

## RESEARCH ARTICLE

# Identifying fine-scale habitat preferences of threatened butterflies using airborne laser scanning

Jan Peter Reinier de Vries<sup>1</sup> | Zsófia Koma<sup>1</sup> | Michiel F. WallisDeVries<sup>2,3</sup> |  
W. Daniel Kissling<sup>1</sup> 

<sup>1</sup>Institute of Biodiversity and Ecosystem Dynamics (IBED), University of Amsterdam, Amsterdam, The Netherlands

<sup>2</sup>De Vlinderstichting/Dutch Butterfly Conservation, Wageningen, The Netherlands

<sup>3</sup>Plant Ecology and Nature Conservation Group, Wageningen University, Wageningen, The Netherlands

## Correspondence

W. Daniel Kissling, Institute of Biodiversity and Ecosystem Dynamics (IBED), University of Amsterdam, Amsterdam, The Netherlands.  
Email: wdkissling@gmail.com

## Funding information

Netherlands eScience Center, Grant/Award Number: ASDI.2016.014

Editor: Reid Tingley

## Abstract

**Aim:** Light Detection And Ranging (LiDAR) is a promising remote sensing technique for ecological applications because it can quantify vegetation structure at high resolution over broad spatial extents. Using country-wide airborne laser scanning (ALS) data, we test to what extent fine-scale LiDAR metrics capturing low vegetation, medium-to-high vegetation and landscape-scale habitat structures can explain the habitat preferences of threatened butterflies at a national extent.

**Location:** The Netherlands.

**Methods:** We applied a machine-learning (random forest) algorithm to build species distribution models (SDMs) for grassland and woodland butterflies in wet and dry habitats using various LiDAR metrics and butterfly presence-absence data collected by a national butterfly monitoring scheme. The LiDAR metrics captured vertical vegetation complexity (e.g., height and vegetation density of different strata) and horizontal heterogeneity (e.g., vegetation roughness, microtopography, vegetation openness and woodland edge extent). We assessed the relative variable importance and interpreted response curves of each LiDAR metric for explaining butterfly occurrences.

**Results:** All SDMs showed a good to excellent fit, with woodland butterfly SDMs performing slightly better than those of grassland butterflies. Grassland butterfly occurrences were best explained by landscape-scale habitat structures (e.g., open patches, microtopography) and vegetation height. Woodland butterfly occurrences were mainly determined by vegetation density of medium-to-high vegetation, open patches and woodland edge extent. The importance of metrics generally differed between wet and dry habitats for both grassland and woodland species.

**Main conclusions:** Vertical variability and horizontal heterogeneity of vegetation structure are key determinants of butterfly species distributions, even in low-stature habitats such as grasslands, dunes and heathlands. The information content of low vegetation LiDAR metrics could further be improved with country-wide leaf-on ALS data or surveys from drones and terrestrial laser scanners at specific sites. LiDAR

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.

© 2021 The Authors. *Diversity and Distributions* published by John Wiley & Sons Ltd.

thus offers great potential for predictive habitat distribution modelling and other studies on ecological niches and invertebrate–habitat relationships.

#### KEYWORDS

active remote sensing, ecological niche, ecosystem structure, environmental heterogeneity, essential biodiversity variables, habitat suitability, insects, landscape ecology, microhabitat

## 1 | INTRODUCTION

Butterflies and other invertebrates have declined severely in recent decades, especially in parts of Europe where structured monitoring schemes have revealed long-term population declines (Hallmann et al., 2017; van Swaay et al., 2006). The specialized niches of many butterflies in terms of habitat and food plant requirements make them vulnerable to ongoing habitat modification and other global change drivers (Thomas et al., 2004). Butterflies are generally a well-studied organism group; they are diverse and often bound to specific habitats and hence a very good indicator and umbrella taxon for invertebrate conservation (van Swaay et al., 2006; Thomas, 2005). Comprehensive survey efforts have especially revealed severe population declines and extinctions of specialist species, for example in the Netherlands (Bos et al., 2006; van Strien et al., 2019), Flanders (Maes & Van Dyck, 2001), Denmark (Eskildsen et al., 2015) and Great Britain (Fox et al., 2015). In the Netherlands, butterflies have declined by 50% since 1992 and over 80% since 1890 (van Strien et al., 2019). The major causes of these declines have been the intensification of human land use, the modification of heterogeneous (semi-)natural landscapes and an increase in habitat fragmentation (e.g., Aguirre-Gutiérrez et al., 2017; Thomas et al., 2004; van Strien et al., 2019; van Swaay et al., 2006). Although a reduction of landscape conversion and an increase in conservation efforts have slowed down butterfly declines since 1990 (Carvalho et al., 2013; van Strien et al., 2016), a large part of the Dutch butterfly species remain highly vulnerable and are still declining (van Swaay, 2019; van Strien et al., 2019). This shows the urgent need of sustaining and increasing efforts to preserve butterflies and their habitats.

The preservation of habitats is of critical importance to prevent further losses and declines of butterflies and other invertebrates (van Swaay et al., 2006; van Strien et al., 2019). As most invertebrates depend on specific habitat elements that provide food resources, nesting sites and shelter, understanding how the fine-scale structure and distribution of habitats determine species distributions is crucial for biodiversity science and conservation (Dennis et al., 2003, 2006; Thomas, 1995). Habitat structure has also many indirect effects on invertebrates, for example by influencing microclimate, light availability and floristic composition (Aguirre-Gutiérrez et al., 2017; Davies & Asner, 2014; Müller et al., 2014). The fine-scale habitat suitability of invertebrates is typically driven by various aspects of vegetation structure, including vertical vegetation complexity (e.g., the density of specific strata), horizontal heterogeneity (e.g., canopy roughness) or the horizontal structure of vegetation at the landscape scale (e.g., the extent of edges and open spaces; Bakx et al., 2019;

Davies & Asner, 2014; Glad et al., 2020; Simonson et al., 2014). Despite many local field studies on butterfly–habitat relationships, the generality of these relationships remains unclear because quantifying vegetation structure across broad spatial extents has often been limited by the difficulty to obtain detailed, high-resolution data in a standardized, comparable and spatially contiguous way (Davies & Asner, 2014; Kissling et al., 2017; Valbuena et al., 2020). Moreover, the development of standardized and spatially contiguous variables and datasets of ecosystem height, cover and vegetation structural complexity covering broad spatial extents is only recently becoming an important focus of biodiversity science and monitoring, for example in the context of essential biodiversity variables (EBVs; Valbuena et al., 2020).

