bioRxiv preprint doi: https://doi.org/10.1101/2023.11.17.567350; this version posted February 27, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Analysing biodiversity observation data collected in continuous time: Should we use discrete- or continuous-time occupancy models?

Léa Pautrel[1,2,*], Sylvain Moulherat[1], Olivier Gimenez[2,**], and Marie-Pierre Etienne[3,**]

[1]OïkoLab, TerrOïko, 2 Place Dom Devic, 81540 Sorèze, France

[2]CEFE, University of Montpellier, CNRS, EPHE, IRD, 1919 route de Mende, 34090 Montpellier, France

[3]IRMAR, Rennes University, Institut Agro Rennes Angers, 65 route de Saint Brieuc, 35042 Rennes, France

[*]Corresponding author: lea.pautrel@terroiko.fr

[**]Co-senior authors

bioRxiv preprint doi: https://doi.org/10.1101/2023.11.17.567350; this version posted February 27, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Abstract

1. Biodiversity monitoring is undergoing a revolution, with fauna observations data being increasingly gathered continuously over extended periods, through sensors like camera traps and acoustic recorders, or via opportunistic observations. These data are often analysed with discrete-time ecological models, requiring the transformation of continuously collected data into arbitrarily chosen non-independent discrete time intervals. To overcome this issue, ecologists are increasingly turning to the existing continuous-time models in the literature. Closer to the real detection process, they are lesser known than discrete-time models, not always easily accessible, and can be more complex. Focusing on occupancy models, a type of species distribution models, we asked ourselves: Should we dedicate time and effort to learning and using these continuous-time models, or can we go on using discrete-time models?

2. We conducted a comparative simulation study using data generated within a continuous-time framework. We assessed the performance of five static occupancy models with varying detection processes: discrete detection/non-detection process, discrete count process, continuous-time Poisson process, and two types of modulated Poisson processes. Our goal was to assess their abilities to estimate occupancy probability with continuously collected data. We applied all models to empirical lynx data as an illustrative example.

3. In scenarios with easily detectable animals, we found that all models accurately estimated occupancy. All models reached their limits with highly elusive animals. Variation in discretisation intervals had minimal impact on the discrete models' capacity to estimate occupancy accurately.

4. Our study underscores that opting for continuous-time models with an increased number of parameters, aiming to get closer to the sensor detection process, may not offer substantial advantages over simpler models when the sole aim is to accurately estimate occupancy. Model choice can thus be driven by practical considerations such as data availability or implementation time. However, occupancy models can encompass goals beyond estimating occupancy probability. Continuous-time models, particularly those considering temporal variations in detection, can offer valuable insights into specific species behaviour and broader ecological inquiries. We hope that our findings offer valuable guidance for researchers and practitioners working with continuously collected data in wildlife monitoring and modelling.

Keywords Camera trap, Continuous-time model, Discrete-time model, Markov Modulated Poisson Process, Occupancy modelling, Poisson Process, Sensors, Wildlife monitoring

bioRxiv preprint doi: https://doi.org/10.1101/2023.11.17.567350; this version posted February 27, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

1 Introduction

The alarming decline of biodiversity has led to a scientific, ethical, and legal need to better understand its drivers in order to protect nature more effectively (IPBES, 2019). With the reinforcement of regulations and recommendations for achieving the objectives of no net loss of biodiversity, the need for wildlife monitoring is growing rapidly (UNECE, 2023). Concurrently, the development of increasingly sophisticated and accessible technologies is leading to a digital revolution. Sensors, such as camera traps or autonomous recording units, are now available to address current ecological challenges (Burton et al., 2015; Potamitis et al., 2014).

Sensors offer many advantages compared to traditional field observations by naturalists. They are non-invasive, often cost-effective, particularly adapted to observe some elusive or shy species, potentially in challenging terrain, and they can improve reproducibility and protocol standardisation (Steenweg et al., 2017; Zwerts et al., 2021). Sensors are therefore good candidates for setting up large-scale monitoring (Oliver et al., 2023) and collaborations such as Biodiversity Observation Networks (Gonzalez et al., 2023). Policies now emphasise the use of sensors, big data and artificial intelligence to improve knowledge and understanding of species and ecosystems, such as the International Union for Conservation of Nature (IUCN) Nature 2030 programme (IUCN, 2021) or the Biodiversa+ European Biodiversity Partnership (Høye et al., 2022; Vihervaara et al., 2023).

We often use ecological models to analyse observation data for monitoring purposes. These models typically assess the presence (Guillera-Arroita, 2017) or abundance (Gilbert et al., 2021) of a species, often while considering the relation with environmental factors. They can be used for a particular species or within a multi-species framework (Pollock et al., 2014). These models produce actionable knowledge about species, influencing our actions and our approach to biodiversity conservation. For example, the area of occupancy, *i.e.* the spatial distribution where a species is present, is one of the criterion used by the IUCN to establish the Red list of Ecosystems (Rodríguez et al., 2015).

In this paper, we focus on occupancy models, a category of ecological models aiming to estimate species presence. Occupancy models, as introduced by MacKenzie et al. (2002), are hierarchical models that include two sub-models. The first sub-model describes the ecological process, occupancy, typically of interest to ecologists. The second sub-model accounts for measurement errors arising from imperfect detection. A site is said occupied when at least one individual went through it (Emmet et al., 2021). At a broader scale, occupancy corresponds to the proportion of sites within a study area that are occupied by the species (MacKenzie et al., 2002). The occupancy model proposed by MacKenzie et al. (2002) uses binary data (0 if the species was not detected, 1 if it was) at each site during each sampling occasion. This model has underpinned numerous occupancy studies in the last two decades, and was refined or adapted by many modelers (Bailey et al., 2014). These adaptations have given rise to new occupancy models, most of them aiming to mirror more closely the expected ecological or detection conditions, impacting the input data required by each model. We here focus on static occupancy models, in which the occupancy state of a site is assumed constant, without

bioRxiv preprint doi: https://doi.org/10.1101/2023.11.17.567350; this version posted February 27, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

extinction-colonisation processes, as opposed to dynamic occupancy models.

Ecological models, including occupancy models, have historically been developed to analyse observation data collected by field operators during one or several short sampling occasions (Bailey et al., 2014). However, the deployment of sensors involves continuous data collection, often over long time periods (*e.g.* Cove et al., 2021; Cusack et al., 2015; Moore et al., 2020). For instance, Kays et al. (2020) recommend deploying sensors for three to five weeks at multiple locations to estimate relative abundance, occupancy, or species richness. Short-term deployments can equate traditional discrete sampling occasions. However, when sensors are stationed at the same location for extended periods, data is often discretised in order to use traditional models in discrete time. We suggest using the term **session** for these discretised time intervals, because they differ from traditional sampling occasions in two respects: *(1)* sampling occasions are determined before the data collection, whereas the discretisation is done after the data has been collected; and *(2)* sessions occur consecutively without any gaps between them, while the traditional sampling occasions are separated by periods of time when the site is not monitored.

Occupancy discrete-time models have been around for 20 years and are commonly used because they are relatively simple to implement. However, continuous-time ecological modelling is not new. The fist mention of a continuous-time model in the capture-recapture literature dates back to Becker (1984). It was not until the advent of sensors, which highlighted the limitations of discrete-time models, that modelers began to turn towards continuous-time models (Kellner et al., 2022; Rushing, 2023; Schofield et al., 2018). Nonetheless, continuous-time models are not a universal cure-all. Each family of models have their pros and cons.

Discretisation simplifies the information. Discretisation is, in other words, an aggregation of data into sessions. This aggregation simplifies the data and blurs the residual variability, which can help in interpreting broad observed trends. But simplification involves information loss. Because accurately estimating occupancy relies on precisely assessing imperfect detection (Kellner & Swihart, 2014; Kéry & Schmidt, 2008), having more information about detection patterns could provide valuable insights, helping to disentangle the observation process from the ecological process of interest. This may lead to more accurate estimations of static occupancy. Dynamic occupancy models could represent the occupancy state in continuous time to go even further, potentially revealing fine patterns with ecological significance.

Discretisation is arbitrary. Researchers usually choose the aggregation period so that the detection probability is not too low, and the occupancy probability is not estimated at its boundaries (close to 0 or 1). Schofield et al. (2018) highlighted that the chosen session length can impact abundance estimates with discrete-time capture-recapture models. Hence, it most likely impacts occupancy models outputs, as capture-recapture and occupancy models are very similar (the individual capture history equates the site "detection history", MacKenzie et al., 2002). Eliminating arbitrary discretisation in occupancy modelling can enhance the method objectivity and reproducibility, and is expected to improve result reliability, at least compared to a non-optimal discretisation.

bioRxiv preprint doi: https://doi.org/10.1101/2023.11.17.567350; this version posted February 27, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Model complexity and data availability. Models with a continuous-time detection process are likely to overcome the limitations mentioned above. Furthermore, harnessing the richness of continuous-time data, researchers can customise models to replicate species-specific observation processes, providing valuable insights into animal behaviour and activity patterns (e.g. Distiller et al., 2020, with continuous-time spatial capture-recapture models). However, the potential drawback is the complexity of such models, that may render them less adapted to derive ecological insights such as occupancy from small data sets. Additionally, if the system is not assumed to be constant over time, continuous-time covariates are necessary for a continuous-time model, and these covariates are often not readily available.

Importance of discretisation versus distribution law. In response to Schofield et al. (2018), Zhang and Bonner (2020) demonstrated that differences in inference with capture-recapture discrete-time models were not attributed to varying data discretisation scales but rather to the choice of the distribution law for modelling the detection process. When dealing with mathematically equivalent models, both continuous- and discrete-time models would yield equivalent outcomes. Consequently, the decision between discrete- and continuous-time models is less significant and impactful than choosing a model with a different distribution. We also note that the choice of distribution law to model detection influences the model parameters. For instance, we find detection rates more practical than detection probabilities, improving the comparability of studies as probabilities are only meaningful at the scale of the discretised period.

Users select an occupancy model depending on a trade-off between model performance and implementation cost. This cost encompasses factors such as model familiarity, programming if necessary, and accessibility to data, all of which can be influenced by the complexity of the model. Existing comparisons between discrete- and continuous-time models are presented in papers introducing new continuous models, focusing on evaluating the new model formulation, and often limited to just two models (Emmet et al., 2021; Guillera-Arroita et al., 2011). Other studies compare models that use time-to-first-detection data with those using repeated measures to improve field survey methods conducted by human observers (Halstead et al., 2021; Henry et al., 2020; Priyadarshani et al., 2022).

