Check for updates

Ecology Letters

WILEY

ECOLOGY LETTERS

METHOD OPEN ACCESS

Leveraging Massive Opportunistically Collected Datasets to Study Species Communities in Space and Time

Maxime Fajgenblat[1,2,3,4] | Robby Wijns[1] | Geert De Knijf[5] | Robby Stoks[4] | Pieter Lemmens[1,5,6] | Marc Herremans[3] | Pieter Vanormelingen[3] | Thomas Neyens[2,7] | Luc De Meester[1,6,8,9]

[1]Laboratory of Freshwater Ecology, Evolution and Conservation, KU Leuven, Leuven, Belgium | [2]Data Science Institute, I-BioStat, Hasselt University, Diepenbeek, Belgium | [3]Natuurpunt Studie, Mechelen, Belgium | [4]Laboratory of Evolutionary Stress Ecology and Ecotoxicology, KU Leuven, Leuven, Belgium | [5]Research Institute for Nature and Forest (INBO), Brussels, Belgium | [6]Leibniz Institute für Gewasserökologie und Binnenfischerei (IGB), Berlin, Germany | [7]Leuven Biostatistics and Statistical Bioinformatics Centre (L-BioStat), KU Leuven, Leuven, Belgium | [8]Institute of Biology, Freie Universität Berlin, Berlin, Germany | [9]Berlin-Brandenburg Institute of Advanced Biodiversity Research (BBIB), Berlin, Germany

Correspondence: Maxime Fajgenblat (maxime.fajgenblat@gmail.com)

Received: 27 August 2024 | **Accepted:** 20 February 2025

Editor: Carl Boettiger

Funding: This work was supported by the European Union's Horizon 2020 research and innovation program under grant agreement No. 869296; the Research Foundation Flanders (11E3222N, G0A3M24N, G0A4121N); KU Leuven (DB/22/007/bm); Onderzoeksraad, KU Leuven (C16/2023/003); and Leibniz-Institut für Gewässerökologie und Binnenfischerei. The computational resources and services used in this work were provided by the VSC (Flemish Supercomputer Center), funded by the Research Foundation Flanders and the Flemish Government.

Keywords: Bayesian hierarchical modelling | citizen science data | joint species distribution modelling | metacommunity ecology | occupancy modelling

ABSTRACT

Online portals have facilitated collecting extensive biodiversity data by naturalists, offering unprecedented coverage and resolution in space and time. Despite being the most widely available class of biodiversity data, opportunistically collected records have remained largely inaccessible to community ecologists since the imperfect and highly heterogeneous detection process can severely bias inference. We present a novel statistical approach that leverages these datasets by embedding a spatiotemporal joint species distribution model within a flexible site-occupancy framework. Our model addresses variable detection probabilities across visits and species by modelling phenological patterns and by extending the use of latent variables to characterise observer-specific detection and reporting behaviour. We apply our model to an opportunistically collected dataset on lentic odonates, encompassing over 100,000 waterbody visits in Flanders (N-Belgium), to show that the model provides insights into biological communities at high resolution, including phenology, interannual trends, environmental associations and spatiotemporal co-distributional patterns in community composition.

1 | Introduction

Understanding how species assemble into communities across space and time and how environmental conditions, biotic interactions, dispersal and anthropogenic pressure shape this process constitutes one of the grand aims of ecology. This gave rise to the field of metacommunity ecology (Leibold and Chase 2017), while simultaneously providing crucial insights for other disciplines such as conservation biology. In addition to theoretical modelling (Lerch et al. 2023; Vanoverbeke et al. 2016) and experimental approaches (Logue et al. 2011), metacommunity ecology predominantly relies on

Maxime Fajgenblat and Robby Wijns: shared first authors.

Thomas Neyens and Luc De Meester: shared senior authors.

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

© 2025 The Author(s). *Ecology Letters* published by John Wiley & Sons Ltd.

multivariate analyses of field-based survey data to generate insights (Leibold et al. 2022; Logue et al. 2011). Over the past decades, developments in statistical ecology have strongly influenced the way high-dimensional community datasets are being analysed, starting with the application of distance-based ordination methods (Legendre and Legendre 1998; Peres-Neto et al. 2006), followed by the more recent development of joint species distribution modelling tools (JSDMs) (Ovaskainen and Abrego 2020; Pollock et al. 2014; Ries et al. 2016; Warton et al. 2015), the latter also revealing potential signatures of interspecific interaction patterns and providing species-specific inferences. Both distance-based ordination and JSDM-based approaches allow linking quantitative signatures in datasets to existing theoretical paradigms of species assembly (Cottenie 2005; Leibold et al. 2022; Ovaskainen et al. 2019).

With some exceptions (e.g., Mungi et al. 2023), most species community datasets are heavily constrained in their geographical, temporal and taxonomic scope and resolution, as rigorously collecting highly standardised data following a predetermined sampling design is expensive and time-consuming. The vast majority of studies treat communities as static by analysing single time points, ignoring interannual and seasonal variation, even though theoretical and conceptual work increasingly highlights the importance of time to understand metacommunity dynamics (Guzman et al. 2022; Wolkovich et al. 2014; Yin and Rudolf 2024).

Over the past decade, the number of citizen science initiatives has been growing rapidly, resulting in large opportunistically collected biodiversity datasets that provide a rich source of data with broad spatiotemporal extent and often high spatiotemporal resolution. For instance, naturalists all over the world are using biodiversity data platforms such as iNaturalist, eBird and Observation.org, generating millions of species records on a weekly basis (Dickinson et al. 2010). While the advantages of these datasets are evident, they feature major drawbacks that hamper straightforward rigorous statistical analyses for metacommunity ecology and conservation research. First, opportunistic data collection typically suffers from spatiotemporal and taxonomic sampling biases due to the absence of predefined sampling guidelines (Bowler et al. 2022; Neyens et al. 2019; Pocock et al. 2023; Shirey et al. 2021). Second, imperfect detection, that is, the near-ubiquitous phenomenon that species tend to go unnoticed even though they are present during biological surveys (MacKenzie et al. 2002), is particularly exacerbated in large-scale opportunistic datasets as the probability of detecting and reporting species strongly varies among naturalists and throughout the season, and as species vary in conspicuousness or attractiveness (Dickinson et al. 2010). Failing to account for imperfect and heterogeneous detection can lead to severe bias in ecological inference and species distribution estimates (Guillera-Arroita et al. 2014; Lahoz-Monfort et al. 2014; Rota et al. 2011).

Often, unstructured opportunistically collected data (e.g., through portals such as iNaturalist or Observation.org) contains more than presence-only information, as records of non-focal species reveal the activity of observers, enabling reconstructing pseudo-visits and checklists (van Strien et al. 2013). In large opportunistically collected datasets, a subset of sites is typically visited repeatedly within a single biological season, producing detection/non-detection histories (Mackenzie et al. 2018). Provided that occupancy remains constant throughout the season (i.e., the closure assumption), the site-occupancy modelling framework (MacKenzie et al. 2002) offers a means to address imperfect and highly heterogeneous detection that is inherent to opportunistically collected data. This has led to several successful applications to opportunistically collected data on single species, often to quantify interannual trends and to address applied conservation research questions (Hochachka et al. 2012; Isaac et al. 2014; van Strien et al. 2010, 2011, 2013; Termaat et al. 2019).

Multispecies site-occupancy models have subsequently been developed to model community data sampled under imperfect detection (Beissinger et al. 2016; Dorazio et al. 2006; Sutherland et al. 2016). Recently, these efforts have been further extended to accommodate large spatial structures (Doser et al. 2023), to integrate multiple population- and community-level data sources and types (Zipkin et al. 2023), to allow for spatially varying coefficients (Doser et al. 2024), to harness the flexibility of neural networks (Joseph 2020) or to facilitate the uptake by ecologists (Doser et al. 2022). While offering much flexibility in many respects, these recently developed methods miss key features of state-of-the-art JSDMs that assume perfect detection (e.g., modelling trait-based and phylogenetic influences; Ovaskainen and Abrego 2020) and they predominantly leave the (spatio)temporal dimension unexplored, in contrast to recently developed cutting-edge machine learning approaches for opportunistically collected data (Cole et al. 2023; Davis et al. 2023; Fink et al. 2023; Smith and Edwards 2021; Teng et al. n.d.). The latter approaches, on the other hand, tend either to ignore imperfect detection or to adjust for search effort using covariates without explicitly disentangling the detection and occupancy process. Furthermore, although the importance of interobserver heterogeneities has long been recognised (Sauer et al. 1994), it has received remarkably limited attention to date (Johnston et al. 2018; Kelling et al. 2015).

In this paper, we develop a Bayesian model to analyse large opportunistically collected datasets in the context of metacommunity ecology by combining a flexible site-occupancy approach with a state-of-the-art JSDM, the Hierarchical Modeling of Species Communities framework (Ovaskainen and Abrego 2020). We present a novel way to parsimoniously map observer-specific detection and reporting behaviour across species by extending the use of the popular latent variable approach in JSDMs from site-species to observer-species relationships. As such, our approach facilitates the use of widely available opportunistically collected biodiversity data to fuel large-scale spatiotemporal metacommunity analyses, while explicitly acknowledging the shortcomings of such data.

2 | Material and Methods

2.1 | Model Structure and Rationale

In the following, we provide a high-level outline of our modelling approach. A full overview, including all model equations, technical details and key assumptions, is provided in the Supporting Information in Section S1.