Active remote sensing techniques such as Light Detection And Ranging (LiDAR) can produce standardized 3D measurements of vegetation structure (e.g., “ecosystem structure EBVs”) at high resolution and over broad spatial extents, with relatively low costs (Davies & Asner, 2014; Kissling et al., 2017; Valbuena et al., 2020). LiDAR data derived from country-wide airborne laser scanning (ALS) are also increasingly becoming available from free and open sources (Valbuena et al., 2020). LiDAR uses short-range laser pulses to measure the x,y,z-coordinates of reflective objects, often from aircrafts. As the exact timing and position of the sensor on the airplane are known, the distance to each point can be calculated and a 3D point cloud with high precision can be derived, from which a large number of vegetation structure parameters can be calculated (Bakx et al., 2019; Davies & Asner, 2014). These parameters—often referred to as LiDAR metrics—are statistical properties of the point cloud describing the mean, variability or proportions of returns for vertical strata. They can capture information on vegetation structure at a local scale (e.g., for a high-resolution grid cell or a radius around a focal observation point) or at the landscape scale (e.g., measuring habitat patches and edges based on grid cells that capture LiDAR-derived vegetation height; Bakx et al., 2019). LiDAR metrics can thus directly be used to quantify ecological niches and habitat requirements of species, for example tree lines and other linear vegetation elements in open landscapes (Lucas et al., 2019) or climatic and other environmental gradients controlled by microtopography and vegetation structure (Zellweger et al., 2019). This makes LiDAR a transformative resource for ecological studies, enabling a detailed understanding of the specific and scale-dependent habitat preferences of species across broad spatial extents, and with direct insights for management, policy and conservation (Davies & Asner, 2014; Moeslund et al., 2019; Müller & Brandl, 2009; Simonson et al., 2014; Valbuena et al., 2020).

**TABLE 1** Summary of the four butterfly species and their habitats, their Red List status and the number of presences, absences and LiDAR metrics included in the statistical analysis

	<i>Boloria selene</i>	<i>Hipparchia semele</i>	<i>Limenitis camilla</i>	<i>Melitaea athalia</i>
Common name	Small pearl-bordered fritillary	Grayling	White admiral	Heath fritillary
Habitat type	Wet grassland	Dry grassland	Moist woodland	Dry woodland
Habitat description	Moist flower-rich grasslands with scattered trees and shrubs, mainly on peat soils	Dry open grasslands with bare sand, low grass tufts and scattered woody vegetation, sandy soils	Moist deciduous woodlands with canopy gaps allowing sunlight penetration, mainly on sandy or loamy (till) soils	Dry and sheltered open spaces with a flower-rich herb vegetation near edges of woody vegetation, sandy soils
Red List status	EN	VU	EN	CR
Presence records	151 (inland: 137; coast: 14)	610 (inland: 181; coast: 429)	384	101
Absence records	92 (inland: 71; coast: 21)	248 (inland: 122; coast: 126)	106	45
Number of LiDAR metrics	7	7	7	6

Note: Information on habitat preferences is derived from ecological field studies (Table S1) and butterfly presences on specific soil types (Table S2). Status on the Dutch Red List of threatened butterflies (van Swaay, 2019) reflects the categories "Endangered" (EN), "Vulnerable" (VU) and "Critically Endangered" (CR). Presence and absence reflect the number of records included in the species distribution models. The specific LiDAR metrics included in the species distribution models of each butterfly species are provided in Table 2.

Few LiDAR studies have so far focussed on invertebrates (Davies & Asner, 2014). Such studies often show that species diversity increases with an increase in vegetation structural diversity and landscape heterogeneity (Davies & Asner, 2014). Moreover, they also show that individual taxa respond differently to specific habitat characteristics and that the performance of habitat models can increase when LiDAR metrics are included (Davies & Asner, 2014; Hess et al., 2013; Vierling et al., 2011). As most LiDAR studies have focussed on forests and woody habitats (Bakx et al., 2019; Davies & Asner, 2014), it remains open to what extent LiDAR can capture vegetation structure of low-stature habitats such as grasslands, dunes and wetlands. Some previous studies show promising results for measuring 3D vegetation structure in grasslands and wetlands (e.g., Alexander et al., 2015; Koma et al., 2020; Zlinszky et al., 2014). However, country-wide LiDAR surveys are often conducted in the leaf-off season to optimize terrain mapping (Reutebuch et al., 2005). As a consequence, such LiDAR data may contain little information for quantifying the vertical structure of vegetation within low-stature habitats (Alexander et al., 2015). On the other hand, measuring vegetation structure with leaf-on data in woodlands can also be challenging because laser returns might predominantly be recorded from the canopy, especially with discrete return data (Anderson et al., 2016). Comparing the explanatory power and information content of a suite of LiDAR metrics in open and woody habitats is thus important to better understand the potential of LiDAR data for ecological research and nature management (Davies & Asner, 2014).

Here, we analyse to what extent specific LiDAR metrics derived from country-wide ALS data can explain the fine-scale habitat preferences of threatened butterflies in the Netherlands. We focus on four species that are all of conservation concern (van Swaay, 2019) and which are bound to specific habitats, representing grassland and woodland habitats in wet or dry conditions (Table 1). We build species distribution models (SDMs) with LiDAR metrics that capture the vertical complexity and horizontal heterogeneity of vegetation, and use species presence-absence data derived from a national butterfly monitoring scheme (van Swaay et al., 2008) as the response variable. We expect that (H1) LiDAR metrics reflecting low vegetation (e.g., forest understorey or grasses and herbs in open habitats) show little importance in explaining habitat preferences of butterflies because of the limitations of LiDAR data in dense forests (due to the low penetrability of the canopy) or in grasslands (due to the leaf-off acquisition of LiDAR data), (H2) metrics reflecting medium-to-high vegetation (e.g., the density or heterogeneity of shrub and tree layers) are especially important to explain habitat preferences of woodland butterflies, and (H3) metrics reflecting landscape-scale habitat structures (e.g., microtopography, woodland edges and vegetation openness) are important to explain habitat preferences of both grassland and woodland butterflies. Our analyses gain new insights into the generality of habitat preferences of butterflies at a national extent, and how LiDAR metrics capturing 3D vegetation structure at local and landscape scales can improve our yet limited knowledge on invertebrate-habitat relationships.