In this paper, we investigate whether continuous-time modelling is beneficial for occupancy estimation using sensor-based observation data and under which circumstances. We conduct a comprehensive comparison of five occupancy models, varying in the complexity of their detection processes. These five models cover a large scope of single-species static occupancy models with no false positives (MacKenzie et al., 2004). We omitted time-to-first-detection models. Although often appropriate to analyse data from time-optimised human-based surveys (Halstead et al., 2021; Henry et al., 2020; Priyadarshani et al., 2022; Priyadarshani et al., 2024), using only the first detection from sensor-based data amounts to discarding lots of informative data. Therefore, we considered only time-to-each-detection for continuous-time models. We also omitted models that consider abundance-induced detection heterogeneity, like the Royle-Nichols model (Royle & Nichols, 2003).

bioRxiv preprint doi: https://doi.org/10.1101/2023.11.17.567350; this version posted February 27, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

We compare the ability of occupancy models to retrieve the simulated occupancy probability using complementary comparison metrics, measuring accuracy, bias, and precision (Liemohn et al., 2021). To fully control the environment, we simulate continuous detection data. This allows us to explore how the rarity and elusiveness of the target species influences the model's ability to retrieve the occupancy. We also simulate extreme cases to refine the models' application limits. As an illustrative example, we used the five compared models to analyse continuously collected empirical lynx (*Lynx lynx*) data observed through camera traps. We aim to offer recommendations for choosing discrete- or continuous-time models and to discuss various considerations that researchers should address when analysing fauna observation data collected through sensors.

2 Material and methods

2.1 Occupancy models

In this section, we describe the five hierarchical occupancy models compared, with an ecological process modelling presence or occupancy, and an observation process addressing imperfect detection. The occupancy sub-model is consistent across all five models, while the detection sub-model differs. Fig. 1 provides an overview of the formulation and input data of the considered models, which are described in detail in the following paragraphs. The mathematical notation is listed in Table 1.

2.1.1 Occupancy sub-model

Across all five models, the occupancy sub-model is identical, assuming that the occupancy state of site i, Z_i, follows a Bernoulli distribution with parameter ψ, the occupancy probability. The sites are assumed independent, regarding both occupancy and detection.

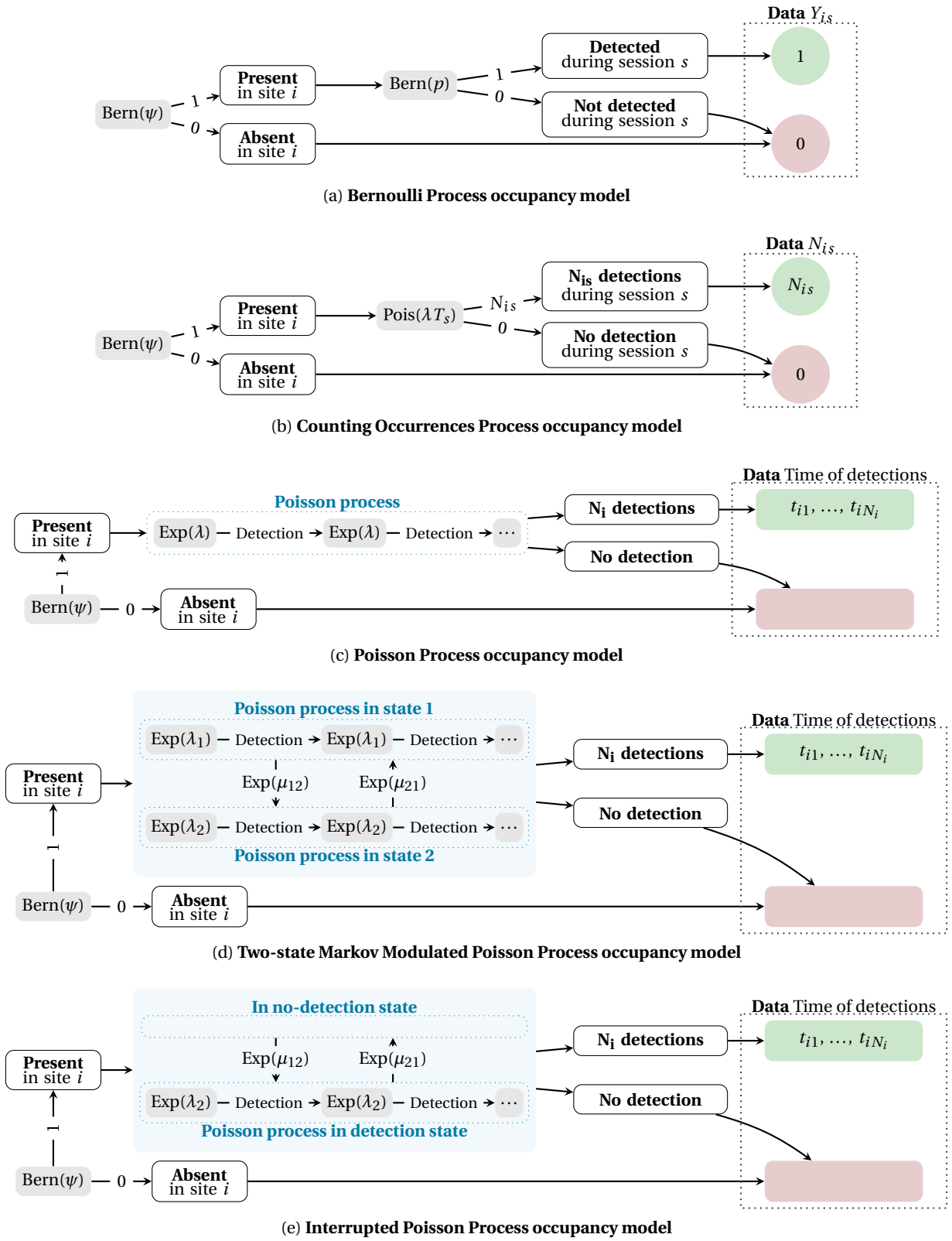
$$Z_i \overset{i.i.d.}{\sim} \text{Bernoulli}(\psi), \quad i = 1, \ldots, I. \tag{1}$$

If the species is detected at least once in a site, that site is considered occupied, assuming no false positives. Temporal changes in occupancy are not considered; for simplicity, we focus on single-season occupancy models with no covariates.

2.1.2 Detection sub-model

Two models rely on the time discretisation of the sensor-based observation data (Bernoulli Process (BP) and Counting Occurrences Process (COP)), while three others consider the detection as the realisation of a continuous-time stochastic process (Poisson Process (PP), Two-state Markov Modulated Poisson Process (2-MMPP) and Interrupted Poisson Process (IPP)). Their growing complexity, associated with an expected closer alignment with reality, influences the input data required for each model. Our primary focus is to

bioRxiv preprint doi: https://doi.org/10.1101/2023.11.17.567350; this version posted February 27, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.



(a) **Bernoulli Process occupancy model**

(b) **Counting Occurrences Process occupancy model**

(c) **Poisson Process occupancy model**

(d) **Two-state Markov Modulated Poisson Process occupancy model**

(e) **Interrupted Poisson Process occupancy model**

Figure 1: **Five occupancy models compared**. With: ψ the occupancy probability; **(a) BP** p the detection probability; Y_{is} the detection/non detection observed in site i during session s; **(b) COP** λ the detection rate; T_s the duration of a session; N_{is} the number of detections in site i during session s; **(c) PP** λ the detection rate; N_i the number of detections in site i; t_{ik} the time of the kth detection in site i; **(d) 2-MMPP** and **(e) IPP** λ_1 the detection rate in state 1; λ_2 the detection rate in state 2; μ_{12} the switching rate from state 1 to state 2; μ_{21} the switching rate from state 2 to state 1; N_i the number of detections in site i; t_{ik} the time of the kth detection in site i.

bioRxiv preprint doi: https://doi.org/10.1101/2023.11.17.567350; this version posted February 27, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Table 1: **Notation**.

I	Number of sites
ψ	Occupancy probability
Z_i	Occupancy state of site i (present = 1, absent = 0)
T_i	Deployment's duration in site i
N_{it}	Number of detections of the species in site i during t
p_t	Probability of detecting at least one individual during t
n_{sim}	Number of simulations per scenario
Discrete-time occupancy models	
T_s	Duration of a discretised session
S	Number of sessions during T_i
Bernoulli Process	
Y_{is}	Species detected in site i during session s (detection = 1, non-detection = 0)
p	Probability of detecting at least one individual during T_s
Counting Occurrences Process	
N_{is}	Number of detections of the species in site i during session s
λ	Detection rate
Continuous-time occupancy models	
N_i	Total number of detections of the species in site i
t_{ik}	Time of the kth detection in site i
Poisson Process	
λ	Detection rate
Two-state Markov Modulated Poisson Process ; Interrupted Poisson Process	
λ_1	Detection rate in state 1, with $\lambda_1 = 0$ for the IPP model
λ_2	Detection rate in state 2
μ_{12}	Switching rate from state 1 to state 2
μ_{21}	Switching rate from state 2 to state 1
π_1	Time-ratio spent in state 1 when the system is stationary
π_2	Time-ratio spent in state 2 when the system is stationary

bioRxiv preprint doi: https://doi.org/10.1101/2023.11.17.567350; this version posted February 27, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

determine if more complex representations of the detection process lead to improved estimates of occupancy probability, with minimised error and bias.

Bernoulli Process (BP) In the classical occupancy model proposed by MacKenzie et al. (2002), the raw data are aggregated and simplified. The continuous data are aggregated into S sessions of duration T_s, and simplified into the observation Y_{is}, which is 1 if at least one detection occurs during session s at site i, and 0 otherwise. Conditionally on the occupancy state Z_i of site i, the model assumes that the distribution of the variable of interest Y depends on p the probability of detecting at least one individual during a session:

$$\begin{cases} Y_{is}|Z_i = 1 & \overset{i.i.d}{\sim} \text{Bernoulli}(p), \quad i = 1, \ldots, I,\ s = 1, \ldots S, \\ Y_{is}|Z_i = 0 & \overset{i.i.d}{\sim} 0 \end{cases} \tag{2}$$

Counting Occurrences Process (COP) In the BP model, detecting few or many individuals during a session leads to the same observation $Y_{is} = 1$, although it corresponds to very different situations. We simplified the approach proposed by Emmet et al. (2021) to avoid references to secondary sessions and to use probability. As a result, its likelihood has been adjusted and is provided in supplementary information.

Although the data is aggregated by session like in the BP model, more information is retained since this approach models N_{is}, the number of individuals seen at site i during session s. Conditionally on the occupancy state Z_i of site i, as it is typical for count data, the COP model assumes that the number of detections N_{is} follows a Poisson distribution of parameter λ the detection rate multiplied by T_s the session duration:

$$\begin{cases} N_{is}|Z_i = 1 & \overset{i.i.d}{\sim} \text{Poisson}(\lambda T_s), \quad i = 1, \ldots, I,\ s = 1, \ldots S, \\ N_{is}|Z_i = 0 & \overset{i.i.d}{\sim} 0 \end{cases} \tag{3}$$

In practical terms, if the time-unit is a day, then when the detection rate $\lambda = 3$, there are on average three individuals detected by day. If each session lasts a week, $T_s = 7$, then there are on average $\lambda T_s = 3 \times 7 = 21$ individuals detected per session. The probability of detecting k individuals during a session is $(\lambda T_s)^k e^{-\lambda T_s} / k!$. With this example, in an occupied site during a session, there is a 8.67% chance of detecting 21 individuals, a 0.35% chance for 10 individuals, and a $7.58e^{-8}$% chance of detecting nothing.