14610248, 2025, 3, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/ele.70094, Wiley Online Library on [17/03/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Typically, JSDMs model whether species have been detected across sites as principal outcome (Guillera-Arroita et al. 2015). Such an approach, however, confounds the actual occupancy process with the detection process (Guillera-Arroita 2017), which can be problematic when analysing opportunistically collected data, especially in the face of sampling biases across sites, years, seasons and species, and with a large number of observers. Therefore, we instead model the detection outcome during individual visits to subsequently infer the latent occupancy process, following the site-occupancy principle (MacKenzie et al. 2002). Specifically, the detection outcome $y_{v,i}$, pertains to whether species i has been reported (0: no; 1: yes) during visit v, which took place on day $d(v)$ by observer $o(v)$. We define a (pseudo-)visit v as an instance where an observer reports at least one sighting of any species on a specific day at a specific site (a grid cell, a transect, a pond ...). Deriving discrete visits is facilitated by the multispecies nature of biodiversity portals such as iNaturalist and Observation.org, since records of non-focal species within a consistent taxonomic scope enable inferring observer activity.

Relying on the site-occupancy framework (MacKenzie et al. 2002), we assume $y_{v,i}$ follows a Bernoulli distribution:

$$y_{v,i} \mid z_{s(v),t(v),i} \sim \text{Bernoulli}\left(p_{v,i}\, z_{s(v),t(v),i}\right),$$

where $p_{v,i}$ is the probability of detecting species i during visit v (if it would be present), and $z_{s,t,i}$, is the latent occupancy status (0: absent; 1: present) for species i in the visited site s and year t. We assume the latent occupancy status $z_{s,t,i}$ follows a Bernoulli distribution as well:

$$z_{s,t,i} \sim \text{Bernoulli}\left(\psi_{s,t,i}\right),$$

where $\psi_{s,t,i}$ is the probability that site s is occupied by species i in year t. Both $p_{v,i}$ and $\psi_{s,t,i}$ can be modelled through a logistic regression approach. Rather than treating $p_{v,i}$ and $\psi_{s,t,i}$ constant across all visits, sites and years, we model both parameters by tailoring their linear predictors to accommodate the inherent complexity of opportunistic data and by harnessing the shared information across space, time and species of the metacommunity.

We model the latent occupancies $z_{s,t,i}$ across sites, years and species on a logit scale through a spatiotemporally modified implementation of HMSC. Following the standard implementation of HMSC, the influence of environmental predictors on the occupancy is modelled efficiently across species through partial pooling, shrinking estimates towards the community mean, and by accounting for traits and phylogeny (Tikhonov et al. 2020). Since opportunistically collected data provides opportunities to model temporal patterns in more detail than conventional applications of HMSC, we model interannual trends parsimoniously using exact, species-specific Gaussian processes, a powerful way of accounting for temporal autocorrelation (Rasmussen and Williams 2006), and we capture yearly deviations from the smooth interannual trends through species-specific, normally distributed (i.i.d.) random effects (Outhwaite et al. 2018). Residual site-species associations are captured by a latent factor approach (Bhattacharya and Dunson 2011; Norberg et al. 2019; Ovaskainen et al. 2016, 2017; Tikhonov et al. 2020), which constitutes a model-based ordination to condense information across sites and species into a finite number of dimensions (Hui et al. 2015). We model the site loadings as a combination of spatially structured and spatially unstructured random effects to capture spatially autocorrelated patterns and noisily distributed patterns, respectively. We model the former using B-splines projected Gaussian processes as an efficient alternative to exact Gaussian processes to cover a large number of locations (Monod et al. 2022), and we use normally distributed (i.i.d.) random effects to model the latter. Finally, residual patterns across sites, years and species not captured by the above-mentioned separable components, are compressed into site-year loadings, modelled using B-splines projected Gaussian processes and a second set of species loadings, capturing regional changes in co-distributional patterns over time. We refer to the Supporting Information in Section S1.2 for all technical details regarding the modelling of the occupancy probabilities.

By integrating the outlined JSDM in a site-occupancy model, non-detections during repeated visits are used to further inform the estimation of occupancy probabilities whenever a species might be present despite not being detected. To ensure the informativeness of individual visits, it is important to model the detection probabilities comprehensively across visits, observers, seasons and species, which we achieve by capitalising on the same statistical techniques used in HMSC. In a similar manner to how environmental associations are modelled in the occupancy part of our model, we use partial pooling, trait data and phylogenetic relatedness (Ovaskainen et al. 2017) to model the influence of detection covariates on detection probabilities across species efficiently. We also introduce a latent factor approach to parsimoniously model interobserver heterogeneity in the propensity to detect and report each species, ordinating observers and species along a limited number of dimensions, akin to how residual patterns among sites and species are modelled in HMSC (Bhattacharya and Dunson 2011). Finally, phenological patterns influencing species detectability are captured by hierarchical cyclic splines, with one common phenological pattern across the metacommunity, from which each species can deviate with a species-specific spline (hierarchical GAM, type GS; Pedersen et al. 2019). For simplicity, we assume individual observer behaviour and phenology to be constant across the study period, though this assumption could easily be relaxed. We refer to the Supporting Information in Section S1.3 for all technical details on the modelling of detection probabilities. We also provide an overview of the considered prior specifications and key modelling assumptions in the Supporting Information in Sections S1.4 and S1.5.

We implemented the model in the probabilistic programming language Stan (Betancourt 2017; Carpenter et al. 2017), relying on the multipath Pathfinder algorithm, a recently developed quasi-Newton variational inference method that has been shown to quickly reach the high probability region of complex target distributions, to initialise dynamic Hamiltonian Monte Carlo (HMC) chains (Zhang et al. 2022). HMC is a gradient-based Markov chain Monte Carlo (MCMC) algorithm that is well suited to estimate high-dimensional models (Monnahan

14610248, 2025, 3, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/ele.70094, Wiley Online Library on [17/03/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

et al. 2017), which further helps in efficiently (often < 100 iterations) reaching the high probability region during the adaptation phase (Hoffman and Ma 2020; Zhang et al. 2022). This combined approach enables sampling using many short chains until the desired level of precision for the intended application is achieved, without wasting computational resources (Margossian and Gelman 2023).

Fully documented Stan code and a complete R-pipeline for data preparation, model running, results extraction and visualisation are available on GitHub: https://github.com/mfajgenblat/opportunisticJSDM.

2.2 | Simulation-Based Model Evaluation

Several methods have been proposed to assess site-occupancy model performance. However, most focus on detection rather than occupancy outcomes (e.g., the MacKenzie-Bailey test or Bayesian posterior predictive checks; MacKenzie and Bailey 2004), or use site-occupancy models as a benchmark for simpler models (Rota et al. 2011). Empirical performance evaluation to infer the occupancy process is challenging because a species' true absence can rarely be confirmed (Miller et al. 2015), precluding computing explanatory power or predictive performance. Therefore, we adopt simulation-based calibration, a stringent and increasingly used method to evaluate the accuracy and reliability of complex Bayesian models and computational algorithms (Cook et al. 2006; Gelman et al. 2020; Modrák et al. 2023; Talts et al. 2018). It is particularly valuable for site-occupancy models, as it can test the model's ability to accurately separate detection probabilities from occupancy probabilities, ensuring that model assumptions and computational methods are correctly specified. Our simulation setting emulates a metacommunity of 100 sites visited by observers over a 10-year period. In addition to calibration, we compare occupancy predictions, environmental associations and interannual trends inferred from our proposed model with those from a simplified version that ignores detection and only models whether each species was ever reported. A detailed description of the model evaluation procedure is provided in the Supporting Information in Section S1.7.

2.3 | Case Study

We illustrate our model and its output through an application to opportunistically collected data of adult lentic dragonflies and damselflies (Odonata) in Flanders (northern Belgium), from the biodiversity data platform Waarnemingen.be – Observations.be, the Belgian subsite of Observation.org. Sightings were collected from 2009 to 2023 by members of the Flemish Dragonfly Society and other naturalists. For their larval development, lentic odonates are bound to bodies of standing water, which are ideal study systems in metacommunity ecology as they represent well-defined patches of suitable habitat surrounded by a terrestrial matrix. The ecology of many species is well studied (Corbet 1999), their taxonomy is stable and odonates are often used as model organisms for ecological and evolutionary research in natural populations (Córdoba-Aguilar et al. 2022) and as sentinels for freshwater conservation and to evaluate restoration processes (Samways 2024).

Each individual waterbody constitutes a site, and the model infers the occupancy state of each waterbody across species and years using environmental information and spatial and (spatio)temporal autocorrelation patterns. As a first step, we linked geotagged sightings of adult odonates to the closest waterbody (max. distance 100 m) using a comprehensive inventory of 93,135 bodies of standing water across the study area (Scheers et al. 2022) and we defined an individual visit as an event where at least one odonate species was sighted near a waterbody by a specific observer on a specific day. We followed the approach of van Strien et al. (2013) to construct detection/non-detection checklists of odonate species from presence-only reports of individual species, inferring non-detections for unreported species. To minimise potential misidentification errors, we only retained visits conducted by experienced odonate 'enthusiasts'—contributors who account for the vast majority of records in the database (see Supporting Information in Section S1.8, for all methodological details). To consider the resulting species checklists under a repeated visit structure, we also made the closure assumption that community occupancy was stationary throughout the year (Doser and Stoudt 2024), and that detection only varies seasonally but not interannually (Supporting Information in Section S1.5).

