a higher portion of random searches in fields and forests. Further, the study region of Oettingen forms a small and homogeneous area. For the Oettingen case study in 2021, the Grid Search algorithm was able to capture the mean of the birth peak in Oettingen based on the parturition of GPS-collared animals best. Estimates of breeding synchrony (standard deviation) also matched the measured standard deviation from field data well. Surprisingly, the trained ML model predicted the mean to be later and similar to the sample's mean  $\bar{x}$  which was confined to mowing. The reason for this presumably poorer performance of ML algorithms can be manifold. One possible pitfall could be the confrontation with a mowing scenario that was not included in the training data set. In that case, the model might have been overfitted to the training data set or trained on an unbalanced data set. As there is no such effect as 'learning' within the Grid Search algorithm, the quality of its estimates does not depend on high-quality and balanced training. Another simulation assumption has not yet been discussed. The probability of a meadow being mowed on a mowing day was assumed to be uniform for all scheduled mowing days, whereas in reality probabilities could be uni- or bimodally distributed or even more complex. This could lead to a more complex function of possible detectable fawns on mowing days. Unfortunately, the specific mowing probability for a field on mowing days is again dependent on weather, machine availability, and work scheduling and thus hardly predictable.

The evaluation of the models for the actual finds in Bavaria in 2021 is more complicated due to a higher uncertainty of the actual birth peak, breeding synchrony and its uniformity across space. The reconstructed mean parturition day for Bavaria was 2.6 days later than in Oettingen (Bavaria  $\bar{x}$ : 134.9, Oettingen  $\bar{x}$ : 132.3), while

the Grid Search and ML algorithms predicted mean birth dates in Oettingen to be later. Again, the Grid Search algorithm seems to yield better results in comparison to the mowing-biased field data. Interpretation of the results for standard deviation must be evaluated carefully within all scenarios. Due to a smaller range of possible true values, results could diverge substantially.

In accordance with Peláez et al. (2020), we were able to relate the variance of parturition timing to plant phenology and elevation across space. Flowering dates of meadow foxtail were a good proxy to describe the environmental conditions of a site (compare Peláez et al., 2020), however, the relationship between parturition dates and elevation was more significant in Bavaria. The flowering dates of meadow foxtail at the approximate site of fawn finds were more synchronized and showed less variation in contrast to parturition days. Yet, flowering dates are based on interpolations and thus mirror phenology with uncertainties and lower (spatial) variability. A clear relationship between parturition dates and flowering dates would have been useful to interpolate parturition dates with plant phenology across space if sampling was confined to mowing.

## 5 | CONCLUSIONS

The results, based on simulated finds of virtual field data with varying mowing dates, have highlighted the drawbacks of non-probabilistic sampling techniques and the issues of inferring moments if searches are confined to a few and consecutive, for example mowing days. The results should raise awareness when dealing with such data and reveal possible ways to retrieve better estimates of underlying distributions. Modelling and understanding skewed and truncated distributions of possible sampling outcomes are essential to leveraging the potential of field data surveyed by non-expert groups like citizen scientists. Simulating field data can help to capture deficiencies of sampling designs but can also show possibilities to trim down costly survey designs while obtaining defensible outcomes. The developed methods can be used for any other species, such as ground-nesting bird species or rodents, and can empower incomplete data for conservation actions related to mowing but also for wildlife studies in general. Potentially, voluntarily collected data from roe deer fawn rescue initiatives can form an integral part to increase data size on regional birth distributions and ultimately help understand how much roe deer are adapting to environmental drivers such as climate warming. Moreover, knowledge of exact birth distributions can lead to more efficient searches during mowing season and ease the thread of mowing deaths among roe deer fawns.

### AUTHOR CONTRIBUTIONS

Johanna Kauffert: conceptualization, data curation, formal analysis, methodology, visualization, writing—original draft and writing—review & editing. Sophie Baur: data curation, methodology, resources, investigation and writing—review & editing. Michael Matiu: methodology and writing—review & editing. Andreas König:

resources and writing—review & editing. Wibke Peters: resources and writing—review & editing. Annette Menzel: methodology, conceptualization, writing—original draft, writing—review & editing, supervision project administration and funding acquisition.

### ACKNOWLEDGEMENTS

We kindly acknowledge the use of meteorological data of the German Meteorological Service (DWD). We thank student assistants, gamekeepers, hunters and drone pilots for their help during fieldwork. We also thank all wildlife volunteers who reported fawns via our citizen science platforms. Open Access funding enabled and organized by Projekt DEAL.

