JYX
JYVÄSKYLÄN YLIOPISTO
UNIVERSITY OF JYVÄSKYLÄ

This is a self-archived version of an original article. This version may differ from the original in pagination and typographic details.

Author(s): Duflot, Rémi; Vähätalo, Anssi V.

Title: Identifying sites with high biodiversity value using filtered species records from a biodiversity information facility

Year: 2024

Version: Published version

Copyright: © 2024 The Author(s). Diversity and Distributions published by John Wiley & Sons

Rights: CC BY 4.0

Rights url: https://creativecommons.org/licenses/by/4.0/

Please cite the original version:

Duflot, R., & Vähätalo, A. V. (2024). Identifying sites with high biodiversity value using filtered species records from a biodiversity information facility. Diversity and Distributions, Early View. https://doi.org/10.1111/ddi.13864

Check for updates

Received: 2 February 2024 | Revised: 4 May 2024 | Accepted: 6 May 2024

DOI: 10.1111/ddi.13864

RESEARCH ARTICLE

Diversity and Distributions WILEY

Identifying sites with high biodiversity value using filtered species records from a biodiversity information facility

Rémi Duflot[1,2] | Anssi V. Vähätalo[1,3]

[1]Department of Biological and Environmental Science, University of Jyväskylä, Jyväskylä, Finland

[2]School of Resource Wisdom, University of Jyväskylä, Jyväskylä, Finland

[3]Department of Forest Sciences, University of Helsinki, Helsinki, Finland

Correspondence
Rémi Duflot, Department of Biological and Environmental Science, University of Jyvaskyla, PO Box 35 (Survontie 9 C), FI-40014 Jyväskylä, Finland.
Email: remi.r.duflot@jyu.fi

Funding information
Koneen Säätiö, Grant/Award Number: 202105759

Editor: Stefano Mammola

Abstract

Aim: Efficient mitigation of the biodiversity crisis requires targeted conservation actions in locations with high species richness, the presence of endangered species and unique species communities. However, prioritising sites remains challenging because of sparse knowledge on biodiversity, limiting the possibility of communicating efficiently with local decision makers. We examine easy-to-replicate, yet robust, methods to identify areas with high conservation values on large spatial scales using data filtering and complementary biodiversity indicators based on species records from a biodiversity information facility.

Location: Finland, Europe.

Methods: We illustrate the protocol by focusing on Lepidoptera in Finnish municipal districts. We mobilised over 3 million species records on 878 native Lepidoptera (2001–2020) from the Finnish Biodiversity Information Facility. We estimated the richness of overall and endangered species using species accumulation curves, as well as the uniqueness of species communities, using measures of local contribution to beta diversity (LCBD). After testing for multiple thresholds and their effect on indicator accuracy, 97 districts with >5000 records were included in the analyses.

Results: Estimated overall species richness was highest on the southern coast and significantly decreased in the North, following a known pattern with Lepidoptera in Finland. Species richness was not the highest in the districts with the greatest number of records and the ranking differed from the raw data, demonstrating the importance of correcting for sampling intensity. The estimated number of endangered species correlated with overall species richness, except in northernmost districts, where the proportion of endangered species was exceptionally high. High LCBD replacement (i.e. unique species communities) was concentrated in the Southwest (hemi-boreal) and North (northern boreal) of the country.

Main Conclusions: We provided an example and interpretations of how scalable biodiversity indicators based on accumulation curves and LCBD analyses, and careful data filtering (thresholds) can be used to identify sites with conservation priorities from multi-sourced species records.

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.
© 2024 The Author(s). *Diversity and Distributions* published by John Wiley & Sons Ltd.

KEYWORDS
Biodiversity, butterfly, citizen science, conservation planning, Lepidoptera, prioritisation, threatened species

1 | INTRODUCTION

Anthropogenic pressures on the biosphere cause unprecedented threats to biodiversity (Butchart et al., 2010; IPBES, 2019). If we are to halt and reverse biodiversity losses, good knowledge and understanding of biodiversity distribution and dynamics are required. Relevant indicators are needed to measure improvements in or worsening of biodiversity and, thus, assess conservation policies' efficiency (Butchart et al., 2007; Pereira et al., 2013). Monitoring biodiversity is also critical to directing conservation actions towards hotspots and the most sensitive areas, a process known as conservation planning (Kukkala & Moilanen, 2013). Identifying priority areas is a key step when expanding protected area networks or implementing sustainable practices (e.g. through farming or forestry subsidised schemes). Ideally, such conservation actions are to be established in landscapes with high biodiversity value. The main challenge is to obtain biodiversity indicators on large spatial scales to compare biodiversity value across potential sites (Margules & Pressey, 2000).

The development and structuring of 'citizen' science is viewed as a powerful way to monitor biodiversity on large geographical scales and overcome the immense task of data collection (Chandler et al., 2017; Johnston et al., 2023; McKinley et al., 2017). Although many countries have implemented structured biodiversity-monitoring programmes (also often based on volunteers), these programmes often target representative landscapes to estimate national trends. Aggregating the available information from monitoring programmes with opportunistic and semi-structured data from expert hobbyists (volunteers) can improve geographical coverage (Shirey et al., 2021). However, such data are not immune from geographical bias either, as participants choose their own sites, and the selection is not random (Callaghan et al., 2022; Johnston et al., 2023). As a result, records tend to concentrate on areas with high human population density (e.g. near urban areas), are easily accessible (e.g. near roads), and attractive (e.g. the presence of protected areas or rare species). Such a pattern exists on global, national and regional scales (e.g. Girardello et al., 2019; Hughes et al., 2021; Matutini et al., 2021; Shirey et al., 2021).

Correcting species records for different sampling efforts in space and time is a long-standing issue in ecological science (e.g. Gotelli & Colwell, 2001). Despite strong geographical biases, adequate statistical methods that account for sampling effort and careful data filtering can provide robust results (Callaghan et al., 2019; Matutini et al., 2021; Shirey et al., 2023). These methods have been focusing on species-level analyses, such as species distribution or occupancy models (or aggregations of species-level models) to provide information on species abundance, distribution or phenology (Johnston et al., 2023; McKinley et al., 2017). Although valuable for cultivating a mechanistic understanding of biodiversity responses, these approaches are missing the potentially useful information revealed when working at the species community level. In particular, these studies often use a subset of species for which sufficient data are available to obtain robust species-level models.

Robust indicators also are needed on the species community level to support conservation in practice by identifying sites with high biodiversity value in general, beyond individual species. First, species richness is viewed widely as a fundamental measure of overall biodiversity and is often used to prioritise areas for protection in conservation planning on various spatial scales (Callaghan et al., 2022; Gotelli & Colwell, 2001). Second, species extinction risk, based on their distribution and/or population status, is viewed as a relevant indicator of biodiversity state (Butchart et al., 2007; Pereira et al., 2013). Therefore, conservation actions often consider the presence of endangered species to identify priority sites (e.g. Mikkonen et al., 2023). Third, beyond the number of species (overall or endangered), the identity of species is also of great importance. Conservation planning aims to preserve the full variety of biodiversity, which is usually achieved by selecting multiple representative and complementary areas (Margules & Pressey, 2000). Beta diversity, which is the variation in the identities of species in space or time (Anderson et al., 2011), can be used to identify sites hosting unusual species composition, that is, hosting species not found elsewhere (Socolar et al., 2016). These sites might not necessarily be the richest.

This study examines easy-to-replicate, yet robust, methods of identifying areas with the highest conservation priorities on large spatial scales using indicators of species richness, species endangered status and species composition derived from multi-sourced species records. We aimed to test whether a combination of data filtering and correction for observation intensity allows for accurate identification of species hotspots, using the latest methods of species accumulation curve and local contribution to beta diversity (LCBD) analyses. Specifically, we focused on identifying municipal districts in Finland (hereafter *districts*) with high conservation values for native Lepidoptera. Lepidoptera is a very well-documented group (the best among insects) and is sensitive to climate and land use changes (Devictor et al., 2012; Montràs-Janer et al., 2024; Zellweger et al., 2017), making them a relevant focal group to study for biodiversity responses to global changes and a proxy for other terrestrial insects (Hällfors et al., 2021; Shirey et al., 2021; Warren et al., 2021) and plants (Tyler, 2020). Many Lepidoptera species are declining in Europe, with some species expanding, leading to biotic homogenisation (Engelhardt et al., 2022; Montràs-Janer et al., 2024; Warren et al., 2021). Finland is home to many threatened species, primarily due to habitat loss and

14724642, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/ddi.13864 by University Of Jyväskylä Library, Wiley Online Library on [22/05/2024]. 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

degradation, as well as climate change (Hyvärinen et al., 2019; Pöyry et al., 2009). We conducted our analyses at the district level, as districts make influential political and administrative decisions regarding land use, which directly impact biodiversity. Local improvements in landscape management—such as extended conservation areas, support for traditional land use practices (e.g. semi-natural grasslands) and extensively managed urban green infrastructure—are efficient tools for maintaining biodiversity, and Lepidoptera in particular (Montràs-Janer et al., 2024; Warren et al., 2021). National conservation programmes could support district hotspots by incentivising local actions. Beyond establishing a district hotspot ranking, we ask the following questions: (i) How much data filtering is required to obtain homogenous and accurate results? (ii) What is the spatial coverage of existing multi-sourced species records in Finland? (iii) How does the ranking of district hotspots vary depending on the diversity indicator considered (i.e. complementarity of indicators)?