In our case study, we considered 12 land cover classes (De Saeger et al. 2020; fractions within a 200 m buffer around each waterbody) and the (logarithmically transformed) waterbody area as environmental variables. We considered six trait variables: body size, good dispersal, affinity for temporal ponds, nutrient level preference, an antipredation index (derived from number and size of larval antipredatory spines) and a temperature optimum proxy (i.e., the species temperature index; STI) (Harabiš and Hronková 2020; De Knijf et al. 2006; Termaat et al. 2019). In an alternative model specification, we did not include any occupancy predictors or traits for unconstrained ordination purposes.

We used categorised list length (i.e., the number of species detected during a visit) as a proxy for search effort, with categories of singleton (one species), short (two–three species) and longer (> three species) lists, by including it as a dummy-coded predictor for detection probabilities, as previous studies have shown this method to account for variation in sampling intensity among visits (Isaac et al. 2014; Szabo et al. 2010).

The model was run on a high-performance computing cluster utilising 36 cores during a period of 7 days. Further methodological details are provided in the Supporting Information in Section S1.8.

3 | Results

3.1 | Simulation-Based Model Evaluation

The histograms of the rank statistics obtained as part of the simulation-based calibration procedure do not indicate important deviations from a uniform distribution (Figure S1),

14610248, 2025, 3, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/ele.70094, Wiley Online Library on [17/03/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

indicating that our model is capable of properly inferring both detection and occupancy parameters in a well-calibrated fashion. Furthermore, simulations indicate that the model outperforms a naive implementation that ignores imperfect detection with respect to predicting occupancy probabilities (92% of simulations), the signs of environmental associations (100% of simulations) as well as retrieving interannual trends (88% of simulations) (Figure S2). Model performance is substantially higher over the full range of detection and occupancy rates, in particular for predicting occupancies, and only becomes comparable when detection is near-perfect (Figures S3 and S4).

3.2 | Metacommunity Insights From Opportunistic Data: Case Study on Lentic Odonates

In our case study, 467,126 of the 756,427 odonate sightings registered in the region between 2009 and 2023 pertain to 50 obligatory or facultative lentic species and could be linked to a waterbody (< 100 m distance). From these sightings, we derived 112,814 discrete waterbody visits, omitting visits by less-experienced observers (defined as having performed < 75 visits throughout the study period) to ensure data quality and waterbodies that have been visited infrequently (< three times throughout the study period). As a result, a total of 6672 waterbodies is considered in the analysis (7.2% of all known waterbodies in the study area), with an average of 16.9 visits performed per waterbody throughout the study period, by a total of 445 selected observers (Figure S5). Due to the large number of individual findings, we restrict ourselves to a non-exhaustive bird's-eye perspective, with the principal aim of showcasing what types of inference our model is capable of.

Overall, we found the probability of detecting and reporting a species (given its presence) during an average visit to be low across all species during their week of peak phenological activity, with posterior medians ranging from 0.0003 for *Leucorrhinia caudalis* to 0.0417 for *Libellula quadrimaculata* for visits with long lists, and even lower probabilities for visits with singleton and short lists (Figure S6). With the exception of *Sympecma fusca* (the only species that overwinters as adult in the study area and, hence, can be detected year-round), all species show a clear seasonal pattern (Figure 1a), strongly influencing detectability. In addition to the time of the year, we also found observer identity to strongly modulate the detection probability, with substantial variation across species (Figure 1b). The model-based ordination of observers and species revealed that the most important axis of variation in observer behaviour pertains to the ease of detection and identification (Figure 1b). At one end of the spectrum, we find observers that mainly report species that are common and easily approachable (facilitating identification), such as several species of damselflies (Zygoptera spp.) and darters (*Sympetrum* spp.). These observers feature strongly lowered propensities to report species that necessitate in-flight identification skills such as many dragonfly species (Anisoptera spp.), which require a well-developed search image (e.g., *Coenagrion lunulatum*) or that are widespread but easily overlooked because of their resemblance to more abundant species (e.g., *Ischnura pumilio*). At the other end of the spectrum, we find observers

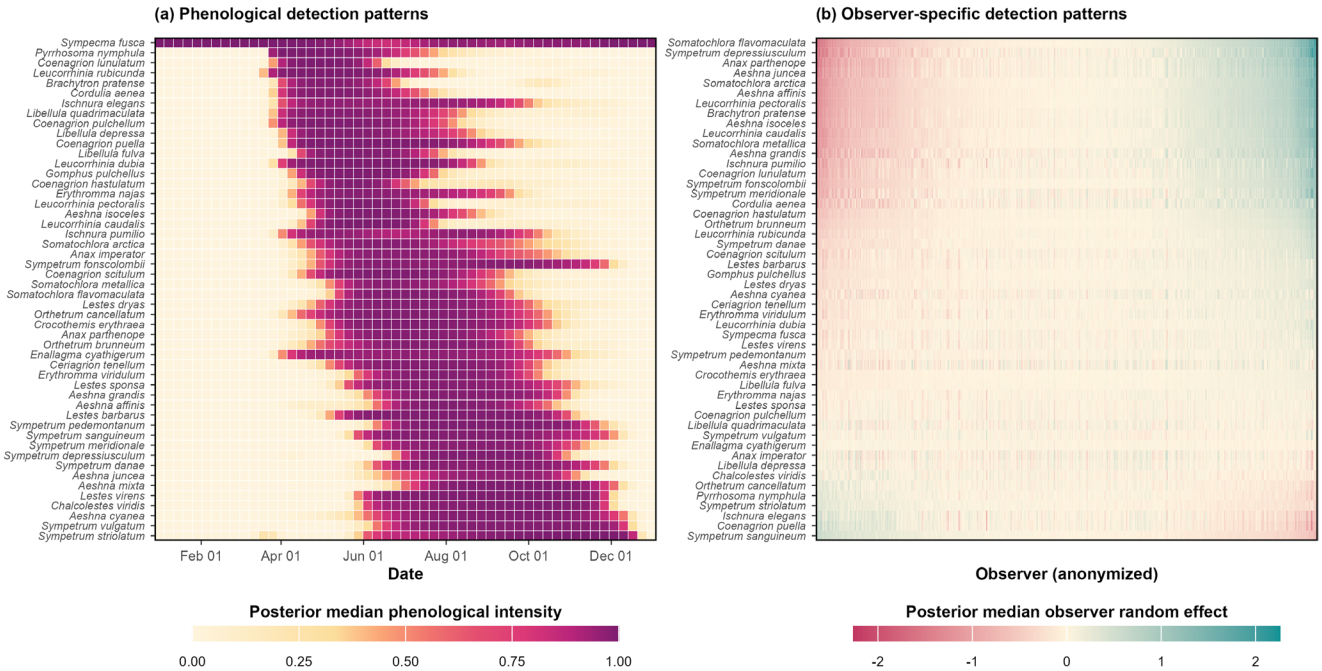

FIGURE 1 | Estimated influences of phenology and observer identity on the detection probabilities of species. (a) Heatmap showing the posterior median phenological activity of each species for every week, with faint yellow colours indicating a very low probability of detecting the species and purple colours indicating a high probability of detecting the species. An inverse logit transformation was applied to the estimated spline values to achieve a range between 0 and 1. Species are ordered by the mode of their phenological curve. (b) Heatmap showing the posterior median effects of observer identity on detecting and reporting the different species (on the logit scale), with red colours indicating a below-average propensity to detect and report a species, and blue-green colours indicating an above-average propensity to detect and report a species. Observers and species are ordered to highlight the most important axis of variation.

14610248, 2025, 3, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/ele.70094, Wiley Online Library on [17/03/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