### FUNDING INFORMATION

This research was funded by the Bavarian State Ministry of Food, Agriculture, and Forestry, grant number: A/19/17.

### CONFLICT OF INTEREST STATEMENT

None declared.

### PEER REVIEW

The peer review history for this article is available at <https://www.webofscience.com/api/gateway/wos/peer-review/10.1002/2688-8319.12225>.

### DATA AVAILABILITY STATEMENT

Code (<https://doi.org/10.6084/m9.figshare.21175441>) and DoYs of parturition of fawns (<https://doi.org/10.6084/m9.figshare.22232734>) are permanently archived on Figshare (Kauffert et al., 2023a, 2023b). Exact fawn locations cannot be published due to reasons of traceability. Phenological grids are freely available from the German Meteorological Service (DWD): [https://opendata.dwd.de/climate\\_environment/CDC/grids\\_germany/annual/phenology/](https://opendata.dwd.de/climate_environment/CDC/grids_germany/annual/phenology/) (Deutscher Wetterdienst, 2023).

### ORCID

Johanna Kauffert  <https://orcid.org/0000-0002-3755-8169>

Sophie Baur  <https://orcid.org/0000-0003-1987-0690>

Michael Matiu  <https://orcid.org/0000-0001-5289-0592>

Andreas König  <https://orcid.org/0000-0002-9232-5026>

Wibke Peters  <https://orcid.org/0000-0002-7105-428X>

Annette Menzel  <https://orcid.org/0000-0002-7175-2512>

### REFERENCES

- Aanes, R., & Andersen, R. (1996). The effects of sex, time of birth, and habitat on the vulnerability of roe deer fawns to red fox predation. *Canadian Journal of Zoology*, 74(10), 1857–1865. <https://doi.org/10.1139/z96-209>
- Alibrahim, H., & Ludwig, S. A. (2021). Hyperparameter optimization: Comparing genetic algorithm against grid search and Bayesian optimization. In *2021 IEEE Congress on Evolutionary Computation (CEC)* (pp. 1551–1559). IEEE. <https://doi.org/10.1109/CEC45853.2021.9504761>
- Awad, M., & Khanna, R. (2015). Support vector regression. In *Efficient learning machines: Theories, concepts, and applications for*