## 2 | METHODS

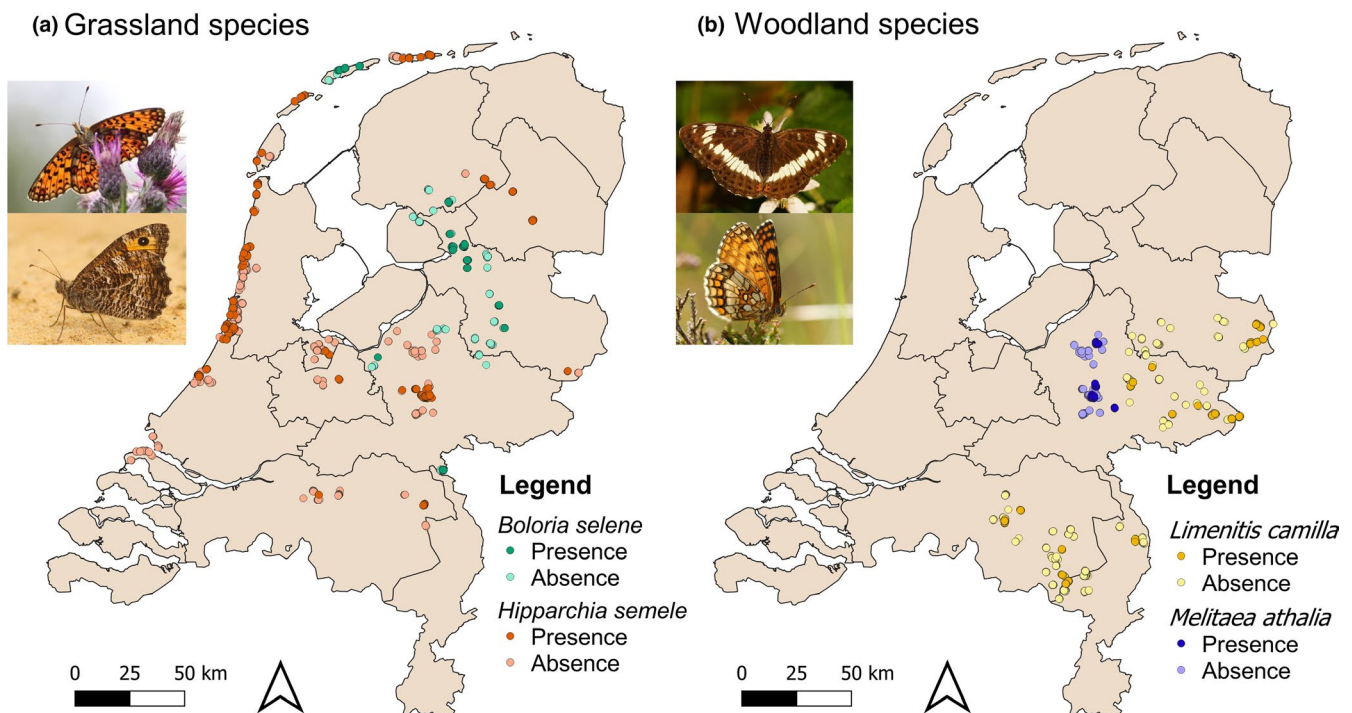
### 2.1 | Butterfly data

We focus on four threatened butterfly species with different habitat preferences (Table 1): (1) the small pearl-bordered fritillary (*Boloria selene*), a specialist of wet grasslands with a low and flower-rich vegetation (Bergman et al., 2008; Bos et al., 2006; Cozzi et al., 2008; van Swaay, 2019); (2) the grayling (*Hipparchia semele*), a species inhabiting dry open habitats with a heterogeneous cover of bare sand, low grasses, nectar plants and scattered woody vegetation (Bos et al., 2006; van Swaay, 2019; Vanreusel et al., 2007); (3) the white admiral (*Limenitis camilla*), a butterfly of moist woodlands with open patches providing sunlight throughfall (Bos et al., 2006; van Swaay, 2019); and (4) the heath fritillary (*Melitaea athalia*), a dry woodland species which is mostly found on sheltered open spaces with a flower-rich herb vegetation near woodland edges (Bergman et al., 2008; Bos et al., 2006; van Swaay, 2019; Warren, 1987a, 1987b). Details on their specific habitat preferences as derived from (mostly local) ecological field studies are summarized in Table S1. All four species have a localized distribution in the Netherlands and have strongly declined over the last century (van Swaay, 2019; van Strien et al., 2019). Two species (*L. camilla* and *M. athalia*) are confined to regions in the east and centre of the Netherlands, whereas the other

two (*B. selene* and *H. semele*) occur on both inland and coastal (dune) locations (Figure 1).

Presence-absence data of all four species were derived from the Dutch butterfly monitoring scheme, which systematically collects butterfly occurrence data by conducting weekly surveys along fixed transect routes throughout the flight season (April to September; van Swaay et al., 2008). Each transect route is about 1 km long and consists of smaller (50 m long) sections, each placed in a homogeneous habitat type. The fieldworkers record all butterflies 2.5 m to their right, 2.5 m to their left, 5 m ahead of them and 5 m above them. We used the nation-wide butterfly monitoring data from 2014 to 2018, in correspondence with the LiDAR data collection period (winter 2014–winter 2019). This comprised a total of >10,000 transect sections across the Netherlands, from which the focal species were recorded in 371 (*B. selene*), 807 (*H. semele*), 369 (*L. camilla*) and 119 (*M. athalia*) sections, respectively. The recorded presence and absence of each species was assigned to the centre point of each 50-m-long transect section and later used as the response variable in the SDMs (see below).

As the number of individuals of all monitored butterfly species is also recorded per transect section, we used this information not only to identify absences but also incidental records. Presence points for which only one individual was observed during all 2014–2018 surveys were excluded, as these records could represent misidentifications or



**FIGURE 1** Spatial distribution of presence and absence points of four grassland and woodland butterflies in the Netherlands. (a) Two grassland species, namely the small pearl-bordered fritillary (*Boloria selene*; green) and the grayling (*Hipparchia semele*; orange). (b) Two woodland species, namely the white admiral (*Limenitis camilla*; yellow) and the heath fritillary (*Melitaea athalia*; blue). Presences are given in bold colours, absences in light colours. Distributional overlap does not occur between the two woodland species but occasionally between the two grassland species, for example on the island of Terschelling (northwest cluster of *B. selene*), where *H. semele* also occurs. Photo credits: (a) top: *Boloria selene* (Michiel F. WallisDeVries), bottom: *Hipparchia semele* (Chris A. van Swaay), (b) top: *Limenitis camilla* (Chris A. van Swaay), bottom: *Melitaea athalia* (Chris A. van Swaay)

TABLE 2 LiDAR metrics capturing different aspects of the vertical complexity and horizontal heterogeneity of vegetation

Metric (unit)	Description	Vegetation	Hypothesis	Inclusion	References <sup>a</sup>
<b>Vertical complexity (25 m radius)</b>					
<0.2 m density (%)	Vegetation density as proportion of returns below 0.2 m relative to all vegetation and ground points	Herb/grass layer	H1 (low vegetation)	All	1; 2; 4; 8; 9; 10; 12; 16
0.2–1 m density (%)	Vegetation density as proportion of returns between 0.2 and 1 m relative to all vegetation and ground points	Tall herbs/low shrubs layer	H1 (low vegetation)	H	2; 4; 8; 9; 12; 16
1–5 m density (%)	Vegetation density as proportion of returns between 1–5 m relative to all vegetation and ground points	Shrub layer	H2 (medium-to-high vegetation)	All	1; 2; 4; 9; 12; 16
5–20 m density (%)	Vegetation density as proportion of returns between 5 and 20 m relative to all vegetation and ground points	Tree layer	H2 (medium-to-high vegetation)	L, M	1; 2; 4; 6; 8; 9; 10; 12; 13; 16
>20 m density (%)	Vegetation density as proportion of returns between >20 m relative to all vegetation and ground points	Canopy trees	H2 (medium-to-high vegetation)	All	2; 4; 6; 9; 13; 16
Height (m)	90 <sup>th</sup> percentile of normalized height of vegetation points	Tall vegetation	H2 (medium-to-high vegetation)	B, H	1; 2; 9; 12; 16; 17; 19
<b>Horizontal heterogeneity</b>					
Total veg. roughness (m)	Roughness of total vegetation DSM (maximum difference in total vegetation height between focal and 8 neighbouring cells, averaged across all 1 m cells in 25 m radius)	Total vegetation	H2 (medium-to-high vegetation)	–	2; 7; 9; 14; 19
Low veg. roughness (m)	Roughness of low vegetation DSM (maximum difference in vegetation height <1 m between focal and 8 neighbouring cells, averaged across all 1 m cells in 25 m radius)	Low vegetation	H1 (low vegetation)	–	2; 7; 9; 14; 19
Slope (°)	Mean slope derived from DTM using maximum ground height difference between focal and 8 neighbouring 1 m cells in 100 m radius	Microtopography	H3 (landscape-scale habitat structure)	All	5; 7; 16; 17; 18
Open area (ha)	Total low vegetation area (cells with 90 <sup>th</sup> percentile height of vegetation <1 m) in the DSM in 100 m radius	Extent of open vegetation	H3 (landscape-scale habitat structure)	B	1; 2; 3; 9; 11; 13; 15; 18
Open patches (count)	Number of patches of connected low vegetation (90 <sup>th</sup> percentile height <1 m) cells separated by other cells in the DSM in 100 m radius	Patchiness of open areas	H3 (landscape-scale habitat structure)	H, L	2; 9; 11; 15; 18
Edge extent (m)	Length of the edges between interfacing low (90 <sup>th</sup> percentile height <1 m) and non-low vegetation cells in the DSM in 100 m radius	Extent of woodland edges	H3 (landscape-scale habitat structure)	B, L, M	2; 3; 9; 10; 11; 15; 18