Poisson Process (PP) Unlike the two previous models which required data discretisation, the PP occupancy model proposed by Guillera-Arroita et al. (2011) uses the time of detections as data, with t_{ij} the time of the jth detection in site i. These raw, unaggregated data retain all of its information. The time of detections are transformed into interdetection times to calculate the likelihood of these data given the model and its parameters. The first interdetection time is usually defined as the time between the deployment beginning and the first detection, the second as the time between the first detection and the second, and so forth. The

bioRxiv preprint doi: https://doi.org/10.1101/2023.11.17.567350; this version posted February 27, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

last value in this vector can be defined as the time between the last detection and the end of deployment. If the time at which the deployment ended is not known, *e.g.* because the battery died, the likelihood can be adapted so that this last value can be the time between the second-to-last detection and the last detection (Guillera-Arroita et al., 2011).

When the site i is occupied, the detection process is modeled as a homogeneous Poisson point process of parameter λ, the detection rate. This means that the interdetection times are exponential variables with rate λ. In practical terms, if the time-unit is a day, then a detection rate $\lambda = 3$ means that on average, three individuals are seen per day. The average time between two detections is $^1/_3$ of a day.

One property of a Poisson process of parameter λ is that the number of detections over a period of time T follows a Poisson distribution with parameter λT. This model is therefore mathematically equivalent to the COP model presented above (Zhang & Bonner, 2020). Nonetheless, using the raw data could enable ecologists to delve deeper and consider the detection rate heterogeneity with the model residuals.

Two-state Markov Modulated Poisson Process (2-MMPP) The 2-MMPP occupancy model was also proposed by Guillera-Arroita et al. (2011) and uses the time of detection events as data, transformed into interdetection times. Unlike the PP occupancy model, which assumes that detection events happen at an homogeneous rate, this model incorporates some hetereogeneity. This approach is likely more representative for many species, considering various ecological processes that can lead to temporal clustering in detection events. Examples include seasonal activity patterns, or, at finer temporal scales, daily activity patterns, and interactions with other species, among others. In the 2-MMPP occupancy model, when the site i is occupied, the detection process is modeled as a system of Poisson processes with two different rates. When the system is in state 1, the detection events are modeled by a Poisson process of parameter λ_1. In state 2, the rate is λ_2. This is a two-state continuous-time Markov chain, where the system switches from one hidden state to the other, with parameters μ_{12} (switching rate from state 1 to state 2) and μ_{21} (switching rate from state 2 to state 1).

With day as the time-unit and a set of parameters of $\lambda_1 = 1$, $\lambda_2 = 5$, $\mu_{12} = {}^1/_{15}$, $\mu_{21} = 1$, this means that:

- State 1 is a low-detection state with 1 detection per day on average (λ_1), State 2 is a high-detection state with 5 detections per day on average (λ_2)
- When the system is in state 1, there is $^1/_{15}$ switch to state 2 per day on average (μ_{12}), corresponding to 15 days spent on average in state 1 before switching to state 2 ($^1/_{\mu_{12}}$). When the system is in state 2, there is 1 switch to state 1 per day on average (μ_{21}), corresponding to 1 days spent on average in state 2 before switching to state 1 ($^1/_{\mu_{21}}$)
- The system is in state 1 for 93.75% of the deployment time on average (π_1 in Equation 4), and in state 2 for 6.25% of the time (π_2 in Equation 4)
- In an occupied site, there are on average 1.25 detections per day (Equation 5) and the variance of the number of daily detections is 4.11 (Equation 6)

bioRxiv preprint doi: https://doi.org/10.1101/2023.11.17.567350; this version posted February 27, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

The proportion of time spent in each state when the system is stationary is the steady-state vector Π of the Markov chain for a 2-MMPP, is presented in Equation 4 (Fischer & Meier-Hellstern, 1993).

$$\Pi = \begin{pmatrix} \pi_1 & \pi_2 \end{pmatrix} = \begin{pmatrix} \frac{\mu_{21}}{\mu_{12}+\mu_{21}} & \frac{\mu_{12}}{\mu_{12}+\mu_{21}} \end{pmatrix} \tag{4}$$

The number of events (here N_{it} the number of detections at site i taking place during an observation time t) of a 2-MMPP is described by its expected value $\mathbb{E}[N_{it}]$ in Equation 5 and by its variance $\mathbb{V}[N_{it}]$ in Equation 6 (see Supplementary Informations and Bhat, 1992).

$$\mathbb{E}[N_{it}] = (\lambda_1\pi_1 + \lambda_2\pi_2)\,T \tag{5}$$

$$\mathbb{V}[N_{it}] = \left(\lambda_1\pi_1 + \lambda_2\pi_2 + \frac{2({\lambda_1}^2 + {\lambda_2}^2)}{{\mu_{12}}^2{\mu_{21}}^2\left(\frac{1}{\mu_{12}} + \frac{1}{\mu_{21}}\right)^3}\right)T \tag{6}$$

The probability of having at least one detection during an observation period of duration T, written p_t, is given in Equation 7, with exp the matrix exponential function (from Guillera-Arroita et al., 2011, section 4.2).

$$p_t = 1 - \Pi \times exp\left[\left(\begin{pmatrix} -\mu_{12} & \mu_{12} \\ \mu_{21} & -\mu_{21} \end{pmatrix} - \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix}\right) \times T\right] \times \begin{pmatrix} 1 \\ 1 \end{pmatrix} \tag{7}$$

MMPPs are a type of Cox processes (Cox, 1955). 2-MMPPs can also be referred to as switched Poisson processes (SPP, Arvidsson and Harris, 1991) or as a doubly stochastic Poisson processes (Bhat, 1992, 1994). For simplicity, we focused on two-state models in this comparison. However, in specific ecological contexts, considering more states may be relevant. For more informations on MMPPs in general, with possibly more than 2 states, see Fischer and Meier-Hellstern (1993), Guillera-Arroita (2012), and Rydén (1994).

Interrupted Poisson Process (IPP) The IPP occupancy model is a special case of a 2-MMPP where there are no detections in one of the two states. This modelling approach is intuitive for ecological settings where we expect periods without detection, such as diurnal species (active and observed during the day, inactive thus unobserved at night) or gregarious species (extended periods with no detections, and at the passage of a herd, shorter periods with numerous detection events). In such contexts, the IPP model, more parcimonious with one less parameter, could be more adapted than the 2-MMPP model, which might still estimate a near-zero detection rate and produce equivalent results. Since usually, $\lambda_1 < \lambda_2$ (Skaug, 2006), an IPP is a 2-MMPP with $\lambda_1 = 0$ constrained.

bioRxiv preprint doi: https://doi.org/10.1101/2023.11.17.567350; this version posted February 27, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

2.2 Simulation study

2.2.1 Continuous detection data simulation

We simulated detection data in $I = 100$ sites, with one deployment per site of $T_i = 100$ time-units. For the sake of simplicity, one time-unit corresponds to one day throughout this article. We simulated data with various occupancy probability and detection parameters. All simulation parameters are described in Table 2. In detection scenarios (a) and (b), we simulated extreme cases of species elusiveness to identify the models' limits and behaviour in extreme situations, even if we expect these to produce insufficient data to perform occupancy modelling. We carried out $n_{sim} = 500$ simulations per simulation scenario.

Table 2: **Simulation parameters**. With p_{100} the probability of having at least one detection during a deployment of $T_i = 100$ days at an occupied site (Equation 7); p_1 the probability of having at least one detection during 1 day (Equation 7); $\mathbb{E}[N_{100}]$ the expected number of detections during a deployment of $T_i = 100$ days at an occupied site (Equation 5); $\mathbb{V}[N_{100}]$ the variance of the number of detections during a deployment of $T_i = 100$ days at an occupied site (Equation 6)

(a) General parameters

I	100 sites
T_i	100 days
n_{sim}	500 simulations per scenario
ψ	0.10, 0.25, 0.50, 0.75, 0.90
T_s	30 (month), 7 (week), 1 (day)

(b) Parameters of the seven detection scenarios

	λ_1	λ_2	μ_{12}	μ_{21}	p_{100}	p_1	$\mathbb{E}[N_{100}]$	$\mathbb{V}[N_{100}]$
(a)	0.00	1.00	$^1/_{15}$	24	0.23	0.003	0.28	0.30
(b)	0.00	5.00	$^1/_{15}$	24	0.68	0.01	1.39	1.96
(c)	0.00	1.00	$^1/_{15}$	1	0.96	0.04	6.25	17.24
(d)	0.25	0.25	$^1/_{15}$	$^1/_{10}$	1.00	0.22	25.00	61.00
(e)	0.00	5.00	$^1/_{15}$	1	1.00	0.09	31.26	306.03
(f)	0.00	1.00	$^1/_{15}$	$^1/_{10}$	1.00	0.26	40.01	327.98
(g)	0.00	5.00	$^1/_{15}$	$^1/_{10}$	1.00	0.42	200.06	7399.34

The occupancy status of each site was determined as the outcome of a Bernoulli trial with probability ψ. The detection process was simulated as a 2-MMPP of parameters $\lambda_1, \lambda_2, \mu_{12}, \mu_{21}$, using R version 4.2.3 (R Core Team, 2023). The state at the beginning of a deployment was drawn according to the stationary distribution, as a random sampling with probability π_1 (*resp.* π_2) of being in state 1 (*resp. 2*). Until the end of the deployment, the time to next event was a draw from an exponential distribution with parameter $\mu_{12} + \lambda_1$ in state 1, and with parameter $\mu_{21} + \lambda_2$ in state 2. In state 1, this event was either a detection with probability $\frac{\lambda_1}{\mu_{12}+\lambda_1}$, or a switch to state 2. In state 2, it was either a detection with probability $\frac{\lambda_2}{\mu_{21}+\lambda_2}$, or a switch to state 1 (Fig. 2).