2 | MATERIALS AND METHODS

2.1 | Species records

The Finnish Biodiversity Information Facility (FinBIF) collects, consolidates and shares (laji.fi) biodiversity datasets from research, state agencies and the public (Schulman et al., 2021). We used the R package FinBIF (Morris, 2021) to retrieve Lepidoptera observations from the main database. We restricted the search to the most intensively studied, large-sized Lepidoptera, that is, the informal group of macro-Lepidoptera (see Supplementary Material [SM] 1.1 for more details on the taxonomy used). We downloaded all records from 1 January 2001 to 31 December 2020 with species names and municipal district of observation (which filters out observations with coarser geographic resolution). More details on how a municipal district is assigned to the reported observations may be found in SM 1.1. Furthermore, we asked FinBIF administrators for non-public observation data on sensitive species (representing 5062 observations of 12 species), which we received. A few instances of subspecies identifications were reclassified at the species level and vagrant species, that is, non-native species occasionally found in Finland, were removed from the analyses. We selected the native species with permanent reproducing populations using the red list (Hyvärinen et al., 2019). Then, observations were aggregated per species and district. The complete list of selected species may be found in SM 2, as well as their recorded number of occurrences in the 10 districts with the highest recorded number of species. The number of observations increased over time, with a minimum in 2001 (61,167 observations) and a maximum in 2015 (205,163 observations; SM 1.2). The temporal variation in the number of observations was unequal across districts, with some districts more evenly sampled over time than others (see SM 1.2 for more details). Finally, we collected the endangered status of all species, also from FinBIF, based on the 2019 red list assessment (Hyvärinen et al., 2019). Analyses on endangered species were conducted on a subset of data, including vulnerable (VU), endangered (EN) and critically endangered (CR) species.

2.2 | Accumulation curve analyses

Comparing biodiversity between locations is sensitive to differences in sampling efforts. Species accumulation curves are commonly used to estimate species richness accurately in samples of different sizes (Chao et al., 2014; Gotelli & Colwell, 2001). This method was recently applied to multi-sourced biodiversity data on various spatial scales (Callaghan et al., 2022; Zattara & Aizen, 2021). A comprehensive statistical framework allows for estimating species richness based on (i) interpolation (down-sample the larger samples until they contain the same number of observations) and (ii) extrapolation (prediction on a larger sample size, guided by an estimated asymptote corresponding to infinite sampling; Chao et al., 2014). Furthermore, accumulation curve analyses provide an estimate of sample completeness for observed data, ranging from 0 to 1, with the value of 1 corresponding to perfect sampling, that is, all species have been detected. Subtracting sample completeness from 1 gives the probability that a new, previously unsampled species would be found if the sample was enlarged by one individual (Chao & Jost, 2012). Species richness estimates for equally large samples, associated confidence intervals and sample completeness of observations were calculated using the R package 'iNEXT' (Hsieh et al., 2016). A graphical example of interpolated and extrapolated accumulation curves can be found in SM 1.3.

To make informed decisions on data filtering, we tested various thresholds of minimum numbers of observations per district and evaluated the results based on estimate uncertainty and sampling completeness derived from accumulation curve analyses. Accumulation curve analyses were conducted on three subsets of districts defined by the number of available observations. We tested three threshold values (1000, 5000 and 10,000 observations) and estimated species richness at 2000, 10,000 and 20,000 observations, respectively, as species richness estimates are accurate up to twice the reference sample size (Chao et al., 2014). Species richness in districts was interpolated or extrapolated to the same sample size depending on whether there were more or fewer observations than the targeted sample size. The same method was applied to observations of endangered species, further filtering districts based on the number of observations of endangered species (threshold values: 50, 100, 150 and 200). This second filter was introduced to reduce uncertainty of estimates.

We aimed to define a threshold that would lead to the inclusion of districts with nonsignificant differences in terms of sampling quality and estimate accuracy. To that end, we compared the sample completeness of observations (raw data) and confidence intervals of richness estimates (as a percentage relative to the estimated value) between the different thresholds, and between districts with interpolated versus extrapolated estimates. We tested for statistical differences using a two-way ANOVA with

14724642, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/ddi.13864 by University Of Jyväskylä Library, Wiley Online Library on [22/05/2024]. 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

the interaction term and a Tukey multi-comparison post hoc test, using the basic functions of R statistical software, Version 4.1.1 (R Core Team, 2021).

2.3 | Geographic drivers of species richness

We investigated the relationship between estimated species richness (total and endangered) and district area and geographic position, that is, latitude and longitude. To do so, we fitted multiple linear models with estimated species richness, the coordinates of the district centroids (including latitude-longitude interaction) and district terrestrial area (i.e. excluding sea and lakes). To investigate the distribution of endangered species independently from total species richness, we also calculated the Red List Index (RLI, see below). As we expected a saturation of species richness with increasing district area, we fitted a log-transformed relationship. Analyses on total species richness were conducted by removing the district of Padasjoki, which was a strong outlier with high uncertainty of estimates (see Figure 2). Analyses of the richness of endangered species and the RLI were conducted by removing the districts of Enontekiö and Kuusamo, which were outliers with higher proportions of endangered species than expected (see Table 2 and Discussion).

We confirmed residual normality in the fitted models using a Shapiro–Wilk test (basic R) and Quantile-Quantile plot (R package '*car*'). We also confirmed the absence of significant spatial autocorrelation of residuals (i.e. spatial autocorrelation unexplained by the predictor variables, R package '*ncf*') despite a tendency for short-distance autocorrelation between Southeast districts for both total and endangered estimated species richness. We examined potential collinearity between all predictors using Pearson correlation coefficients (basic *R*). In both total and endangered species datasets, districts tend to have a larger area at northern latitudes ($r = 0.74$ and 0.85 respectively), following a well-known pattern for Finland.

2.4 | Species composition indicators

In addition to the estimated number of endangered species, we calculated the Red List Index (RLI, Butchart et al., 2007) to identify districts with a species community that has a high level of extinction risks (Equation 1). RLI species weights depend on their red-list status: Critically Endangered = 4; Endangered = 3; Vulnerable = 2; Near Threatened = 1; Least Concern = 0 (recommended 'equal step' weights). Weights are aggregated at the community level and compared with the Extinction status (weight = 5). The RLI is independent of species richness and takes values between zero and one: the closer the value is to zero, the closer the set of species is to extinction, and if the value is one, all species are of Least Concern:

$$RLI = 1 - \frac{1}{n \times W_{EXT}} \sum_{i=1}^{n} W_i \tag{1}$$

in which n is the number of species in a district, W_i is the status weight for species i and W_{EXT} is the weight given to extinction status.

We used the beta diversity framework of Legendre and De Cáceres (2013) and related indices of local contributions to beta diversity (LCBD) to identify districts with atypical species communities. LCBD indicators represent the ecological uniqueness of the sites, in which high values for a given site indicate high dissimilarity between the local community and other sites in the region. Such differences in species composition (beta diversity) can be characterised by species overlap (shared species), species replacement (turnover) and richness difference (nestedness, Legendre, 2014). For each selected district, we calculated total LCBD based on the Jaccard dissimilarity index, as well as its replacement and richness difference components (R package '*adespatial*'). We reported the normalised values of LCBD (i.e. centred on the mean and scaled deviation of 1) in a table and map to highlight the sites with LCBD values that are higher (positive) or lower (negative) than the mean (Legendre & De Cáceres, 2013).