**Note:** For each metric, a description, the vegetation part, relation to hypotheses and inclusion in the modelling is provided. Vertical complexity metrics were directly derived from LiDAR point clouds within a 25 m radius. Horizontal heterogeneity metrics were derived in a 25 m radius (vegetation roughness) or 100 m radius (landscape-scale microtopography or vegetation structure), either from a 1 m resolution digital surface model (DSM) based on the 90th percentile height of vegetation (vegetation structure metrics), or from a 1 m digital terrain model (DTM) based on the minimum height of the ground points in each cell (slope). Normalization of the point cloud was carried out on a local surface based on an inverse-distance weighting (IDW) interpolation of the 10 nearest ground points. The inclusion column indicates which metrics were finally included in the species distribution models for each species (All = all species; B = *B. selene*; H = *H. semele*; L = *L. camilla*; M = *M. athalia*). References 1–10 refer to ecological field studies and 11–18 to LiDAR studies.

<sup>a</sup> References: 1 = Bergman et al. (2008), 2 = Bos et al. (2006), 3 = Cozzi et al. (2008), 4 = Dennis et al. (2006), 5 = Karlsson and Wiklund (2005), 6 = Lederer (1960), 7 = Maes et al. (2006), 8 = Vanreusel et al. (2007), 9 = van Swaay (2019), 10 = Warren (1987a), 11 = Warren (1987b), 12 = Aguirre-Gutiérrez et al. (2017), 13 = Glad et al. (2020), 14 = Graham et al. (2019), 15 = Hesselbarth et al. (2019), 16 = Moeslund et al. (2019), 17 = Müller and Brandl (2009), 18 = Zellweger et al. (2013), 19 = Zlinszky et al. (2014).



wandering individuals. For the analyses, we only included absences in a 10 km buffer around presence points to account for the limited mobility of the species (Essens et al., 2017). This selection was done using QGIS 3.4 (QGIS Development Team, 2009). We further identified the soil type of each transect section—a key determinant of vegetation and thus an indirect driver of butterfly distributions—using a national soil classification (Wösten et al., 1988). This soil classification distinguishes 21 soil types and additionally includes three classes that contain either information on land cover (“water” and “urban”) or no information (“zero,” i.e., information not available). We used the geospatial layer of this national soil classification (provided by Wösten et al., 2013) and included, for each butterfly species, absence points from those soil types that also hosted presence points (Table S2). We further excluded the land cover classes “water” and “urban” as well as undefined grid cells (“zero”) because they do not represent specific soil types or key habitats of the focal species. Selecting absence points that are in principle reachable for the focal species (10 km radius) and potentially suitable given the abiotic environment (soil conditions), but yet remain unoccupied, enables to identify the effect of vegetation structure on the presence-absence of the species (Zellweger et al., 2013). We note, however, that vegetation and soils might not be the only drivers of species absence, especially for rare species (e.g., *M. athalia*) where a low population size may not allow individuals to occupy all suitable habitats.

To reduce the spatial clustering of data points induced by the transect sampling design, we discarded presence and absence points that were located within 100-m distance from their nearest neighbour (compare Zielewska-Büttner et al., 2018). A 100-metre distance was chosen because it roughly represents the home range of the focal species (Bos et al., 2006; Essens et al., 2017). We used the thinning optimization algorithm “spThin” (Aiello-Lammens et al., 2015) in R 3.5.3 (R Core Team, 2019) with 1,000 repetitions per species to derive the maximum number of data points given the 100 m distance criterion. This resulted in a final sample size of 248 presence and 610 absence points for *H. semele*, 106 presence and 384 absence points for *L. camilla*, 92 presence and 151 absence points for *B. selene* and 45 presence and 101 absence points for *M. athalia* (Figure 1).

## 2.2 | LiDAR data

We used LiDAR data from the third country-wide ALS campaign (AHN3) in the Netherlands (see <https://ahn.arcgisonline.nl/ahnviewer>), conducted in the years 2014–2019 in leaf-off conditions (Northern Hemisphere winter, December–March). The data have an average point density of 6–10 points per m<sup>2</sup>, an overall point cloud accuracy of 10 cm and a vertical standard deviation of 5 cm (<https://ahn.nl/kwaliteitsbeschrijving>). Further details—including the scanner type, pulse repetition frequency, flight lines and flight elevations—are not provided with the published dataset. Information on uncalibrated intensity and the number of returns is provided, but as the intensity data are not radiometrically corrected, their use is limited because of the potential influence of the flight pattern and

laser scanner type. Ground points, buildings and water are pre-classified, which enables (1) vegetation points to be distinguished from ground points and (2) to exclude infrastructure and water as non-vegetation elements. We downloaded the LiDAR data in a 1,000 m radius around each transect, corresponding to 483 LiDAR tiles, from which point clouds of a 100 m radius around the transect section centroids were extracted.

From the LiDAR point clouds, we derived 12 LiDAR metrics that captured the vertical complexity and horizontal heterogeneity of the vegetation (Table 2). This was done using the R package “lidR” (Roussel & Auty, 2019). Each metric was chosen to reflect vegetation structure-related habitat preferences of the focal species as reported in the ecological literature, either from field or from LiDAR studies (see Table 2). A total of six LiDAR metrics reflected the vertical complexity of vegetation. Those were directly derived from the LiDAR point cloud using a 25 m radius around each centroid (Table 2). This scale matches the length of a transect section ( $2 \times 25 \text{ m} = 50 \text{ m}$ ) and was chosen to describe the local habitat conditions which mobile invertebrates such as butterflies are exposed to. We deliberately chose not to extract LiDAR metrics for polygons of the  $5 \text{ m} \times 50 \text{ m}$  transect sections because the transect routes for the two woodland species often run along forest edges (but in the open habitat). Extracting LiDAR metrics for  $5 \text{ m} \times 50 \text{ m}$  polygons would thus neglect (or at least underestimate) the importance of adjacent woodland habitat (e.g., the need for trees as shelter or the availability of climbing host plants such as *Lonicera* sp.). In addition to the six vertical complexity metrics (25 m radius), we further extracted six metrics reflecting the horizontal heterogeneity of vegetation in either 25 m (vegetation roughness) or 100 m (landscape-scale microtopography or vegetation structure) around each section centroid. The 100-m scale reflects the home range scale of the butterflies (Bos et al., 2006; Maes et al., 2006; Warren, 1987b). Vegetation roughness and landscape-scale vegetation structure metrics were derived from the variability of a digital surface model (DSM), based on LiDAR-derived vegetation height (90th percentile of  $z$ ) within 1-m resolution grid cells, using the R package “landscapemetrics” (Hesselbarth et al., 2019). Microtopography was quantified by calculating the mean slope of the terrain with a 1 m resolution digital terrain model (DTM) based on the minimum height of ground points in each cell. To test whether the vertical accuracy of the LiDAR data ( $\pm 5 \text{ cm}$ ) may introduce bias when separating vegetation returns from ground returns, we recalculated all LiDAR metrics by additionally excluding all non-ground returns  $< 5 \text{ cm}$  above ground. This sensitivity analysis showed that LiDAR metrics excluding points  $< 5 \text{ cm}$  above ground were highly correlated ( $r \geq 0.99$ ) with LiDAR metrics including all vegetation points (Figure S1). We therefore used the LiDAR metrics including all vegetation points in all statistical analyses below.