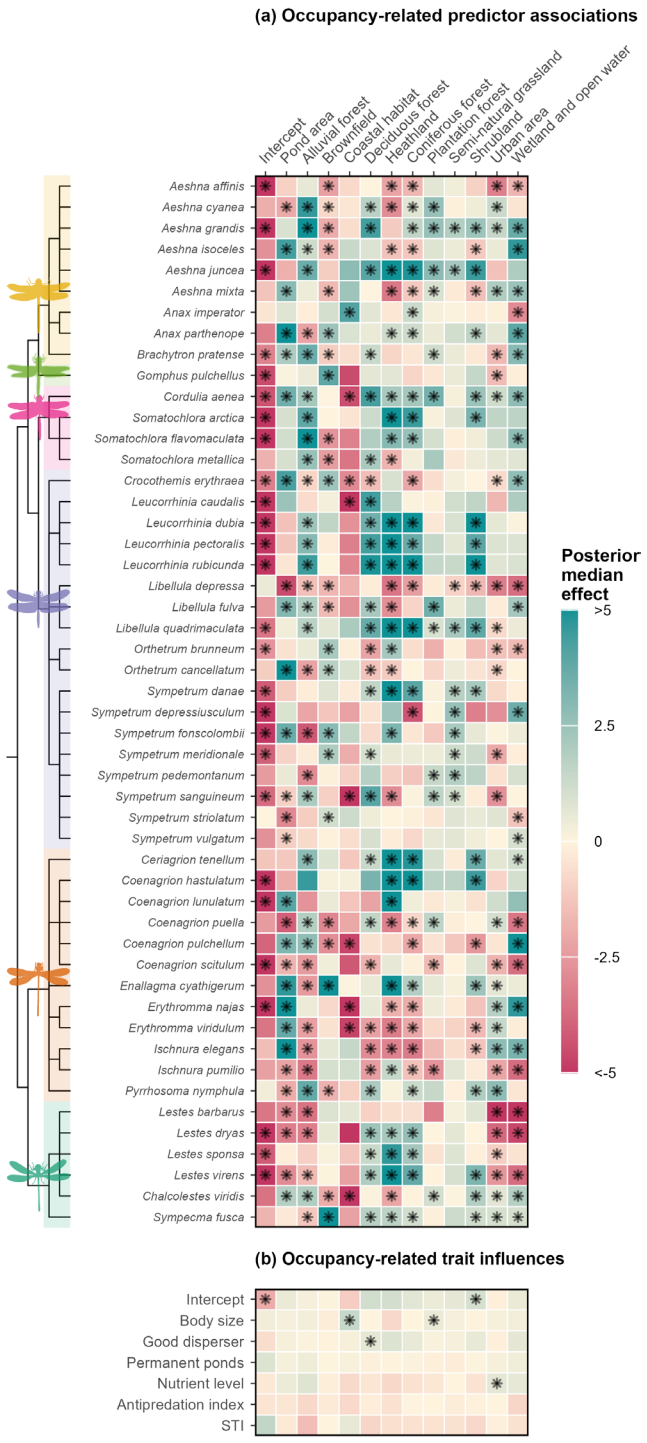

FIGURE 2 | Associations between environmental variables and the occupancy of species, and the influence of traits thereon. (a) Heatmap showing the posterior median regression coefficients for the influence of the considered environmental predictors on the occupancy probability (on the logit scale), for each species. For clarity, colour scales have been truncated to the [−5,5] interval, with more extreme values being represented by the ends of the colour spectrum. Associations with high statistical support (> 95% posterior probability) are indicated by a star in the centre of the heatmap cell. Species are ordered taxonomically, with a taxonomic tree displayed at the left of the heatmap, shaded by family (yellow: Aeshnidae; green: Gomphidae; pink: Corduliidae; purple: Libellulidae; orange: Coenagrionidae; blue-green: Lestidae). (b) Heatmap showing the posterior median regression coefficients for the influence of the considered species traits on the environmental associations. Associations with high statistical support (> 95% posterior probability) are indicated by a star (*) in the centre of the heatmap cell.

displaying opposite behaviour, primarily focusing on the rare and challenging species while being less likely to report easily detected and identified species (Figure 1b).

We identified strong statistical support for associations between the considered environmental predictor variables (pond area and land cover variables) and the occupancy of species (Figure 2a). Some of these associations are related to the considered species traits (Figure 2b), and we found that environmental associations show a relatively weak phylogenetic structure, as reflected by a posterior median phylogenetic signal of 11.3% (95% credible interval [2.8, 34.2]). Among the land cover variables, we found the fraction of heathland to be the most important occupancy predictor for the metacommunity, with 22 species featuring high statistical support (> 95% posterior probability) for a positive association, including *Leucorrhinia rubicunda* and *Coenagrion hastulatum* (Figure 2a). When considering patterns of species richness along gradients of the environmental predictor variables (Ovaskainen and Abrego 2020), the posterior predicted species richness increases most along the fraction of heathland surrounding water bodies, while the opposite holds for agricultural land (Figure S7).

Residual co-distributional associations among species across waterbodies that are not explained by the considered environmental factors are captured by the latent factor component of the model, constituting a constrained, model-based ordination. The model can accommodate both structured (i.e., spatially autocorrelated) and unstructured (i.e., noisy) patterns across space and weights their importance through a spatial signal parameter for each latent dimension. Spatially structured patterns seem to be the most important, with an average posterior median spatial signal of 75.3% across latent dimensions. The species loadings can be used to derive an interspecific association matrix, reflecting which pairs of species tend to co-occur more often than expected by chance alone. Upon comparison to an interspecific association matrix derived from an alternative model without any environmental predictors (i.e., unconstrained ordination), we can evaluate the extent to which the environmental predictors are able to explain apparent (co-)distributional patterns. In our case study, a considerable amount of co-distributional variation remains present (Figure 3). At least some of the patterns seem to be driven by a shared preference for early successional, shallow waterbodies, as indicated by the high correlations among the species *Lestes barbarus*, *Sympetrum fonscolombii*, *Lestes dryas* and *Ischnura pumilio* (among others).

For each species, pure interannual variation in occupancy patterns is captured by temporal random effects. In analogy with the spatial co-distributional associations, we can derive spatio-temporal co-distributional associations, as the model also accounts for regional differences in interannual trends within the species communities. These results suggest that communities

14610248, 2025, 3, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/ele.70094, Wiley Online Library on [17/03/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

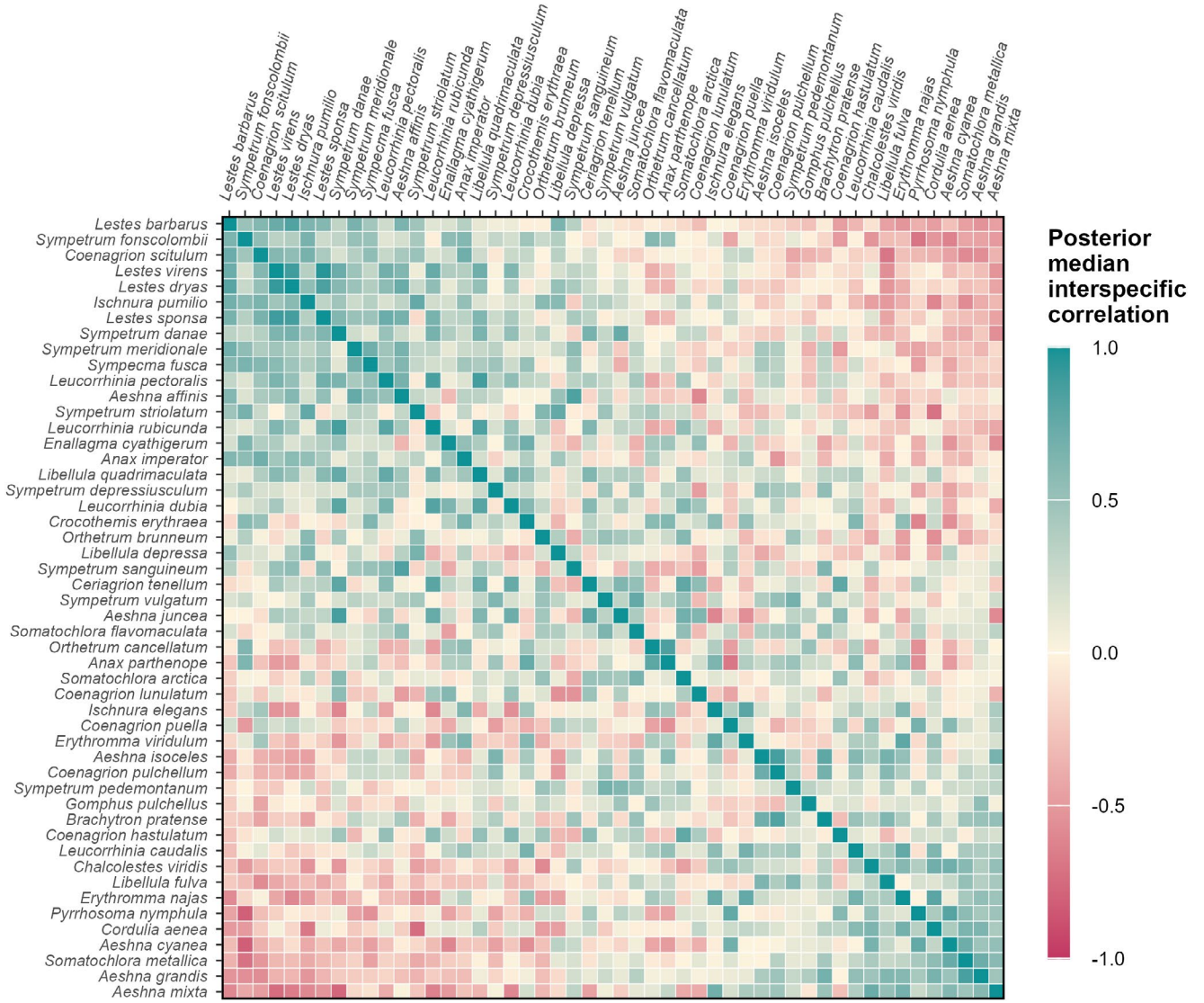

FIGURE 3 | Spatial interspecific association matrix. Each cell indicates whether a pair of species features a higher (blue-green) or lower (red) probability of co-occurring across sites. The (above-diagonal) upper triangle shows the posterior median associations obtained under the main model, that corresponds to a constrained ordination (i.e., conditioned on the environmental predictor variables). The (below-diagonal) lower triangle shows the posterior median associations obtained under an alternative model that does not include environmental predictor variables and, hence, constitutes an unconstrained ordination. Species are ordered to highlight the most important axis of variation.

dominated by generalist and thermophilic species (e.g., *Ischnura elegans*, *Anax imperator* and *Coenagrion scitulum*) display regionally opposite interannual trends compared to species associated with oligotrophic and cooler conditions (e.g., *L. rubicunda* and *Leucorrhinia dubia*), hinting at alternative stable communities that displace each other (Figure S8).