3 | RESULTS

3.1 | Ranking of municipal districts based on reported observations

Our data comprised 3,023,681 observations of 878 native Lepidoptera species in Finland, compiled in the FinBIF between 1 January 2001 and 31 December 2020. The district with the highest reported species richness was Kemiönsaari (758 species, 115,319 observations), followed by Raasepori and Virolahti (756 species each; 59,904 and 162,429 observations respectively; SM 2 and 1.4). The 10 districts with the highest reported species richness were located on the southern coast of Finland, including the capital, Helsinki, which ranked ninth, with 716 species and 47,877 observations (SM 1.4). The number of observations varied greatly among districts, ranging from more than 160,000 observations in Virolahti and Rääkkylä to less than 50 in Humppila, Kannonkoski and Kihniö (Figure 1a; SM 1.4), where realistic estimates of species richness in communities comprising hundreds of species are not possible. Overall, 29,963 observations of 113 endangered species were reported in 282 districts and the number of observations per district varied from 0 to over 1000 (Figure 2a; SM 1.5). The highest reported number of endangered species was found in the districts of Virolahti and Hanko (49 species), followed by Kemiönsaari (48 species; Table 2; SM 1.5).

3.2 | Uncertainty in species richness estimates depends on observation data threshold

Sample completeness of observed data (including all species) and confidence intervals of total species richness estimates were both affected significantly by the estimation method (interpolation vs.

14724642, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/ddi.13864 by University Of Jyväskylä Library, Wiley Online Library on [22/05/2024]. 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

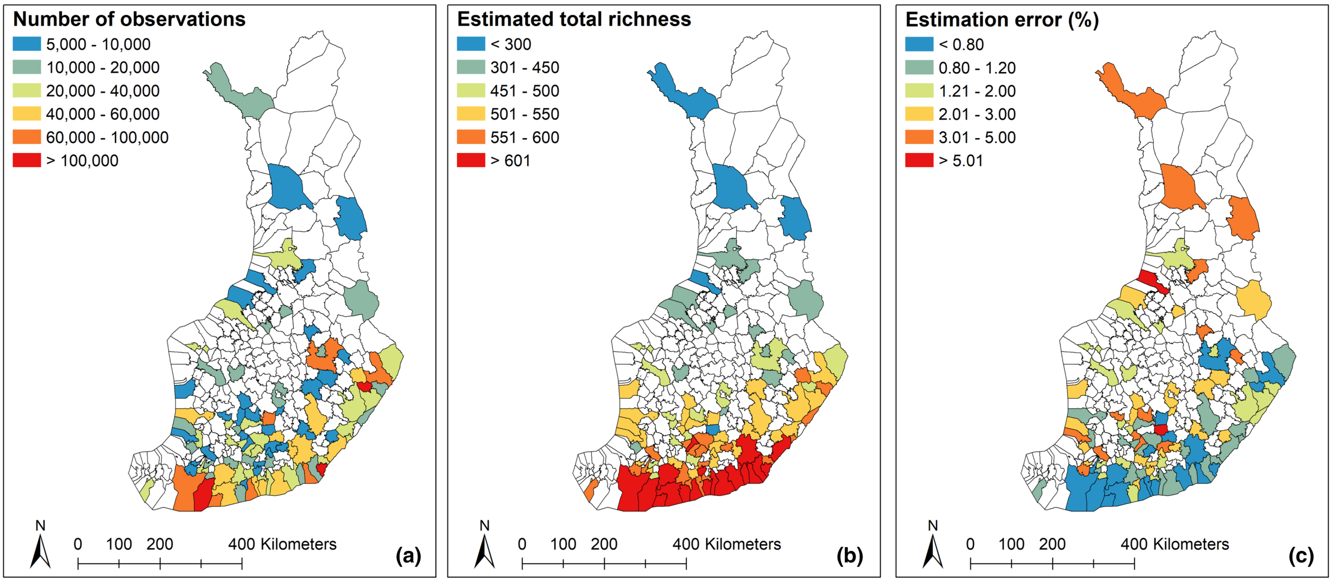

FIGURE 1 Maps of the 97 selected municipal districts with >5000 observations (all species): (a) total number of observations in 2001–2020; (b) estimated species richness of native Lepidoptera and (c) associated error (95% confidence intervals as a percentage of estimated value), extrapolated or interpolated at 10,000 observations. Districts with white fill had less than 5000 observations and, thus, were not included in the analyses.

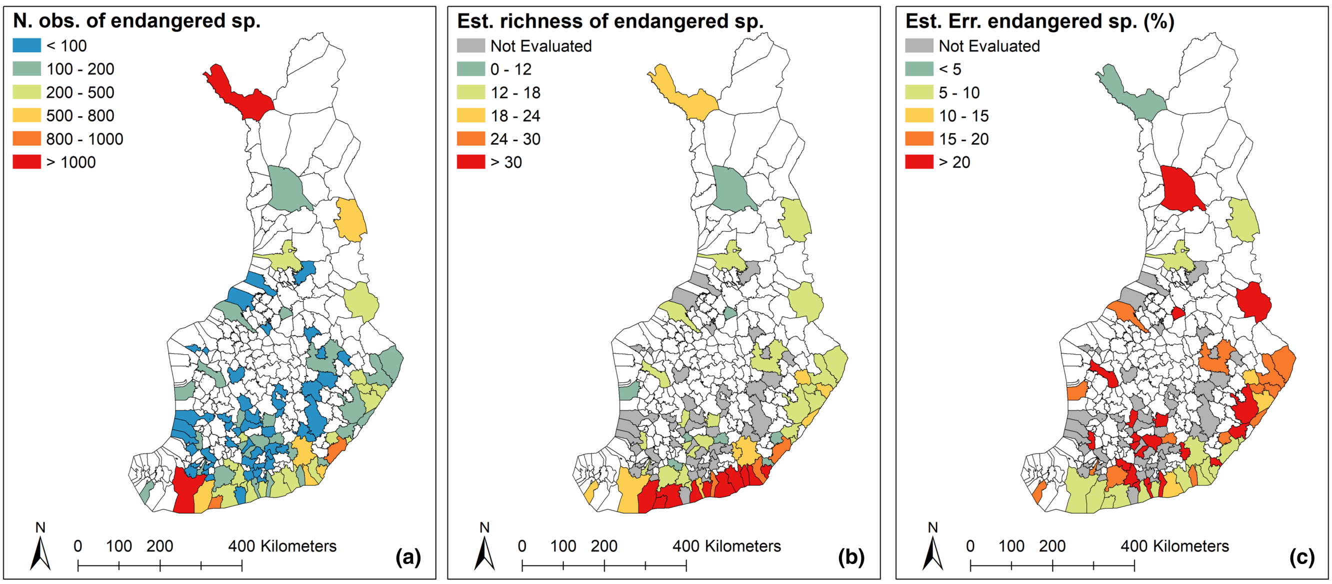

FIGURE 2 Maps of the 51 municipal districts with >5000 observations (all species) and >100 observation of endangered species: (a) number of observations of endangered species in 2001–2020; (b) estimated species richness of endangered lepidopteran species and (c) associated error (95% confidence intervals as a percentage), extrapolated or interpolated at 200 observations. Districts with white fill had <5000 observations, districts with grey fill had >5000 observations (all species) and <100 observations of endangered species; thus, these were not included in the analyses.

extrapolation) and the threshold in the number of observations to select districts (two-way ANOVA, including interaction term, *p*-values <.001; Figure 3). Logically, species richness estimates had much smaller confidence intervals when obtained through interpolation than extrapolation. Sample completeness and confidence intervals were not statistically different between the 5000 and 10,000 thresholds (i.e. including 97 or 68 districts respectively; Figure 3). However, these measures were significantly higher when using the threshold of 1000 observations (189 districts), indicating lower accuracy of estimates. As the estimation of species richness was not more uncertain with a 5000 threshold than a 10,000 threshold, the districts (out of 310) with more than 5000 observations were included in the analyses (Figure 3a).