## 2.3 | Statistical analysis

We build species distribution models (SDMs) to analyse whether and how specific LiDAR metrics (Table 2) can explain the presence-absence

of the four butterfly species. We used the ODMAP protocol (Zurell et al., 2020) to document the modelling objectives and decisions (Supporting Information ODMAP protocol). We carefully explored multicollinearity among the LiDAR metrics with Spearman rank correlations (Figure S2). Metrics that showed high pairwise Spearman rank correlations ( $r > |0.70|$ ) were discarded in the SDMs by first removing the metric with the largest variance inflation factor (VIF) using the function *vifcor* in the R package “usdm” (Naimi et al., 2014). For conceptually related metrics that were highly collinear with each other or other metrics (e.g., open area, open patches and edge extent), we kept the metric that best reflected the ecology and habitat preferences of a specific species (compare Table S1). All metrics in the final SDMs were not strongly correlated ( $r < |0.70|$ ) and had  $VIF < 3$ , as suggested for model implementation (Naimi et al., 2014).

We initially tested three different SDM algorithms for modelling butterfly species distributions and habitat suitability: Random Forest (RF), generalized linear models (GLM) and maximum entropy (Maxent; Breiman, 2001; Naimi & Araújo, 2016; Phillips et al., 2006). To implement the SDMs, we used the R package “sdm” (Naimi & Araújo, 2016) which depends on the R package “stats” for GLMs, the R package “randomForest” for RF and the Java software for Maxent (“maxent.jar”). Model accuracy (mean  $\pm$  SD across 100 model runs) was examined with the area under curve (AUC; Brotons et al., 2004; Pearce & Ferrier, 2000) and the true skill statistic (TSS; Allouche et al., 2006) and visualized using receiver operating characteristics (ROC) curves (Pearce & Ferrier, 2000). As RF outperformed the other two SDM algorithms for all species in terms of AUC and TSS (see results below, and Table S3), we mainly focus on the results of the RF. The RF algorithm is a machine-learning method which implicitly deals with nonlinear relationships. We build RF models using 500 decision trees and default settings (e.g.,  $\text{maxnodes} = 20$ ). Model calibration was performed on 100 random bootstrap subsets of 70% of the data, and predictive performance was then validated with the remaining 30% of the data in each run. Deviance  $D$  was calculated with a loss function (for binomial data) that represents the loss in predictive performance due to a suboptimal model, as implemented in the R package “sdm” (Naimi & Araújo, 2016). In addition to the RF results, we also present in the appendix the ROC curves for all SDM algorithms (Figure S3–S5) and the response curves averaged across the three SDM algorithms (Figure S6–S8). The latter was done using the “rcurve” function in the R package “sdm” which calculates the mean and confidence interval over the individual responses from the three fitted models (RF, GLM, Maxent).

To test our three hypotheses, we assessed whether and to what extent specific LiDAR metrics reflecting low vegetation (H1), medium-to-high vegetation (H2) and landscape-scale habitat structure (H3) can explain the presence-absence of the four butterfly species (Table 2). Specifically, we used the relative variable importance and the response curves of each metric as implemented in the R package “sdm” (mean  $\pm$  SD over 100 model runs for each species) to interpret the role of LiDAR metrics in explaining butterfly habitat preferences. Relative variable importance—measured by AUC improvements of model performance due to inclusion of the focal

variable—was obtained using the function “getVarImp” (Naimi & Araújo, 2016). Response curves for the RF algorithm were visualized using the function “getResponseCurve” (Naimi & Araújo, 2016). The response curves show the species-specific responses (i.e., probability of occurrence) along the gradient of vegetation structure (i.e., a specific LiDAR metric) while keeping all other predictor variables in the model at their mean (following the method described in Elith et al., 2005). As the two grassland species occur in both coastal (dune) and inland habitats, we additionally implemented separate RF models for coastal and inland populations to explore whether habitat relationships differ between these habitats. For inland populations, this included 71 and 137 (*B. selene*) and 122 and 181 (*H. semele*) presence and absence points, respectively. For coastal populations, sample size was only sufficient for *H. semele* (126 presence and 429 absence points, respectively) whereas records for *B. selene* were too limited (21 presence and 14 absence points, respectively).

### 3 | RESULTS

#### 3.1 | Metrics selection

Spearman rank correlations were high ( $r = 0.7\text{--}0.9$ ) between several LiDAR metric pairs (Figure S2). For instance, total vegetation roughness and low vegetation roughness and most density metrics of adjacent strata were discarded from the models based on the VIF. Vegetation height and the 5–20 m vegetation density were highly correlated with each other but also with landscape-scale habitat structure metrics (particularly open area). Vegetation height was selected for both grassland species as it best reflects the ecological conditions in grasslands (e.g., shelter), whereas the 5–20 m density was selected for both woodland species to capture forest cover and woody vegetation. Microtopography (i.e., slope) was only weakly correlated with other metrics and thus kept in all models. Landscape-scale vegetation structure metrics were highly correlated with each other. Open area and edge extent were selected for *B. selene*, reflecting shelter and open vegetation in wet grasslands. The number of open patches was selected for *H. semele* reflecting whether patches of low vegetation in dry habitats (i.e., grasslands and heathlands) are separated by high vegetation or not. Open patches and edge extent were selected for *L. camilla*, reflecting canopy gaps in moist woodlands. Edge extent was selected for *M. athalia* to reflect edges of woody vegetation along patches of open vegetation in dry woodlands. The final selection of LiDAR metrics partly differed among species due to their specific habitat preferences and comprised seven (*B. selene*, *H. semele* and *L. camilla*) and six (*M. athalia*) LiDAR metrics, respectively (Table 2), overall avoiding high correlations (all  $r < 0.7$  and  $VIF < 3$ ).