By modelling the occupancies of species across space and time by means of environmental, interannual, spatial and spatio-temporal components, as well as by accounting for imperfect detection during individual visits and by assuming a MAR missingness mechanism (Bowler et al. 2024), we can produce yearly estimates of the spatial distribution of each individual species across the study area (see Figure S9 for a single example species) and derive yearly trends in the fraction of occupied waterbodies (Figure 4). Upon linearisation of the interannual trends by ordinary least squares, we find high statistical support (>95% posterior probability) for 25 species declining and 14 species increasing in occupancy across years. An exploratory analysis of the relationship between the linearised trend slopes and the six considered traits suggests that species with a low thermal optimum (STI) show particularly stark declines (Figure S10).

By summing posterior occupancy probabilities across individual species, we can visualise how patterns in species richness develop across space and time (Figure 5a). In addition to visualising the averaged species richness across the study area over the study period (Figure 5b), we can also visualise computed species richness slopes for all waterbodies within the study area (Figure 5c). The Campine region (northeastern Flanders) harbours the highest species richness but simultaneously faces the strongest absolute declines, with estimated (posterior median) losses of over three species in many waterbodies over the 15-year study period. Conversely, a considerable increase in species richness can be observed in the northern part of the province of East Flanders (northwest of Brussels) (Figure 5c).

14610248, 2025, 3, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/ele.70094, Wiley Online Library on [17/03/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

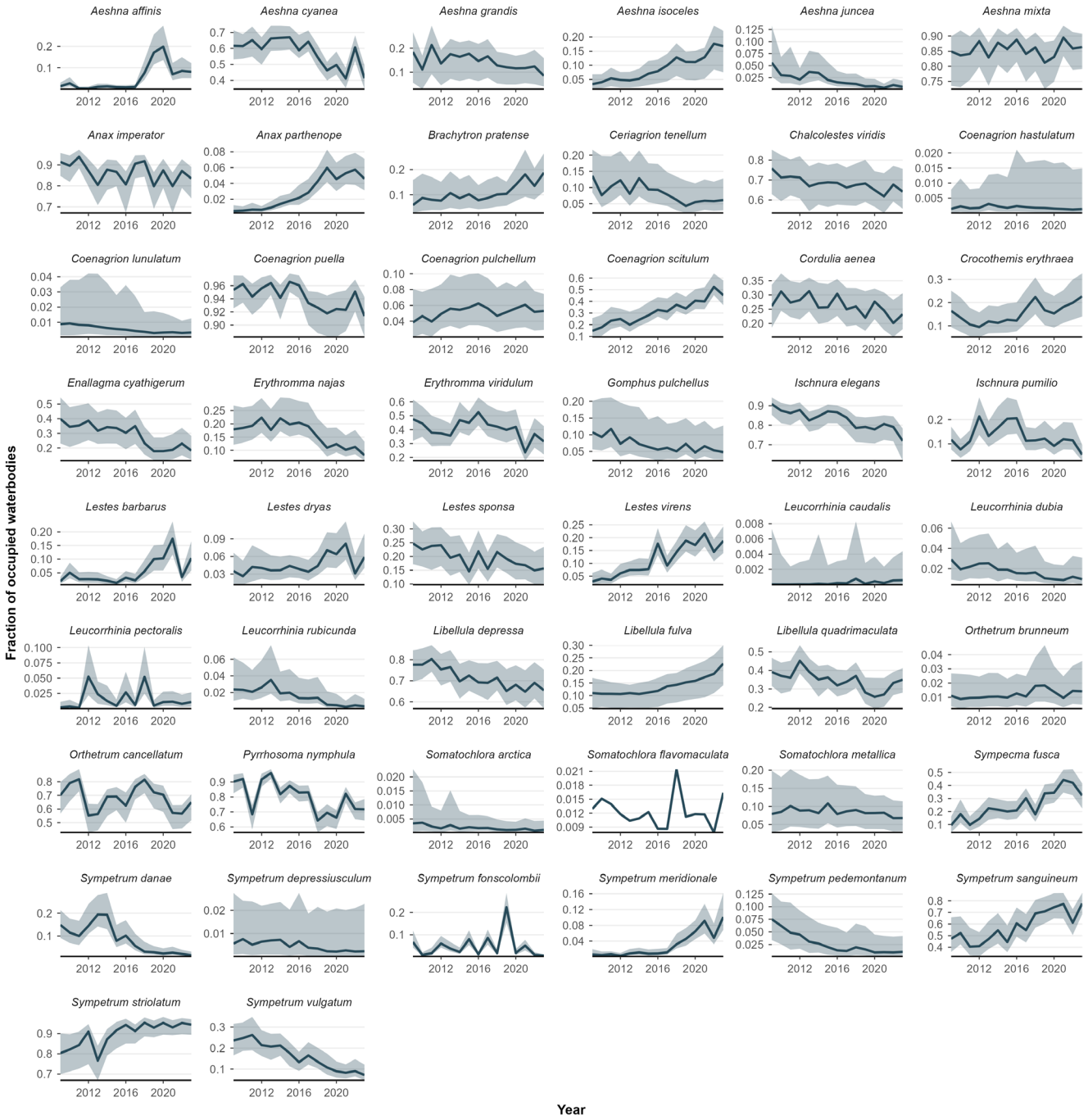

FIGURE 4 | Yearly trends in the estimated proportion of occupied waterbodies by each species in the study area. The full line indicates the posterior median trend, while the shaded area indicates 95% credible intervals. Species are ordered alphabetically.

4 | Discussion

Opportunistically collected biodiversity data are notoriously challenging to handle, leaving them underutilised in disciplines such as metacommunity ecology and conservation research. Our approach leverages the vast amount of data to explicitly and flexibly address the complexities underlying the detection process, extracting valuable biological and ecological insights into the distribution of species in time and space. As long as detection/non-detection checklists can reliably be inferred and reasonable assumptions regarding closure among repeated visits can be made (Doser and Stoudt 2024; Hochachka et al. 2023; van Strien et al. 2013), it provides a complementary method to standardised survey data. Through our case study on lentic odonates, we demonstrate how to infer phenology, observer behaviour patterns, environmental associations, trait and phylogenetic influences, interannual trends and spatial and spatiotemporal co-distributional patterns, collectively describing the metacommunity and how it is observed. Furthermore, our simulation results indicate that high-quality insights can also be derived from moderately sized datasets (100 sites visited over 10 years). The wealth of results can serve and benefit a wide range of applications, from fuelling metacommunity ecology research to informing

14610248, 2025, 3, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/ele.70094, Wiley Online Library on [17/03/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

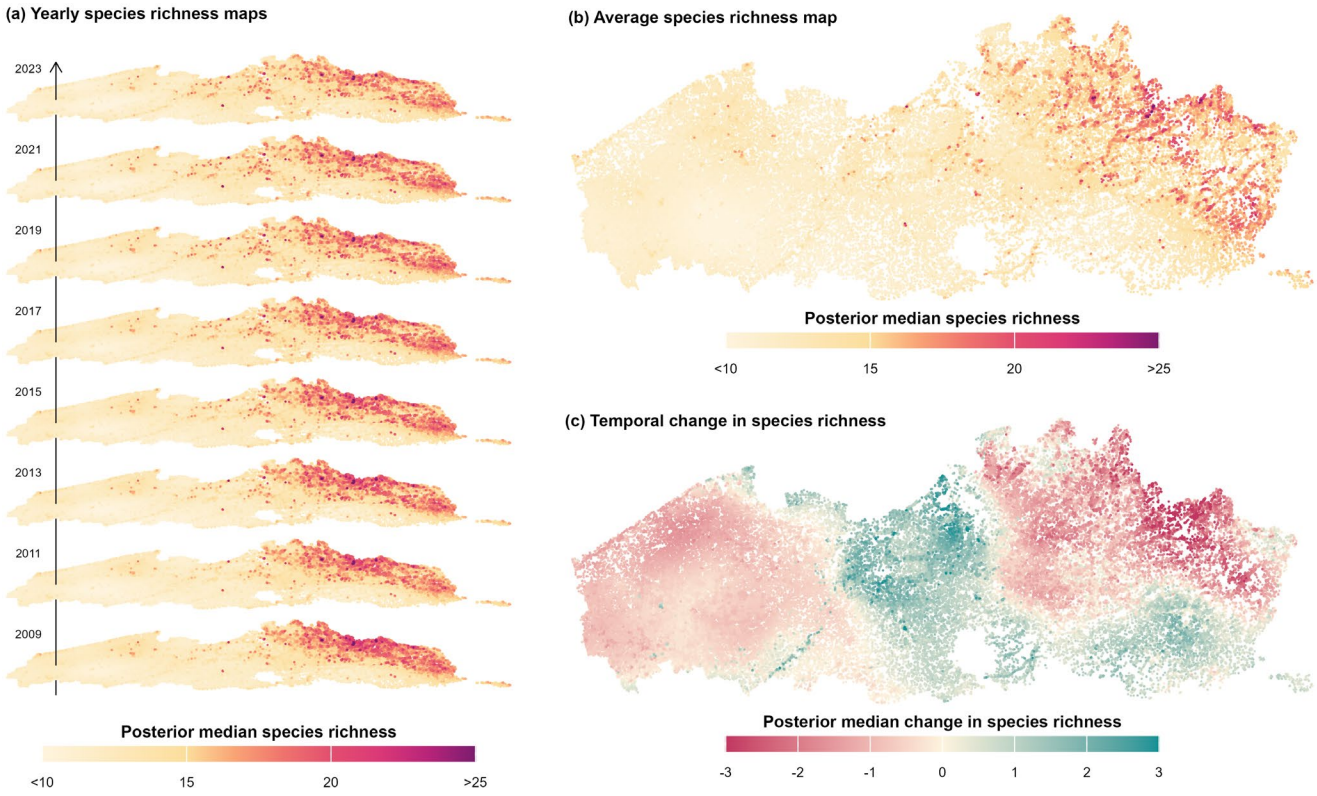

FIGURE 5 | Patterns of species richness in space and time. (a) Yearly maps of posterior median estimated species richness across all 93,135 known waterbodies in Flanders. For clarity, colour scales have been truncated to the [−5,5] interval, with faint yellow colours corresponding to a richness of 10 (or less) species and purple corresponding to a richness of 25 (or more) species. Maps are shown biennially (every 2 years) and presented in a slightly tilted, stacked perspective to conserve space, with the year printed next to each map. (b) Map of the posterior median estimated species richness across all 93,135 known waterbodies in Flanders, averaged over all years. An identical colour scale has been used as for panel (a). (c) Map of the posterior median slope of the species richness throughout the 15-year study period all 93,135 known waterbodies in Flanders. The slopes have been obtained by performing ordinary least squares on the (untransformed) yearly species richness estimates, for each posterior iteration and by subsequently computing the posterior median slope. By using the untransformed yearly species richness estimates, the values can be interpreted as the number of species lost or gained over the 15-year period. Red colours correspond to three (or more) species lost, while blue-green colours correspond to three (or more) species gained.

conservation and policymaking, while simultaneously valorising efforts of naturalists.