For the endangered species, the sample completeness of observations and the confidence interval of richness estimates were always lower when extrapolated than when interpolated

14724642, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/ddi.13864 by University Of Jyväskylä Library, Wiley Online Library on [22/05/2024]. 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

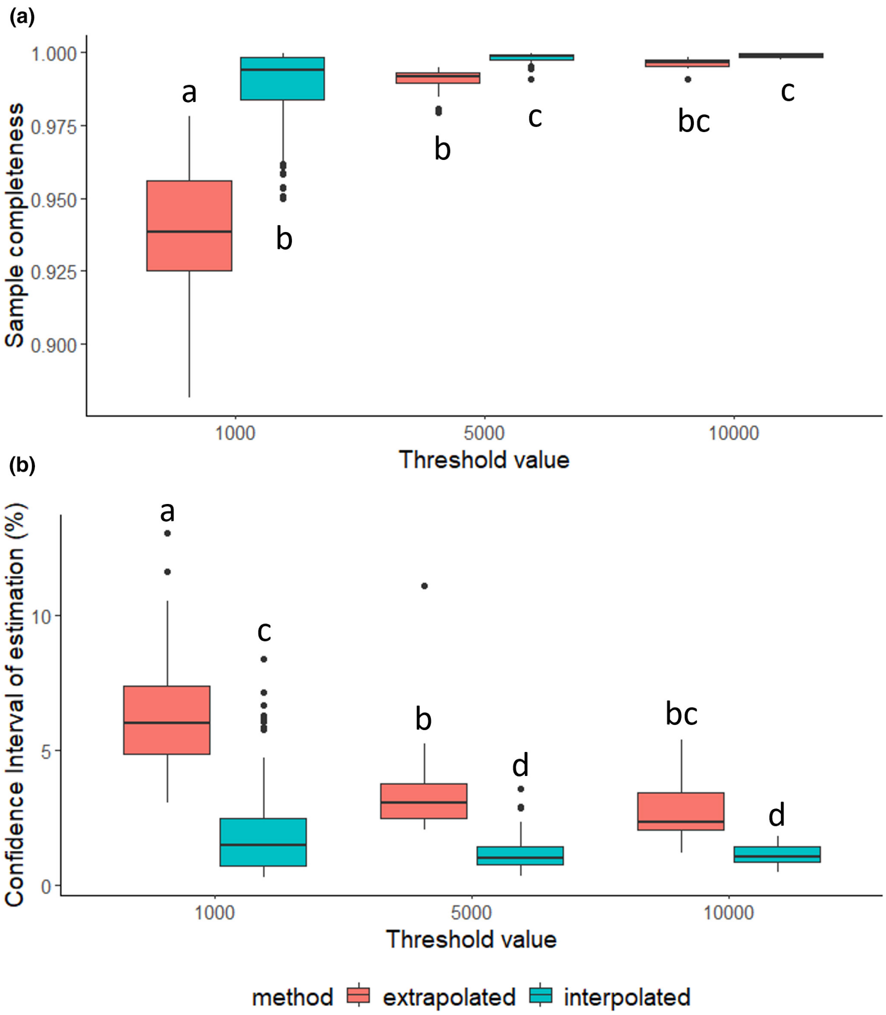

FIGURE 3 Sample completeness of observations (a) and confidence intervals of total species richness estimates of native Lepidoptera at the targeted sample size (b) for the three thresholds of minimum number of observations per municipal district (1000, 5000 and 10,000), and the extrapolated versus interpolated districts. Confidence intervals are expressed as percentages of estimated richness. Letters indicate significant differences from a Tukey multiple comparison post hoc test ($\alpha = 5\%$).

(two-way ANOVA p-values <.001, non-significant interaction; SM 1.6). However, differences between the tested thresholds (50, 100, 150 and 200 observations) were mostly non-significant, except that sample completeness was lower for the 50-observation threshold than for the 150 and 200 thresholds (two-way ANOVA p-value <.01, Tukey multiple comparison post hoc test p-values <.05; SM 1.6). As the accumulation curve analyses were not more uncertain using a 100-observation threshold than when using a higher threshold, we selected that value for further analyses. Altogether, 51 districts registered >5000 observations, including all species, and >100 observations of endangered species. By comparison, the total was 72 districts when using the 50-observation threshold.

3.3 | Total species richness of native Lepidoptera

Using the 97 districts with >5000 observations, the estimated total species richness (at 10,000 observations) ranged from 684 ± 3 species in Kotka to 168 ± 6 species in Enontekiö (Table 1 and SM 1.4). The estimated species richness was similar (i.e. within confidence intervals) in Kotka, Hamina and Pyhtää, but significantly higher in these districts than in Kemiönsaari or Virolahti (Table 1). Estimated total species richness was unevenly distributed geographically over the 97 districts (Figure 1b), with the top 30 districts concentrated at the southern coast of Finland (Figure 4a). Estimated total species richness significantly decreased along the latitudinal gradient (p-value <0.001; SM 1.7 and SM 1.8a), with the lowest value observed in the northernmost district of Enontekiö. To a lesser extent, estimated total species richness significantly increased in districts with larger terrestrial area (p-value <.01; SM 1.7 and 1.8b). Together, latitude, longitude (*not significant*), their interaction (*not significant*) and district terrestrial area explained 78% of the variability of estimated species richness (adj. R-squared) in a highly significant multiple linear model (p-value <.001; SM 1.7). The uncertainty of species richness estimates (95% confidence intervals) was relatively small, for example, <1% in 33 districts and >5% in one district, that is, Raahe, at 5.2% (Figure 1c), indicating one outlier, that is, the district of Padasjoki, at 11.1% (which was removed from further analyses). The uncertainty was irregularly distributed geographically (Figure 1c), depending on number of observations (comparison of Figure 1a,c; see Figure 3b: interpolated vs. extrapolated at the 5000-observation threshold).

14724642, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/ddi.13864 by University Of Jyväskylä Library, Wiley Online Library on [22/05/2024]. 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

TABLE 1 Observed (S) and estimated (Est. S) species richness of native Lepidoptera for the 20 Finnish municipal districts with the highest estimated richness at 10,000 observations (sorted by estimated species richness).

	Obs. data		Estimates at 10,000 obs.			
District	*n*	S	Est. S	Lower CI	Upper CI	Significant
Kotka	60,703	745	684	681	688	a
Hamina	18,332	708	682	675	689	ab
Pyhtää	19,629	710	680	674	687	ab
Hanko	61,097	750	677	673	680	b
Raasepori	59,904	756	676	672	680	b
Loviisa	37,160	733	674	669	679	b
Kemiönsaari	115,319	758	655	652	658	c
Parainen	77,118	731	647	643	650	d
Virolahti	162,429	756	643	640	645	de
Kirkkonummi	74,506	740	636	632	640	f
Sipoo	28,076	695	635	629	641	efg
Inkoo	14,158	647	628	620	637	fgh
Porvoo	48,148	715	626	621	631	gh
Salo	45,397	679	619	615	624	hi
Helsinki	47,877	716	619	615	624	hi
Kouvola	59,853	682	613	610	617	i
Hattula	5302	557	608	584	632	fghijkl
Nurmijärvi	6238	593	601	586	615	i kl
Lappeenranta	49,132	702	601	596	606	l
Espoo	48,003	680	594	589	598	lm

Note: Number of observations (*n*) and 95% lower and upper confidence intervals (CI) for species richness estimates are also provided. Significant differences (i.e. non-overlapping CIs) are indicated with different letters. Districts referred to in the main text are in bold. Results for the 97 districts included in the analysis are presented in Supplementary Material S2.4.

3.4 | Species richness of native endangered Lepidoptera

Using the 51 districts with >5000 observations (all species) and >100 observations of endangered species, the estimated species richness of endangered species (at 200 observations) ranged from 39 ± 3 species in Raasepori to 8 ± 2 species in Rovaniemi (Table 2a and SM 1.5). Due to higher uncertainty (95% confidence intervals) compared with total species richness (Figure 2c), the 10 districts with the highest species richness of endangered species had mostly non-significant differences (Table 2a). However, these districts had significantly higher estimated numbers of endangered species than districts with lower rankings (SM 1.5). All these districts were in the 20 districts with the highest estimated total species richness (Table 1 and 2a). Indeed, the estimated species richness of endangered species was significantly dependent on estimated total species richness (p-value <.001; adj. R-squared = 0.59) and, therefore, followed similar geographical patterns (Figures 1b and 2b). However, RLI, which is independent of the total number of species, was not significantly influenced by any geographical drivers (p-value >.10; SM 1.9). The northern districts of Enontekiö and Kuusamo (excluded from regression analyses) were clearly atypical cases, as their RLI values were much higher than expected with respect to total or endangered species richness (Table 2b). The RLI map confirmed this pattern with high aggregated extinction risks in Enontekiö and Kuusamo and the southern coastal districts (Figure 4b).