#### 3.2 | Model performance

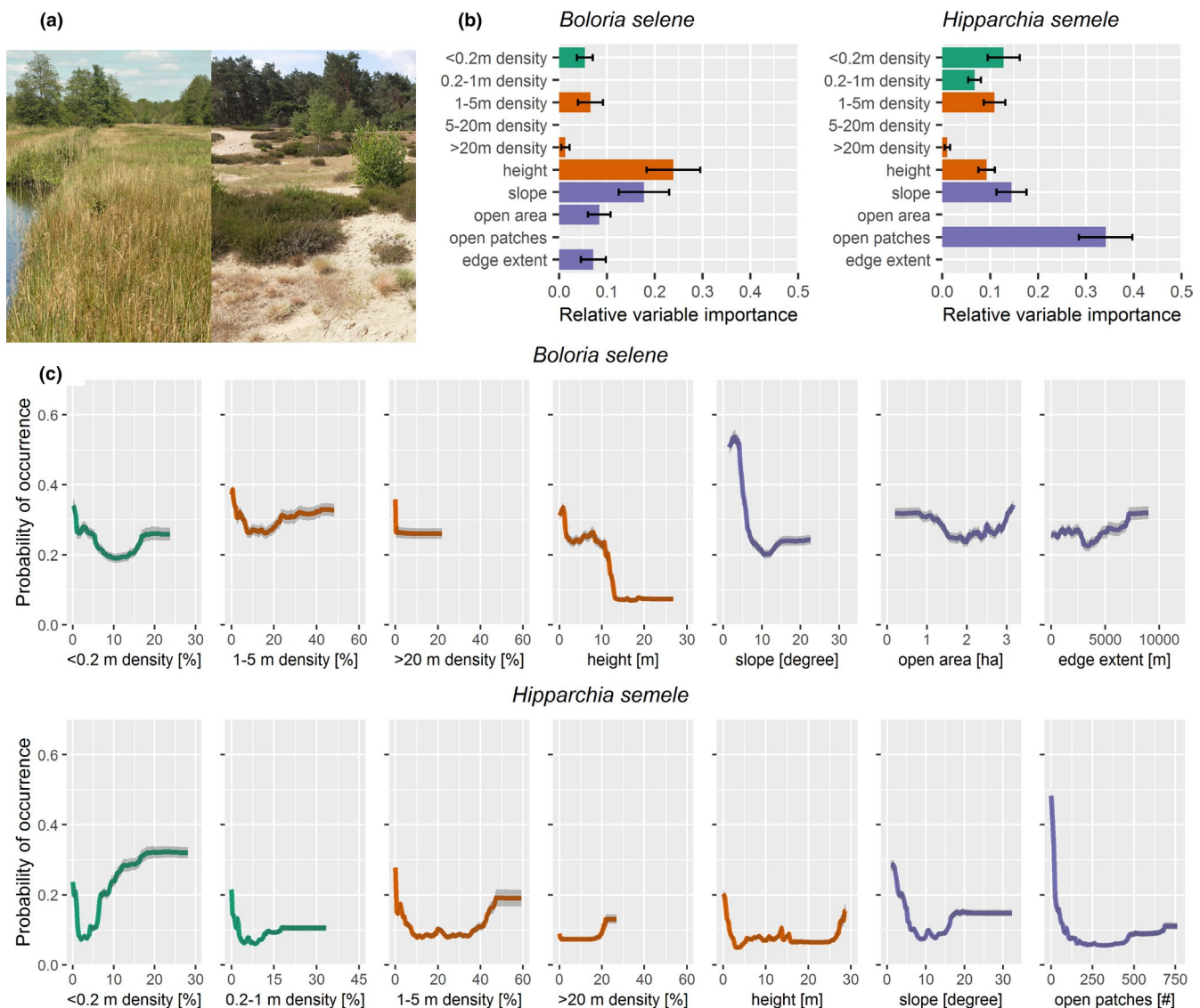
The RF algorithm outperformed the GLMs and Maxent models for all species in terms of AUC, TSS and  $D$  (Table S3). The ROC curves of

the RF models (Figure S3) revealed a good fit of the test data for *B. selene*, *H. semele* and *M. athalia* (AUC = 0.88–0.90; TSS = 0.67–0.70) and an excellent fit for *L. camilla* (AUC = 0.96, TSS = 0.84; Table S3). The deviance  $D$  between RF model repetitions was acceptable for *B. selene*, *H. semele* and *M. athalia* ( $D = 0.79$ ,  $0.74$  and  $0.78$ ) and low for *L. camilla* ( $D = 0.41$ ). This provided a robust basis for interpreting the contributions of predictor variables in the RF models. Moreover, the response curves of the RF model (see below) were generally consistent with the averaged response curves across the three SDM algorithms (Figure S6–S8). For the grassland species *B. selene*, model performance was better when its inland distribution was modelled separately from the coastal population (AUC = 0.91, TSS = 0.73,  $D = 0.75$ ). For the other grassland species, *H. semele*,

model performance in coastal habitats was higher than model performance in inland habitats (Table S3).