Our approach provides means for researchers in metacommunity ecology to exploit opportunistically collected data, now by far the most widely available biodiversity data on Earth. It enables analyses at unprecedented taxonomic and spatial scales and offers opportunities to include the temporal dimension, meeting recent calls for an increased emphasis on the role of time in metacommunity ecology (Guzman et al. 2022; Yin and Rudolf 2024). Despite a high degree of standardisation, typically used datasets in metacommunity ecology might be prone to phenological biases, as communities are rarely sampled simultaneously across all sites, let alone during the phenological peaks of each individual species. Our model, on the other hand, spans the entire yearly phenological cycle, fully acknowledging seasonally varying detection probabilities. Our model can easily be extended to capture interannual phenological variation, further improving the estimation of detection probabilities and providing meaningful biological insights on phenological change in the process. Other model developments such as the inclusion of hidden Markov model components, distributed lag structures, spatially varying environmental associations or temporally varying interspecific associations might allow the detection of more complex (spatio)temporal dynamics in metacommunity structure, such as regime shifts and historical legacy effects (Doser et al. 2024; Fukami 2015; Gasparrini et al. 2010; McClintock et al. 2020).

While properly executed standardised surveys of species (abundances) always constitute the gold standard approach for biodiversity monitoring (Reynolds et al. 2011), our model aids these efforts by providing yearly trend estimates for the entire species community through the efforts of naturalists and citizen scientists, without coercing them into following a restrictive protocol. More specifically, our approach can fill gaps for species, sites or time windows for which rigorously collected data is absent (Bowler et al. 2024). Ongoing efforts in the field of data fusion will facilitate the combined use of both types of monitoring paradigms (Pacifici et al. 2016; Theobald et al. 2015; Zipkin et al. 2023).

Our approach shares some inherent limitations of HMSC and similar joint species distribution models. For instance, its correlative nature does not eliminate the possibility of confounding, and the omission of relevant environmental predictors can impede the interpretation of interspecific associations as biotic interactions (Ovaskainen and Abrego 2020). These limitations may be

14610248, 2025, 3, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/ele.70094, Wiley Online Library on [17/03/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

further compounded in large-scale applications like ours, where the choice of environmental predictors is constrained by their availability (e.g., land cover data). In our case study, the inclusion of more detailed local environmental predictors (e.g., water physicochemical variables) was rendered impossible due to the spatial resolution, but it could have further improved our understanding of the metacommunity. Further model extensions using Bayesian imputation techniques (Gelman et al. 2014) offer perspectives to utilise partially available physicochemical data on a fraction of waterbodies, thereby exploiting species as bioindicators, as well as statistical relationships between local physicochemical variables and land cover. Likewise, the availability of detection metadata is often limited in opportunistically collected biodiversity datasets. For instance, the use of list length in our application might be a suboptimal proxy for search effort as it is sensitive to species richness patterns through space and time. Biodiversity portals such as Observation.org now enable observers to record their GPS tracks during visits, which could offer substantial benefits for the analysis of opportunistically collected datasets. An additional limitation arises from acknowledging imperfect detection, as the true absence of a species cannot be established with certainty, complicating the straightforward assessment of predictive performance. A final limitation of our approach is that it is restricted to inferring occupancy, whereas incorporating species abundances could further enhance the understanding of metacommunities.

Since imperfect and heterogeneous detection is a near-ubiquitous phenomenon in biology, our approach can cater applications beyond opportunistically sampled biodiversity data in the strict sense. For instance, Fountain-Jones et al. (2024) recently outlined the exciting opportunities that both joint species distribution models and site-occupancy models could offer for the analysis of microbiomes, and Allaband et al. (2024) present compelling evidence that the time of sample collection is critical for the replicability of microbiome analyses, with the effects of time exceeding those of experimental interventions, leading to altered conclusions. Our approach offers a one-step solution, as diurnal patterns in detection can be accounted for in the same way as phenological patterns. Other potential applications include eDNA sampling and camera trapping.

We hope the present framework provides a venue for further methodological developments in joint species distribution modelling while accounting for imperfect detection, and that its application to the large amount of available biodiversity data will increasingly contribute to our understanding of the processes shaping the assembly of species across time and space.

Author Contributions

M.F., R.W., M.H. and L.D.M. conceived the initial idea; G.D.K., R.S., P.L., P.V. and T.N. contributed substantially to further conceptualisation; M.F., R.W. and T.N. designed the methodology; M.F. and R.W. performed the statistical analysis and led the writing of the manuscript. All authors contributed significantly to the drafts and gave final approval for publication.

Acknowledgements

We thank the Flemish Dragonfly Society and the many naturalists for contributing their sightings to Waarnemingen.be – Observations.be, as well as the experts of the platform's data validation team. M.F., T.N. and L.D.M. acknowledge funding by Research Foundation Flanders (FWO, grant numbers 11E3222N, G0A4121N and G0A3M24N). R.W. acknowledges a KU Leuven PhD scholarship (grant number DB/22/007/bm). L.D.M. and R.S. acknowledge financial support from a KU Leuven research project (number C16/2023/003). L.D.M. acknowledges support by an IGB starting fund. M.F., R.W., P.L. and L.D.M. acknowledge support from the PONDERFUL project, funded by the European Union's Horizon 2020 research and innovation program under grant agreement No. 869296. The computational resources and services used in this work were provided by the VSC (Flemish Supercomputer Center), funded by the Research Foundation Flanders (FWO) and the Flemish Government.

Conflicts of Interest

The authors declare no conflicts of interest.

Data Availability Statement

Fully documented Stan code and a complete R-pipeline for data preparation, model running, results extraction and visualisation are available on GitHub: https://github.com/mfajgenblat/opportunisticJSDM. Through these scripts, interested readers can replicate analyses and apply the model to their own datasets.

References

Allaband, C., A. Lingaraju, S. Flores Ramos, et al. 2024. "Time of Sample Collection Is Critical for the Replicability of Microbiome Analyses." *Nature Metabolism* 6, no. 7: 1282–1293. https://doi.org/10.1038/s42255-024-01064-1.

Beissinger, S. R., K. J. Iknayan, G. Guillera-Arroita, et al. 2016. "Incorporating Imperfect Detection Into Joint Models of Communities: A Response to Warton et al." *Trends in Ecology & Evolution* 31, no. 10: 736–737. https://doi.org/10.1016/j.tree.2016.07.009.

Betancourt, M. 2017. "A Conceptual Introduction to Hamiltonian Monte Carlo." ArXiv.

Bhattacharya, A., and D. B. Dunson. 2011. "Sparse Bayesian Infinite Factor Models." *Biometrika* 98: 291–306.

Bowler, D. E., R. J. Boyd, C. T. Callaghan, R. A. Robinson, N. J. B. Isaac, and M. J. O. Pocock. 2024. "Treating Gaps and Biases in Biodiversity Data as a Missing Data Problem." *Biological Reviews* 100: 50–67.

Bowler, D. E., C. T. Callaghan, N. Bhandari, et al. 2022. "Temporal Trends in the Spatial Bias of Species Occurrence Records." *Ecography* 2022: e06219.

Carpenter, B., A. Gelman, M. Hoffman, et al. 2017. "Stan: A Probabilistic Programming Language." *Journal of Statistical Software* 76: 1–32.

Cole, E., G. Van Horn, C. Lange, et al. 2023. "Spatial Implicit Neural Representations for Global-Scale Species Mapping." In *Proceedings of the 40th International Conference on Machine Learning*.

Cook, S. R., A. Gelman, and D. B. Rubin. 2006. "Validation of Software for Bayesian Models Using Posterior Quantiles." *Journal of Computational and Graphical Statistics* 15: 675–692.

Corbet, P. S. 1999. *Dragonflies: Behavior and Ecology of Odonata*. Comstock Pub. Associates.

Córdoba-Aguilar, A., C. Beatty, and J. Bried. 2022. *Dragonflies and Damselflies: Model Organisms for Ecological and Evolutionary Research*. Oxford University Press.

Cottenie, K. 2005. "Integrating Environmental and Spatial Processes in Ecological Community Dynamics." *Ecology Letters* 8: 1175–1182.