3.5 | Beta diversity among districts

In the 96 districts (i.e. excluding Padajoki), LCBD of lepidopteran communities was due mostly to species richness differences (64%) and to a lesser extent, species replacement (36%). The highest total LCBD, indicating unique lepidopteran communities, was found in the district of Enontekiö, followed by Raahe, Kuusamo and Rovaniemi. All four are located in northern Finland (Table 3a and SM 1.10). In the districts of Enontekiö and Kuusamo, both the LCBD richness difference and the LCBD replacement values were higher than average (Table 3a; SM 1.10). However, the districts of Raahe and Rovaniemi also had specific species communities, but only because of species richness differences, while their species replacement was lower than average (Table 3a), that is, they hosted a subset of species that can be found elsewhere. The districts of Valkeakoski and Naantali had the largest LCBD replacement (Table 3b), although the species richness of these two cities species richness was, on average, similar to that of other districts (LCBD richness difference close to zero). Except for Enontekiö and

14724642, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/ddi.13864 by University Of Jyväskylä Library, Wiley Online Library on [22/05/2024]. 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

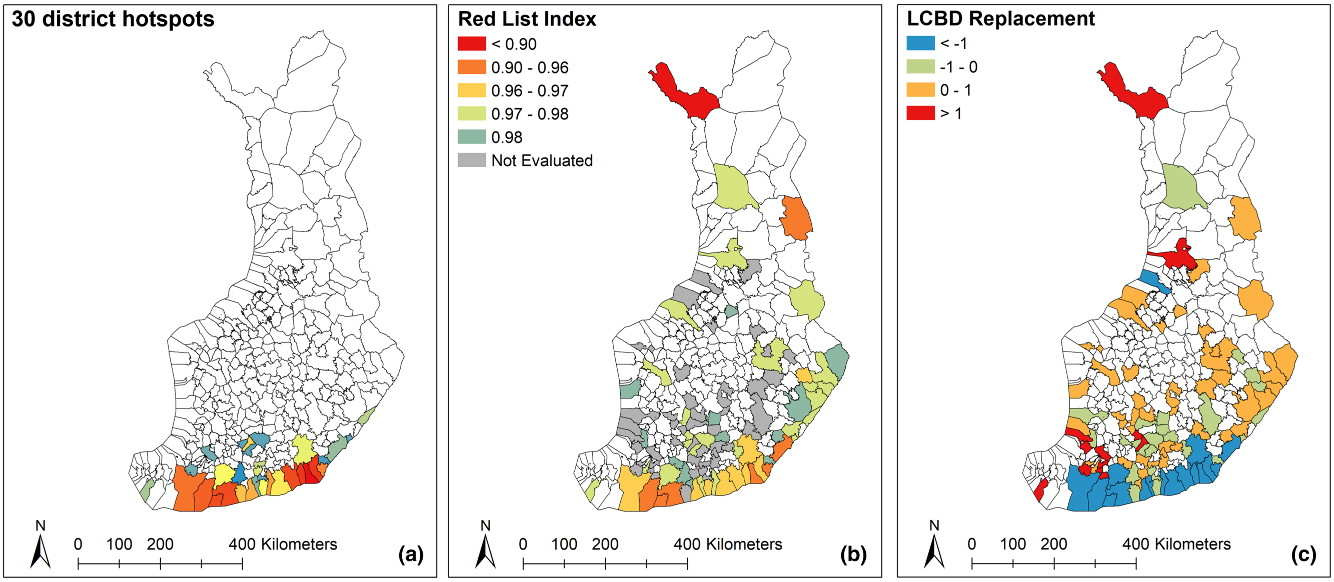

FIGURE 4 Three criteria for the conservation values of municipal districts for native Lepidoptera, where warm colours indicate the highest priorities: (a) the 30 districts with the highest estimated total species richness, grouped (colours) according to significant differences (see Table 1); (b) red list index, indicating districts with high community-level extinction risks (based on species red-list status; SM 1.5); (c) replacement component of local contribution to beta diversity (LCBD), indicating districts with the highest species turnover (normalised values: SM 1.10). (b and c) Indicators are provided for districts with >5000 observations (all species). White fill indicates districts with <5000 observations. In (b), the Red List Index was not calculated for districts having <100 observations of endangered species (grey fill).

Oulu in northern Finland, the 10 cities with the highest species replacement are in Åland (Southwest archipelagos in the Baltic Sea) and the Southeast (Figure 4c and Table 3b).

4 | DISCUSSION

4.1 | Ecological interpretation of the case study

Our case study on native Lepidoptera in Finnish municipal districts demonstrated that a set of complementary indicators is needed to inform where conservation actions should be prioritised. Estimated total and endangered species richness (based on accumulation curves), red list index (RLI, based on the red-list species status) and uniqueness of species communities (based on local contribution to beta diversity, LCBD) highlighted some priority districts, which were not the same depending on the indicator. Estimations of total species richness pointed to coastal southern districts, particularly in the Southeast (e.g. the districts of Kotka, Hamina and Pyhtää) and in the Southwest (e.g. Hanko, Raasepori and Kemiönsaari). Total species richness decreased consistently towards the North following a well-known pattern for Lepidoptera in Finland (Antão et al., 2020; Huldén et al., 2000; Leinonen et al., 2016).

The estimated number of endangered species followed a pattern similar to that of total richness, suggesting that in most cases, the conservation of Lepidoptera endangered species will also maintain high overall species richness. However, such a relationship might not always hold across taxa and regions (Fraixedas et al., 2022; Howard et al., 2020); thus, monitoring and accounting for overall biodiversity in conservation decision-making are important as well. Northern districts (e.g. Enontekiö and Kuusamo) deviated from the general trend and had a high RLI despite low species richness, indicating a high overall extinction risk in their species communities. Therefore, the RLI may represent a better indicator than the total number of endangered species in our case, indicating the relative proportion of endangered species in the total community independently from the total number of species (Butchart et al., 2007). This result aligns with numerous previous findings that global warming is affecting northern communities most severely (e.g. Pöyry et al., 2009).

Finally, LCBD replacement was an important complementary indicator in our study, identifying districts with a unique species composition (i.e. hosting species often not found elsewhere). The southern coastal districts had a very high number of species, including those found elsewhere (low replacement). However, districts in northern Finland, such as Enontekiö and Kuusamo, had a much smaller number of species, but a high number of endangered species and a specific group of species. Similarly, LCBD replacement highlighted districts in the hemi-boreal zone (Southwest) with an average species richness and relatively few endangered species, such as Valkeakoski and Naantali. This suggests that LCBD replacement is a good indicator to identify sites with unique species composition outside of the richest sites and that beta diversity should be used to prioritise conservation efforts (Gossner et al., 2013). LCBD was highly sensitive to differences in species richness due to significant large differences among districts. Therefore, the use of the replacement component of LCBD is more relevant to identify locations that host many species usually not found in other districts, indicating conservation priorities (Legendre, 2014; Montràs-Janer et al., 2024).

14724642, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/ddi.13864 by University Of Jyväskylä Library, Wiley Online Library on [22/05/2024]. 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

TABLE 2 Observed (S) and estimated (Est. S) species richness of native endangered Lepidoptera for the ten Finnish municipal districts with (a) the highest estimated species richness at 200 observations (sorted in decreasing order) and (b) the lowest Red List Index (sorted in increasing order).

		Obs. Data			Estimation at 200 obs.			
	Districts	n	S	RLI	Est. S	Lower CI	Upper CI	Significant
(a)	**Raasepori**	365	46	0.959	39	36	42	a
	Kirkkonummi	206	39	0.960	39	35	42	ab
	Virolahti	790	49	0.955	37	35	39	ab
	Kemiönsaari	755	48	0.955	36	34	38	ab
	Helsinki	250	37	0.963	35	32	38	ab
	Loviisa	290	38	0.962	35	31	38	ab
	Kotka	621	42	0.959	34	32	36	b
	Hanko	933	49	0.953	34	32	36	b
	Pyhtää	149	30	0.967	33	27	40	ab
	Porvoo	251	35	0.964	33	29	37	ab
(b)	**Enontekiö**	6810	30	0.892	23	22	23	–
	Hanko	933	49	0.953	34	32	36	–
	Virolahti	790	49	0.955	37	35	39	–
	Kemiönsaari	755	48	0.955	36	34	38	–
	Kuusamo	737	18	0.955	16	15	16	–
	Lappeenranta	842	42	0.958	27	25	29	–
	Raasepori	365	46	0.959	39	36	42	–
	Kotka	621	42	0.959	34	32	36	–
	Kirkkonummi	206	39	0.960	39	35	42	–
	Parainen	1302	37	0.962	20	19	22	–

Note: Number of observations (*n*) and 95% lower and upper confidence intervals (ci) for species richness estimates are also provided. Significant differences (i.e. non-overlapping CIs) are indicated with different letters in Table 2a, but are not applicable (–) for the Red List Index in Table 2b. The districts referred to in the main text are in bold. Results for the 51 districts included in the analysis are presented in Supplementary Material S2.5.

Presenting these indicators on maps allows for clearly identifying districts with high biodiversity conservation value, as well as geographical gaps. Indeed, despite the use of a multi-sourced biodiversity information facility, large geographical gaps in biodiversity knowledge remained, as only a third of Finnish districts could be included in the analyses (and even less—one-sixth—for endangered species). Kotka, Hamina and Pyhtää have not been recognised earlier as biodiversity hotspots because the observation activity has concentrated on other districts. Decision-makers in these districts (and regions) must recognise that these areas host the highest species richness in the country, so they need to maintain the landscapes in a state that will support high biodiversity in the future.