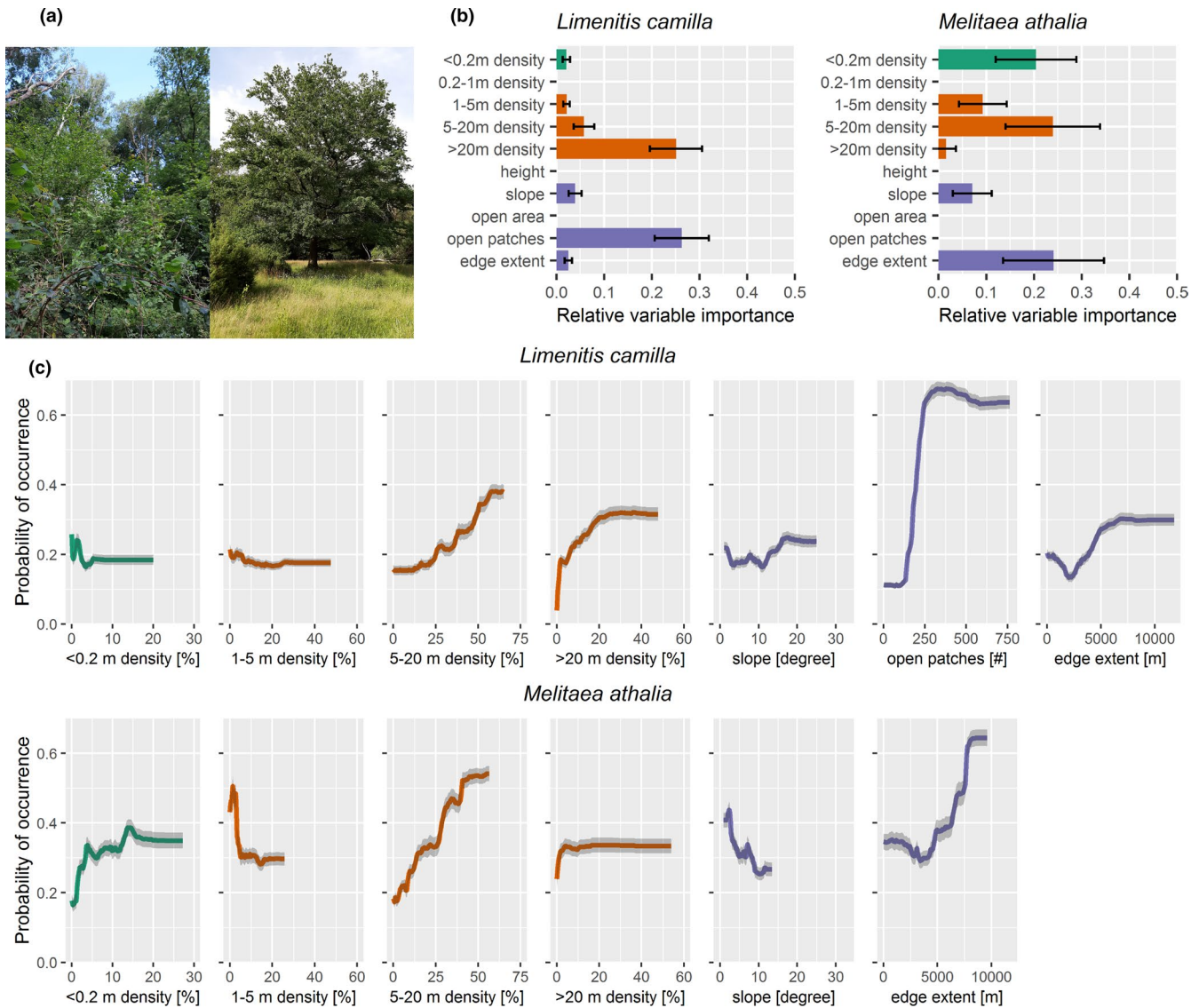
### 3.3 | Effects of low vegetation

Low vegetation metrics were of minor importance in most SDMs (green coloured bars and lines in Figures 2 and 3). *B. selene* showed a weak response to the density <0.2 m in wet grasslands. The probability of occurrence of *H. semele* increased with the density of low vegetation (<0.2 m density; green coloured line in Figure 2c), that is with higher vegetation biomass (captured by non-ground LiDAR returns) in low vegetation strata of dry open habitats. This effect



**FIGURE 2** Associations of two grassland butterflies with LiDAR metrics. (a) Typical habitats of the small pearl-bordered fritillary (*Boloria selene*; a wet grassland species, left) and the grayling (*Hipparchia semele*; a dry grassland species, right). (b) Relative variable importance, showing the contribution of each LiDAR metric to explain butterfly distributions by the mean and standard deviation of 100 random forest (RF) model runs (empty rows are metrics discarded from the RF). (c) Response curves of included LiDAR metrics, showing how they are associated with the species' probability of occurrence by the mean and confidence interval of 100 RF model runs. In (b) and (c), colours indicate LiDAR metrics related to low vegetation (green), medium-to-high vegetation (orange) and landscape-scale habitat structure (purple). Photo credits: J. P. Reinier de Vries





**FIGURE 3** Associations of two woodland butterflies with LiDAR metrics. (a) Typical habitats of the white admiral (*Limenitis camilla*; a moist woodland species, left) and the heath fritillary (*Melitaea athalia*; a dry woodland species, right). (b) Relative variable importance, showing the contribution of each LiDAR metric to explain butterfly distributions by the mean and standard deviation of 100 random forest (RF) model runs (empty rows are metrics discarded from the RF). (c) Response curves of included LiDAR metrics, showing how they are associated with the species' probability of occurrence by the mean and confidence interval of 100 RF model runs. In (b) and (c), colours indicate LiDAR metrics related to low vegetation (green), medium-to-high vegetation (orange) and landscape-scale habitat structure (purple). Photo credits: J. P. Reinier de Vries

was particularly pronounced in inland populations, but not in coastal habitats (Figure S9). The response of *H. semele* to vegetation density between 0.2 and 1 m (reflecting tall herbs and low shrubs) was generally weak (green coloured bars in Figures 2b and S9). For woodland butterflies, vegetation density <0.2 m in moist woodlands was unimportant for *L. camilla* (green coloured bar in Figure 3b), but important for *M. athalia* in dry woodlands (green coloured bar in Figure 3b).

### 3.4 | Medium-to-high vegetation

Medium-to-high vegetation metrics were of major importance for both woodland species (orange coloured bars and lines in Figure 3b,c) and

for the wet grassland species *B. selene* (orange coloured bars and lines in Figure 2b,c). *L. camilla* was associated with a high density of >20 m tall trees whereas *M. athalia* was most strongly associated with a high 5–20 m vegetation density (orange coloured lines in Figure 3c). For the wet grassland species *B. selene*, vegetation height <10 m strongly increased its probability of occurrence (orange coloured line in Figure 2c). The dry grassland species *H. semele* was only weakly associated with medium-to-high vegetation metrics (orange coloured bars and lines in Figure 2b,c), but the importance of vegetation height increased in coastal habitats (Figure S9). The density of 1–5 m shrubs was unimportant in all SDMs and weakly associated with the species' probability of occurrence, but *M. athalia* and *H. semele* in inland habitats responded positively to a low density of 1–5 m tall vegetation (Figures 3 and S9).

### 3.5 | Landscape-scale habitat structure

Metrics related to landscape-scale habitat structure were of key importance in all SDMs, but the specific metrics partly differed among species (purple coloured bars and lines in Figures 2b,c and 3b,c). *B. selene*, *H. semele* and *M. athalia* mainly occurred in flat terrain (mean slope  $<10^\circ$ ), and this effect was particularly pronounced in inland populations of *B. selene* and *H. semele* (Figure S9). In addition to microtopography (i.e., slope), *H. semele* was strongly associated with a low number of open patches in dry habitats (purple coloured line in Figure 2c), reflecting wide open landscapes with large patches of low vegetation. *L. camilla* was most strongly associated with a high number of open vegetation patches in moist woodlands (purple coloured bar and line in Figure 3b,c). Both, *B. selene* (in wet grasslands) and *L. camilla* (in moist woodlands) responded positively to a high edge extent (purple coloured bars and lines in Figures 2b,c and 3b,c, respectively), and *M. athalia* (in dry woodlands) showed a strong preference for high edge extents (purple coloured line in Figure 3c).

## 4 | DISCUSSION

Our LiDAR-based analysis of four threatened butterfly species in the Netherlands provides detailed insights into how vegetation structure shapes invertebrate–habitat relationships at a national extent and thus generalizes beyond local and landscape-scale studies that only analyse invertebrate habitat use at a particular study site. Using high-quality butterfly presence–absence data derived from a national monitoring scheme, we show that landscape-level habitat structures are especially important for both grassland and woodland species and that medium-to-high vegetation structures (and to some extent low vegetation density) are crucial for woodland species. LiDAR metrics capturing low vegetation structure were generally of minor importance across all four species.

