Landsc Ecol (2026) 41:203
https://doi.org/10.1007/s10980-026-02443-7

RESEARCH ARTICLE

Including linear landscape elements improves fine-scale species distribution models of bees and hoverflies

Merijn Moens · Jacobus C. Biesmeijer · Marialuce Paladini · Gregorio Calderoni · Hens A. C. Runhaar · Saskia Klumpers · Joris van de Ven · Petra Schoon · Leon Marshall

Received: 25 November 2025 / Accepted: 15 August 2026
© The Author(s) 2026

Abstract

Context Species distribution models combine environmental and occurrence data to map and predict habitat suitability for a species. Predicting species distributions in agricultural regions can be challenging due to historical under-sampling in those areas and the presence of small landscape elements or non-productive land adjacent to the agricultural fields.

Objectives Here we model bees and hoverflies in Dutch agricultural regions at a 10 m resolution to measure the improvement in habitat suitability predictions when landscape elements are included as model parameters.

Methods We compared models excluding landscape elements with those incorporating them in two ways, as distance to the nearest element and as percentage local cover, and further assessed the latter against null models with randomized landscape variables. We evaluated the models using both withheld training data and independent field-collected data from three distinct agricultural regions.

Results Our results demonstrate that integrating linear landscape elements into species distribution models significantly improves model performance by correcting the overprediction in agricultural fields and under prediction in field edges. Herbaceous landscape elements were especially important for modelling the distribution of pollinators, showing statistically significant model improvements for the withheld test data and independently collected data.

Conclusions Integrating landscape elements into distribution models is becoming both increasingly relevant and more feasible. It is gaining relevance as European policies promote the expansion of semi-natural habitats within agricultural landscapes, while advances in the availability of high-resolution spatial data are making such integration more achievable. We recommend that modelers integrate these elements to more accurately predict biodiversity in agricultural habitats. The resulting maps provide a more robust estimate of habitat suitability, underpinning conservation and management strategies.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s10980-026-02443-7.

M. Moens (✉) · J. C. Biesmeijer · M. Paladini · G. Calderoni · S. Klumpers · J. van de Ven · L. Marshall
Naturalis Biodiversity Center, Leiden, 2333 CR Leiden, The Netherlands
e-mail: merijn.moens@naturalis.nl; merijnmoens@gmail.com

M. Moens · J. C. Biesmeijer
Leiden University, 2333 CC Leiden, The Netherlands

H. A. C. Runhaar
Utrecht University, 3584 CB Utrecht, The Netherlands

P. Schoon
Cobra Adviseurs B.V., 5434 SJ Vianen, The Netherlands

Published online: 28 August 2026

Keywords Pollinators · Species distribution modelling · Agricultural landscapes · Field edges · Ground-truthing · Semi-natural habitats

Introduction

In the fields of ecology and conservation researchers frequently apply species distribution models (SDMs) and the use of these models has seen a massive increase since 1970 (Zurell et al. 2020; Asch et al. 2022). SDMs combine environmental data with species occurrence or abundance data (Elith and Leathwick 2009) to model and map the distribution of a species. A species' distribution is shaped by biotic factors, abiotic conditions and the accessibility of suitable areas for dispersal (Soberon and Peterson 2005). Modelling these relationships allows researchers to identify suitable habitats, anticipate range shifts and direct conservation efforts (Guisan and Zimmermann 2000; Johnson and Gillingham 2005; Frans et al. 2022). Such models can also be used to identify regions at risk of invasive species spread (Guillera-Arroita et al. 2015; Melo-Merino et al. 2020) and to project the effects of climate change on future distributions (Guillera-Arroita et al. 2015; Melo-Merino et al. 2020). By identifying suitable habitats, SDMs can help guide the implementation of conservation measures. In some cases, predicted habitat suitability may also be correlated to species abundance (Weber et al. 2017), potentially providing additional information to improve their application in conservation.

The accuracy of SDMs depends not only on ecological and dispersal factors, but also on the quality of the underlying occurrence data (Kramer-Schadt et al. 2013; Costa et al. 2015; Liu et al. 2019). Uncertainty in model outcomes can result from common sampling biases in the occurrence data, such as spatial and temporal biases. Spatial biases arise from preferential sampling in certain habitats (Kramer-Schadt et al. 2013; Aiello-Lammens et al. 2015), when remote areas lack data (Meyer et al. 2015) or uneven sampling between countries (Beck et al. 2014). Temporal biases occur when methods of collection change over time (e.g. from historical expert naturalists to modern citizen scientists) (Boakes et al. 2010) or due to preferential sampling related to daily temperature and day of the week (Rosario et al. 2024). In the case of few occurrence records SDMs generally become less accurate (e.g. Liu et al. 2019; Wang & Jackson 2023). Factors besides abundance and distribution of a species that influence the number of occurrence points include preferential recording of certain taxonomic groups (Troudet et al. 2017), imperfect detectability leading to fewer available records and inaccurate models (Lahoz-Monfort et al. 2014; Wang and Jackson 2023) or misidentification (Costa et al. 2015). Misidentification leads not only to less occurrences of the species itself but also introduces erroneous data for the species it was incorrectly identified as (Costa et al. 2015). Models with the incorrect data may lead to expansion, shrinkage or shift of the predicted distributions (Costa et al. 2015).

Preferential sampling in certain land-use types over others can introduce significant biases into species distribution models. Opportunistic collection data tends to favor easily accessible areas, such as urban areas near roads, or areas with the potential to encounter specific species or higher biodiversity, such as protected areas (Boakes et al. 2010; Johnston et al. 2020). Agricultural landscapes may consequently be under sampled compared to other land use types (Geldmann et al. 2016). For example, a recent study found that, by the end of 2024, approximately 60% of the global cultivated area had no species recorded in the previous year (Remelgado et al. 2025). Limited access to privately owned land and the prioritization of other environments may contribute to undersampling in agricultural areas (Remelgado et al. 2025). Research on the extent to which agricultural areas are undersampled remains limited.

In addition to these sampling issues, models may also suffer from biases of thematic resolution and the thematic resolution itself can affect model performance, predicted habitat ranges and variable importance (Marshall et al. 2021). Biodiversity assessments generally omit or oversimplify land use data (Dou et al. 2021). For example, when agricultural landscapes are represented as single homogenous land use categories, the contribution and importance of fine-scale semi-natural features interspersing agricultural fields are overlooked.

Therefore, another factor that may affect the performance of models in agricultural areas is whether up-to-date information on semi-natural landscape elements is integrated into the models. The agricultural landscape is interspersed with landscape elements (LEs) that are often linear at the edge of the fields and

that consist of non-productive areas, such as hedgerows, tree lines, herbaceous vegetation at field edges, ditches and rivers. These linear landscape elements or so-called 'green–blue veining' in the landscape are part of local, national and even international policies aimed at bending the curve on biodiversity loss in the countryside and may enhance the diversity of many species, including pollinators (Schweiger et al. 2006; Musters et al. 2009; Cole et al. 2020; von Königslöw et al. 2022). Semi-natural habitats with herbaceous vegetation support pollinators and other taxa in intensively managed agricultural landscapes (Cole et al. 2017; Bartual et al. 2019). In addition to providing habitat and foraging resources, LEs may influence species occurrence through several mechanisms, such as microclimatic conditions (Kozelová et al. 2020; Vanneste et al. 2020), the presence of foraging resources (Garratt et al. 2017), and enhanced connectivity to natural areas (Garratt et al. 2017). Semi-natural landscape elements may be especially important in intensively managed agricultural landscapes, where low habitat heterogeneity can limit the availability of suitable habitats and resources for pollinators (Prendergast et al. 2022).

Several SDMs or spatially explicit models have incorporated LEs as their predictors (Morelli 2013; Bellamy et al. 2020; De Vries et al. 2021; Jungandreas et al. 2022; Roilo et al. 2023; Egloff et al. 2025). The predictors were introduced as percentage cover of the specific LE type (Jungandreas et al. 2022; Roilo et al. 2023), distance to the specific LE type (Bellamy et al. 2020; Jungandreas et al. 2022; Roilo et al. 2023), length of the LE (Sullivan et al. 2017; De Vries et al. 2021) or a density measure of the specific LE type (Bellamy et al. 2020; De Vries et al. 2021; Jungandreas et al. 2022). Recommending their further integration into future SDMs and their downstream applications requires a better understanding of how LEs used as predictors affect model performance and alter spatial predictions.

This requires thorough validation, both internally (to assess model fit) and with independently collected datasets that can test performance across different spatial contexts (Johnson and Gillingham 2005; Marshall et al. 2015; Roberts et al. 2017; Lee-Yaw et al. 2022).

Building on this rationale we test whether the explicit inclusion of LEs can enhance wild bee and hoverfly SDM performance in agricultural systems in the Netherlands. These insect groups are diverse in numbers of species (Reverté et al. 2023) and are sensitive to land use changes (De Palma et al. 2015) while also providing important pollination services, making them relevant for both biodiversity conservation and ecosystem functioning. As many SDMs do not integrate LEs within the modelling framework, we expect that they will overgeneralize predicted habitat suitability in agricultural regions, missing the fine-scale contrast in species richness between landscape elements and the surrounding fields. By incorporating LEs, we aim to make more accurate predictions of habitat suitability in agricultural landscapes, that better reflect resources availability. The study area encompasses various agricultural regions in the Netherlands, which serve as a relevant context given that agricultural land use accounts for 54% of the country's territory (Central Bureau of Statistics (CBS) 2015). Much of this land is privately owned and remains largely inaccessible to most observers, highlighting the relevance and importance of accurate, model-based approaches for identifying key habitat and conservation resources in intensively managed landscapes.

Materials and methods

Species occurrences used for model building

The bee and hoverfly occurrence data used in this study were provided by the European Invertebrate Survey in the Netherlands (European Invertebrate Survey (EIS) 2020). The dataset comprised occurrences sourced from a wide array of contributors, ranging from professionals to amateur experts and citizen scientists, including opportunistic collections as well as structured monitoring and scientific surveys. Consequently, the data is likely subject to several sampling biases, including spatial bias toward more accessible or frequently visited areas (Meyer et al. 2015), temporal bias arising from uneven survey effort across years and seasons (Boakes et al. 2010; Rosario et al. 2024) and taxonomic bias toward conspicuous or easily identifiable species (Troudet et al. 2017; Díaz-Calafat et al. 2024). Misidentification by the observer is expected to be lower than unsupervised data, as the citizen science records are validated through automated image recognition and

expert review. We chose the period from 2008 to 2022, reducing the effect of seasonal and yearly survey effort. Narrowing the time span to less than fifteen years would lead to insufficient records for some species, and we argue that, even though tenancy of different parcels may vary significantly over this fifteen-year period, the location of the parcels and their adjacent landscape elements would remain relatively stable.