Discretisation into sessions For the two models that required discretisation into sessions, we used three levels of discretisation: monthly, weekly, and daily. Incomplete sessions are deemed invalid and will be excluded from the analysis. Consequently, when the data is discretised into months, there are three sessions consisting of 30 days each, and the detection data of the last 10 days of each deployment is disregarded. Similarly, when the data is discretised into weeks, there are 14 sessions of 7 days each, the last 2 days of each deployment is discarded.

bioRxiv preprint doi: https://doi.org/10.1101/2023.11.17.567350; this version posted February 27, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

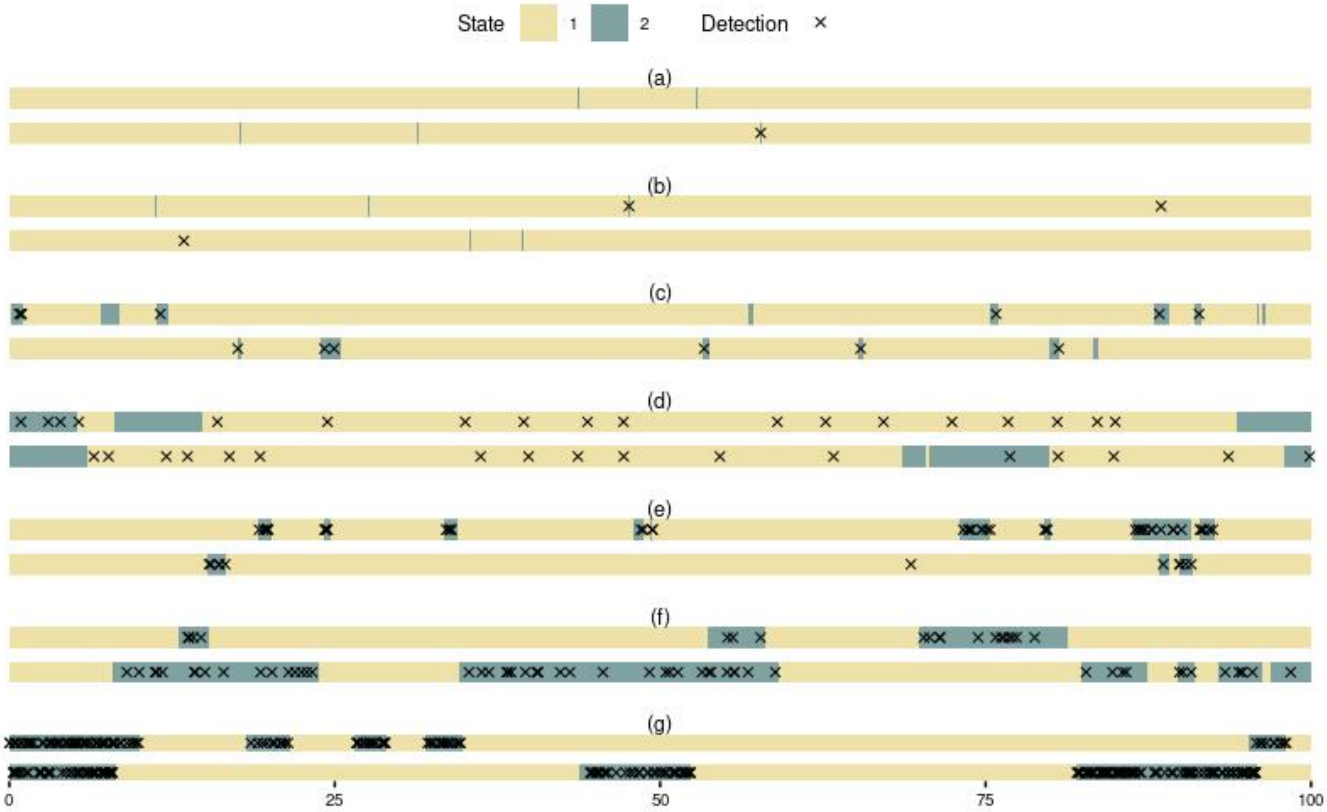


Figure 2: **Simulated detection data**. To help understand the impact of the detection parameters, two examples are given per detection scenario. With scenarios (a to g) described in Table 2. The detection process is simulated in an occupied site during 100 days.

2.2.2 Frequentist parameter estimation

We estimated models parameters by maximum likelihood estimation and implemented it in R version 4.2.3 (R Core Team, 2023). For the COP, PP, 2-MMPP and IPP models, we used the `optim` function from the `stats` package (R Core Team, 2023) to maximise the log-likelihood. For the BP model, we used the function `occu` from the `unmarked` package version 1.3.2 (Fiske & Chandler, 2011), which calls the same `optim` function. We used the `Nelder-Mead` algorithm to maximise the likelihood. To reduce the optimisation time, we used the simulated parameters as the initial parameters to start the optimisation algorithm. The likelihood maximisation methodology was equivalent for the 5 models, making their results comparable. In order to perform unconstrained optimisation, we applied a logit transformation to the probabilities (ψ, p) and a log transformation to rates (λ, λ_1, λ_2, μ_{12} and μ_{21}). In addition, we fitted the models with the BFGS optimisation algorithm. The results are not shown here but presented in supplementary information.

2.2.3 Performance comparison for occupancy probability estimation

For each simulation scenario, we calculated the Root Mean Square Error (RMSE, Equation 8) as an error metric, measuring the absolute difference between the models' point estimates of occupancy probability ($\hat{\psi}$) and the ground-truth occupancy probability (ψ), used to simulate data sets of this simulation scenario.

bioRxiv preprint doi: https://doi.org/10.1101/2023.11.17.567350; this version posted February 27, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

$$\text{RMSE} = \sqrt{\frac{1}{n_{sim}} \sum_{j=1}^{n_{sim}} \left(\hat{\psi}_j - \psi\right)^2} = \sqrt{\overline{\left(\hat{\psi} - \psi\right)^2}} \tag{8}$$

To complete this metric, we calculated absolute bias (AB, Equation 9) to better understand if this error was due to under-estimation or over-estimation of ψ.

$$\text{AB} = \frac{1}{n_{sim}} \sum_{j=1}^{n_{sim}} (\hat{\psi}_j - \psi) = \overline{\left(\hat{\psi} - \psi\right)} \tag{9}$$

To compare the distributions of the point estimates $\hat{\psi}$ of the five different models and the different discretisations for BP and COP, we performed a Kruskal-Wallis test for each simulated scenario. We also conducted Wilcoxon tests with Bonferroni correction and visualised the distribution of $\hat{\psi}$.

We calculated for each inference the 95% confidence interval (CI) of the occupancy probability. To summarise this information for all the n_{sim} simulations by model in each simulation scenario, we used two metrics, the coverage (Equation 10) and the average range of the confidence interval (ARCI, Equation 11). We note CI_l and CI_u the lower and upper bounds of the 95% confidence interval of the estimated occupancy probability.

Coverage is the proportion of simulations for which the true simulated occupancy probability (ψ) is within the 95% CI of the estimated occupancy probability. In other words, coverage can be interpreted as the percentage of good predictions of the occupancy probability by a model.

$$\text{Coverage} = \frac{1}{n_{sim}} \sum_{j=1}^{n_{sim}} I(CI_l \leq \psi \leq CI_u) = \overline{I(CI_l \leq \psi \leq CI_u)} \tag{10}$$

The average range of the 95% confidence interval measures the precision of the estimation, with the width of the confidence interval. It completes coverage, since even a model with poor performances can have a coverage of 100%: If its range is 1, it means that this model predicts that the occupancy probability is between 0 and 1.

$$\text{ARCI} = \frac{1}{n_{sim}} \sum_{j=1}^{n_{sim}} CI_u - CI_l = \overline{CI_u - CI_l} \tag{11}$$

2.3 Empirical comparison with lynx detection data

We applied the five occupancy models to empirical data from Gimenez et al. (2022). We analyzed data from 11 camera-trap locations in Ain county, France, monitored from February 2017 to May 2019. Monitoring durations per site were heterogeneous, ranging from 148 to 801 days (Fig. S8), totaling 5396 camera-trap days across all 11 sites. We here focus on lynx (*Lynx lynx*) occupancy. Lynx were detected in 9 sites out of 11, with 203 detections in total, ranging from 1 to 59 detections per site (Fig. S8). For the two discrete models, BP and COP, we discretised the data similarly to the simulation study: into monthly, weekly, and daily intervals (Fig. S9). We discarded

bioRxiv preprint doi: https://doi.org/10.1101/2023.11.17.567350; this version posted February 27, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

data from incomplete sessions, during which a site was monitored only partially and not throughout the entire session.

We estimated parameters following the method described in Section 2.2.2, consistent with the simulation study approach, with the exception of the optimisation algorithm starting points. We used intuitive starting points, such as the ratio of sites with at least one detection for ψ. Starting points for the models' different detection parameters are detailled in our code. We did not include covariates in the analysis. For each parameter of each model, we retrieved its point estimate and derived its 95% and 50% confidence intervals from the Hessian matrix.

3 Results

3.1 Simulation study

For easily detectable species (detection scenarios d, e, f, and g), all models retrieve well the simulated occupancy probability. Bias ranges from -0.0094 to 0.0025 (Fig. 3) and RMSE are less than 0.060 (Fig. S2). With those detection parameters, the Kruskal-Wallis tests indicate that there are no statistically significant differences in the distribution of $\hat{\psi}$ between models, except with simulation parameters (e) and $\psi = 0.1$, (e) and $\psi = 0.9$ and (f) and $\psi = 0.9$ (Table S3). The Wilcoxon tests indicate that there is no difference in medians with (e) and $\psi = 0.1$ (Fig. S3). With (e) and $\psi = 0.9$ and (f) and $\psi = 0.9$, only the BP model with daily sessions differs from the others, with a slight underestimation of ψ (Fig. 3).

For elusive species (detection scenario c), the BP model's ability to retrieve the simulated occupancy probability is slightly inferior to other models, with a RMSE ranging from 0.057 to 0.121 while the RMSE of other models are still less than 0.060. (Fig. S2). The Wilcoxon tests (Fig. S3) indicate differences between BP and the other models, and this difference depends on the discretised session duration. The distribution of $\hat{\psi}$ with BP is wider than for the other models with the same simulated data (Fig. S1).

For highly elusive species (detection scenarios a and b), all five models reach their limits. BP, COP and PP tend to overestimate ψ, whereas IPP and 2-MMPP tend to underestimate ψ (Fig. 3). BP tend to estimate ψ at 0 or most often at 1 (Fig. S1). COP and PP point estimates of ψ have similar distributions, both are widely spread (Fig. S1). IPP and 2-MMPP tend to underestimate ψ, with a tighter distribution for its point estimate, which often does not include the simulated value of ψ (Fig. S1).

It was not always possible to calculate the confidence interval (CI) of the occupancy probability estimate, when the Hessian matrix was not invertible. This occurred in two main cases in our study: when there were not many sessions with detections in the BP model, or when λ_1 was estimated to zero in the 2-MMPP model. As a result, the 2-MMPP CIs are not interpretable with detection scenarios other than (d), where data were simulated as an IPP.