4.2 | Methodological vigilance points

Species richness estimated using accumulation curves was not necessarily the highest in the most investigated districts and the rankings differed from reported observations, demonstrating the importance of accounting (or correcting) for differences in sampling intensity. Based on the raw data, species richness was highest in the districts of Raasepori, Kemiönsaari and Virolahti, which are known Lepidoptera hotspots in Finland, attracting many observers and, thus, accumulating a large number of records. Although these were in the top districts based on estimated values, they were not in the top three districts anymore. Indeed, these districts were some of the most intensively sampled districts, introducing a sampling effort bias. Such a discrepancy between sampling effort and species richness was most stringent in the district of Rääkylä, which had the second highest number of observations over the study period. However, it did not rank as a hotspot of estimated total or endangered species richness (ranked 76th out of 97 and 43rd out of 51 respectively), nor provide a specific species composition. This illustrates the risk of unnecessary sampling intensity, in which many volunteers sample the same set of accessible and attractive sites (Callaghan et al., 2022). However, this does not negate the usefulness of recurrent sampling over time, as it allows for evaluating temporal trends.

To ensure the reliability of the obtained results, we tested different data-filtering thresholds based on the number of available observations, assuming that the quality of estimated indices depends on this criterion. Accumulation curves are a well-established method designed for comparing species richness between sites, but an insufficient number of observations would inflate the confidence interval of estimates. For both overall and endangered species, it was possible to determine a threshold beyond which the confidence interval

14724642, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/ddi.13864 by University Of Jyväskylä Library, Wiley Online Library on [22/05/2024]. 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

TABLE 3 Total, replacement (Repl.) and richness difference (Rich. Diff.) components of local contribution to beta diversity (LCBD) for the 10 Finnish municipal districts with (a) the highest Total LCBD and (b) the highest LCBD Repl.

	Districts	LCBD total	LCBD Repl.	LCBD rich. Diff.
(a)	**Enontekiö**	5.72	2.34	4.69
	Raahe	3.71	−1.06	4.38
	Kuusamo	3.43	0.76	3.14
	Rovaniemi	2.89	−0.65	3.32
	Valkeakoski	1.45	3.14	−0.13
	Kuhmo	1.37	0.38	1.21
	Kaavi	1.36	−0.47	1.65
	Utajärvi	1.10	0.05	1.11
	Kalajoki	1.01	0.30	0.89
	Kärsämäki	0.92	0.22	0.84
(b)	**Valkeakoski**	1.45	3.14	−0.13
	Naantali	0.68	2.55	−0.63
	Enontekiö	5.72	2.34	4.69
	Sauvo	0.31	1.86	−0.65
	Pöytyä	0.23	1.80	−0.70
	Hattula	0.00	1.40	−0.73
	Laitila	0.60	1.22	−0.02
	Lemland	0.16	1.17	−0.44
	Oulu	0.74	1.16	0.16
	Kaarina	−0.12	1.14	−0.72

Note: All 96 districts with >5000 observations were included (excluding the municipality of Padasjoki, an outlier). Districts referred to in the main text are in bold. Results for the 96 districts included in the analysis are presented in Supplementary Material S2.10.

of estimates did not change significantly, as well as the ranking of the top districts (data not shown). A lower threshold than the one used here could increase spatial coverage, but this would be detrimental to the accuracy of estimates (note that threshold values might be study-specific). Furthermore, estimated uncertainty was much higher for endangered species than for total species richness, in relation to the much lower number of observations available. This result aligns with previous findings demonstrating that a higher sampling effort is required for an adequate estimation of rare species using data from citizen science (Callaghan et al., 2022). We highlighted the importance of reporting confidence intervals alongside estimates of species richness, as they were variable across districts. Besides, the RLI based on species conservation status and the LCBD measure of species replacement requires high certainty in the observed species composition. We included all selected districts for these indicators because observed sample completeness for both total and endangered species was very high, that is, most species have been detected. In cases in which sampling completeness is more variable among studied sites, further data selection would be necessary, excluding sites in which species composition is only partially known. Alternatively, one could replace the RLI by the ratio of estimated endangered/total species richness, as they correlate strongly (results not shown).

Ideally, the data collection methodologies should be the same in each district, both in terms of observation methods and reporting consistency, which is often an issue in multi-sourced biodiversity data (Johnston et al., 2023). Lepidoptera species are mainly observed using light and bait traps for moths and butterfly nets for day-active Lepidoptera, leading to a relatively homogenous observation method across sources and observers. Consistency in observation method is particularly critical for analysing species composition, as differences in trapping methods may influence which species are caught and, therefore, may inflate indicators of species turnover. However, reporting activity can vary greatly, as some observers report all their observations, while others report only 'interesting' species (i.e. generating incomplete species lists). This is the main difference between systematic monitoring, incomplete trap reporting from expert hobbyists and opportunistic observations from amateur naturalists, all of which are included in the multi-sourced data used in this study. These sources also have different geographical coverage, potentially creating bias in data types across districts. Such reporting bias can lead to an overrepresentation of rare and uncommon species at the expense of common species (Johnston et al., 2023) and can pose consequences for species accumulation curves. If common species are reported infrequently, estimated species richness may be overestimated. Reporting differences and the absence of information on sampling duration also made it impossible to assess differences in abundance across districts properly, which can be problematic for detecting biodiversity change beyond species richness and extinction risks (Fraixedas et al., 2022). However, information on species abundance may be available in some monitoring programmes through more standardised sampling (Callaghan et al., 2019).

4.3 | Future applications

Our results suggest that Lepidoptera hotspots in Finland are driven by interaction between climate and habitat gradients. Indeed, total species richness followed a latitudinal gradient that correlates strongly with climate drivers, such as growing season length or temperature sum. Also, the LCBD indicator highlighted typical species composition in transitional (hemi-boreal) and atypical (North boreal) climatic areas. However, districts at similar latitudes often had significantly different species richness (particularly in the South), and the relationship between species richness and district area was weak, suggesting that other important drivers were in play, such as land use, habitat diversity or the presence of protected areas. For example, this is the case for the southern districts with the highest estimated total species richness (Kotka, Hamina and Pyhtää), which are rather small in terrestrial area, but include rich archipelagos and valuable coastal habitats (including a large part of the Gulf of Finland National Park), as well as inland boreal forests and large protected bog areas (including the Valkmusa National Park).

14724642, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/ddi.13864 by University Of Jyväskylä Library, Wiley Online Library on [22/05/2024]. 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

Such climate and land use/habitat gradients and their interactions are common drivers of biodiversity and operate from global (e.g. Pinkert et al., 2022) to local scales. For example, on a national scale, a combination of climate and land use/vegetation types explained lepidopteran species distribution and composition in Switzerland (Zellweger et al., 2017), as well as biodiversity changes in the United Kingdom (Montràs-Janer et al., 2024). On a local scale, the interaction between elevation (~ temperature) and vegetation influenced Lepidoptera response to climate change (Álvarez et al., 2024). Community-level indicators derived from multi-sourced species records successfully detected biodiversity variation in space and, therefore, may be useful in investigating biodiversity patterns in relation to climate and land use, in addition to detecting biodiversity hotspot on multiple scales.

Our study was conducted at the national level using municipal districts as spatial units, but the biodiversity indicators that we used are applicable on multiple scales and resolutions, including continental, national or regional (Butchart et al., 2007; Socolar et al., 2016). Estimating species richness by using multiple nested scales may be particularly useful to better describe and measure biotic homogenisation phenomena (Blowes et al., 2024). They can also be used to infer significant temporal trends by comparing different time periods because confidence intervals are provided for species richness estimates. For example, such analyses would allow for identifying locations with higher species richness changes than others, that is, sites in which biodiversity is at higher risk, as well as relating biodiversity changes with climate or land use changes (e.g. Montràs-Janer et al., 2024), offering the possibility of better anticipating future changes and mitigating them.

5 | CONCLUSION

Our study demonstrates how a combination of data filtering and correction for observation intensity can effectively provide complementary biodiversity indicators and identify areas with high conservation values. These community-level indicators are easily understandable and can be communicated efficiently to decision makers through maps. However, one must keep in mind the potential for reporting, sampling and geographical biases, which may affect comparisons among districts. Nevertheless, a combination of data filtering (minimum number of observations) and evaluations of sampling completeness based on accumulation curves can mitigate these biases. We also must emphasise the importance of reporting uncertainty associated with such biodiversity estimates, as any method is inherently subject to uncertainty, as well as recognise explicitly incomplete knowledge (Fraixedas et al., 2022; Margules & Pressey, 2000). We propose using community-level biodiversity indicators to 'make the most' of multi-sourced species records and identify locations with high conservation values beyond species-level analyses. Provided that many national or international databases (e.g. the Global Biodiversity Information Facility) have consolidated millions of species records, they hold enormous potential to identify areas with high conservation priorities.