To further reduce spatial biases, the occurrences were spatially thinned to 1 km resolution using the 'spThin' package version 0.2.0 (Aiello-Lammens et al. 2015). Spatial thinning has been shown to reduce model overfitting and spatial biases (Boria et al. 2014). Recent studies suggest that spatial thinning does not necessarily improve model performance or reduce spatial bias (Ten Caten and Dallas 2023) and may be beneficial primarily for large datasets with pronounced sampling biases (Lamboley and Fourcade 2024), as may be the case for our dataset. All spatial and statistical analyses were performed in R version 4.5.0 or later (R Core Team 2024). The occurrences were then stratified by land use type (urban, agricultural, natural or mixed) based on the proportion of land use in each 1 km cell to reduce the effect of sampling bias in certain areas, such as urban areas and nature areas (Boakes et al. 2010; Geldmann et al. 2016; Johnston et al. 2020). Land use types that were overrepresented to their actual coverage in the Netherlands were randomly subsampled to align the proportion of occurrences with the proportion of each land use in the study area. The land use classification was based on a 50% threshold of land cover; the mixed land use category was used when neither of the other three categories (urban, agricultural, natural) exceeded 50%.

Occurrence data were located proportionally 3.87 times less in agricultural areas, 1.75 times more often in urban areas, 2.65 times more often in nature areas than the background distribution of land use in the Netherlands. Other grid cells with a mixture of land uses were similar (1.01 times more often for occurrence data compared to the background distribution). To avoid propagating the land use bias in the models, we proceeded with the occurrence data scaled to the proportion of land use data in the background points in the SDMs.

For modeling, we retained only bee species that had more than 28 occurrences (maximum number of predictors) and were detected during the independent field sampling in 2023. This selection resulted in 82 species (Table S1 in Supplementary material), with the number of spatially thinned occurrence points ranging between 29 and 2825. To reduce overfitting, we checked the predictors for collinearity and compared different regularization multipliers (see section "Environmental data used in model building" and "Species distribution models").

Environmental data used in model building

Climate

Daily climate data was extracted from the KNMI weather API (KNMI 2023) to retrieve average monthly maximum and minimum temperature and precipitation over the years 2008–2022. Using the 'dismo' package version 1.3.14 (Hijmans et al. 2023), we calculated the bioclimatic variables for this period (with the same formula as the calculation of the bioclimatic variables in WorldClim) (Fick and Hijmans 2017). The resulting climate data was at a resolution of 1 km and the data was resampled using linear interpolations to a resolution of 10 m to match the resolution of the land use and LE data. The fine-scale resampling of the climate data is justified because climate mainly varies at a large scale, as opposed to land use data, which shows greater variability at a small-scale (Martin et al. 2013; Santos et al. 2021). The climate data, together with the land use data and LEs data were tested for collinearity, variables with a Spearman's correlation coefficient higher than 0.7 were removed, following (Dormann et al. 2013). Twelve bioclimatic variables were included in the species distribution models (see Table S2 in Supplementary material).

Land use

Land use data was sourced from the National Land Use of the Netherlands database (LGN; Hazeu et al. 2020) at a resolution of 5 m. The spatial data was aggregated into lower thematic resolution land use classifications that were uncorrelated (see Table S2 in Supplementary material): agriculture, fruit farms, tulip fields, coniferous forest, deciduous forest, intensive grassland, extensive grassland, urban gray, urban green, freshwater, salt water, heather, peat, sandy

soil vegetation, shrubs, and wetlands. This aggregation followed the broader land-use classes used by Marshall (2018) for modelling bee distributions in the Netherlands. Although derived from LGN land-use data, these classes broadly correspond to European habitat or land-cover categories, including EUNIS habitat groups (Thuiller et al. 2025) such as heathlands, wetlands, forests, grasslands, and inland waters, and CORINE Land Cover classes (Bossard et al. 2000) such as green urban areas, coniferous forest, broad-leaved forest, wetlands, and water bodies. Each variable was included in the model as a measure of proportional cover (0, 25, 50, or 100%) at 10 m × 10 m.

Linear landscape elements

Data for the blue and woody LEs were retrieved from Cobra Adviseurs B.V. (Cobra adviseurs B.V. 2024) and the herbaceous vegetation cover was retrieved from the Atlas Natuurlijk Kapitaal (RIVM 2022). The woody elements included shrubs and tree species, and the blue elements represent the network of ditches, (river)banks and dry ditches. These elements have been mapped through topological operations, combining data from *Basisregistratie Grootschalige Topografie (BGT)* (Kadaster 2020) and the Tree Monitor dataset from *Co*bra Adviseurs B.V that contains the location of every tree in the Netherlands (Cobra adviseurs B.V. 2024). The detailed mapping procedure is confidential; however, additional information is available at https://cobra-groeninzicht.nl/. Other spatial datasets on landscape elements were less suitable for our study for their lower spatial resolution data, such as 1 km2 (Van Der Zanden et al. 2013), other focal geographic regions (O'Connell et al. 2015), or specific types of landscape elements with predefined dimensional criteria (Copernicus Land Monitoring Service 2021). The minimum distance between different points and polygons were calculated with the 'sf' package version 1.0.15 (Pebesma 2018). The LE data preparation can be found in Fig. 1 (see also Textbox S1 in Supplementary material). The type of LE input data consisted of "distance to" and "percentage cover" of an LE. The type of LE included herbaceous, woody, blue and any LE and only those LEs adjacent to agriculture were selected (see Fig. 1). For the calculation of the "distance to" category, the "any" LE category, used in this study, consisted of all polygons joint together and represented the minimum distance towards any of those polygons. For the percentage cover of any LE, we first scaled the percentage cover of woody and blue elements and the herbaceous elements relative to their minimum and maximum value resulting in a value ranging from 0 to 1. This gave both components equal potential weight in the combined index, regardless of differences in their original abundance and ranges. The two scaled variables were then summed, resulting in a combined index ranging from 0 to 2.

Although the land use and climate predictors did not show a correlation coefficient higher than 0.7 with any of the other predictors, the LE variables showed correlations between several categories: distance to herbaceous LEs and distance to any LEs (Spearman's $\rho = 0.88$), distance to blue LEs and distance to any LEs ($\rho = 0.85$) and percentage cover of herbaceous LE and any LE ($\rho = 0.88$) (Figure S1 in Supplementary material). For this reason, we built separate models for each LE variable. Spearman's correlation between other LE variables and other variables in the models did not exceed 0.7 (Dataset 1 in Supplementary material).

Additionally, to test whether the inclusion of LE variables genuinely improved model performance we employed a null variable approach. For each LE type (blue, woody, herbaceous and any) both 'percentage cover' and 'distance to' variables were randomized to create null predictors with the same structure but without ecological meaning. Testing against models containing these randomized variables allowed us to assess whether any observed increase in model performance by adding the real LE variables was greater than could be expected from adding any additional variable with the same range of values.

Species distribution models

We used maximum entropy modelling (MaxEnt; Phillips et al. 2006), to construct the SDMs. The MaxEnt algorithm predicts a species' distribution from predictors but tends toward a more uniform distribution, maximizing entropy (Phillips et al. 2006). It is suitable for presence-only data (Kaky et al. 2020) and performs especially well with small sample sizes among other SDM algorithms (Aguirre-Gutiérrez et al. 2013). The distributions of the bees and hoverflies were modelled in the 'ENMeval' package version 2.0

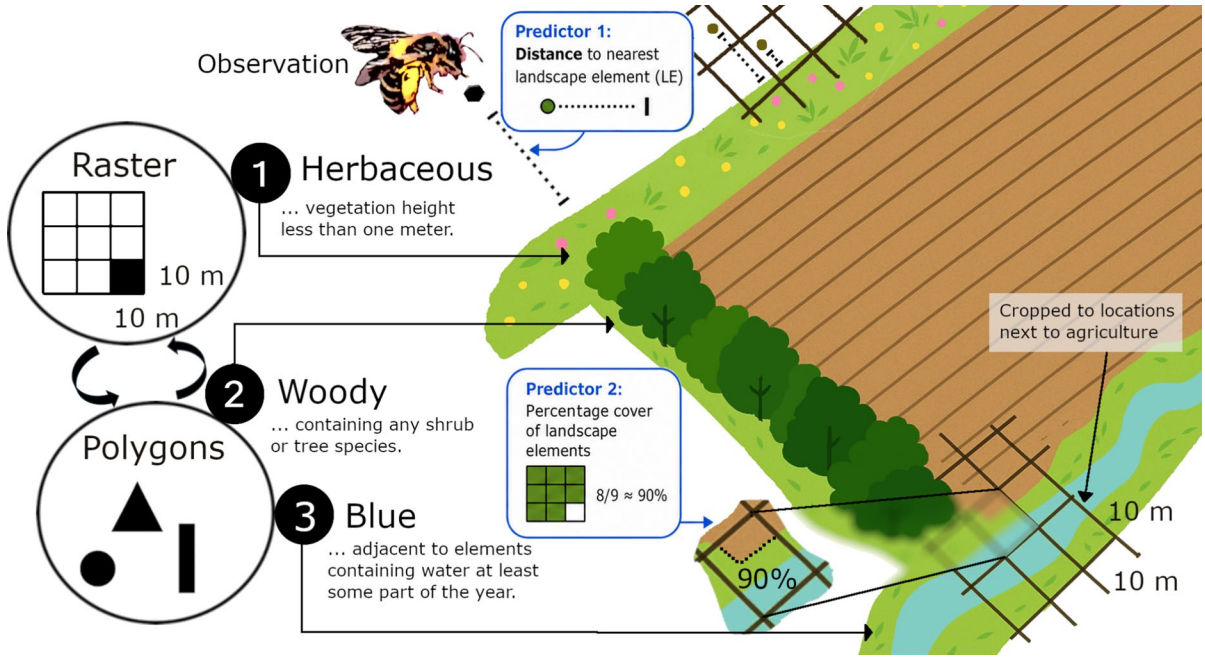


Fig. 1 Overview of the preparation of the landscape element (LE) data for input as predictors in the species distribution models in the Netherlands. The different categories consist of herbaceous, woody and blue LEs. The raster data (herbaceous LEs; RIVM 2022) and polygon data (woody and blue LEs; Cobra adviseurs B.V., 2024) are prepared in two categories of predictors for the species distribution models: distance to the nearest LE (calculated from polygon data; predictor 1 in Fig. 1) and percentage cover of the LE (calculated from raster data; predictor 2 in Fig. 1). These two categories represent the types of input variables into the models in raster format at a resolution of 10 m. Only locations adjacent to grid cells with at least some percentage cover of agriculture were included. The background data of the distance to the nearest LE is calculated based on the distance to a random point in each grid cell to the nearest LE polygon