For easily detectable species (detection scenarios e, f and g), all models have similar coverages (Fig. S4) and

bioRxiv preprint doi: https://doi.org/10.1101/2023.11.17.567350; this version posted February 27, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

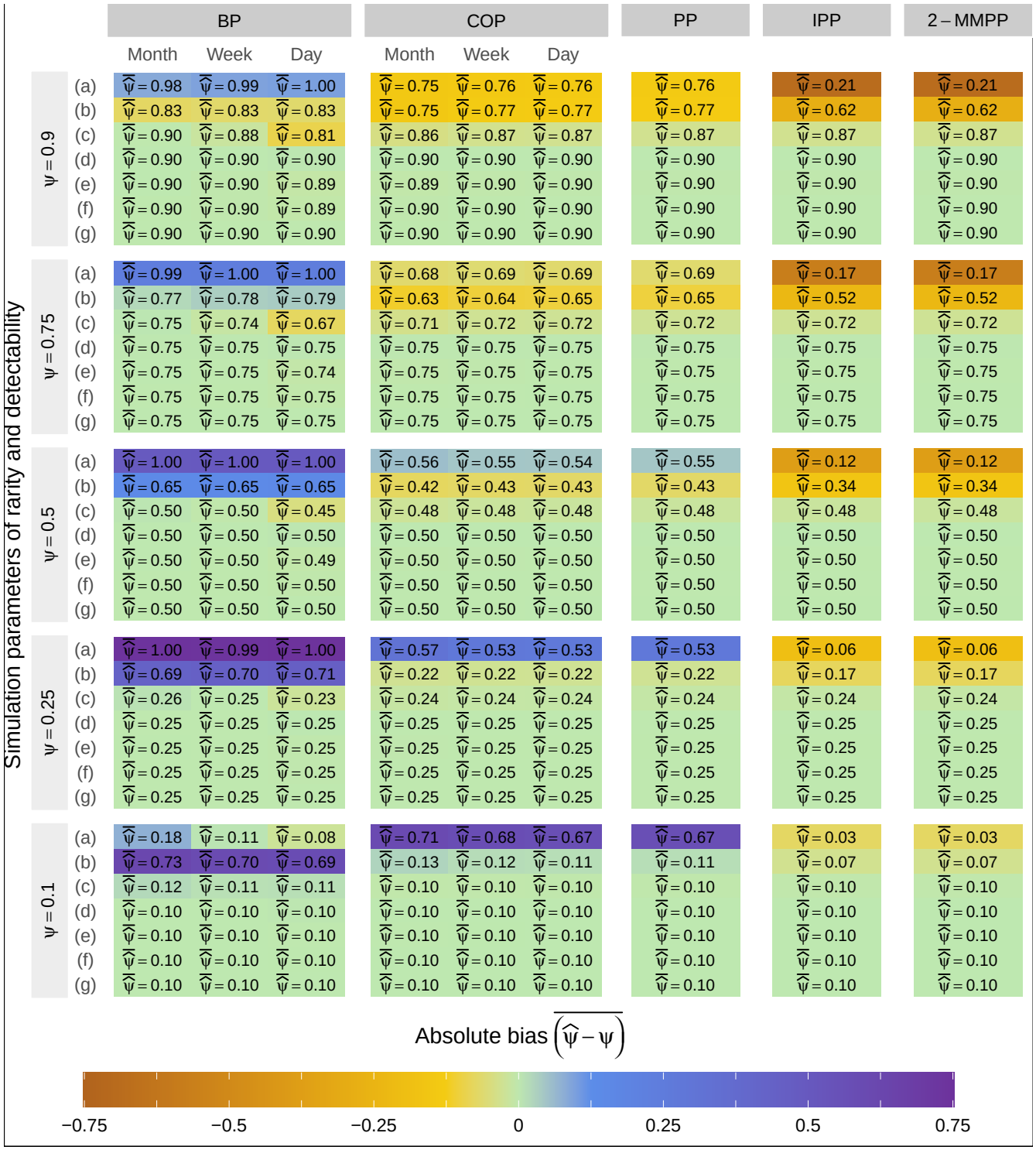


Figure 3: **Absolute bias of the occupancy probability point estimate**. Depending on ψ the simulated occupancy probability and detection scenarios as described in Table 2. The average value of the occupancy probability point estimate $\bar{\hat{\psi}}$ is inside each cell. For two scenarios characterised by low occupancy and detection probabilities, certain repetitions failed to yield any data. With no detection within any of the sites, it was impossible to infer parameters. With detection parameters (a) and $\psi = 0.25$, 494 simulations were used to estimate the models' ability to retrieve the simulation parameters. With detection parameters (a) and $\psi = 0.1$, only 423 simulations were used.

bioRxiv preprint doi: https://doi.org/10.1101/2023.11.17.567350; this version posted February 27, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

occupancy probability CI ranges (Fig. S5). As detectability decreases, the CIs widens for BP, COP and PP, although this is more marked and quicker for BP than for COP and PP (Fig. S5). The IPP CIs do not widen, but the coverage drops (Fig. S4).

3.2 Application to lynx occupancy

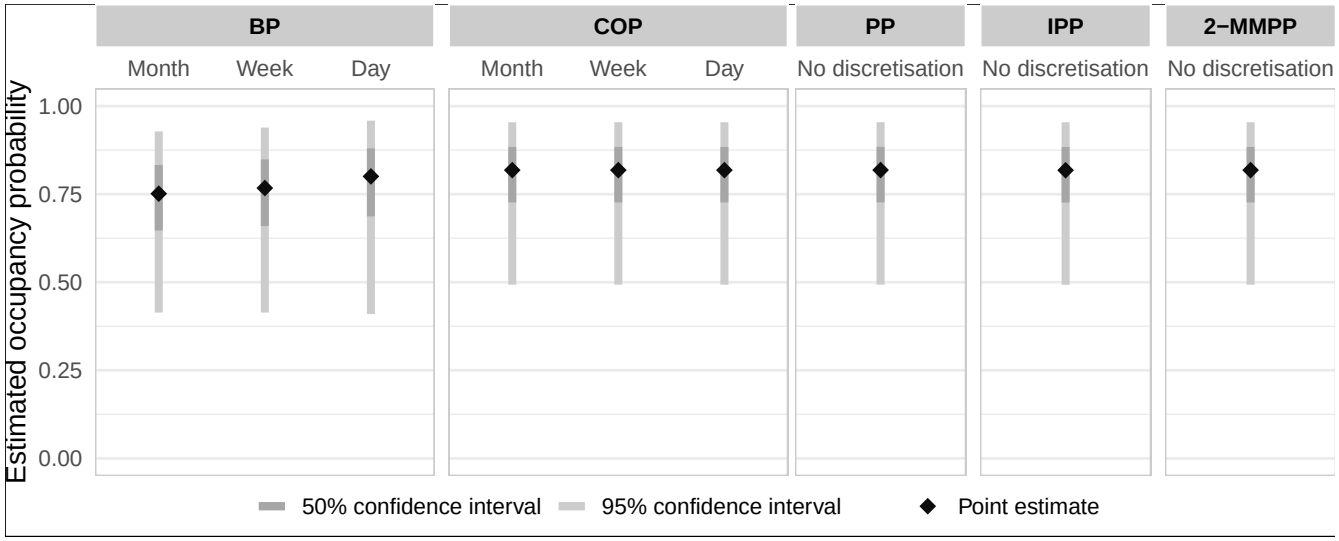


Figure 4: **Occupancy probability estimation with the lynx data set from Gimenez et al. (2022) in the Ain County, France**.

Lynx occupancy estimates are similar across all five model (Fig. 4). The COP model, regardless of the session duration, and the three continuous models (PP, IPP, and 2-MMPP), converge to an estimated occupancy probability of 0.82, with identical confidence intervals (95% CI: 0.49 to 0.95, 50% CI: 0.73 to 0.88) (Fig. 4, Table S4). Although the BP model produced a slightly different point estimate, ranging between 0.75 and 0.80 depending on the session duration, it remains in close proximity to the estimates of the other four models (Fig. 4, Table S4).

Regarding detectability, the BP model with daily sessions estimates the probability of detecting at least one lynx in a day at $p_1 = 0.008$ (95% CI: 0.006 to 0.012, Table S4). Models assuming homogeneous detection rates (COP, PP) estimate 0.045 detection events per day (95% CI: 0.039 to 0.052, Table S4), resulting in an expected average of $\mathbb{E}[N_{100}] = 0.045 \times 100 = 4.5$ detection events in 100 days. In comparison to detectability in our simulation scenarios, lynx elusiveness likely falls between detection scenario (b) ($p_1 = 0.01$) and scenario (c) ($\mathbb{E}[N_{100}] = 6.25$) (Table 2).

The 2-MMPP estimates a low detection rate in state 1, close to 0 ($\widehat{\lambda_1} < 0.01$, 95% CI: 0.001 to 0.06, Table S4), suggesting that lynx detection events can be modeled by the the IPP model. The IPP model estimates that only 0.31% of the time is spent in the state with detections, state 2, (Equation 4, with $\widehat{\mu_{21}} = 0.16$ and $\widehat{\mu_{21}} = 50.70$, Table S4). The detection rate is estimated at $\widehat{\lambda_2} = 14.22$ detection events per day in this state (95% CI: 9.24 to 21.89, Table S4).

bioRxiv preprint doi: https://doi.org/10.1101/2023.11.17.567350; this version posted February 27, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

4 Discussion

The focal ecological parameter of interest is the occupancy probability ψ, which is represented similarly in all the five models compared. However, the precision of the occupancy estimation is impacted by the quality of the estimation for the detection process (Kellner & Swihart, 2014; Kéry & Schmidt, 2008). In this study, we focused on cases in which data is collected continuously, for example with sensors or opportunistic data. We aimed to evaluate whether modelling the detection process in continuous-time could enhance the precision of the estimated probability of occupancy.

In line with the concept of operating models, commonly used for assessing management strategies (Butterworth, 1999; Punt et al., 2016), we simulated data under models that aim to get more in line with our expectations regarding the detection process when observing fauna through sensors. Specifically, we considered special cases of 2-MMPP, consisting of six scenarios with detections simulated under an IPP framework (scenarios a, b, c, e, f, g) and one scenario simulated under a PP framework (scenario d). We intentionally simulated occupancy very simply, not incorporating covariates or variations in abundance, aiming to focus solely on the detection process for sensor-collected data. Subsequently, we aimed to recover the simulation parameters, focusing on occupancy probability, using these complex models, as well as simpler models well-known and widely used by ecologists. By simplifying the information and the detection process, we asked the question of whether these models are sufficient to estimate the ecological parameter of interest in a situation that we expect to be close to reality.

We expected that continuous models would outperform discrete models in accurately retrieving the simulated occupancy probability, since detection data simulation aligned with the framework of the continuous models of our comparison. Moreover, data aggregation by discretisation leads to a loss of information. We expected this loss to result in less accurate estimates of detection parameters, consequently introducing bias in the occupancy estimate, given the relationship between occupancy and detection parameters. However, in the majority of cases where detectability was sufficiently high (with a minimum expectation of 25 detections in occupied sites throughout the entire deployment), all models produced equivalent results, all were able to retrieve the occupancy probability well, with little bias and error.