- engineers and system designers (pp. 67–80). Apress. <https://doi.org/10.1007/978-1-4302-5990-9>
- Bock, A., Sparks, T. H., Estrella, N., & Menzel, A. (2013). Changes in the timing of hay cutting in Germany do not keep pace with climate warming. *Global Change Biology*, 19(10), 3123–3132. <https://doi.org/10.1111/gcb.12280>
- Boob, M., Elsaesser, M., Thumm, U., Hartung, J., & Lewandowski, I. (2019). Harvest time determines quality and usability of biomass from lowland hay meadows. *Agriculture (Switzerland)*, 9(9), 1–15. <https://doi.org/10.3390/agriculture9090198>
- Breiman, L. (1996). Bagging predictors. *Machine Learning*, 24, 123–140. <https://doi.org/10.3390/risks8030083>
- Brown, E. D., & Williams, B. K. (2019). The potential for citizen science to produce reliable and useful information in ecology. *Conservation Biology*, 33(3), 561–569. <https://doi.org/10.1111/cobi.13223>
- Brus, D. J., & De Groot, J. J. (2003). A method to combine non-probability sample data with probability sample data in estimating spatial means of environmental variables. *Environmental Monitoring and Assessment*, 83(3), 303–317. <https://doi.org/10.1023/A:1022618406507>
- Caughley, G., & Caughley, J. (1974). Estimating median date of birth. *The Journal of Wildlife Management*, 38(3), 552–556.
- Christen, N., Janko, C., & Rehnus, M. (2018). The effect of environmental gradients on the bed site selection of roe deer (*Capreolus capreolus*). *Mammal Research*, 63(1), 83–89. <https://doi.org/10.1007/s13364-017-0343-z>
- Conn, P. B., Moreland, E. E., Regehr, E. V., Richmond, E. L., Cameron, M. F., & Boveng, P. L. (2016). Using simulation to evaluate wildlife survey designs: Polar bears and seals in the Chukchi Sea. *Royal Society Open Science*, 3, 150561. <https://doi.org/10.1098/rsos.150561>
- Conn, P. B., Thorson, J. T., & Johnson, D. S. (2017). Confronting preferential sampling when analysing population distributions: Diagnosis and model-based triage. *Methods in Ecology and Evolution*, 8(11), 1535–1546. <https://doi.org/10.1111/2041-210X.12803>
- Cukor, J., Bartoška, J., Rohla, J., Sova, J., & Machálek, A. (2019). Use of aerial thermography to reduce mortality of roe deer fawns before harvest. *PeerJ*, 7, e6923. <https://doi.org/10.7717/peerj.6923>
- Dambly, L. I., Jones, K. E., Boughey, K. L., & Isaac, N. J. (2021). Observer retention, site selection and population dynamics interact to bias abundance trends in bats. *Journal of Applied Ecology*, 58(2), 236–247. <https://doi.org/10.1111/1365-2664.13760>
- Deák, G., Katona, K., & Biró, Z. (2021). Exploring the use of a carcass detection dog to assess mowing mortality in Hungary. *Journal of Vertebrate Biology*, 69(3), 20089–1. <https://doi.org/10.25225/jvb.20089>
- Deutscher W. (2023). *Index of climate\_environment/CDC/grids\_germany/annual/phenology/*. Deutscher Wetterdienst Open Data. [https://opendata.dwd.de/climate\\_environment/CDC/grids\\_germany/annual/phenology/](https://opendata.dwd.de/climate_environment/CDC/grids_germany/annual/phenology/)
- Dicks, L. V., Ashpole, J. E., Dänhardt, J., James, K., Jönsson, A., Randall, N., Showler, D. A., Smith, R. K., Turpie, S., Williams, D. R., & Sutherland, W. J. (2020). Farmland conservation. In W. J. Sutherland, L. V. Dicks, S. O. Petrovan, & R. K. Smith (Eds.), *What works in conservation 2020* (pp. 283–321). Open Book Publishers.
- Driehuis, F., Wilkinson, J. M., Jiang, Y., Ogunade, I., & Adesogan, A. T. (2018). Silage review: Animal and human health risks from silage. *Journal of Dairy Science*, 101(5), 4093–4110. <https://doi.org/10.3168/jds.2017-13836>
- Elith, J., Leathwick, J. R., & Hastie, T. (2008). A working guide to boosted regression trees. *Journal of Animal Ecology*, 77(4), 802–813. <https://doi.org/10.1111/j.1365-2656.2008.01390.x>
- Ellenberg, H. (1978). Zur Populationsökologie des Rehes (*Capreolus capreolus* L., Cervidae) in Mitteleuropa. *Zoolog. Staatssammlung*, 2, 1–212.
- Elliott, M. R. (2009). Combining data from probability and non-probability samples using pseudo-weights. *Survey Practice*, 2(6), 1–7. <https://doi.org/10.29115/SP-2009-0025>