(Kass et al. 2021) in R (R Core Team 2024). We compared the following different combinations of feature types: (i) linear, (ii) linear and quadratic, (ii) hinge, (iv) linear, quadratic and hinge (v) linear, quadratic, hinge and product and (vi) linear, quadratic, hinge and threshold. The features represent functions of the input variables (Phillips and Dudík 2008) and can infer the relation between the input variable and the distribution of a species. A more complex set of features can increase model complexity and model overfitting (Merow et al. 2013) on one hand, while on the other hand increasing model performance. A series of penalties for model overfitting (regularization multipliers; Merow et al. 2013) were compared, ranging between one to five. The training and test data (or withheld data) were selected with a spatial block validation (Muscarella et al. 2014), that divides the study area into four geographically distinct parts based on the occurrence data. In this approach, the 10,000 background points are partitioned as well and the background points are selected randomly across the study area, i.e. the area within the administrative borders of the Netherlands.

The SDMs were categorized into models with LEs, with randomized LEs and without LEs. The comparison between the models with LEs and randomized LEs is similar to comparing against null models (Raes and ter Steege 2007).

Field sampling for independent model validation

We collected an independent field dataset that was not used for model fitting, calibration, or selection, but was retained as a separate dataset for model evaluation. We sampled three to six linear LEs and fields of 25 farmers in agricultural regions encompassing tulip fields near Lisse, peat meadows in the Alblasserwaard, and mixed farming areas in the Ooijpolder (see Fig. 2). The total number of locations, including both LEs and the individual agricultural fields,

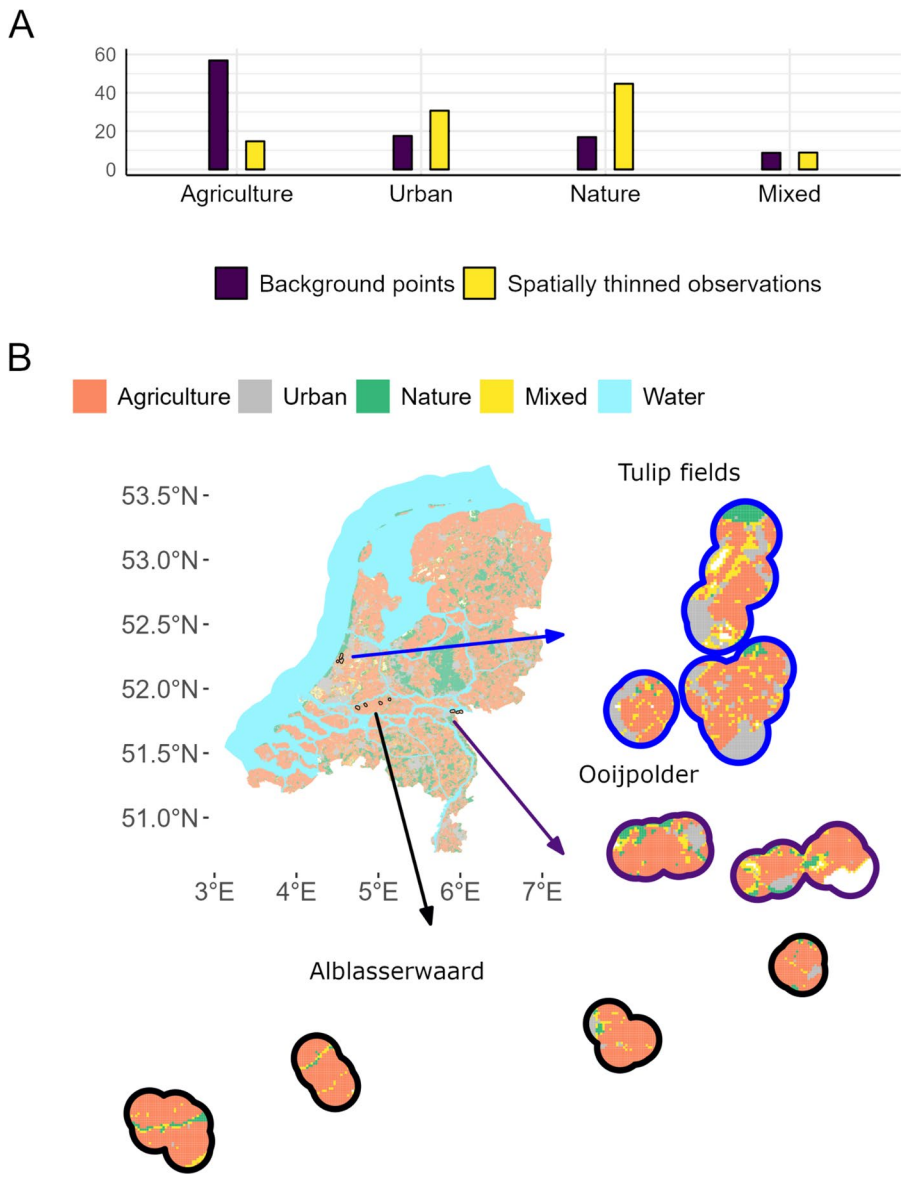


Fig. 2 The distribution of occurrence data in areas with more than 50% agricultural land, urban area, natural area and a mixture of land use types at 1 km resolution (A). The figure shows that agriculture is the dominant land-use type in the Netherlands, covering 54% of the country's territory (Central Bureau of Statistics (CBS) 2015). The occurrence data is spatially thinned at a 1 km resolution. A map of the Netherlands with the three different case studies in the Alblasserwaard, the Ooijpolder and the tulip fields near Lisse, and the land use share of agricultural, natural, urban and mixed land use (B)

totaled 101 unique locations. The landscape elements were chosen randomly within the farm, but we picked at least one landscape element near ditches or waterways and one with shrubs and trees, when possible (Textbox S2 in Supplementary material). The coordinates are left anonymous to protect the privacy of the farmers, but an overview of the study regions can be found in Fig. 2.

At every farm and in all locations, we sampled both within the field and next to the LEs during three rounds in 2023. The first sampling round was performed in May and April, the second sampling round in June and the third sampling round in July and August. The sampling methods included pan traps and transects (Textbox S2 in Supplementary material).

All bees and hoverflies specimens were identified by *names hidden for anonymized peer review* with reference material (Falk 2019). The samples that could not be identified were barcoded, using the manufacturer's materials and protocol (ON Technologies 2023, 2024). For every barcoded species, the match with at least 99% sequence identity was selected and for multiple matches, we chose the match with the

highest sequence identity. The total number of identified bees and hoverflies with coordinates totaled 2863. The specimens are stored within the NL Biodiversity and Society collection at Naturalis Biodiversity Center.

Statistical analyses

Evaluation measures

We evaluated model performance of the different SDMs with the continuous Boyce index (CBI; Hirzel et al. 2006) and the area under the curve (AUC) of the receiver operating characteristic. These evaluation measures were calculated for the models with LEs, with randomized LEs and without LEs and the difference in model performance for these groups of models was calculated with the aforementioned measures. The permutation importance measure was calculated to estimate variable importance of the LEs and this measure was calculated for the models with LEs. The Corrected Akaike Information Criterion (AICc; Burnham & Anderson 2002) was applied to decide on the optimal regularization multiplier for all the different models.

The AUC is a threshold independent measure of binary classification problems that can be applied to species distribution models (Bradley 1997; Phillips et al. 2006). The AUC values of the models with randomized LEs were subtracted from the models with LEs for the withheld data and the independent data for any, blue, woody and herbaceous LEs. However, the AUC weighs the omission and commission rate equally (Peterson et al. 2008), even though omission errors or occurrence points predicted as absences, are generally more problematic in presence-only distribution modelling (Anderson et al. 2003). The relative importance of different prediction errors depends on the aim of the modelling exercise. For example, false positives may become more problematic when conservation resources are limited and management actions need to be targeted to specific locations, because overprediction could lead to inefficient allocation of effort (Guisan et al. 2013). For this reason, we also included another evaluation measure that does not directly use omission and commission rates but instead correlates occurrence points to a continuous output, such as habitat suitability: the continuous Boyce index (CBI) (Hirzel et al. 2006).

To optimize model parameterization, we selected combinations of regularization multipliers and feature classes that minimized the corrected Akaike Information Criterion (AICc), following the approach described by Kass et al. (2021). If the difference in AICc was less than two, we selected the model with the lowest evaluation AUC. For the importance of the individual variables, we compared permutation importance between models. Permutation importance in SDMs measures the importance of a variable by randomizing its values at both presence locations and background points while holding all other variables constant (Kalle et al. 2013; Jarnevich et al. 2016).

While the evaluation measures quantify the importance of the LE input variables within the models, the response curves illustrate the relationships between these variables and the bee and hoverfly species, providing valuable insights into how the LE inputs influence model dynamics. The response curves result from different linear and non-linear features found in the models (Merow et al. 2013) and the marginalized response curves show the habitat suitability in relation to individual variables by keeping the other variables at their mean value.

For the comparison of proportion sampling and background distribution in different land uses, we applied a Fisher's goodness of fit test (Fisher 1922). For the comparisons of model output, we applied a Wilcoxon's test (Wilcoxon 1992) with a Benjamini–Hochberg correction (Benjamini and Hochberg 1995) for multiple comparisons.