For models requiring discrete data, we expected that different discretisations would impact the models outputs (Schofield et al., 2018), but in most simulated scenarios, that was not the case. Our results indicate that estimation of ψ with BP is more impacted by the session duration's choice than with COP. Since COP is mathematically equivalent to PP (Zhang & Bonner, 2020), minor variations in the occupancy estimates between session lengths for COP are likely due to data discarding. Our comparison framework could be reused to further test the impact of discretisation, by choosing more diverse session durations that reuse exactly the same data - rather than dropping data of incomplete sessions as we did.

The BP model, as noted by MacKenzie et al. (2002), tends to produce estimates of ψ close to one for rare and elusive species. Our findings align with this observation, suggesting however that elusiveness has a more

bioRxiv preprint doi: https://doi.org/10.1101/2023.11.17.567350; this version posted February 27, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

pronounced impact on this limit than rarity.

The COP model was adapted from the model proposed by Emmet et al. (2021). Their model differs from the one presented here mainly because they considered site use. However, they compared their counting model with its detection/non-detection equivalent from Bled et al. (2013), much like we compared COP with BP. Their model estimated occupancy probability with either equivalent or smaller bias compared to the equivalent detection/non-detection model, which aligns with our results.

In a simulation study, Guillera-Arroita et al. (2011) evaluated BP and PP using data generated within a PP framework. They reported that both models provided reasonably unbiased estimates of occupancy, except for rare and elusive species. In these cases, BP exhibited greater bias and variance, particularly with larger discretisation intervals and fewer sessions, which matches our results. They also compared PP and 2-MMPP using clustered detection data generated within an IPP framework. They noted negative bias in the occupancy estimates with the PP model, which was not observed in our results. In our study, both models performed similarly for easily detectable species. However, for elusive species, the 2-MMPP and IPP models exhibited more pronounced negative bias than the PP and COP models.

All five models consistently estimated lynx occupancy in our empirical example, with COP, PP, IPP, and 2-MMPP estimating occupancy probability at 0.82, while BP provided slightly lower but still very close estimates. Lynx detectability falls between simulation scenarios (b) and (c), where models began to approach their limits due to elusiveness in our simulations. As different models produced significantly varied occupancy estimates when they reached their limits in simulations, the consistent occupancy estimates across models suggest that they have not yet reached their limits, indicating reliable occupancy estimates.

To better define the limitations of these models, we could perform additional comparisons using simulation scenarios with various detection parameters. Given the impossibility of exhaustively covering all potential scenarios, we encourage modelers to compare models when encountering borderline cases of occupancy models applicability, such as potentially insufficient monitoring time in view of the species elusiveness. Our code is available to use as a base to conduct simulations with parameters adapted to a specific study context and compare models to choose the best model. Alternatively, researchers and practitioners can analyse their data with multiple models to ensure the consistency of results across different modelling approaches, as we did with the lynx data set.

4.1 Choosing the appropriate model

4.1.1 Occupancy modelling for easily detectable species

When the species is easily detectable and thus enough observation data have been obtained, all models accurately estimate the occupation probability. Under these conditions, if the sole aim of a study is to accurately estimate occupancy, selecting any of these models essentially amounts to choosing the right one. Therefore, the choice can be guided by other considerations, to find the right balance between performance

bioRxiv preprint doi: https://doi.org/10.1101/2023.11.17.567350; this version posted February 27, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

and execution costs.

Learning and implementation costs Continuous-time models may be unfamiliar to ecologists, potentially requiring a steep learning curve to become proficient with these seemingly complex models. For models that are not readily available, the implementation costs can be substantial for a study. The rise of sensor-based monitoring and the growing interest have led to efforts to make time-to-detection occupancy models more accessible for ecologists, such as through R packages like `unmarked` (Kellner et al., 2023). The costs shifts from fully implementing a model to using existing functions, a faster alternative. We are currently working on adding the COP, PP, IPP and 2-MMPP models to `unmarked`.

Study objectives If the primary goal is to estimate the occupancy of the target species, any of the models can be employed effectively. Simple models, such as BP, COP or PP, require the estimation of only two parameters: one for occupancy and one for detection. Choosing such a model can enhance interpretability and provide a greater statistical power than models with more parameters. This is especially advantageous when incorporating several spatial and temporal covariates into the analysis. Conversely, if the aim is to conduct a detailed analysis of the target species detection timeline to better understand a species behaviour, analysing continuous-time data is preferable, as it data holds valuable information that is increasingly lost as data are more and more aggregated over time. Models that accommodate the detection process in multiple states could be particularly adapted to unravel animal behaviours, such as temporal activity patterns. The hidden state could be reconstructed to further enhance the interpretability of these models' parameters.

Temporal auto-correlation Unlike sampling occasions, consecutive discretised sessions are not temporally independent (Bailey et al., 2014), and there may be significant temporal auto-correlation (Neilson et al., 2018). Therefore, discretised session data does not meet the discrete-time model assumption of independence. However, the PP model has the exact same drawback when considering a constant detection rate, since the number of events on two disjoints time intervals are independent. In this study, we did not thoroughly examine the influence of time dependence on occupancy estimates. However, two-state models, introduced for clustered observation data (Guillera-Arroita et al., 2011), incorporate some time dependence through two homogeneous detection rates that differ conditional on state. Future studies could explore the impact of time dependence on occupancy estimates and consider various approaches to account for non-constant detection rates. These may include Cox processes, where the detection rate is a random variable (Cox, 1955), time-dependent regression splines (Distiller et al., 2020), or Hawkes processes, a form of self-exciting point process where the detection rate increases temporarily after a detection event (Hawkes, 1971; Rushing, 2023).

Calculation time All models were fairly fast to fit, so calculation time should probably not be the main reason for choosing a model for most studies. We have not robustly evaluated the optimisation time for each model, as we used different computers with varying characteristics. However, the two-state models seemed significantly longer to fit than the other models. BP, COP and PP all took less than 6 seconds to fit, even in the simulation

bioRxiv preprint doi: https://doi.org/10.1101/2023.11.17.567350; this version posted February 27, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

scenario with most detections, in which there was 200 detections on average in occupied sites. IPP and 2-MMPP often took more than a minute, up to 28 minutes.

Detection rate A detection probability per discretised session, as in the BP model, is relevant only at the discretisation scale. This is not the case with a detection rate, as used in the discrete-time COP model or in continuous-time models. We argue that using a detection rate instead of a detection probability would enhance the comparability among studies. This could simplify the process of experimental design, especially concerning observation duration, by using the insights from existing literature on the target species. Another advantage of using a detection rate, as opposed to a probability, is the flexibility to accommodate sessions of different durations. For instance, in our lynx example, monthly sessions varied in length (28, 30, or 31 days). We specified this in the COP model to estimate the number of detection per day, an explicit time unit. We could not do so in the BP model, in which the detection probability is derived per session.

4.1.2 Occupancy modelling for highly elusive species

When the species is highly elusive, the five models we considered provided inaccurate estimates of its presence probability, exhibiting high bias, error, and a low precision or coverage. The BP model's limits became apparent at lower species elusiveness compared to the other models. This could be because valuable information gets lost when simplifying the data into detection and no detection. The 2-MMPP and IPP models showed larger errors in estimating ψ compared to the COP and PP models. This might be due to the higher number of parameters in the 2-MMPP and IPP models (5 and 4, respectively, versus 2 for COP and PP), which would require more data to fit them correctly. COP and PP models appear to strike a good balance between simplification and realism. One is discrete, while the other is continuous, but both perform similarly, which is consistent with the demonstration of Zhang and Bonner (2020) that a Poisson process in continuous time is equivalent to a classical model with discretisation where the detection process is not a Bernoulli trial but a Poisson distribution draw.

However, if the species' high elusiveness resulted in the collection of insufficient observation data, the best course of action probably is to collect more data by extending the monitoring period (Kays et al., 2020). In cases where it is expected that the species will be challenging to detect, conducting simulations and comparing different models with expected detection and occupancy parameters could assist in fine-tuning the study design and model choice.

If obtaining more data is not feasible, it might be best to refrain from running an occupancy model, or at least approach the results with caution, regardless of the chosen model. In this case, we recommend fitting different models, particularly when using the two-state models. For these models, our findings indicate that with highly elusive species, the confidence interval of the estimated ψ can be narrow but substantially different from the actual ψ. This can potentially lead to a misleading perception of model reliability.

bioRxiv preprint doi: https://doi.org/10.1101/2023.11.17.567350; this version posted February 27, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

4.2 Implications for continuous monitoring frameworks

The advanced processors available today offer great computing power, enabling the fast development of artificial intelligence (AI). Recognising species automatically is becoming more common, on camera-trap images (Le Borgne & Bouget, 2023), ARUs recordings (Potamitis et al., 2014), or even with sensors networks (Wägele et al., 2022). AI combined with sensors offers the potential to fully automate the analysis workflow (Gimenez et al., 2022; Lahoz-Monfort & Magrath, 2021). Overall, sensors and AI have led to a paradigm shift in the conditions and capabilities of biodiversity monitoring (Besson et al., 2022; Tuia et al., 2022; Zwerts et al., 2021). With our comparison, we found that modelling occupancy with a continuous-time detection sub-model is not necessary to estimate occupancy accurately. Therefore, in operational conditions, the necessary trade-off between accuracy and ease of implementation may turn in favour of discrete-time models, with easily available data for temporal covariates.

Focusing primarily on the temporal aspect of the detection process modelling, we explored a simple version of occupancy models: single-species, static, with no positive and no abundance-induced heterogeneity. There are numerous avenues for further investigation. With the emergence of collaborative platforms aggregating and providing sensor data across large spatiotemporal scales (e.g. Oliver et al., 2023), developing dynamic occupancy models with continuous-time occupancy sub-models has the potential to enhance our understanding of extinction-colonisation processes, to get the most out of continuous-time data. Additionally, we did not consider abundance as a factor influencing detection heterogeneity, although it is a particularly intuitive consideration for species with highly variable abundance across sites (Royle & Nichols, 2003). Further research into these models, in conjunction with continuous data, is called for: In N-mixture models, time-to-first-detection and time-to-each-detection could offer improved estimates compared to binary detection/non-detection and count data, respectively (Haines et al., 2023; Priyadarshani et al., 2024). Furthermore, our study did not account for false-positives, and sensor data (e.g., images from camera traps) can be prone to incorrect species identification. Potential solution, such as using AI confidence scores (Rhinehart et al., 2022), could be explored further with continuous-time data to consider the complete workflow, from sensor data to inference.