Davis, C. L., Y. Bai, D. Chen, et al. 2023. "Deep Learning With Citizen Science Data Enables Estimation of Species Diversity and Composition at Continental Extents." *Ecology* 104: e4175.

14610248, 2025, 3, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/ele.70094, Wiley Online Library on [17/03/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

De Knijf, G., A. Anselin, P. Goffart, and M. Tailly. 2006. *De Libellen (Odonata) van België: verspreiding - evolutie - habitats*. Libellenwerkgroep Gomphus i.s.m. Instituut voor Natuur- en Bosonderzoek.

De Saeger, S., R. Guelinckx, P. Oosterlynck, et al. 2020. "Biologische Waarderingskaart en Natura 2000 Habitatkaart, Uitgave 2020."

Dickinson, J. L., B. Zuckerberg, and D. N. Bonter. 2010. "Citizen Science as an Ecological Research Tool: Challenges and Benefits." *Annual Review of Ecology, Evolution, and Systematics* 41: 149–172.

Dorazio, R. M., J. A. Royle, B. Söderström, and A. Glimskär. 2006. "Estimating Species Richness and Accumulation by Modeling Species Occurrence and Detectability." *Ecology* 87: 842–854.

Doser, J. W., A. O. Finley, and S. Banerjee. 2023. "Joint Species Distribution Models With Imperfect Detection for High-Dimensional Spatial Data." *Ecology* 104: e4137.

Doser, J. W., A. O. Finley, M. Kéry, and E. F. Zipkin. 2022. "spOccupancy: An R Package for Single-Species, Multi-Species, and Integrated Spatial Occupancy Models." *Methods in Ecology and Evolution* 13: 1670–1678.

Doser, J. W., A. O. Finley, S. P. Saunders, M. Kéry, A. S. Weed, and E. F. Zipkin. 2024. "Modeling Complex Species-Environment Relationships Through Spatially-Varying Coefficient Occupancy Models." *Journal of Agricultural, Biological, and Environmental Statistics*. https://link.springer.com/article/10.1007/s13253-023-00595-6#citeas.

Doser, J. W., and S. Stoudt. 2024. "Fractional Replication in Single-Visit Multi-Season Occupancy Models: Impacts of Spatiotemporal Autocorrelation on Identifiability." *Methods in Ecology and Evolution* 15: 358–372.

Fink, D., A. Johnston, M. Strimas-Mackey, et al. 2023. "A Double Machine Learning Trend Model for Citizen Science Data." *Methods in Ecology and Evolution* 14: 2435–2448.

Fountain-Jones, N. M., T. Giraud, L. Zinger, H. Bik, S. Creer, and E. Videvall. 2024. "Molecular Ecology of Microbiomes in the Wild: Common Pitfalls, Methodological Advances and Future Directions." *Molecular Ecology* 33: e17223.

Fukami, T. 2015. "Historical Contingency in Community Assembly: Integrating Niches, Species Pools, and Priority Effects." *Annual Review of Ecology, Evolution, and Systematics* 46: 1–23.

Gasparrini, A., B. Armstrong, and M. G. Kenward. 2010. "Distributed Lag Non-Linear Models." *Statistics in Medicine* 29: 2224–2234.

Gelman, A., J. B. Carlin, H. S. Stern, D. B. Dunson, A. Vehtari, and D. B. Rubin. 2014. *Bayesian Data Analysis*. 3rd ed. Chapman and Hall/CRC.

Gelman, A., A. Vehtari, D. Simpson, et al. 2020. "Bayesian Workflow." ArXiv.

Guillera-Arroita, G. 2017. "Modelling of Species Distributions, Range Dynamics and Communities Under Imperfect Detection: Advances, Challenges and Opportunities." *Ecography* 40: 281–295.

Guillera-Arroita, G., J. J. Lahoz-Monfort, J. Elith, et al. 2015. "Is My Species Distribution Model Fit for Purpose? Matching Data and Models to Applications." *Global Ecology and Biogeography* 24: 276–292.

Guillera-Arroita, G., J. J. Lahoz-Monfort, D. I. MacKenzie, B. A. Wintle, and M. A. McCarthy. 2014. "Ignoring Imperfect Detection in Biological Surveys Is Dangerous: A Response to "Fitting and Interpreting Occupancy Models"." *PLoS One* 9, no. 7: e99571.

Guzman, L. M., P. L. Thompson, D. S. Viana, et al. 2022. "Accounting for Temporal Change in Multiple Biodiversity Patterns Improves the Inference of Metacommunity Processes." *Ecology* 103: e3683.

Harabiš, F., and J. Hronková. 2020. "European Database of the Life-History, Morphological and Habitat Characteristics of Dragonflies (Odonata)." *European Journal of Entomology* 117: 302–308.

Hochachka, W. M., D. Fink, R. A. Hutchinson, D. Sheldon, W. K. Wong, and S. Kelling. 2012. "Data-Intensive Science Applied to Broad-Scale Citizen Science." *Trends in Ecology & Evolution* 27: 130–137.

Hochachka, W. M., V. Ruiz-Gutierrez, and A. Johnston. 2023. "Considerations for Fitting Occupancy Models to Data From eBird and Similar Volunteer-Collected Data." *Ornithology* 140: 1–13.

Hoffman, M., and Y. Ma. 2020. "Black-Box Variational Inference as a Parametric Approximation to Langevin Dynamics." In *Proceedings of the 37th International Conference on Machine Learning, Proceedings of Machine Learning Research*, edited by H. D. III and A. Singh, 4324–4341. PMLR.

Hui, F. K. C., S. Taskinen, S. Pledger, S. D. Foster, and D. I. Warton. 2015. "Model-Based Approaches to Unconstrained Ordination." *Methods in Ecology and Evolution* 6: 399–411.

Isaac, N. J. B., A. J. van Strien, T. A. August, M. P. de Zeeuw, and D. B. Roy. 2014. "Statistics for Citizen Science: Extracting Signals of Change From Noisy Ecological Data." *Methods in Ecology and Evolution* 5: 1052–1060.

Johnston, A., D. Fink, W. M. Hochachka, and S. Kelling. 2018. "Estimates of Observer Expertise Improve Species Distributions From Citizen Science Data." *Methods in Ecology and Evolution* 9: 88–97.

Joseph, M. B. 2020. "Neural Hierarchical Models of Ecological Populations." *Ecology Letters* 23: 734–747.

Kelling, S., A. Johnston, W. M. Hochachka, et al. 2015. "Can Observation Skills of Citizen Scientists Be Estimated Using Species Accumulation Curves?" *PLoS One* 10: 1–20.

Lahoz-Monfort, J. J., G. Guillera-Arroita, and B. A. Wintle. 2014. "Imperfect Detection Impacts the Performance of Species Distribution Models." *Global Ecology and Biogeography* 23: 504–515.

Legendre, P., and L. Legendre. 1998. *Numerical Ecology*. Elsevier Science.

Leibold, M. A., and J. M. Chase. 2017. *Metacommunity Ecology*. Princeton University Press.

Leibold, M. A., F. J. Rudolph, F. G. Blanchet, et al. 2022. "The Internal Structure of Metacommunities." *Oikos* 2022: e08618.

Lerch, B. A., A. Rudrapatna, N. Rabi, J. Wickman, T. Koffel, and C. A. Klausmeier. 2023. "Connecting Local and Regional Scales With Stochastic Metacommunity Models: Competition, Ecological Drift, and Dispersal." *Ecological Monographs* 93: e1591.

Logue, J. B., N. Mouquet, H. Peter, and H. Hillebrand. 2011. "Empirical Approaches to Metacommunities: A Review and Comparison With Theory." *Trends in Ecology & Evolution* 26, no. 9: 482–491. https://doi.org/10.1016/j.tree.2011.04.009.

MacKenzie, D. I., and L. L. Bailey. 2004. "Assessing the Fit of Site-Occupancy Models." *Journal of Agricultural, Biological, and Environmental Statistics* 9: 300–318.

MacKenzie, D. I., J. D. Nichols, G. B. Lachman, S. Droege, A. A. Royle, and C. A. Langtimm. 2002. "Estimating Site Occupancy Rates When Detection Probabilities Are Less Than One." *Ecology* 83: 2248–2255.

Mackenzie, D. I., J. D. Nichols, J. A. Royle, K. H. Pollock, L. L. Bailey, and J. E. Hines. 2018. *Occupancy Estimation and Modeling*. 2nd ed. Academic Press.

Margossian, C. C., and A. Gelman. 2023. "For How Many Iterations Should We Run Markov Chain Monte Carlo?" ArXiv.

McClintock, B. T., R. Langrock, O. Gimenez, et al. 2020. "Uncovering Ecological State Dynamics With Hidden Markov Models." *Ecology Letters* 23, no. 12: 1878–1903. https://doi.org/10.1111/ele.13610.

Miller, D. A. W., L. L. Bailey, E. H. C. Grant, B. T. Mcclintock, L. A. Weir, and T. R. Simons. 2015. "Performance of Species Occurrence Estimators

14610248, 2025, 3, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/ele.70094, Wiley Online Library on [17/03/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

When Basic Assumptions Are Not Met: A Test Using Field Data Where True Occupancy Status Is Known." *Methods in Ecology and Evolution* 6: 557–565.

Modrák, M., A. H. Moon, S. Kim, et al. 2023. "Simulation-Based Calibration Checking for Bayesian Computation: The Choice of Test Quantities Shapes Sensitivity." *Bayesian Analysis* 1–28.