ACKNOWLEDGEMENTS

The authors thank the expert hobbyists who reported >3,000,000 observations to the databases and the Finnish Biodiversity Information Facility for managing the databases. RD was supported by the Kone Foundation (application 202105759). Vuokon Luonnonsuojelusäätiö supported the presentation of this study at the 22nd European Congress of Lepidopterology. We thank Jari-Pekka Kaitila for clarifying the list of lepidopteran species. The authors thank two anonymous reviewers for their constructive comments on earlier version of the manuscript.

CONFLICT OF INTEREST STATEMENT

The authors declare no conflicts of interest.

PEER REVIEW

The peer review history for this article is available at https://www.webofscience.com/api/gateway/wos/peer-review/10.1111/ddi.13864.

DATA AVAILABILITY STATEMENT

The analyses presented in this manuscript are based on data retrieved from the Finnish Biodiversity Information Facility (FinBIF, www.laji.fi). The data that support the findings of this study are available in the supplementary material of this article.

ORCID

Rémi Duflot https://orcid.org/0000-0002-5070-903X
Anssi V. Vähätalo https://orcid.org/0000-0003-3434-5915

REFERENCES

Álvarez, H. A., Walker, E., Mingarro, M., Ursul, G., Cancela, J. P., Bassett, L., & Wilson, R. J. (2024). Heterogeneity in habitat and microclimate delay butterfly community tracking of climate change over an elevation gradient. *Biological Conservation*, *289*, 110389. https://doi.org/10.1016/j.biocon.2023.110389

Anderson, M. J., Crist, T. O., Chase, J. M., Vellend, M., Inouye, B. D., Freestone, A. L., Sanders, N. J., Cornell, H. V., Comita, L. S., Davies, K. F., Harrison, S. P., Kraft, N. J. B., Stegen, J. C., & Swenson, N. G. (2011). Navigating the multiple meanings of beta diversity: A roadmap for the practicing ecologist. *Ecology Letters*, *14*(1), 19–28.

Antão, L. H., Pöyry, J., Leinonen, R., & Roslin, T. (2020). Contrasting latitudinal patterns in diversity and stability in a high-latitude species-rich moth community. *Global Ecology and Biogeography*, *29*(5), 896–907. https://doi.org/10.1111/geb.13073

Blowes, S. A., McGill, B., Brambilla, V., Chow, C. F. Y., Engel, T., Fontrodona-Eslava, A., Martins, I. S., McGlinn, D., Moyes, F., Sagouis, A., Shimadzu, H., van Klink, R., Xu, W.-B., Gotelli, N. J., Magurran, A., Dornelas, M., & Chase, J. M. (2024). Synthesis reveals approximately balanced biotic differentiation and homogenization. *Science. Advances*, *10*(8), eadj9395. https://doi.org/10.1126/sciadv.adj9395

Butchart, S. H. M., Akçakaya, H. R., Chanson, J., Baillie, J. E. M., Collen, B., Quader, S., Turner, W. R., Amin, R., Stuart, S. N., & Hilton-Taylor, C. (2007). Improvements to the red list index. *PLoS One*, *2*(1), e140. https://doi.org/10.1371/journal.pone.0000140

14724642, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/ddi.13864 by University Of Jyväskylä Library, Wiley Online Library on [22/05/2024]. 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

Butchart, S. H. M., Walpole, M., Collen, B., van Strien, A., Scharlemann, J. P. W., Almond, R. E. A., Baillie, J. E. M., Bomhard, B., Brown, C., Bruno, J., Carpenter, K. E., Carr, G. M., Chanson, J., Chenery, A. M., Csirke, J., Davidson, N. C., Dentener, F., Foster, M., Galli, A., … Watson, R. (2010). Global biodiversity: Indicators of recent declines. *Science*, *328*(5982), 1164–1168. https://doi.org/10.1126/science.1187512

Callaghan, C. T., Bino, G., Major, R. E., Martin, J. M., Lyons, M. B., & Kingsford, R. T. (2019). Heterogeneous urban green areas are bird diversity hotspots: Insights using continental-scale citizen science data. *Landscape Ecology*, *34*(6), 1231–1246. https://doi.org/10.1007/s10980-019-00851-6

Callaghan, C. T., Bowler, D. E., Blowes, S. A., Chase, J. M., Lyons, M. B., & Pereira, H. M. (2022). Quantifying effort needed to estimate species diversity from citizen science data. *Ecosphere*, *13*(4), e3966. https://doi.org/10.1002/ecs2.3966

Chandler, M., See, L., Copas, K., Bonde, A. M. Z., López, B. C., Danielsen, F., Legind, J. K., Masinde, S., Miller-Rushing, A. J., Newman, G., Rosemartin, A., & Turak, E. (2017). Contribution of citizen science towards international biodiversity monitoring. *Biological Conservation*, *213*, 280–294. https://doi.org/10.1016/j.biocon.2016.09.004

Chao, A., Gotelli, N. J., Hsieh, T. C., Sander, E. L., Ma, K. H., Colwell, R. K., & Ellison, A. M. (2014). Rarefaction and extrapolation with hill numbers: A framework for sampling and estimation in species diversity studies. *Ecological Monographs*, *84*(1), 45–67. https://doi.org/10.1890/13-0133.1

Chao, A., & Jost, L. (2012). Coverage-based rarefaction and extrapolation: Standardizing samples by completeness rather than size. *Ecology*, *93*(12), 2533–2547. https://doi.org/10.1890/11-1952.1

Devictor, V., van Swaay, C., Brereton, T., Brotons, L., Chamberlain, D., Heliölä, J., Herrando, S., Julliard, R., Kuussaari, M., Lindström, Å., Reif, J., Roy, D. B., Schweiger, O., Settele, J., Stefanescu, C., Van Strien, A., Van Turnhout, C., Vermouzek, Z., WallisDeVries, M., … Jiguet, F. (2012). Differences in the climatic debts of birds and butterflies at a continental scale. *Nature Climate Change*, *2*(2), 121–124. https://doi.org/10.1038/nclimate1347

Engelhardt, E. K., Biber, M. F., Dolek, M., Fartmann, T., Hochkirch, A., Leidinger, J., Löffler, F., Pinkert, S., Poniatowski, D., Voith, J., Winterholler, M., Zeuss, D., Bowler, D. E., & Hof, C. (2022). Consistent signals of a warming climate in occupancy changes of three insect taxa over 40 years in central Europe. *Global Change Biology*, *28*(13), 3998–4012. https://doi.org/10.1111/gcb.16200

Fraixedas, S., Roslin, T., Antão, L. H., Pöyry, J., & Laine, A.-L. (2022). Nationally reported metrics can't adequately guide transformative change in biodiversity policy. *Proceedings of the National Academy of Sciences*, *119*(9), e2117299119. https://doi.org/10.1073/pnas.2117299119

Girardello, M., Chapman, A., Dennis, R., Kaila, L., Borges, P. A. V., & Santangeli, A. (2019). Gaps in butterfly inventory data: A global analysis. *Biological Conservation*, *236*, 289–295. https://doi.org/10.1016/j.biocon.2019.05.053

Gossner, M. M., Getzin, S., Lange, M., Pašalić, E., Türke, M., Wiegand, K., & Weisser, W. W. (2013). The importance of heterogeneity revisited from a multiscale and multitaxa approach. *Biological Conservation*, *166*, 212–220. https://doi.org/10.1016/j.biocon.2013.06.033

Gotelli, N. J., & Colwell, R. K. (2001). Quantifying biodiversity: Procedures and pitfalls in the measurement and comparison of species richness. *Ecology Letters*, *4*(4), 379–391.

Hällfors, M. H., Pöyry, J., Heliölä, J., Kohonen, I., Kuussaari, M., Leinonen, R., Schmucki, R., Sihvonen, P., & Saastamoinen, M. (2021). Combining range and phenology shifts offers a winning strategy for boreal Lepidoptera. *Ecology Letters*, *24*(8), 1619–1632. https://doi.org/10.1111/ele.13774

Howard, C., Flather, C. H., & Stephens, P. A. (2020). A global assessment of the drivers of threatened terrestrial species richness. *Nature Communications*, *11*(1), 993. https://doi.org/10.1038/s41467-020-14771-6

Hsieh, T. C., Ma, K. H., & Chao, A. (2016). iNEXT: An R package for rarefaction and extrapolation of species diversity (Hill numbers). *Methods in Ecology and Evolution*, *7*(12), 1451–1456. https://doi.org/10.1111/2041-210X.12613

Hughes, A. C., Orr, M. C., Ma, K., Costello, M. J., Waller, J., Provoost, P., Yang, Q., Zhu, C., & Qiao, H. (2021). Sampling biases shape our view of the natural world. *Ecography*, *44*(9), 1259–1269. https://doi.org/10.1111/ecog.05926

Huldén, L., Albrecht, A., Itämies, J., Malinen, P., & Wettenhovi, J. (2000). *Suomen suurperhosatlas: Finlands storfjärilsatlas = Atlas of Finnish macrolepidoptera*. Suomen perhostutkijain seura ry.