Results

Occurrence data in agriculture

There was a strong discrepancy between the sampling background of spatially thinned occurrence data and the percentage of land use categories in the Netherlands (see Fig. 2A). Distribution of occurrences amongst agricultural, urban, natural and mixed land uses was not equal to background distribution of those land uses in the Netherlands ($\chi^2 = 233{,}520$; df = 3; $p<0.001$). Especially sampling in agricultural areas was low compared to the background distribution of urban, nature and mixed areas.

Landscape elements in species distribution models

Model improvement was most apparent when integrating herbaceous or any LEs, though improvements were also observed for blue and woody LEs (Fig. 3A). Detailed results and test statistics are presented in Table S3 in the Supplementary material. Both the 'distance to' and 'percentage cover' of the different LEs resulted in a statistically significant model improvement. These model improvements were found when testing the model against both the withheld evaluation data and the independently collected data, when model performance with the LEs was compared to a randomized null variable.

Models without LE predictors had a mean AUC for the withheld evaluation data of 0.682, with substantial variation among species, ranging from 0.468 to 0.918 (Figure S2–S5 and Dataset 2 in Supplementary material). Model performance tended to be lower for species with relatively small sample sizes (Figure S2–S5 in Supplementary material). For the independently collected data the AUC of the models without

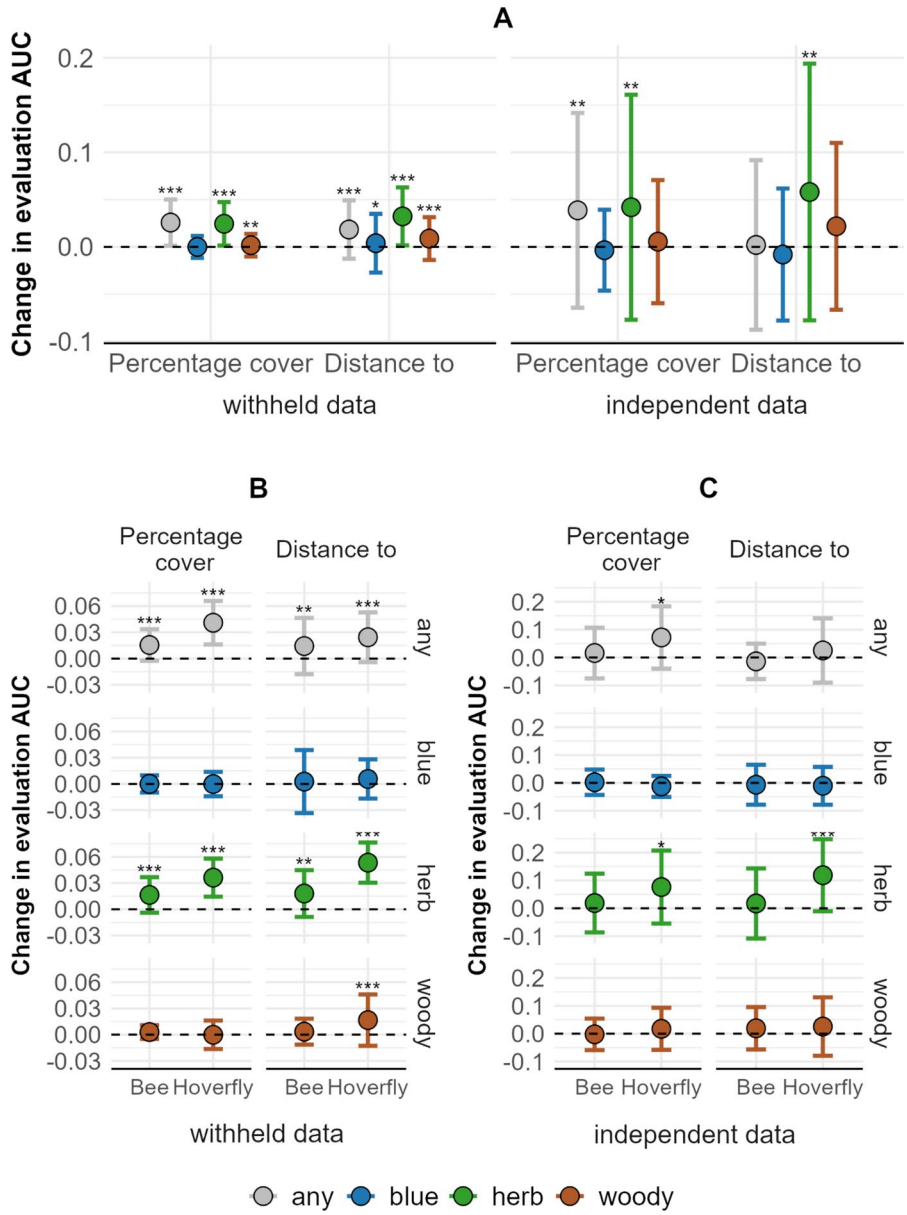


Fig. 3 Model performance with landscape elements (LEs) subtracted by model performance with randomized variables, averaged for all modelled species (A) and for bees and hoverflies separated (B, C). Model performance is measured with the area under the curve (AUC) of the receiver operating characteristic. Sample size is the number of species modelled (n = 82). The LEs are subdivided in any LE (the nearest), herbaceous LEs, blue LEs and woody LEs. Model improvement is grouped by withheld test data and independently collected data, and the categories are calculated as 'percentage cover' and as 'distance to' the nearest landscape element. The symbols above the standard deviations represent the significance levels of a Wilcoxon's test with Benjamin-Hochberg adjustment (Benjamini & Hochberg 1995) for multiple comparisons. Exact differences and p-values can be found in the Supplementary material (Dataset 2 in Supplementary material)

LE predictors averaged at 0.666 and ranged between 0.500 and 0.964 per species (Figure S2–S5 and Dataset 2 in Supplementary material). Most AUC values are sensitive to prevalence and the spatial extent of the study area (Lobo et al. 2008). Moreover, because presence-only data do not provide information on true absences, comparisons of absolute evaluation values amongst models or species can be difficult to interpret, particularly because omission and commission errors are weighted equally (Peterson et al. 2008; Jiménez and Soberón 2020). As our research contains widespread species with different sample sizes and presence-only data, we focused our analyses on species-specific changes in model performance, while reporting the absolute evaluation values in the Supplementary material (Figure S2–S5 and Table S4 in Supplementary material).

We found that model improvement (change in evaluation AUC) of the withheld test data was statistically significant for percentage cover of any LEs ($p_{adj} < 0.001$), for herbaceous LEs ($p_{adj} < 0.001$) and for woody LEs ($p_{adj} = 0.006$). The 'distance to' variables showed a model improvement for distance to any LEs ($p_{adj} < 0.001$), distance to herbaceous LEs ($p_{adj} < 0.001$), distance to blue LEs ($p_{adj} = 0.011$) and distance to woody LEs ($p_{adj} < 0.001$). We found a similar trend for the model improvement of the independently collected data with percentage cover of any LEs ($p_{adj} = 0.0027$) and herbaceous LEs ($p_{adj} = 0.0033$) showing a statistically significant improvement. Distance to herbaceous LEs showed a significant model improvement for the independently collected data ($p_{adj} = 0.001$). Although differences in evaluation measures between the distance to and percentage cover of herbaceous LEs and a randomized version were higher for the independently collected data compared to the withheld data, statistical significance was lower (Table S3 in Supplementary material). A likely reason for this discrepancy is the difference in average sample size per species of the withheld data (118.25 ± 63.50 S.E.M.) and the independently collected data (25.54 ± 6.44), potentially leading to more variable outcomes in the evaluation measures for the independently collected data.

The CBI compares how well higher predicted suitability values correspond with presences. We found a similar trend when comparing the CBI (see Figure S6 in Supplementary material), compared to the AUC. When comparing variable importance, adding the LE information as 'distance to' instead of 'percentage cover', showed a statistically clear increase in permutation importance for blue LE ($p_{adj}<0.001$), herbaceous LE ($p_{adj}<0.001$) and woody LE ($p_{adj}<0.001$) and a decrease for any LE ($p_{adj}<0.001$) (Table S5 in Supplementary material).

After subdividing the results in bees and hoverflies, both groups showed similar trends in terms of the change in evaluation AUC (Fig. 3B and Fig. 3C; see Table S3 in Supplementary material), particularly when using withheld data to test the models. However, lower sample size may have contributed to some differences in the independently collected data not reaching statistical significance, particularly for bees. The hoverflies showed a statistically significant model improvement for distance to woody LEs in the withheld data ($p_{adj} < 0.001$) in contrast to the bees ($p_{adj} =1$). For the withheld data bees improved with the percentage cover of any LE ($p_{adj} <0.001$), the distance to any LE ($p_{adj}=0.0081$), the percentage cover of herbaceous LEs ($p_{adj}<0.001$) and the distance to herbaceous LEs ($p = 0.0015$). The independently collected data showed a similar trend, but none of the differences were statistically significant for the bees (Fig. 3C). The withheld data showed a significant model improvement for the hoverflies for the percentage cover of any LE ($p_{adj} < 0.001$), the distance to any LE ($p_{adj} < 0.001$), the percentage cover of herbaceous LE ($p_{adj} < 0.001$) and the distance to herbaceous LE ($p_{adj} < 0.001$). The independently collected data confirmed these results with a statistically clear improvement for the hoverflies of the percentage cover of any LE ($p_{adj} = 0.0015$), the percentage cover of herbaceous LEs ($p_{adj} = 0.0038$) and the distance to herbaceous LEs ($p_{adj} < 0.001$).