- Etikan, I., Musa, S., & Alkassim, R. (2016). Comparison of convenience sampling and purposive sampling. *American Journal of Theoretical and Applied Statistics*, 5(1), 1–4. <https://doi.org/10.11648/j.aj-tas.20160501.11>
- Fink, A. (2003). *How to sample in surveys* (2nd ed.). Sage.
- Friedman, J. (2001). Greedy function approximation: A gradient boosting machine. *The Annals of Statistics*, 29(5), 1189–1232.
- Gaillard, J. M., Delorme, D., Jullien, J. M., & Tatin, D. (1993). Timing and synchrony of births in roe deer. *Journal of Mammalogy*, 74(3), 738–744.
- Gaillard, J. M., Mark Hewison, A. J., Klein, F., Plard, F., Douhard, M., Davison, R., & Bonenfant, C. (2013). How does climate change influence demographic processes of widespread species? Lessons from the comparative analysis of contrasted populations of roe deer. *Ecology Letters*, 16(Suppl.1), 48–57. <https://doi.org/10.1111/ele.12059>
- Galli, A., Focardi, S., Iseppi, M., & Piasentier, E. (2008). Age determination in fallow deer *Dama dama* neonates. *Wildlife Biology*, 14(3), 386–390. [https://doi.org/10.2981/0909-6396\(2008\)14\[386:ADIFD D\]2.0.CO;2](https://doi.org/10.2981/0909-6396(2008)14[386:ADIFD D]2.0.CO;2)
- Gilbert, S. L., Lindberg, M. S., Hundertmark, K. J., & Person, D. K. (2014). Dead before detection: Addressing the effects of left truncation on survival estimation and ecological inference for neonates. *Methods in Ecology and Evolution*, 5(10), 992–1001. <https://doi.org/10.1111/2041-210X.12234>
- Goszczyński, J., Misiorowska, M., & Juszko, S. (2008). Changes in the density and spatial distribution of red fox dens and cub numbers in Central Poland following rabies vaccination. *Acta Theriologica*, 53(2), 121–127. <https://doi.org/10.1007/BF03194245>
- Hagen, R., Ortman, S., Elliger, A., & Arnold, J. (2021). Advanced roe deer (*Capreolus capreolus*) parturition date in response to climate change. *Ecosphere*, 12(11), e03819. <https://doi.org/10.1002/ecs2.3819>
- Harris, C. R., Millman, K. J., van der Walt, S. J., Gommers, R., Virtanen, P., Cournapeau, D., Wieser, E., Taylor, J., Berg, S., Smith, N. J., Kern, R., Picus, M., Hoyer, S., van Kerkwijk, M. H., Brett, M., Haldane, A., del Río, J. F., Wiebe, M., Peterson, P., ... Oliphant, T. E. (2020). Array programming with NumPy. *Nature*, 585(7825), 357–362. <https://doi.org/10.1038/s41586-020-2649-2>
- Israel, M. (2011). A UAV-based roe deer fawn detection system. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 38, 51–55. <https://doi.org/10.5194/isprsarchives-xxxviii-1-c22-51-2011>
- Jarnemo, A. (2002). Roe deer *Capreolus capreolus* fawns and mowing—Mortality rates and countermeasures. *Wildlife Biology*, 8(3), 211–218. <https://doi.org/10.2981/wlb.2002.035>
- Jarnemo, A. (2004). Predation processes: Behavioural interactions between red fox and roe deer during the fawning season. *Journal of Ethology*, 22(2), 167–173. <https://doi.org/10.1007/s10164-004-0118-2>
- Jarnemo, A., & Liberg, O. (2005). Red fox removal and roe deer fawn survival—A 14-year study. *Journal of Wildlife Management*, 69(3), 1090–1098. [https://doi.org/10.2193/0022-541x\(2005\)069\[1090:rfrard\]2.0.co;2](https://doi.org/10.2193/0022-541x(2005)069[1090:rfrard]2.0.co;2)
- Jarnemo, A., Liberg, O., Lockowandt, S., Olsson, A., & Wahlström, K. (2004). Predation by red fox on European roe deer fawns in relation to age, sex, and birth date. *Canadian Journal of Zoology*, 82(3), 416–422. <https://doi.org/10.1139/z04-011>
- Johnson, D. S., Barry, R. P., & Bowyer, R. T. (2004). Estimating timing of life-history events with coarse data. *Journal of Mammalogy*, 85(5), 932–939. <https://doi.org/10.1644/BFW-009>
- Jullien, J. M., Delorme, D., & Gaillard, J. M. (1992). Détermination de l'âge chez le faon de chevreuil (*Capreolus capreolus*) dans son premier mois de vie. *Mammalia*, 56(2), 307–311. <https://doi.org/10.1515/mamm-1992-0221>
- Kałuźniński, J. (1982). Roe deer mortality due to mechanization of work in agrocenoses. *Acta Theriologica*, 27(30), 449–455. <https://doi.org/10.4098/at.arch.82-39>