Our results do not only concern sensor data, but all continuously collected data. Opportunistic data, collected at non-defined and irregular intervals, can be considered as continuously collected and thus pose some of the same challenges as sensor data (Altwegg & Nichols, 2019; Hsing, 2019). Some studies analyse opportunistic data using with discrete-time models, discretising data into long sessions (*e.g.*, by year, as in van Strien et al., 2013), while other are developing new continuous-time frameworks adapted to this particular data (*e.g.* Choquet et al., 2017, using continuous-time capture-recapture for individualised opportunistic data). Our comparison suggests that for future studies aiming to estimate occupancy with unmarked opportunistic data, both discrete- and continuous-time occupancy models could produce accurate occupancy estimates, provided that challenges associated with opportunistic data, such as highly variable observation effort, are effectively addressed.

bioRxiv preprint doi: https://doi.org/10.1101/2023.11.17.567350; this version posted February 27, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Acknowledgements

LP benefits from a French government CIFRE grant for PhD students. This work is part of the PSI-BIOM project granted by the French PIA 3 under grant number 2182D0406-A.

We thank the Federations of Hunters from the Jura and Ain counties for sharing the lynx camera trap data. Data collection was carried out through the Lynx Predator Prey Program, which was funded by Auvergne-Rhone-Alpes Region, Ain and Jura departmental Councils, the French National Federation of Hunters, French Environmental Ministry based in Auvergne-Rhone-Alpes and Bourgogne-Franche-Comte Region and the French Office for Biodiversity.

Conflict of Interest statement

The authors declare no conflict of interest.

Author Contributions

- LP: Formal analysis, Investigation, Methodology, Visualisation, Writing – original draft.
- SM, OG, MPE: Validation, Writing – review and editing.
- All authors: Conceptualisation.

Code availability

Code is available at:

https://oikolab.terroiko.fr:10001/publications/occupancy-modelling-comparison-discrete-continuous

bioRxiv preprint doi: https://doi.org/10.1101/2023.11.17.567350; this version posted February 27, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

References

Altwegg, R., & Nichols, J. D. (2019). Occupancy models for citizen-science data. *Methods in Ecology and Evolution*, *10*(1), 8–21. https://doi.org/10.1111/2041-210X.13090

Arvidsson, A., & Harris, R. (1991). Performance comparison of models of individual and merged bursty traffics. *Sixth Australian Teletraffic Research Seminar*.

Bailey, L. L., MacKenzie, D. I., & Nichols, J. D. (2014). Advances and applications of occupancy models (E. Cooch, Ed.). *Methods in Ecology and Evolution*, *5*(12), 1269–1279. https://doi.org/10.1111/2041-210X.12100

Becker, N. G. (1984). Estimating Population Size from Capture-Recapture Experiments in Continuous Time. *Australian Journal of Statistics*, *26*(1), 1–7. https://doi.org/10.1111/j.1467-842X.1984.tb01261.x

Besson, M., Alison, J., Bjerge, K., Gorochowski, T. E., Høye, T. T., Jucker, T., Mann, H. M. R., & Clements, C. F. (2022). Towards the fully automated monitoring of ecological communities. *Ecology Letters*, *25*(12), 2753–2775. https://doi.org/10.1111/ele.14123

Bhat, V. N. (1992). Switched poisson process/g/1 queue with service interruptions. *Computers & Operations Research*, *19*(8), 751–756. https://doi.org/10.1016/0305-0548(92)90014-V

Bhat, V. N. (1994). Renewal Approximations of the Switched Poisson Processes and Their Applications to Queueing Systems. *Journal of the Operational Research Society*, *45*(3), 345–353. https://doi.org/10.1057/jors.1994.47

Bled, F., Nichols, J. D., & Altwegg, R. (2013). Dynamic occupancy models for analyzing species' range dynamics across large geographic scales. *Ecology and Evolution*, *3*(15), 4896–4909. https://doi.org/10.1002/ece3.858

Burton, A. C., Neilson, E., Moreira, D., Ladle, A., Steenweg, R., Fisher, J. T., Bayne, E., & Boutin, S. (2015). REVIEW: Wildlife camera trapping: A review and recommendations for linking surveys to ecological processes. *Journal of Applied Ecology*, *52*(3), 675–685. https://doi.org/10.1111/1365-2664.12432

Butterworth, D. (1999). Experiences in the evaluation and implementation of management procedures. *ICES Journal of Marine Science*, *56*(6), 985–998. https://doi.org/10.1006/jmsc.1999.0532

Choquet, R., Garnier, A., Awuve, E., & Besnard, A. (2017). Transient state estimation using continuous-time processes applied to opportunistic capture–recapture data. *Ecological Modelling*, *361*, 157–163. https://doi.org/10.1016/j.ecolmodel.2017.08.001

Cove, M. V., Kays, R., Bontrager, H., Bresnan, C., Lasky, M., Frerichs, T., Klann, R., Lee Jr., T. E., Crockett, S. C., Crupi, A. P., Weiss, K. C. B., Rowe, H., Sprague, T., Schipper, J., Tellez, C., Lepczyk, C. A., Fantle-Lepczyk, J. E., LaPoint, S., Williamson, J., … McShea, W. J. (2021). SNAPSHOT USA 2019: A coordinated national camera trap survey of the United States. *Ecology*, *102*(6), e03353. https://doi.org/10.1002/ecy.3353

Cox, D. R. (1955). Some Statistical Methods Connected with Series of Events. *Journal of the Royal Statistical Society: Series B (Methodological)*, *17*(2), 129–157. https://doi.org/10.1111/j.2517-6161.1955.tb00188.x

Cusack, J. J., Dickman, A. J., Rowcliffe, J. M., Carbone, C., Macdonald, D. W., & Coulson, T. (2015). Random versus Game Trail-Based Camera Trap Placement Strategy for Monitoring Terrestrial Mammal Communities (R. Guralnick, Ed.). *PLOS ONE*, *10*(5), e0126373. https://doi.org/10.1371/journal.pone.0126373

bioRxiv preprint doi: https://doi.org/10.1101/2023.11.17.567350; this version posted February 27, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Distiller, G. B., Borchers, D. L., Foster, R. J., & Harmsen, B. J. (2020). Using Continuous-Time Spatial Capture–Recapture models to make inference about animal activity patterns. *Ecology and Evolution*, *10*(20), e6822. https://doi.org/10.1002/ece3.6822

Emmet, R. L., Long, R. A., & Gardner, B. (2021). Modeling multi-scale occupancy for monitoring rare and highly mobile species. *Ecosphere*, *12*(7), e03637. https://doi.org/10.1002/ecs2.3637

Fischer, W., & Meier-Hellstern, K. (1993). The Markov-modulated Poisson process (MMPP) cookbook. *Performance Evaluation*, *18*(2), 149–171. https://doi.org/10.1016/0166-5316(93)90035-S

Fiske, I., & Chandler, R. (2011). Unmarked: An R Package for Fitting Hierarchical Models of Wildlife Occurrence and Abundance. *Journal of Statistical Software*, *43*, 1–23. https://doi.org/10.18637/jss.v043.i10

Gilbert, N. A., Clare, J. D. J., Stenglein, J. L., & Zuckerberg, B. (2021). Abundance estimation of unmarked animals based on camera-trap data. *Conservation Biology: The Journal of the Society for Conservation Biology*, *35*(1), 88–100. https://doi.org/10.1111/cobi.13517

Gimenez, O., Kervellec, M., Fanjul, J.-B., Chaine, A., Marescot, L., Bollet, Y., & Duchamp, C. (2022). Trade-off between deep learning for species identification and inference about predator-prey co-occurrence. *Computo*. https://doi.org/10.57750/yfm2-5f45

Gonzalez, A., Vihervaara, P., Balvanera, P., Bates, A. E., Bayraktarov, E., Bellingham, P. J., Bruder, A., Campbell, J., Catchen, M. D., Cavender-Bares, J., Chase, J., Coops, N., Costello, M. J., Dornelas, M., Dubois, G., Duffy, E. J., Eggermont, H., Fernandez, N., Ferrier, S., … Wright, E. (2023). A global biodiversity observing system to unite monitoring and guide action. *Nature Ecology & Evolution*, 1–5. https://doi.org/10.1038/s41559-023-02171-0

Guillera-Arroita, G. (2012). *Occupancy modelling : Study design and models for data collected along transects* (Doctoral dissertation). University of Kent. https://doi.org/10.22024/UniKent/01.02.86472

Guillera-Arroita, G. (2017). Modelling of species distributions, range dynamics and communities under imperfect detection: Advances, challenges and opportunities. *Ecography*, *40*(2), 281–295. https://doi.org/10.1111/ecog.02445

Guillera-Arroita, G., Morgan, B. J. T., Ridout, M. S., & Linkie, M. (2011). Species Occupancy Modeling for Detection Data Collected Along a Transect. *Journal of Agricultural, Biological, and Environmental Statistics*, *16*(3), 301–317. https://doi.org/10.1007/s13253-010-0053-3

Haines, L. M., Altwegg, R., & Borchers, D. L. (2023). Exact likelihoods for N-mixture models with time-to-detection data. *Australian & New Zealand Journal of Statistics*, *n/a*(n/a). https://doi.org/10.1111/anzs.12401

Halstead, B. J., Rose, J. P., & Kleeman, P. M. (2021). Time-to-detection occupancy methods: Performance and utility for improving efficiency of surveys. *Ecological Applications*, *31*(3), e2267. https://doi.org/10.1002/eap.2267

Hawkes, A. G. (1971). Spectra of some self-exciting and mutually exciting point processes. *Biometrika*, *58*(1), 83–90. https://doi.org/10.1093/biomet/58.1.83

bioRxiv preprint doi: https://doi.org/10.1101/2023.11.17.567350; this version posted February 27, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Henry, D. A. W., Lee, A. T. K., & Altwegg, R. (2020). Can time-to-detection models with fewer survey replicates provide a robust alternative to traditional site-occupancy models? *Methods in Ecology and Evolution*, *11*(5), 643–655. https://doi.org/10.1111/2041-210X.13379

Høye, T. T., Groom, Q., Juslén, A., Mandon, C., Dinesen, L., Rosenberg, A., Hendriks, R. J., Eggermont, H., & Vihervaara, P. (2022). *Biodiversity monitoring knowledge gaps and research & innovation priorities* (tech. rep. Biodiversa+ D2.1).

Hsing, P.-Y. (2019). *Monitoring the UK's wild mammals: A new grammar for citizen science engagement and ecology* (Doctoral dissertation). Durham University.

IPBES. (2019). *Global assessment report on biodiversity and ecosystem services of the Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services* (tech. rep.). Zenodo. https://doi.org/10.5281/zenodo.6417333

IUCN. (2021). *Nature 2030 : One nature, one future : A programme for the Union 2021-2024* (tech. rep.). IUCN.