We found that SDMs excluding LE information predicted comparatively higher species richness in areas with only agriculture, and comparatively lower species richness in areas with herbaceous LEs, than models that incorporated LE information. Including information on landscape elements revealed a "diversity hotspot" or increased number of predicted species on the contours of the agricultural landscape, where the LEs are located (Fig. 4). In absence of LE information, models predict 56.47 ± 12.93 s.d. species in locations with only agriculture and 67.04 ± 10.40 species in locations with exclusively herbaceous cover. When the herbaceous LEs are integrated as percentage cover, models predict 49.68 ± 14.11

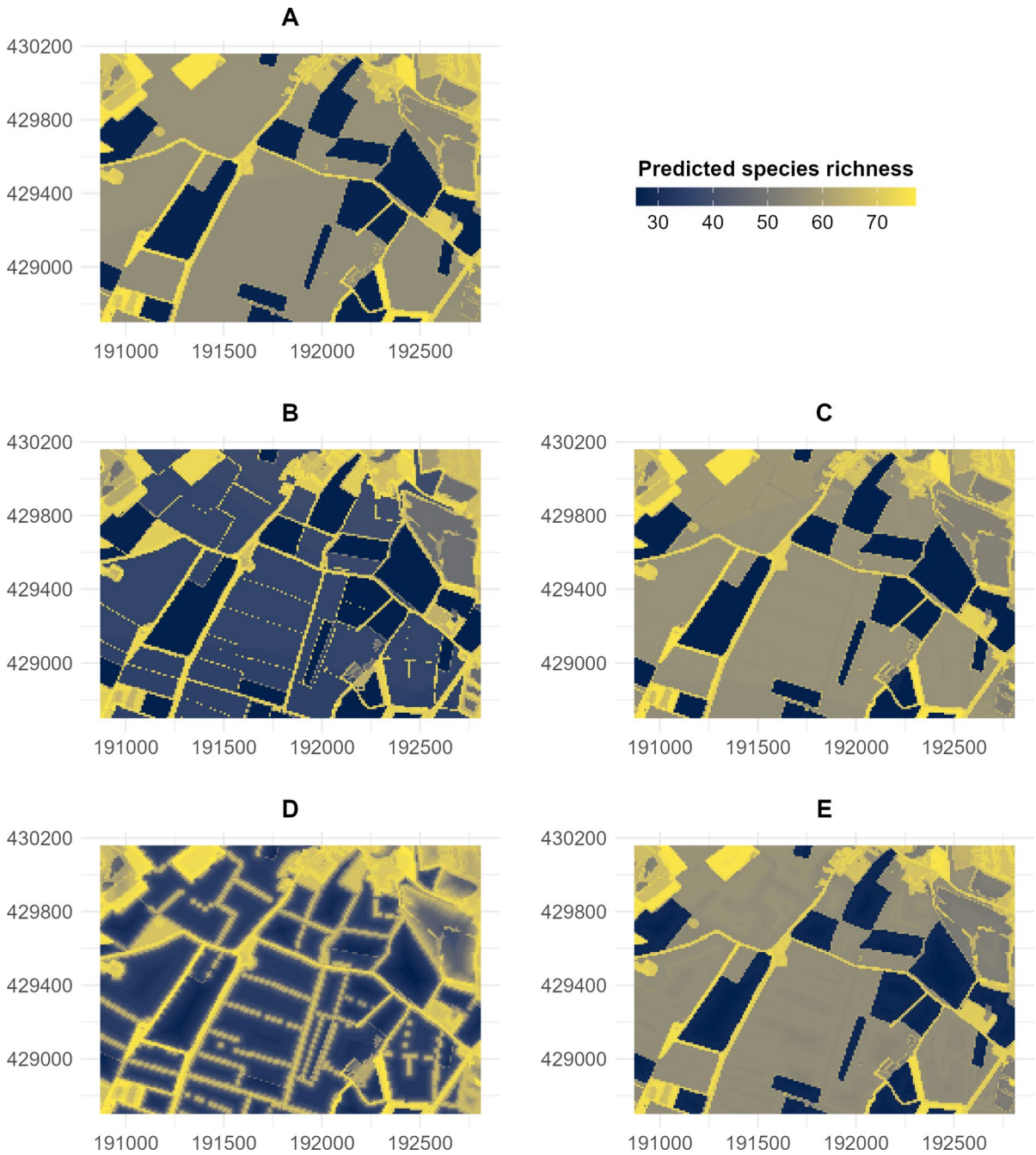


Fig. 4 The predicted species richness for an agricultural area in one of the case studies (Ooijpolder). Higher values represent a larger number of predicted bee and hoverfly species for models without any LEs information (A), models with the percentage cover of LEs (B), models with a randomized control of percentage cover of LEs (C), models with distance to any LEs (D) and models with a randomized control of distance towards any LEs (E)

species in exclusively agriculture and 76.11 ± 4.55 species at locations with only herbaceous LEs. When the herbaceous LEs are introduced as distance to the nearest LE, models predict 51.34 ± 17.87 species in exclusively agriculture and 74.66 ± 5.48 species at locations with only herbaceous cover. Besides the

control of models without the LE variable, there was another type of model that functioned as a second control, with a randomized version of either percentage cover of LE or distance to the nearest LE (Fig. 4C and Fig. 4E). These models showed a similar pattern in terms of species richness to that of the control without information on LEs, ruling out the possibility of increased model performance by adding any (or a random) variable.

In general, the features identified in the models showed that habitat suitability decreased gradually with decreasing percentage cover of herbaceous LE and dropped sharply as the distance to the nearest herbaceous LE increased (see Figure S7 in Supplementary material). On average, decreasing percentage cover of the herbaceous LE from 100% to 50.48% ± 9.79% s.d. (bees) or 46.8% ± 8.72% (hoverflies) halved the relative habitat suitability in those places for those species. Increasing distance to the nearest herbaceous LE halved the relative habitat suitability at 33.08 m ± 20.55 m for bees and at 20.55 m ± 18.97 m for hoverflies.

Discussion

Main findings

Although linear landscape elements form an integral part of the agricultural landscape and have been included in some SDMs (Bellamy et al. 2020; De Vries et al. 2021; Jungandreas et al. 2022; Roilo et al. 2023), their specific effects as fine-scale predictors on model performance and spatial predictions remain poorly understood. By explicitly comparing models with and without LE predictors, this study builds on previous SDM work by testing whether these fine-scale features improve both model performance and spatial predictability. In this study, we hypothesized that SDMs in agricultural areas lack essential information when linear LEs are not included, and resulting maps are likely to generalize predicted habitat suitability in agricultural regions. The addition of LE information improved many of the models in this study, indicating that LEs may be important for understanding the spatial distribution of pollinators in agricultural fields and for improving spatial predictions. This suggests that conventional SDMs based mainly on broader land-use and climatic predictors may miss ecologically relevant fine-scale variation in agricultural landscapes, particularly the resources, shelter, connectivity and microclimatic conditions associated with LEs.

Among the different types of LEs, herbaceous vegetation stood out as the most consistently important predictor, which aligns with field-based evidence that flower-rich herbaceous elements can support wild bee and hoverfly diversity. The more context-dependent effects of woody and blue LEs may suggest that these elements provide resources or habitat functions that operate at broader spatial scales and are therefore better captured by distance-based predictors. Alternatively, they may be less relevant for this set of pollinators. Models incorporating this information show statistically significant improvements, including when validated with independently collected field data and when compared against randomized controls. However, we only found a statistically significant improvement in model performance with the independently collected data, when the distance to and percentage cover of herbaceous LEs or percentage cover of any LEs were integrated. By adding information on LEs, the resulting prediction maps become more nuanced, indicating lower predicted species richness within the agricultural fields and higher species richness at the edge of the fields, where the LEs are located. We found that woody and blue LEs can enhance model performance when considered as distance to rather than percentage cover, highlighting their significance in the broader landscape context. Altogether, these findings are valuable from both nature conservation and a modelling perspective, because they show that incorporating fine-scale landscape structures can improve model performance and provide a more nuanced spatial representation of pollinator habitats within agricultural landscapes.

Scientific implications

The observed bias highlights the need for more data in the under sampled agricultural regions and the overrepresentation of occurrence data in urban areas is in line with other studies. Opportunistic occurrence data in the Netherlands is spatially biased towards urban and natural land uses, as was also found in other studies (Husby et al. 2021; Bowler et al. 2022). Occurrences in privately owned land in agriculture, such as the data collected in this study, could improve

our knowledge on the distribution of different species in those areas. Citizen science could play a role in providing data in under sampled regions (Callaghan et al. 2021). In agricultural regions, examples include initiatives involving farmers themselves (Billaud et al. 2021) or structured monitoring programs supported by governmental bodies (Andrade et al. 2021). Such approaches may help overcome access constraints on privately owned farmland while also enabling biodiversity observations to be linked to agricultural practices (Andrade et al. 2021; Billaud et al. 2021). A targeted sampling approach and stratified sampling may improve the data collected by citizen scientists, as this data often contains spatial and temporal biases (Backstrom et al. 2024) and citizen scientists may record less species than professional scientists (Kremen et al. 2011). To reduce collection bias based on the proportion of different land uses, a stratified sampling approach has been suggested by several other studies (Husby et al. 2021; Scherber et al. 2021).

The ecological significance of LEs is well-documented for different taxa from field studies (Tóth et al. 2019; Cole et al. 2020; Pustkowiak et al. 2021) and our results confirm especially the importance of herbaceous landscape elements. In another study, wild bee abundance and diversity were higher in flower strips than hedges despite shrubs offering resources earlier in the season (von Königslöw et al. 2022). This is in line with our results, showing more often a significant model improvement for herbaceous LEs than other LEs. In another study in Great Britain several grassland types produced the most nectar from different sources per unit area as opposed to arable land (Baude et al. 2016). Broadleaved woodland showed similar values in terms of nectar content and the diversity of sources it comes from, as different grassland types (Baude et al. 2016), suggesting the potential of tree lines to provide resources for pollinators. Although water and woody vegetation LEs may provide fewer floral resources than herbaceous elements, they can be crucial habitats for hoverfly larvae. For example, saproxylic hoverfly species develop in decaying wood and aquatic hoverflies, such as species in the genus *Eristalis*, depend on water bodies for larval development (Reemer et al. 2009). Nevertheless, adult hoverflies are not found in the blue elements themselves, but on the herbaceous LEs next to them. However, different taxa, including wild bees and hoverflies, exhibit a trend where after a certain threshold of increasing the proportion of LEs, the habitat quality of these areas becomes a more important determinant of their abundance (Bishop et al. 2025). Future research could explore whether incorporating habitat quality and heterogeneity within LEs improves model predictive performance within agricultural landscapes. Blue LEs did not show a clear model improvement in our study in the Netherlands, the abundance of water in the Netherlands may reduce its role as a limiting factor. By contrast, in drier areas, water availability might be a more critical variable and water related variables can be important predictors for species distribution models of different types of organisms (Mod et al. 2016; Bradie and Leung 2017). Besides drought, agricultural practice may influence the results, when LEs are modelled in other regions. The effect of distance to the nearest element could show different trends based on agricultural practices. Dutch agriculture is among the most intensive in Europe (Bos et al. 2013). In more extensive agricultural systems, where fields may provide more resources and shelter (Langlois et al. 2020), the added value of LE predictors may be smaller or expressed through different LE types. The increased model performance attributed to LEs may also partly reflect local microclimatic conditions that were not captured by the interpolated climatic variables used in this study. While these variables represent broad-scale climatic patterns, linear LEs can create fine-scale variation in temperature, humidity, wind exposure and shelter, thereby providing suitable microhabitats for pollinators (Kozelová et al. 2020; Vanneste et al. 2020).