- Katuziński, J., & Pielowski, Z. (1976). *The effect of technical agricultural operations on the hare population* (pp. 205–211). Ecology and Management of European Hare Populations, Pastwowe Wydawnictwo Rolnicze i Lesne.
- Kauffert, J. M., Baur, S., Matiu, M., König, A., Peters, W., & Menzel, A. (2023a). Code for “fawn birthdays—From opportunistically sampled fawn rescue data to true wildlife demographic parameters”. *Figshare*. <https://doi.org/10.6084/m9.figshare.21175441.v1>
- Kauffert, J. M., Baur, S., Matiu, M., König, A., Peters, W., & Menzel, A. (2023b). Fawn data for “fawn birthdays—From opportunistically sampled fawn rescue data to true wildlife demographic parameters”. *Figshare*. <https://doi.org/10.6084/m9.figshare.22232734.v1>
- Kittler, L. (1979). Wildverluste durch den Einsatz landwirtschaftlicher Maschinen nach einer Erhebung aus dem Jagdjahr 1976/77 in Nordrhein-Westfalen. *Zeitschrift für Jagdwissenschaft*, 25(1), 22–32.
- König, P., Tautenhahn, S., Cornelissen, J. H. C., Kattge, J., Bönsch, G., & Römermann, C. (2018). Advances in flowering phenology across the northern hemisphere are explained by functional traits. *Global Ecology and Biogeography*, 27(3), 310–321. <https://doi.org/10.1111/geb.12696>
- Kurt, F. (1968). Zusammenhänge zwischen Verhalten und Fortpflanzungsleistung beim Reh (*Capreolus capreolus* L.). *Zeitschrift für Jagdwissenschaft*, 14, 97–106.
- Leiner, D. J. (2016). Our research's breadth lives on convenience samples a case study of the online respondent pool “SoSci panel”. *Studies in Communication|Media*, 5(4), 367–396. <https://doi.org/10.5771/2192-4007-2016-4-367>
- Lent, C. L. (1974). Mother-infant relationships in ungulates. In V. Geist & F. Walther (Eds.), *The behaviour of ungulates and its relation to management* (pp. 14–55). International Union for Conservation of Nature and Nature Resources.
- LimeSurvey Project Team. (2022). LimeSurvey: An open source survey tool.
- Linnell, J. D. (1994). *Reproductive tactics and parental care in Norwegian roe deer* [doctoral dissertation]. NUI, Department of Zoology, UCC.
- Linnell, J. D., & Andersen, R. (1998). Timing and synchrony of birth in a hider species, the roe deer *Capreolus capreolus*. *Journal of Zoology*, 244(4), 497–504. <https://doi.org/10.1017/S0952836998004038>
- Linnell, J. D., Nijhuis, P., Teurlings, I., & Andersen, R. (1999). Selection of bed-sites by roe deer *Capreolus capreolus* fawns in a boreal forest landscape. *Wildlife Biology*, 5(1), 225–231. <https://doi.org/10.1007/BF03192512>
- Linnell, J. D., Nilsen, E. B., & Reidar, A. (2004). Selection of bed-sites by roe deer *Capreolus capreolus* fawns in an agricultural landscape. *Acta Theriologica*, 49(1), 103–111. <https://doi.org/10.2981/wlb.1999.027>
- Marchand, P., Garel, M., Morellet, N., Benoit, L., Chaval, Y., Itty, C., Petit, E., Cargnelutti, B., Hewison, A. J., & Loison, A. (2021). A standardised biologging approach to infer parturition: An application in large herbivores across the hider-follower continuum. *Methods in Ecology and Evolution*, 12, 1017–1030. <https://doi.org/10.1111/2041-210X.13584>
- Menke, W. (2012). Nonlinear inverse problems. In W. Menke (Ed.), *Geophysical data analysis: Discrete inverse theory* (3rd ed., pp. 163–188). Academic Press chapter 9.
- Moeller, R. B., Jr., & Puschner, B. (2007). Botulin in cattle—A review. *The Bovine Practitioner*, 41(1), 54–59. <https://doi.org/10.21423/bovine-vol41no1p54-58>
- Moussus, J.-P., Julliard, R., & Jiguet, F. (2010). Featuring 10 phenological estimators using simulated data. *Methods in Ecology and Evolution*, 1(2), 140–150. <https://doi.org/10.1111/j.2041-210x.2010.00020.x>
- Panzacchi, M., Herfindal, I., Linnell, J. D., Odden, M., Odden, J., & Andersen, R. (2010). Tradeoffs between maternal foraging and fawn predation risk in an income breeder. *Behavioral Ecology and Sociobiology*, 64(8), 1267–1278. <https://doi.org/10.1007/s00265-010-0941-0>
- Panzacchi, M., Linnell, J. D., Odden, M., Odden, J., & Andersen, R. (2009). Habitat and roe deer fawn vulnerability to red fox predation. *Journal of Animal Ecology*, 78(6), 1124–1133. <https://doi.org/10.1111/j.1365-2656.2009.01584.x>
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, E. (2011). Scikit-learn: Machine learning in python. *Journal of Machine Learning Research*, 12, 2825–2830.