Kays, R., Arbogast, B. S., Baker-Whatton, M., Beirne, C., Boone, H. M., Bowler, M., Burneo, S. F., Cove, M. V., Ding, P., Espinosa, S., Gonçalves, A. L. S., Hansen, C. P., Jansen, P. A., Kolowski, J. M., Knowles, T. W., Lima, M. G. M., Millspaugh, J., McShea, W. J., Pacifici, K., … Spironello, W. R. (2020). An empirical evaluation of camera trap study design: How many, how long and when? *Methods in Ecology and Evolution*, *11*(6), 700–713. https://doi.org/10.1111/2041-210X.13370

Kellner, K. F., Parsons, A. W., Kays, R., Millspaugh, J. J., & Rota, C. T. (2022). A Two-Species Occupancy Model with a Continuous-Time Detection Process Reveals Spatial and Temporal Interactions. *Journal of Agricultural, Biological and Environmental Statistics*, *27*(2), 321–338. https://doi.org/10.1007/s13253-021-00482-y

Kellner, K. F., Smith, A. D., Royle, J. A., Kéry, M., Belant, J. L., & Chandler, R. B. (2023). The unmarked R package: Twelve years of advances in occurrence and abundance modelling in ecology. *Methods in Ecology and Evolution*, *n/a*(n/a). https://doi.org/10.1111/2041-210X.14123

Kellner, K. F., & Swihart, R. K. (2014). Accounting for Imperfect Detection in Ecology: A Quantitative Review. *PLOS ONE*, *9*(10), e111436. https://doi.org/10.1371/journal.pone.0111436

Kéry, M., & Schmidt, B. (2008). Imperfect detection and its consequences for monitoring for conservation. *Community Ecology*, *9*(2), 207–216. https://doi.org/10.1556/comec.9.2008.2.10

Lahoz-Monfort, J. J., & Magrath, M. J. L. (2021). A Comprehensive Overview of Technologies for Species and Habitat Monitoring and Conservation. *BioScience*, *71*(10), 1038–1062. https://doi.org/10.1093/biosci/biab073

Le Borgne, H., & Bouget, C. (2023). Suivis de biodiversité par la reconnaissance automatique des espèces sur photographies : Perspectives et défis. *Naturae*, (6). https://doi.org/10.5852/naturae2023a6

Liemohn, M. W., Shane, A. D., Azari, A. R., Petersen, A. K., Swiger, B. M., & Mukhopadhyay, A. (2021). RMSE is not enough: Guidelines to robust data-model comparisons for magnetospheric physics. *Journal of Atmospheric and Solar-Terrestrial Physics*, *218*, 105624. https://doi.org/10.1016/j.jastp.2021.105624

bioRxiv preprint doi: https://doi.org/10.1101/2023.11.17.567350; this version posted February 27, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

MacKenzie, D. I., Nichols, J. D., Lachman, G. B., Droege, S., Andrew Royle, J., & Langtimm, C. A. (2002). Estimating Site Occupancy Rates When Detection Probabilities Are Less Than One. *Ecology*, *83*(8), 2248–2255. https://doi.org/10.2307/3072056

MacKenzie, D. I., Royle, J. A., Brown, J. A., & Nichols, J. D. (2004). Occupancy Estimation and Modeling for Rare and Elusive Populations. *Sampling Rare or Elusive Species: Concepts, Designs, and Techniques for Estimating Population Parameters* (pp. 149–172). Island Press.

Moore, H. A., Champney, J. L., Dunlop, J. A., Valentine, L. E., Nimmo, D. G., Moore, H. A., Champney, J. L., Dunlop, J. A., Valentine, L. E., & Nimmo, D. G. (2020). Spot on: Using camera traps to individually monitor one of the world's largest lizards. *Wildlife Research*, *47*(4), 326–337. https://doi.org/10.1071/WR19159

Neilson, E. W., Avgar, T., Burton, A. C., Broadley, K., & Boutin, S. (2018). Animal movement affects interpretation of occupancy models from camera-trap surveys of unmarked animals. *Ecosphere*, *9*(1), e02092. https://doi.org/10.1002/ecs2.2092

Oliver, R. Y., Iannarilli, F., Ahumada, J., Fegraus, E., Flores, N., Kays, R., Birch, T., Ranipeta, A., Rogan, M. S., Sica, Y. V., & Jetz, W. (2023). Camera trapping expands the view into global biodiversity and its change. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *378*(1881), 20220232. https://doi.org/10.1098/rstb.2022.0232

Pollock, L. J., Tingley, R., Morris, W. K., Golding, N., O'Hara, R. B., Parris, K. M., Vesk, P. A., & McCarthy, M. A. (2014). Understanding co-occurrence by modelling species simultaneously with a Joint Species Distribution Model (JSDM). *Methods in Ecology and Evolution*, *5*(5), 397–406. https://doi.org/10.1111/2041-210X.12180

Potamitis, I., Ntalampiras, S., Jahn, O., & Riede, K. (2014). Automatic bird sound detection in long real-field recordings: Applications and tools. *Applied Acoustics*, *80*, 1–9. https://doi.org/10.1016/j.apacoust.2014.01.001

Priyadarshani, D., Altwegg, R., Lee, A. T. K., & Hwang, W.-H. (2022). What can occupancy models gain from time-to-detection data? *Ecology*, *n/a*(n/a), e3832. https://doi.org/10.1002/ecy.3832

Priyadarshani, D., Huynh, H.-D., Altwegg, R., & Hwang, W.-H. (2024). A unified framework for time-to-detection occupancy and abundance models. *Methods in Ecology and Evolution*, *n/a*(n/a). https://doi.org/10.1111/2041-210X.14296

Punt, A. E., Butterworth, D. S., De Moor, C. L., De Oliveira, J. A. A., & Haddon, M. (2016). Management strategy evaluation: Best practices. *Fish and Fisheries*, *17*(2), 303–334. https://doi.org/10.1111/faf.12104

R Core Team. (2023). *R: A language and environment for statistical computing*. Manual. R Foundation for Statistical Computing. Vienna, Austria.

Rhinehart, T. A., Turek, D., & Kitzes, J. (2022). A continuous-score occupancy model that incorporates uncertain machine learning output from autonomous biodiversity surveys. *Methods in Ecology and Evolution*, *13*(8), 1778–1789. https://doi.org/10.1111/2041-210X.13905

bioRxiv preprint doi: https://doi.org/10.1101/2023.11.17.567350; this version posted February 27, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Rodríguez, J. P., Keith, D. A., Rodríguez-Clark, K. M., Murray, N. J., Nicholson, E., Regan, T. J., Miller, R. M., Barrow, E. G., Bland, L. M., Boe, K., Brooks, T. M., Oliveira-Miranda, M. A., Spalding, M., & Wit, P. (2015). A practical guide to the application of the IUCN Red List of Ecosystems criteria. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *370*(1662), 20140003. https://doi.org/10.1098/rstb.2014.0003

Royle, J. A., & Nichols, J. D. (2003). Estimating Abundance from Repeated Presence–Absence Data or Point Counts. *Ecology*, *84*(3), 777–790. https://doi.org/10.1890/0012-9658(2003)084[0777:EAFRPA]2.0.CO;2

Rushing, C. S. (2023). An ecologist's introduction to continuous-time multi-state models for capture-recapture data. *Journal of Animal Ecology*, *n/a*(n/a). https://doi.org/10.1111/1365-2656.13902

Rydén, T. (1994). Parameter estimation for Markov modulated poisson processes. *Communications in Statistics. Stochastic Models*, *10*(4), 795–829. https://doi.org/10.1080/15326349408807323

Schofield, M. R., Barker, R. J., & Gelling, N. (2018). Continuous-time capture–recapture in closed populations. *Biometrics*, *74*(2), 626–635. https://doi.org/10.1111/biom.12763

Skaug, H. J. (2006). Markov Modulated Poisson Processes for Clustered Line Transect Data. *Environmental and Ecological Statistics*, *13*(2), 199–211. https://doi.org/10.1007/s10651-005-0006-0

Steenweg, R., Hebblewhite, M., Kays, R., Ahumada, J., Fisher, J. T., Burton, C., Townsend, S. E., Carbone, C., Rowcliffe, J. M., Whittington, J., Brodie, J., Royle, J. A., Switalski, A., Clevenger, A. P., Heim, N., & Rich, L. N. (2017). Scaling-up camera traps: Monitoring the planet's biodiversity with networks of remote sensors. *Frontiers in Ecology and the Environment*, *15*(1), 26–34. https://doi.org/10.1002/fee.1448

Tuia, D., Kellenberger, B., Beery, S., Costelloe, B. R., Zuffi, S., Risse, B., Mathis, A., Mathis, M. W., van Langevelde, F., Burghardt, T., Kays, R., Klinck, H., Wikelski, M., Couzin, I. D., van Horn, G., Crofoot, M. C., Stewart, C. V., & Berger-Wolf, T. (2022). Perspectives in machine learning for wildlife conservation. *Nature Communications*, *13*(1), 792. https://doi.org/10.1038/s41467-022-27980-y

UNECE. (2023). *Guidelines for Developing National Biodiversity Monitoring Systems*. United Nations. https://doi.org/10.18356/9789210024006

van Strien, A. J., van Swaay, C. A., & Termaat, T. (2013). Opportunistic citizen science data of animal species produce reliable estimates of distribution trends if analysed with occupancy models. *Journal of Applied Ecology*, *50*(6), 1450–1458. https://doi.org/10.1111/1365-2664.12158

Vihervaara, P., Lipsanen, A., Suni, T., Mandon, C., Eggermont, H., Body, G., Basille, M., Naeslund, M., Silva Del Pozo, M., Basset, A., Onen Tarantini, S., Dinesen, L., Hoye, T., Hendriks, R., Heck, A., Eichenberg, D., McCallum, I., Liquete, C., Maes, J., … Borg, J. (2023). *Biodiversa+ strategic biodiversity monitoring governance document (Phase 1)* (tech. rep. Biodiversa+ D2.8).

Wägele, J. W., Bodesheim, P., Bourlat, S. J., Denzler, J., Diepenbroek, M., Fonseca, V., Frommolt, K.-H., Geiger, M. F., Gemeinholzer, B., Glöckner, F. O., Haucke, T., Kirse, A., Kölpin, A., Kostadinov, I., Kühl, H. S., Kurth, F., Lasseck, M., Liedke, S., Losch, F., … Wildermann, S. (2022). Towards a multisensor station for automated biodiversity monitoring. *Basic and Applied Ecology*, *59*, 105–138. https://doi.org/10.1016/j.baae.2022.01.003

bioRxiv preprint doi: https://doi.org/10.1101/2023.11.17.567350; this version posted February 27, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Zhang, W., & Bonner, S. J. (2020). On continuous-time capture-recapture in closed populations. *Biometrics*, *76*(3), 1028–1033. https://doi.org/10.1111/biom.13185

Zwerts, J. A., Stephenson, P. J., Maisels, F., Rowcliffe, M., Astaras, C., Jansen, P. A., van der Waarde, J., Sterck, L. E. H. M., Verweij, P. A., Bruce, T., Brittain, S., & van Kuijk, M. (2021). Methods for wildlife monitoring in tropical forests: Comparing human observations, camera traps, and passive acoustic sensors. *Conservation Science and Practice*, *3*(12), e568. https://doi.org/10.1111/csp2.568