Even though LEs have been integrated within the SDM framework (Morelli 2013; Bellamy et al. 2020; De Vries et al. 2021; Jungandreas et al. 2022; Roilo et al. 2023; Egloff et al. 2025), none of these studies compared model performance and predictability with and without the integration of LEs within the modelling framework. From a modeler's perspective, incorporating fine-scale information on LEs can improve our understanding of their importance to different taxa. The observed model improvement should be interpreted as the combined effect of several fine-scale ecological characteristics of LEs, rather than as the effect of a single mechanism. LEs may simultaneously represent habitat structure, resource availability, connectivity, shelter and local microclimatic conditions (Garratt et al. 2017). Comparing LE predictors

with fine-scale microclimatic or other high-resolution environmental predictors in future studies could further our knowledge on the contribution of these different components.

For other modelers, we recommend including information on LEs, especially when the aim is to map the distribution of the species within agriculture and if LEs are ecologically meaningful for the particular species of interest. While the data and results are specific to the Netherlands and pollinators, the general approach of integrating fine-scale LE information into SDMs could be transferable to other regions and species. However, the significance and direction of LE effects are likely to depend on regional agricultural intensity, landscape context, water availability and the habitat requirements of the modelled organism.

In recent years, data availability for LEs has increased, including the locations of woody LEs (European Environment Agency 2021) and canopy height from LiDAR data for mapping hedgerows and tree lines (Moudrý et al. 2026). Without available LE information, the edges of agricultural fields could potentially serve as a proxy for the location of LEs or, at the very least, for areas of lower land use intensity. In addition to indicating potential LE locations, field size can act as a variable related to land use intensity and has been suggested as an explanatory variable in SDMs (Roilo et al. 2024). Smaller fields, with their greater abundance of field edges and potential LEs, have been associated with higher farmland biodiversity, although these effects may vary depending on landscape context and region (Clough et al. 2020). For several EU countries, harmonized parcel-level datasets such as EuroCrops v2.0 (Claverie et al. 2026) provide information on field size. However, Roilo et al. (2024) showed that the choice of land use intensity variable, including field size, can influence model results. Therefore, field size should ideally be considered alongside other spatial variables that may indicate the presence of LEs.

Additionally, different spatial modelling methods, such as spatial autocorrelation-based methods, can be applied to map the distribution of different LE types (Van Der Zanden et al. 2013), providing input data for SDM studies in other regions. For modelers integrating the LEs at a coarser resolution, we recommend exploring using the distance to nearest LEs, as this variable is scale independent.

The usefulness of 10 m-resolution LE predictors for other taxonomic groups will depend on the mobility and habitat requirements of the species. For less mobile or habitat-specialist species, fine-scale LE information may directly capture relevant habitat features. For more mobile vertebrates, such as birds or mammals, individual LEs may be less informative as local predictors, but they may still indicate important foraging, breeding, shelter or movement structures within agricultural landscapes. Additionally, life cycle stages, dispersal and habitat requirements could be integrated within the modelling framework (Zurell 2017). In general, including movement for mobile species may improve predictions (Miller and Holloway 2015; Schank et al. 2019; Duan et al. 2026) and improve the relevance of fine-scale predictors.

Societal implications

Few studies evaluate SDM predictions with ground truthing (Johnson and Gillingham 2005; Marshall et al. 2015; Roberts et al. 2017; Lee-Yaw et al. 2022) and confirming model performance with an additional independently collected data set increases the reliability of the results and therefore the downstream applicability of the results by stakeholders. However, SDMs are inherently correlative, and we can never be certain about causal relationships. Collecting independent data is often constrained by the number of locations that can be sampled and potential biases. Therefore, it is important for model developers to effectively communicate the different uses of the models for stakeholders dependent on the methodology (Sofaer et al. 2019).

The model improvement after integration of information on LEs supports recent EU policies aimed at increasing semi-natural habitats in agricultural fields, such as the Common Agricultural Policy from 2023 to 2027 (Cagliero et al. 2023). Additionally, the resulting maps provide spatial information on the relation between LEs and pollinators, supporting integration of SDMs within ecological impact assessments (Baker et al. 2021). We found an increase in habitat suitability with a higher percentage cover of LEs and a smaller distance towards different types of LEs in the response curves of the models. A meta-analysis found that habitat suitability often correlates with abundance, especially when multiple types of environmental variables are included (Weber et al.

2017). Increasing the distance from the nearest herbaceous LE to 33 m for bees and 21 m for hoverflies reduced habitat suitability by half. These thresholds could serve as useful rules of thumb for policymakers when planning the spatial arrangement of LEs. The selection of modelled species included those that were already present in agricultural fields, and further research is needed to understand what the requirements are for accommodating more threatened bees and hoverflies in the agricultural landscape. Increasing the total length of seminatural linear landscape structures within the agricultural landscape can increase wild bee species richness (Sydenham et al. 2025) and may or may not be able to accommodate different threatened wild bees. Still, these results are highly relevant for pollination services, as most of the crop pollination is carried out by a small number of common pollinator species (Kleijn et al. 2015).

Accurate SDMs have the potential to benefit stakeholders, such as municipalities/ provinces, farmer collectives and nature organizations, although the effectiveness of their implementation depends on different factors. Distribution maps can provide spatially explicit information that can showcase key locations for landscape elements and inform strategic landscape planning. The integration of LEs within the SDMs could be extended to urban areas, where there is an overrepresentation of occurrence data compared to agricultural land use, and the resulting maps could play a role in planning urban green infrastructure. Additionally, predictions under different scenarios can motivate stakeholders to implement conservation measures. The maps are flexible in interpretation, while being recognizable by stakeholders: characteristics that facilitate communication between stakeholders (Shaw et al. 2022). An important aspect for the implementation of these modelling studies is a participatory approach (Wassen et al. 2011). An example would be to invite stakeholders of the study regions and showcase the distribution maps of the bee and hoverflies within the region. Based on the current problems and comments from the stakeholders, the models could be adapted to represent target species of nature conservation organisations or investigate relationships that are relevant for policies. As prediction maps of SDMs present different types of uncertainties, it is important to communicate clearly to the stakeholders the exact measure that is uncertain, the underlying causes of this uncertainty and a means to express the level of uncertainty (Van Der Bles et al. 2019).

Conclusion

This study integrates LEs with woody and herbaceous vegetation and LEs with the presence of water (blue LEs) within SDMs. An improvement in model performance was found, when different types of LEs were included within the modelling framework. We found that SDMs that leave out information on LEs, can lead to inaccurate maps with overprediction of species presence within the agricultural fields and underprediction of species presence in the borders of the fields with LEs. To improve the accuracy of predictive distribution maps and SDMs, we recommend that modellers focusing on pollinators integrate herbaceous LEs, as these showed a statistically significant improvement in most cases. The blue and woody LEs could be integrated as distance to and their effect may be in a broader context of the landscape. If data on LEs is lacking, we suggest integrating the edges of the field as an approximate for the presence of LEs. This SDM study underlines the importance of integrating information on LEs in agricultural areas for model performance and more nuanced predictive maps, facilitating downstream application by different stakeholders.

Author contribution MM, ML, and JCB conceived the ideas and designed methodology; MM, PM, CG, KS, Vde VJ and PS collected the data; MM. analysed the data; MM, ML, JCB and RHAC led the writing of the manuscript. All authors contributed critically to the drafts and gave final approval for publication.

Funding The research was supported by the MAPPY project, which focuses on the multisectoral analysis of climate and land use change impacts on pollinators, plant diversity, and crop yields. MAPPY is part of AXIS, an ERA-NET established by JPI Climate. It receives funding from various agencies, including FFG (Austria), F.R.S.-FNRS and BELSPO (Belgium), DLR/BMBF (Germany), NWO (Netherlands), and AEI (Spain), with additional co-funding from the European Union under grant no. 776608. L.M. was supported by the projectVI. Veni.222.141 from the Dutch Research Council (NWO).

Data availability The datasets generated during and/or analysed during the current study are available in the Zenodo repository, [10.5281/zenodo.21870444] (https://doi.org/10.5281/zenodo.21870444).