Monnahan, C. C., J. T. Thorson, and T. A. Branch. 2017. "Faster Estimation of Bayesian Models in Ecology Using Hamiltonian Monte Carlo." *Methods in Ecology and Evolution* 8: 339–348.

Monod, M., A. Blenkinsop, A. Brizzi, et al. 2022. "Regularised B-Splines Projected Gaussian Process Priors to Estimate Time-Trends in Age-Specific COVID-19 Deaths." *Bayesian Analysis* 18, no. 3: 957–987.

Mungi, N. A., Y. V. Jhala, Q. Qureshi, E. le Roux, and J. C. Svenning. 2023. "Megaherbivores Provide Biotic Resistance Against Alien Plant Dominance." *Nature Ecology & Evolution* 7, no. 10: 1645–1653. https://doi.org/10.1038/s41559-023-02181-y.

Neyens, T., P. J. Diggle, C. Faes, N. Beenaerts, T. Artois, and E. Giorgi. 2019. "Mapping Species Richness Using Opportunistic Samples: A Case Study on Ground-Floor Bryophyte Species Richness in the Belgian Province of Limburg." *Scientific Reports* 9: 1–11.

Norberg, A., N. Abrego, F. G. Blanchet, et al. 2019. "A Comprehensive Evaluation of Predictive Performance of 33 Species Distribution Models at Species and Community Levels." *Ecological Monographs* 89: e01370.

Outhwaite, C. L., R. E. Chandler, G. D. Powney, B. Collen, R. D. Gregory, and N. J. B. Isaac. 2018. "Prior Specification in Bayesian Occupancy Modelling Improves Analysis of Species Occurence Data." *Ecological Indicators* 93: 333–343.

Ovaskainen, O., and N. Abrego. 2020. *Joint Species Distribution Modelling*. Cambridge University Press.

Ovaskainen, O., D. B. Roy, R. Fox, and B. J. Anderson. 2016. "Uncovering Hidden Spatial Structure in Species Communities With Spatially Explicit Joint Species Distribution Models." *Methods in Ecology and Evolution* 7: 428–436.

Ovaskainen, O., J. Rybicki, and N. Abrego. 2019. "What Can Observational Data Reveal About Metacommunity Processes?" *Ecography* 42: 1877–1886.

Ovaskainen, O., G. Tikhonov, A. Norberg, et al. 2017. "How to Make More out of Community Data? A Conceptual Framework and Its Implementation as Models and Software." *Ecology Letters* 20: 561–576.

Pacifici, K., B. J. Reich, R. M. Dorazio, and M. J. Conroy. 2016. "Occupancy Estimation for Rare Species Using a Spatially-Adaptive Sampling Design." *Methods in Ecology and Evolution* 7: 285–293.

Pedersen, E. J., D. L. Miller, G. L. Simpson, and N. Ross. 2019. "Hierarchical Generalized Additive Models in Ecology: An Introduction With Mgcv." *PeerJ* 2019: e6876.

Peres-Neto, P. R., P. Legendre, S. Dray, and D. Borcard. 2006. "Variation Partitioning of Species Data Matrices: Estimation and Comparison of Fractions." *Ecology* 87: 2614–2625.

Pocock, M. J. O., M. Logie, N. J. B. Isaac, R. Fox, and T. August. 2023. "The Recording Behaviour of Field-Based Citizen Scientists and Its Impact on Biodiversity Trend Analysis." *Ecological Indicators* 151: 110276.

Pollock, L. J., R. Tingley, W. K. Morris, et al. 2014. "Understanding Co-Occurrence by Modelling Species Simultaneously With a Joint Species Distribution Model (JSDM)." *Methods in Ecology and Evolution* 5: 397–406.

Rasmussen, C. E., and C. K. I. Williams. 2006. *Gaussian Processes for Machine Learning*. MIT Press.

Reynolds, J. H., W. L. Thompson, and B. Russell. 2011. "Planning for Success: Identifying Effective and Efficient Survey Designs for Monitoring." *Biological Conservation* 144: 1278–1284.

Ries, L., E. F. Zipkin, M. D. Scheuerell, et al. 2016. "Joint Dynamic Species Distribution Models: A Tool for Community Ordination and Spatio-Temporal Monitoring." *Global Ecology and Biogeography* 25: 1144–1158.

Rota, C. T., R. J. Fletcher, J. M. Evans, and R. L. Hutto. 2011. "Does Accounting for Imperfect Detection Improve Species Distribution Models?" *Ecography* 34: 659–670.

Samways, M. J. 2024. *Freshwater Assessment and Monitoring Using Dragonflies*, 331–438. CABI Books.

Sauer, J. R., B. G. Peterjohn, and W. A. Link. 1994. "Observer Differences in the North American Breeding Bird Survey." *Auk* 111: 50–62.

Scheers, K., V. Smeekens, C. Wils, J. Packet, A. Leyssen, and L. Denys. 2022. "Watervlakken Versie 1.2: Polygonenkaart van Stilstaand Water in Vlaanderen Uitgave 2022."

Shirey, V., M. W. Belitz, V. Barve, and R. Guralnick. 2021. "A Complete Inventory of North American Butterfly Occurrence Data: Narrowing Data Gaps, but Increasing Bias." *Ecography* 44, no. 4: 537–547.

Smith, A. C., and B. P. M. Edwards. 2021. "North American Breeding Bird Survey Status and Trend Estimates to Inform a Wide Range of Conservation Needs, Using a Flexible Bayesian Hierarchical Generalized Additive Model." *Ornithological Applications* 123: 1–16.

Sutherland, C., M. Brambilla, P. Pedrini, and S. Tenan. 2016. "A Multiregion Community Model for Inference About Geographic Variation in Species Richness." *Methods in Ecology and Evolution* 7: 783–791.

Szabo, J. K., P. A. Vesk, P. W. J. Baxter, and H. P. Possingham. 2010. "Regional Avian Species Declines Estimated From Volunteer-Collected Long-Term Data Using List Length Analysis." *Ecological Applications* 20: 2157–2169.

Talts, S., M. Betancourt, D. Simpson, A. Vehtari, and A. Gelman. 2018. "Validating Bayesian Inference Algorithms With Simulation-Based Calibration." ArXiv.

Teng, M., A. Elmustafa, B. Akera, Y. Bengio, H. Radi, and A. Mila. n.d. "SatBird: Bird Species Distribution Modeling with Remote Sensing and Citizen Science Data."

Termaat, T., A. J. van Strien, R. H. A. van Grunsven, et al. 2019. "Distribution Trends of European Dragonflies Under Climate Change." *Diversity and Distributions* 25: 936–950.

Theobald, E. J., A. K. Ettinger, H. K. Burgess, et al. 2015. "Global Change and Local Solutions: Tapping the Unrealized Potential of Citizen Science for Biodiversity Research." *Biological Conservation* 181: 236–244.

Tikhonov, G., Ø. H. Opedal, N. Abrego, et al. 2020. "Joint Species Distribution Modelling With the r-Package Hmsc." *Methods in Ecology and Evolution* 11: 442–447.

van Strien, A. J., C. A. M. Swaay, and M. Kéry. 2011. "Metapopulation Dynamics in the Butterfly Hipparchia Semele Changed Decades Before Occupancy Declined in The Netherlands." *Ecological Applications* 21: 2510–2520.

van Strien, A. J., T. Termaat, D. Groenendijk, V. Mensing, and M. Kéry. 2010. "Site-Occupancy Models May Offer New Opportunities for Dragonfly Monitoring Based on Daily Species Lists." *Basic and Applied Ecology* 11: 495–503.

van Strien, A. J., C. A. M. Van Swaay, and T. Termaat. 2013. "Opportunistic Citizen Science Data of Animal Species Produce Reliable Estimates of Distribution Trends if Analysed With Occupancy Models." *Journal of Applied Ecology* 50: 1450–1458.

Vanoverbeke, J., M. C. Urban, and L. De Meester. 2016. "Community Assembly Is a Race Between Immigration and Adaptation: Eco-Evolutionary Interactions Across Spatial Scales." *Ecography* 39: 858–870.

14610248, 2025, 3, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/ele.70094, Wiley Online Library on [17/03/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Warton, D. I., F. G. Blanchet, R. B. O'Hara, et al. 2015. "So Many Variables: Joint Modeling in Community Ecology." *Trends in Ecology & Evolution* 30, no. 12: 766–779. https://doi.org/10.1016/j.tree.2015.09.007.

Wolkovich, E. M., B. I. Cook, K. K. McLauchlan, and T. J. Davies. 2014. "Temporal Ecology in the Anthropocene." *Ecology Letters* 17: 1365–1379.

Yin, H., and V. H. W. Rudolf. 2024. "Time Is of the Essence: A General Framework for Uncovering Temporal Structures of Communities." *Ecology Letters* 27: e14481.

Zhang, L., B. Carpenter, A. Gelman, and A. Vehtari. 2022. "Pathfinder: Parallel Quasi-Newton Variational Inference." *JMLR* 23: 1–49.

Zipkin, E. F., J. W. Doser, C. L. Davis, W. Leuenberger, S. Ayebare, and K. L. Davis. 2023. "Integrated Community Models: A Framework Combining Multispecies Data Sources to Estimate the Status, Trends and Dynamics of Biodiversity." *Journal of Animal Ecology* 92: 2248–2262.

Supporting Information

Additional supporting information can be found online in the Supporting Information section.

14610248, 2025, 3, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/ele.70094, Wiley Online Library on [17/03/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License