- Peláez, M., Gaillard, J. M., Bollmann, K., Heurich, M., & Rehnus, M. (2020). Large-scale variation in birth timing and synchrony of a large herbivore along the latitudinal and altitudinal gradients. *Journal of Animal Ecology*, 89(8), 1906–1917. <https://doi.org/10.1111/1365-2656.13251>
- Plard, F., Gaillard, J. M., Bonenfant, C., Hewison, A. J., Delorme, D., Cargnelutti, B., Kjellander, P., Nilsen, E. B., & Coulson, T. (2013). Parturition date for a given female is highly repeatable within five roe deer populations. *Biology Letters*, 9(1), 20120841. <https://doi.org/10.1098/rsbl.2012.0841>
- Plard, F., Gaillard, J. M., Coulson, T., Hewison, A. J., Delorme, D., Warnant, C., & Bonenfant, C. (2014). Mismatch between birth date and vegetation phenology slows the demography of roe deer. *PLoS Biology*, 12(4), e1001828. <https://doi.org/10.1371/journal.pbio.1001828>
- Raganella-Pelliccioni, E., Scremin, M., & Toso, S. (2007). Phenology and synchrony of roe deer breeding in northern Italy. *Acta Theriologica*, 52(1), 95–100. <https://doi.org/10.1007/BF03194204>
- Rehnus, M., Arnold, J., Elliger, A., & Reimoser, F. (2018). Ear-marking of roe deer fawns (*Capreolus capreolus*): Results of long-term studies in Central Europe. *Beiträge Zur Jagd & Wildforschung*, 43, 71–90. <https://www.researchgate.net/publication/329559991>
- Rehnus, M., Peláez, M., & Bollmann, K. (2020). Advancing plant phenology causes an increasing trophic mismatch in an income breeder across a wide elevational range. *Ecosphere*, 11(6), e03144. <https://doi.org/10.1002/ecs2.3144>
- Rehnus, M., & Reimoser, F. (2014). Rehkitzmarkierung—Nutzen für Praxis und Forschung. *FaunaFocus*, 9 (September), 1–16.
- Ruiz-Gutierrez, V., Hooten, M. B., & Campbell Grant, E. H. (2016). Uncertainty in biological monitoring: A framework for data collection and analysis to account for multiple sources of sampling bias. *Methods in Ecology and Evolution*, 7, 900–909. <https://doi.org/10.1111/2041-210X.12542>
- Sams, M. G., Lochmiller, R. L., Helgren, E. C., Warde, W. D., & Varner, L. W. (1996). Morphometric predictors of neonatal age for white-tailed deer. *Wildlife Society Bulletin*, 24(1), 53–57.
- Soe, E., Davison, J., Süld, K., Valdmann, H., Laurimaa, L., & Saarma, U. (2017). Europe-wide biogeographical patterns in the diet of an ecologically and epidemiologically important mesopredator, the red fox *Vulpes vulpes*: A quantitative review. *Mammal Review*, 47, 198–211. <https://doi.org/10.1111/mam.12092>
- Speak, A., Escobedo, F. J., Russo, A., & Zerbe, S. (2018). Comparing convenience and probability sampling for urban ecology applications. *Journal of Applied Ecology*, 55(5), 2332–2342. <https://doi.org/10.1111/1365-2664.13167>
- Stamm, A., Mosler, E., & Rehnus, M. (2017). Ausnahme oder Regel—Merkmale zur Altersbestimmung von Rehkitzen. *Jagd & Natur*, 5, 30–32.
- Tyler, G. A., Green, R. E., & Casey, C. (1998). Survival and behaviour of corncrake *Crex crex* chicks during the mowing of agricultural grassland. *Bird Study*, 45(1), 35–50. <https://doi.org/10.1080/00063659809461076>
- Vapnik, V., Golowich, S. E., & Smola, A. (1996). Support vector method for function approximation, regression estimation, and signal processing. *Advances in Neural Information Processing Systems*, 9, 281–287.
- Weiser, E. L., Diffendorfer, J. E., Lopez-Hoffman, L., Semmens, D., & Thogmartin, W. E. (2020). Challenges for leveraging citizen science to support statistically robust monitoring programs.

*Biological Conservation*, 242, 108411. <https://doi.org/10.1016/j.biocon.2020.108411>

Yuan, Y., Härer, S., Ottenheim, T., Misra, G., Lüpke, A., Estrella, N., & Menzel, A. (2021). Maps, trends, and temperature sensitivities—Phenological information from and for decreasing numbers of volunteer observers. *International Journal of Biometeorology*, 65, 1377–1390. <https://doi.org/10.1007/s00484-021-02110-3>

## SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

**Figure S1.** Results of the sensitivity analysis with an increasing number of observations (50–2400) per sample in the analysis

(sampling was performed based on assumptions of scenario 2: maximal age at find: 25 days + age estimation error).

**How to cite this article:** Kauffert, J., Baur, S., Matiu, M., König, A., Peters, W., & Menzel, A. (2023). Fawn birthdays: From opportunistically sampled fawn rescue data to true wildlife demographic parameters. *Ecological Solutions and Evidence*, 4, e12225. <https://doi.org/10.1002/2688-8319.12225>