Declarations

Competing interests All authors declare that they have no conflicts of interest related to this study.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

References

Aguirre-Gutiérrez J, Carvalheiro LG, Polce C et al (2013) Fit-for-purpose: species distribution model performance depends on evaluation criteria - Dutch hoverflies as a case study. PLoS ONE. https://doi.org/10.1371/journal.pone.0063708

Aiello-Lammens ME, Boria RA, Radosavljevic A et al (2015) spThin: an R package for spatial thinning of species occurrence records for use in ecological niche models. Ecography 38:541–545

Anderson RP, Lew D, Peterson AT (2003) Evaluating predictive models of species' distributions: criteria for selecting optimal models. Ecol Modell 162:211–232

Andrade C, Villers A, Balent G et al (2021) A real-world implementation of a nationwide, long-term monitoring program to assess the impact of agrochemicals and agricultural practices on biodiversity. Ecol Evol 11:3771–3793

Asch RG, Sobolewska J, Chan K (2022) Assessing the reliability of species distribution models in the face of climate and ecosystem regime shifts: small pelagic fishes in the California Current System. Front Mar Sci 9:711522

Backstrom LJ, Callaghan CT, Worthington H et al (2024) Estimating sampling biases in citizen science datasets. Ibis. https://doi.org/10.1111/ibi.13343

Baker DJ, Maclean IMD, Goodall M, Gaston KJ (2021) Species distribution modelling is needed to support ecological impact assessments. J Appl Ecol 58:21–26

Bartual AM, Sutter L, Bocci G et al (2019) The potential of different semi-natural habitats to sustain pollinators and natural enemies in European agricultural landscapes. Agric Ecosyst Environ 279:43–52

Baude M, Kunin WE, Boatman ND et al (2016) Historical nectar assessment reveals the fall and rise of floral resources in Britain. Nature 530:85–88

Beck J, Böller M, Erhardt A, Schwanghart W (2014) Spatial bias in the GBIF database and its effect on modeling species' geographic distributions. Ecol Inform 19:10–15

Bellamy C, Boughey K, Hawkins C et al (2020) A sequential multi-level framework to improve habitat suitability modelling. Landsc Ecol 35:1001–1020

Benjamini Y, Hochberg Y (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing. J R Stat Soc Ser B Stat Methodol 57:289–300

Billaud O, Vermeersch R, Porcher E (2021) Citizen science involving farmers as a means to document temporal trends in farmland biodiversity and relate them to agricultural practices. J Appl Ecol 58:261–273

Bishop GA, Kleijn D, Albrecht M et al (2025) Critical habitat thresholds for effective pollinator conservation in agricultural landscapes. Science 389(6767):1314–1319

Boakes EH, McGowan PJK, Fuller RA et al (2010) Distorted views of biodiversity: spatial and temporal bias in species occurrence data. PLoS Biol 8:e1000385

Boria RA, Olson LE, Goodman SM, Anderson RP (2014) Spatial filtering to reduce sampling bias can improve the performance of ecological niche models. Ecol Modell 275:73–77

Bos JFFP, Smit ABL, Schröder JJ (2013) Is agricultural intensification in the Netherlands running up to its limits? NJAS Wageningen J Life Sci 66:65–73

Bossard M, Feranec J, Otahel J (2000) CORINE land cover technical guide: addendum 2000. European Environment Agency, Copenhagen

Bowler DE, Bhandari N, Repke L et al (2022) Decision-making of citizen scientists when recording species observations. Sci Rep 12:11069

Bradie J, Leung B (2017) A quantitative synthesis of the importance of variables used in MaxEnt species distribution models. J Biogeogr 44:1344–1361

Bradley AP (1997) The use of the area under the ROC curve in the evaluation of machine learning algorithms. Pattern Recognit 30:1145–1159

Burnham KP, Anderson DR (2002) Model selection and multimodal inference: a practical information-theoretic approach., 2nd edition. Springer, New York

Cagliero R, Vassallo M, Pierangeli F et al (2023) The Common Agricultural Policy 2023–2027. How do member states implement the new delivery model? Italian Review of Agricultural Economics (REA) 78:49–66

Callaghan CT, Poore AGB, Mesaglio T et al (2021) Three frontiers for the future of biodiversity research using citizen science data. Bioscience 71:55–63

Central Bureau of Statistics (CBS) (2015) Hoe wordt de Nederlandse bodem gebruikt? In: Retrieved from: https://longreads.cbs.nl/nederland-in-cijfers-2020/hoe-wordt-de-nederlandse-bodem-gebruikt/

Claverie M, Chan A, See L et al (2026) EuroCrops v2. 0: Multi-annual harmonized parcel level crop type data linked to European Union-wide survey, statistical and Earth Observation products. Earth System Science Data Discussions 2026:1–31

Clough Y, Kirchweger S, Kantelhardt J (2020) Field sizes and the future of farmland biodiversity in European landscapes. Conserv Lett 13:e12752

Cobra adviseurs B.V. (2024) Data on landscape elements of the Netherlands.
Cole LJ, Brocklehurst S, Robertson D et al (2017) Exploring the interactions between resource availability and the utilisation of semi-natural habitats by insect pollinators in an intensive agricultural landscape. Agric Ecosyst Environ 246:157–167
Cole LJ, Kleijn D, Dicks LV et al (2020) A critical analysis of the potential for EU Common Agricultural Policy measures to support wild pollinators on farmland. J Appl Ecol 57:681–694
Copernicus Land Monitoring Service (2021) Small Woody Features 2018 and Small Woody Features Changes 2015–2018
Costa H, Foody GM, Jiménez S, Silva L (2015) Impacts of species misidentification on species distribution modeling with presence-only data. ISPRS Int J Geoinf 4:2496–2518
De Palma A, Kuhlmann M, Roberts SPM et al (2015) Ecological traits affect the sensitivity of bees to land-use pressures in European agricultural landscapes. J Appl Ecol 52:1567–1577
De Vries JPR, Koma Z, WallisDeVries MF, Kissling WD (2021) Identifying fine-scale habitat preferences of threatened butterflies using airborne laser scanning. Divers Distrib 27:1251–1264
Díaz-Calafat J, Jaume-Ramis S, Soacha K et al (2024) Revealing biases in insect observations: a comparative analysis between academic and citizen science data. PLoS ONE 19:e0305757
Dormann CF, Elith J, Bacher S et al (2013) Collinearity: a review of methods to deal with it and a simulation study evaluating their performance. Ecography 36:27–46
Dou Y, Cosentino F, Malek Z et al (2021) A new European land systems representation accounting for landscape characteristics. Landsc Ecol 36:2215–2234
Duan M, Xing J, Zheng X et al (2026) Enhancing species distribution models by considering dispersal ability. Ecol Evol 16:e73425
Egloff GB, Schenker L, Riverendo S et al (2025) The role of small structures for stoats in agricultural landscapes. J Nat Conserv 85:126865
Elith J, Leathwick JR (2009) Species distribution models: ecological explanation and prediction across space and time. Annu Rev Ecol Evol Syst 40:677–697. https://doi.org/10.1146/annurev.ecolsys.110308.120159
European Environment Agency (2021) High Resolution Layer Small Landscape Features
European Invertebrate Survey (EIS) (2020) https://www.eis-nederland.nl/
Falk S (2019) Field guide to the bees of Great Britain and Ireland. Bloomsbury Publishing
Fick SE, Hijmans RJ (2017) WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas. Int J Climatol 37:4302–4315
Fisher RA (1922) On the interpretation of χ 2 from contingency tables, and the calculation of P. J R Stat Soc 85:87–94
Frans VF, Augé AA, Fyfe J et al (2022) Integrated SDM database: enhancing the relevance and utility of species distribution models in conservation management. Methods Ecol Evol 13:243–261
Garratt MPD, Senapathi D, Coston DJ et al (2017) The benefits of hedgerows for pollinators and natural enemies depends on hedge quality and landscape context. Agric Ecosyst Environ 247:363–370
Geldmann J, Heilmann-Clausen J, Holm TE et al (2016) What determines spatial bias in citizen science? Exploring four recording schemes with different proficiency requirements. Divers Distrib 22:1139–1149
Guillera-Arroita G, Lahoz-Monfort JJ, Elith J et al (2015) Is my species distribution model fit for purpose? Matching data and models to applications. Glob Ecol Biogeogr 24:276–292
Guisan A, Tingley R, Baumgartner JB et al (2013) Predicting species distributions for conservation decisions. Ecol Lett 16:1424–1435
Guisan A, Zimmermann NE (2000) Predictive habitat distribution models in ecology. Ecol Modell 135:147–186
Hazeu GW, Vittek M, Schuiling R, et al (2020) LGN2018: a new representation of land use in the Netherlands.
Hijmans RJ, Phillips S, Leathwick J, Elith J (2023) dismo: Species distribution modeling. R Package Version 3:1
Hirzel AH, Le Lay G, Helfer V et al (2006) Evaluating the ability of habitat suitability models to predict species presences. Ecol Modell 199:142–152
Husby M, Hoset KS, Butler S (2021) Non-random sampling along rural–urban gradients may reduce reliability of multi-species farmland bird indicators and their trends. Ibis 163:579–592
Jarnevich C, Holcombe T, Grisham B et al (2016) Assessing range-wide habitat suitability for the Lesser Prairie-Chicken. Avian Conserv Ecol 11(1):2. https://doi.org/10.5751/ACE-00807-110102
Jiménez L, Soberón J (2020) Leaving the area under the receiving operating characteristic curve behind: an evaluation method for species distribution modelling applications based on presence-only data. Methods Ecol Evol 11:1571–1586
Johnson CJ, Gillingham MP (2005) An evaluation of mapped species distribution models used for conservation planning. Environ Conserv 32:117–128
Johnston A, Moran N, Musgrove A et al (2020) Estimating species distributions from spatially biased citizen science data. Ecol Modell 422:108927
Jungandreas A, Roilo S, Strauch M et al (2022) Response of endangered bird species to land-use changes in an agricultural landscape in Germany. Reg Environ Change 22:19
Kadaster (2020) Basisregistratie Grootschalige Topografie (BGT) . Publieke Dienstverlening Op de Kaart (PDOK)
Kaky E, Nolan V, Alatawi A, Gilbert F (2020) A comparison between ensemble and MaxEnt species distribution modelling approaches for conservation: a case study with Egyptian medicinal plants. Ecol Inform 60:101150
Kalle R, Ramesh T, Qureshi Q, Sankar K (2013) Predicting the distribution pattern of small carnivores in response to environmental factors in the Western Ghats. PLoS ONE 8:e79295
Kass JM, Muscarella R, Galante PJ et al (2021) ENMeval 2.0: redesigned for customizable and reproducible