Hyvärinen, E., Juslén, A. K., Kemppainen, E., Uddström, A., & Liukko, U.-M. (2019). *Suomen lajien uhanalaisuus 2019—Punainen kirja: The 2019 Red List of Finnish Species*. Ympäristöministeriö & Suomen ympäristökeskus.

IPBES. (2019). *Global assessment report on biodiversity and ecosystem services of the Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services. E. S. Brondizio, J. Settele, S. Díaz, and H. T. Ngo (editors)*. IPBES secretariat.

Johnston, A., Matechou, E., & Dennis, E. B. (2023). Outstanding challenges and future directions for biodiversity monitoring using citizen science data. *Methods in Ecology and Evolution*, *14*(1), 103–116. https://doi.org/10.1111/2041-210X.13834

Kukkala, A. S., & Moilanen, A. (2013). Core concepts of spatial prioritisation in systematic conservation planning. *Biological Reviews*, *88*(2), 443–464. https://doi.org/10.1111/brv.12008

Legendre, P. (2014). Interpreting the replacement and richness difference components of beta diversity. *Global Ecology and Biogeography*, *23*(11), 1324–1334. https://doi.org/10.1111/geb.12207

Legendre, P., & De Cáceres, M. (2013). Beta diversity as the variance of community data: Dissimilarity coefficients and partitioning. *Ecology Letters*, *16*(8), 951–963. https://doi.org/10.1111/ele.12141

Leinonen, R., Pöyry, J., Söderman, G., & Tuominen-Roto, L. (2016). Suomen yöperhosseuranta (Nocturna) 1993–2012. Suomen ympäristökeskus. https://helda.helsinki.fi/handle/10138/161221

Margules, C. R., & Pressey, R. L. (2000). Systematic conservation planning. *Nature*, *405*(6783), 243–253. https://doi.org/10.1038/35012251

Matutini, F., Baudry, J., Pain, G., Sineau, M., & Pithon, J. (2021). How citizen science could improve species distribution models and their independent assessment. *Ecology and Evolution*, *11*(7), 3028–3039. https://doi.org/10.1002/ece3.7210

McKinley, D. C., Miller-Rushing, A. J., Ballard, H. L., Bonney, R., Brown, H., Cook-Patton, S. C., Evans, D. M., French, R. A., Parrish, J. K., Phillips, T. B., Ryan, S. F., Shanley, L. A., Shirk, J. L., Stepenuck, K. F., Weltzin, J. F., Wiggins, A., Boyle, O. D., Briggs, R. D., Chapin, S. F., … Soukup, M. A. (2017). Citizen science can improve conservation science, natural resource management, and environmental protection. *Biological Conservation*, *208*, 15–28. https://doi.org/10.1016/j.biocon.2016.05.015

Mikkonen, N., Leikola, N., Lehtomäki, J., Halme, P., & Moilanen, A. (2023). National high-resolution conservation prioritisation of boreal forests. *Forest Ecology and Management*, *541*, 121079. https://doi.org/10.1016/j.foreco.2023.121079

Montràs-Janer, T., Suggitt, A. J., Fox, R., Jönsson, M., Martay, B., Roy, D. B., Walker, K. J., & Auffret, A. G. (2024). Anthropogenic climate and land-use change drive short- and long-term biodiversity shifts across taxa. *Nature Ecology & Evolution*, *8*(4), 739–751. https://doi.org/10.1038/s41559-024-02326-7

Morris, W. K. (2021). *Finbif* [computer software]. *Zenodo*. https://doi.org/10.5281/zenodo.5414464

Pereira, H. M., Ferrier, S., Walters, M., Geller, G. N., Jongman, R. H. G., Scholes, R. J., Bruford, M. W., Brummitt, N., Butchart, S. H. M., Cardoso, A. C., Coops, N. C., Dulloo, E., Faith, D. P., Freyhof, J.,

14724642, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/ddi.13864 by University Of Jyväskylä Library, Wiley Online Library on [22/05/2024]. 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

Gregory, R. D., Heip, C., Höft, R., Hurtt, G., Jetz, W., … Wegmann, M. (2013). Essential biodiversity variables. *Science*, *339*(6117), 277–278. https://doi.org/10.1126/science.1229931

Pinkert, S., Barve, V., Guralnick, R., & Jetz, W. (2022). Global geographical and latitudinal variation in butterfly species richness captured through a comprehensive country-level occurrence database. *Global Ecology and Biogeography*, *31*(5), 830–839. https://doi.org/10.1111/geb.13475

Pöyry, J., Luoto, M., Heikkinen, R. K., Kuussaari, M., & Saarinen, K. (2009). Species traits explain recent range shifts of Finnish butterflies. *Global Change Biology*, *15*(3), 732–743. https://doi.org/10.1111/j.1365-2486.2008.01789.x

R Core Team. (2021). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. https://www.R-project.org/

Schulman, L., Lahti, K., Piirainen, E., Heikkinen, M., Raitio, O., & Juslén, A. (2021). The Finnish biodiversity information facility as a best-practice model for biodiversity data infrastructures. *Scientific Data*, *8*(1), 137. https://doi.org/10.1038/s41597-021-00919-6

Shirey, V., Belitz, M. W., Barve, V., & Guralnick, R. (2021). A complete inventory of north American butterfly occurrence data: Narrowing data gaps, but increasing bias. *Ecography*, *44*(4), 537–547. https://doi.org/10.1111/ecog.05396

Shirey, V., Khelifa, R., M'Gonigle, L. K., & Guzman, L. M. (2023). Occupancy–detection models with museum specimen data: Promise and pitfalls. *Methods in Ecology and Evolution*, *14*(2), 402–414. https://doi.org/10.1111/2041-210X.13896

Socolar, J. B., Gilroy, J. J., Kunin, W. E., & Edwards, D. P. (2016). How should Beta-diversity inform biodiversity conservation? *Trends in Ecology & Evolution*, *31*(1), 67–80. https://doi.org/10.1016/j.tree.2015.11.005

Tyler, T. (2020). Relationship between moth (night active Lepidoptera) diversity and vegetation characteristics in southern Sweden. *Journal of Insect Conservation*, *24*(6), 1005–1015. https://doi.org/10.1007/s10841-020-00270-y

Warren, M. S., Maes, D., van Swaay, C. A. M., Goffart, P., Dyck, H. V., Bourn, N. A. D., Wynhoff, I., Hoare, D., & Ellis, S. (2021). The decline of butterflies in Europe: Problems, significance, and possible solutions. *Proceedings of the National Academy of Sciences*, *118*(2). https://doi.org/10.1073/pnas.2002551117

Zattara, E. E., & Aizen, M. A. (2021). Worldwide occurrence records suggest a global decline in bee species richness. *One Earth*, *4*(1), 114–123. https://doi.org/10.1016/j.oneear.2020.12.005

Zellweger, F., Roth, T., Bugmann, H., & Bollmann, K. (2017). Beta diversity of plants, birds and butterflies is closely associated with climate and habitat structure. *Global Ecology and Biogeography*, *26*(8), 898–906. https://doi.org/10.1111/geb.12598

BIOSKETCHES

Rémi Duflot's research interests are in the broad field of landscape ecology. He studies the effects of landscape heterogeneity and connectivity on biodiversity in human-dominated environments (agricultural, urban, forested). His research focuses on biodiversity conservation in relation to land use, landscape planning and management.

Anssi V. Vähätalo is an avid lepidopterologist. He has reported together with Leo Vähätalo >45,000 records of >1800 lepidopteran species into the Finnish Biodiversity Information Facility. He is currently investigating lepidopteran biodiversity in the Caucasus region using DNA barcoding technique.

Author contributions: Rémi Duflot: Conceptualisation; data curation; formal analysis; visualisation; writing—original draft. Anssi V. Vähätalo: Conceptualisation; data curation; writing—review & editing.

SUPPORTING INFORMATION

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

How to cite this article: Duflot, R., & Vähätalo, A. V. (2024). Identifying sites with high biodiversity value using filtered species records from a biodiversity information facility. *Diversity and Distributions*, *00*, e13864. https://doi.org/10.1111/ddi.13864

14724642, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/ddi.13864 by University Of Jyväskylä Library, Wiley Online Library on [22/05/2024]. 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