modeling of species' niches and distributions. Methods Ecol Evol 12:1602–1608
Kleijn D, Winfree R, Bartomeus I et al (2015) Delivery of crop pollination services is an insufficient argument for wild pollinator conservation. Nat Commun 6:1–9
KNMI (2023) KNMI Data Platform: https://dataplatform.knmi.nl/
Kozelová I, Špulerová J, Miklósová V et al (2020) The role of artificial ditches and their buffer zones in intensively utilized agricultural landscape. Environ Monit Assess 192:656
Kramer-Schadt S, Niedballa J, Pilgrim JD et al (2013) The importance of correcting for sampling bias in MaxEnt species distribution models. Divers Distrib 19:1366–1379
Kremen C, Ullman KS, Thorp RW (2011) Evaluating the quality of citizen-scientist data on pollinator communities. Conserv Biol 25:607–617
Lahoz-Monfort JJ, Guillera-Arroita G, Wintle BA (2014) Imperfect detection impacts the performance of species distribution models. Glob Ecol Biogeogr 23:504–515
Lamboley Q, Fourcade Y (2024) No optimal spatial filtering distance for mitigating sampling bias in ecological niche models. J Biogeogr 51:1783–1794
Langlois A, Jacquemart A-L, Piqueray J (2020) Contribution of extensive farming practices to the supply of floral resources for pollinators. InSects 11:818
Lee-Yaw JA, McCune JL, Pironon S, Sheth SN (2022) Species distribution models rarely predict the biology of real populations. Ecography 2022:e05877
Liu C, Newell G, White M (2019) The effect of sample size on the accuracy of species distribution models: considering both presences and pseudo-absences or background sites. Ecography 42:535–548
Lobo JM, Jiménez-Valverde A, Real R (2008) AUC: a misleading measure of the performance of predictive distribution models. Glob Ecol Biogeogr 17:145–151
Marshall L (2018) Wild bee diversity across space and time: the role of land use/land cover and climate
Marshall L, Beckers V, Vray S et al (2021) High thematic resolution land use change models refine biodiversity scenarios: a case study with Belgian bumblebees. J Biogeogr 48:345–358
Marshall L, Carvalheiro LG, Aguirre-Gutiérrez J et al (2015) Testing projected wild bee distributions in agricultural habitats: predictive power depends on species traits and habitat type. Ecol Evol 5:4426–4436
Martin Y, Van Dyck H, Dendoncker N, Titeux N (2013) Testing instead of assuming the importance of land use change scenarios to model species distributions under climate change. Glob Ecol Biogeogr 22:1204–1216
Melo-Merino SM, Reyes-Bonilla H, Lira-Noriega A (2020) Ecological niche models and species distribution models in marine environments: a literature review and spatial analysis of evidence. Ecol Modell 415:108837
Merow C, Smith MJ, Silander JA (2013) A practical guide to MaxEnt for modeling species' distributions: what it does, and why inputs and settings matter. Ecography 36:1058–1069
Meyer C, Kreft H, Guralnick R, Jetz W (2015) Global priorities for an effective information basis of biodiversity distributions. Nat Commun 6:1–8
Miller JA, Holloway P (2015) Incorporating movement in species distribution models. Prog Phys Geogr 39:837–849
Mod HK, Scherrer D, Luoto M, Guisan A (2016) What we use is not what we know: environmental predictors in plant distribution models. J Veg Sci 27:1308–1322
Morelli F (2013) Relative importance of marginal vegetation (shrubs, hedgerows, isolated trees) surrogate of HNV farmland for bird species distribution in Central Italy. Ecol Eng 57:261–266
Moudrý V, Remelgado R, Forkel M et al (2026) Spaceborne canopy height products should be complemented with airborne laser scanning data: toward a European canopy height model. Earth Space Sci 13:e2025EA004544
Muscarella R, Galante PJ, Soley-Guardia M et al (2014) ENM eval: an R package for conducting spatially independent evaluations and estimating optimal model complexity for Maxent ecological niche models. Methods Ecol Evol 5:1198–1205
Musters CJM, Van Alebeek F, Geers R et al (2009) Development of biodiversity in field margins recently taken out of production and adjacent ditch banks in arable areas. Agric Ecosyst Environ 129:131–139
O'Connell J, Bradter U, Benton TG (2015) Wide-area mapping of small-scale features in agricultural landscapes using airborne remote sensing. ISPRS J Photogramm Remote Sens 109:165–177
ON Technologies (2023) Native Barcoding Kit 24 V14 SQK-NBD114.24. In: Accessed from: https://store.nanoporetech.com/eu/native-barcoding-kit-24-v14.html
ON Technologies (2024) Ligation sequencing DNA V14 - dual barcoding . In: Accessed from: https://nanoporetech.com/document/ligation-sequencing-dual-barcoding-v14
Pebesma EJ (2018) Simple features for R: standardized support for spatial vector data. R J 10:439
Peterson AT, Papeş M, Soberón J (2008) Rethinking receiver operating characteristic analysis applications in ecological niche modeling. Ecol Modell 213:63–72
Phillips SJ, Anderson RP, Schapire RE (2006) Maximum entropy modeling of species geographic distributions. Ecol Modell 190:231–259
Phillips SJ, Dudík M (2008) Modeling of species distributions with Maxent: new extensions and a comprehensive evaluation. Ecography 31:161–175
Prendergast KS, Dixon KW, Bateman PW (2022) A global review of determinants of native bee assemblages in urbanised landscapes. Insect Conserv Divers 15:385–405
Pustkowiak S, Kwieciński Z, Lenda M et al (2021) Small things are important: the value of singular point elements for birds in agricultural landscapes. Biol Rev Camb Philos Soc 96:1386–1403
R Core Team (2024) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. Version 4.5. 0
Raes N, ter Steege H (2007) A null-model for significance testing of presence-only species distribution models. Ecography 30:727–736
Reemer M, Renema W, Van Steenis W, et al (2009) De Nederlandse Zweefvliegen (Diptera: Syrphidae). Nederlandse

Fauna 8. Nationaal Natuurhistorisch Museum Naturalis, Leiden
Remelgado R, Levers C, Cord AF (2025) Cultivated lands: blind spots in global biodiversity data. Trends Ecol Evol. https://doi.org/10.1016/j.tree.2025.09.017
Reverté S, Miličić M, Ačanski J et al (2023) National records of 3000 European bee and hoverfly species: a contribution to pollinator conservation. Insect Conserv Divers 16:758–775
RIVM (2022) Digitale Atlas Natuurlijk Kapitaal: Lage vegetatie in Nederland
Roberts DR, Bahn V, Ciuti S et al (2017) Cross-validation strategies for data with temporal, spatial, hierarchical, or phylogenetic structure. Ecography 40:913–929
Roilo S, Engler JO, Václavík T, Cord AF (2023) Landscape-level heterogeneity of agri-environment measures improves habitat suitability for farmland birds. Ecol Appl 33:e2720
Roilo S, Paulus A, Alarcón-Segura V et al (2024) Quantifying agricultural land-use intensity for spatial biodiversity modelling: implications of different metrics and spatial aggregation methods. Landsc Ecol 39:55
Rosario I, Tiago P, Chozas S, Capinha C (2024) Drivers of temporal bias in biodiversity recording by citizen scientists. bioRxiv 2024–2027
Santos MJ, Smith AB et al (2021) The role of land use and land cover change in climate change vulnerability assessments of biodiversity: a systematic review. Landsc Ecol 36(12):3367–3382
Schank CJ, Cove MV, Kelly MJ et al (2019) A sensitivity analysis of the application of integrated species distribution models to mobile species: a case study with the endangered Baird's Tapir. Environ Conserv 46:184–192
Scherber C, Beduschi T, Tscharntke T (2021) A grid-based sampling approach to insect biodiversity monitoring in agricultural landscapes. Exploring and optimizing agricultural landscapes. Springer, pp 415–424
Schweiger O, Dormann CF, Bailey D, Frenzel M (2006) Occurrence pattern of *Pararge aegeria* (Lepidoptera: Nymphalidae) with respect to local habitat suitability, climate and landscape structure. Landsc Ecol 21:989–1001
Shaw A, Steelman T, Bullock R (2022) Evaluating the efficacy of GIS maps as boundary objects: unpacking the limits and opportunities of Indigenous knowledge in forest and natural resource management. J Cult Geogr 39:90–116
Soberon J, Peterson AT (2005) Interpretation of models of fundamental ecological niches and species' distributional areas
Sofaer HR, Jarnevich CS, Pearse IS et al (2019) Development and delivery of species distribution models to inform decision-making. Bioscience 69:544–557
Sullivan MJP, Pearce-Higgins JW, Newson SE et al (2017) A national-scale model of linear features improves predictions of farmland biodiversity. J Appl Ecol 54:1776–1784
Sydenham MAK, Nielsen A, Dupont YL et al (2025) Role of forest edges and other seminatural linear landscape features in structuring wild bee habitat connectivity in intensively managed landscapes. Conserv Biol. https://doi.org/10.1111/cobi.70152
Ten Caten C, Dallas T (2023) Thinning occurrence points does not improve species distribution model performance. Ecosphere 14:e4703
Thuiller W, Mucher S, Hennekens S, Si-Moussi S (2025) EUNIS Habitat Maps: Enhancing Thematic and Spatial Resolution for Europe through Machine Learning. Sci Data
Tóth R, Czeglédi I, Kern B, Erős T (2019) Land use effects in riverscapes: diversity and environmental drivers of stream fish communities in protected, agricultural and urban landscapes. Ecol Indic 101:742–748
Troudet J, Grandcolas P, Blin A et al (2017) Taxonomic bias in biodiversity data and societal preferences. Sci Rep 7:9132
Van Der Bles AM, Van Der Linden S, Freeman ALJ et al (2019) Communicating uncertainty about facts, numbers and science. R Soc Open Sci 6:181870
Van Der Zanden EH, Verburg PH, Mücher CA (2013) Modelling the spatial distribution of linear landscape elements in Europe. Ecol Indic 27:125–136
Vanneste T, Govaert S, Spicher F et al (2020) Contrasting microclimates among hedgerows and woodlands across temperate Europe. Agric for Meteorol 281:107818
von Königslöw V, Fornoff F, Klein A (2022) Wild bee communities benefit from temporal complementarity of hedges and flower strips in apple orchards. J Appl Ecol 59:2814–2824
Wang L, Jackson DA (2023) Effects of sample size, data quality, and species response in environmental space on modeling species distributions. Landsc Ecol 38:4009–4031
Wassen MJ, Runhaar H, Barendregt A, Okruszko T (2011) Evaluating the role of participation in modeling studies for environmental planning. Environ Plann B Plann des 38:338–358
Weber MM, Stevens RD, Diniz-Filho JAF, Grelle CEV (2017) Is there a correlation between abundance and environmental suitability derived from ecological niche modelling? A meta-analysis. Ecography 40:817–828
Wilcoxon F (1992) Individual comparisons by ranking methods. Springer
Zurell D (2017) Integrating demography, dispersal and interspecific interactions into bird distribution models. J Avian Biol 48:1505–1516
Zurell D, Franklin J, König C et al (2020) A standard protocol for reporting species distribution models. Ecography 43:1261–1277

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.