The weak association of low vegetation LiDAR metrics with grassland butterflies confirms our hypothesis (H1) that low vegetation elements are difficult to capture with leaf-off ALS data. Ecological field studies show the critical importance of such low vegetation elements for butterflies in grassland habitats, for example in terms of food plants for both adults and larvae (Bos et al., 2006; van Swaay, 2019). This indicates that the leaf-off conditions in which many airborne LiDAR data are captured are not well suited to capture the seasonal structure of annual herbs and grasses in grassland habitats. Leaf-on ALS surveys at national extents or LiDAR data captured from unmanned aerial vehicles (UAV) and terrestrial laser scanners (TLS) at specific study sites could improve the information content of metrics capturing bare ground and low vegetation, especially  $<0.2$  m. Low vegetation structure was especially difficult to measure in wet grassland habitats of *B. selene*, which are nearly all mown in winter (van Swaay, 2019). In dry open habitats of *H. semele*, structures of low vegetation height such as low grasses and bare ground were difficult to quantify with leaf-off ALS data, but perennial dwarf shrubs (i.e., short woody plants) such as heather (*Calluna vulgaris*) are

mostly well captured because they maintain their woody material in winter. This was reflected in the metric of vegetation density  $<0.2$  m which played an important role to explain the presence–absence of *H. semele* in dry open habitats, especially in inland populations (Figure S9). Heather is abundant and an important nectar source in these habitats (mainly heathlands), and a patchy cover is preferred over monotonous heather fields (Bos et al., 2006; van Swaay, 2019; Vanreusel et al., 2007). The response curves of the SDM indicated that a vegetation density  $<0.2$  m between 10% and 30% promotes the occurrence of *H. semele*. The density of perennial dwarf shrubs probably also drives the strong response of *M. athalia* to the  $<0.2$  m vegetation density, reflecting its dependence on open patches with low vegetation in dry woodlands (Bos et al., 2006; van Swaay, 2019). Heather and other perennial dwarf shrubs (e.g., *Vaccinium*) are typically abundant in these places. The weak responses of *L. camilla* to understorey vegetation strata (both  $<0.2$  and 1–5 m) may reflect the difficulty to capture understorey vegetation structure from discrete return LiDAR data in dense, moist woodlands, as initially expected (H1). In contrast, the dry woodlands of *M. athalia* are relatively open, providing a high potential of discrete return LiDAR data to capture low vegetation structure. While *L. camilla* mainly flies in high vegetation strata, understorey vegetation such as bramble (*Rubus* sp.) is important as a nectar source (Bos et al., 2006).

We expected that LiDAR metrics reflecting medium-to-high vegetation (e.g., density or heterogeneity of shrub and tree layers) are especially important to explain habitat preferences of woodland butterflies (H2). This may be particularly important in woodland habitats if species differ in their vertical habitat niches and use of different vegetation strata. As a moist woodland species, the white admiral (*Limenitis camilla*) is thought to primarily use the forest canopy layer (Bos et al., 2006), for example as gathering places of territorial males or as resting places (Lederer, 1960). Our analysis supports the importance of the forest canopy for this species by showing that sites with  $>20\%$  density of tall ( $>20$  m high) trees are preferred, while sites where trees of this height are absent are avoided. In contrast, the heath fritillary (*Melitaea athalia*), a dry woodland species, flies low to the ground and needs trees to provide sheltered conditions and to support its main host plant, the parasitic cow-wheat (*Melampyrum*; Bos et al., 2006; Warren, 1987a). This association is reflected by a strong preference for sites with a high vegetation density of 5–20 m tall shrubs and trees as derived from LiDAR. Dense scrub vegetation is unfavourable for *M. athalia* and may even act as a barrier to its dispersal (Warren, 1987b), reflected in its preference of a low vegetation density within 1–5 m height. Both grassland butterflies (*B. selene* and *H. semele*) use open habitats with preferably few tall vegetation elements (Bos et al., 2006; van Swaay, 2019). This is reflected in our results by a preference for low vegetation height in *B. selene* and by a low number of open patches in *H. semele*, capturing large open areas in dry open habitats.

Our results confirmed that LiDAR metrics reflecting landscape-scale habitat structures (e.g., microtopography, woodland edges and vegetation openness) are particularly important for explaining the habitat preferences of both grassland and woodland butterflies (H3).

The structural heterogeneity at the landscape scale can be captured with LiDAR metrics that quantify microtopography, or the extent and patchiness of open vegetation (defined by vegetation height) or the length of woodland edges (Table 2). The most important effect of microtopography (i.e., slope) was found for *B. selene*, which prefers flat terrain, especially on inland locations that have marshy habitat conditions. Both *B. selene* and *H. semele* occur on steeper terrain in dune habitats, in which microtopography is also positively correlated with the roughness of low vegetation (Figure S2). For *H. semele*, this might reflect the amount of bare sand patches in these habitats (Maes et al., 2006; van Swaay, 2019). Another important landscape-scale LiDAR metric was the number of open patches which showed a contrasting response for *H. semele* and *L. camilla*, reflecting their different habitats (dry grasslands vs. moist woodlands). Whereas *H. semele* was associated with low-stature landscapes that have only few open patches (i.e., large areas with vegetation height <1 m), *L. camilla* prefers woodlands with many small patches of open vegetation (roughly >70 patches of >2 m<sup>2</sup>/ha) where sunlight can penetrate through the woodland canopy (Bos et al., 2006; van Swaay, 2019). This shows that the same LiDAR metric can capture different characteristics of the landscape depending on which species or habitat is considered. The extent of woodland edges was especially important for the dry woodland species *M. athalia* which reflects its association with woodland edges, where it finds shelter and host plant habitats (van Swaay, 2019; Warren, 1978a, 1978b). Woodland edges can also provide suitable habitat features for *L. camilla* (e.g., sunny conditions within woodland) and shelter for inland populations of *B. selene* (Bos et al., 2006; van Swaay, 2019), but effects were only weakly reflected in the SDMs.

Most SDMs use macroclimate, land cover and topography to map species distributions over broad spatial extents (Guisan et al., 2017). However, our study shows that vertical complexity (e.g., vegetation density, cover and height) and horizontal heterogeneity (e.g., vegetation openness and woodland edge extent) are key determinants of species distributions that need to be taken into account when predicting the probability of invertebrate occurrences at fine resolution over broad spatial extents. This has important implications for assessing organismal responses to climate change because microclimates are often shaped by fine-scale topography and vegetation structure and may allow organisms to persist despite shifts in macroclimates, for example through local temperature buffering or microrefugia (Zellweger et al., 2019). Our derived set of LiDAR metrics shows that microtopography (i.e., slope) is associated with the variability of vegetation height <1 m in grasslands (i.e., roughness of low vegetation) and with landscape-scale habitat structure (i.e., amount of open area, number of patches and edge extent) in woodland habitats (Figure S2). Such habitat structures may allow these butterflies (or their larval host plants) to persist, despite general predictions of strong northward range shifts based on macroclimate (Settele et al., 2008). Country-wide or regional ALS datasets derived from massive LiDAR point clouds (Meijer et al., 2020) will thus provide valuable information for modelling and mapping species distributions under climate change because they provide spatially

contiguous, fine-scale information to quantify microtopography, vegetation structure and microclimate (Zellweger et al., 2019).

The SDMs of woodland species achieved higher performances than those of grassland species. This probably reflects the fact that LiDAR data obtained under leaf-off conditions are better suited to capture the physical structure of tall woody vegetation compared to low-stature vegetation in open habitats such as grasslands. This means that several structural habitat characteristics that are potentially preferred by grassland butterflies may be insufficiently captured by country-wide airborne LiDAR data that are captured in the leaf-off season. Nevertheless, further developments in LiDAR technology (e.g., full-waveform data instead of discrete echoes, provisioning of high-density point clouds, information for calibrating intensity data, complementary data from UAV or TLS) and LiDAR flight campaigns during leaf-on conditions could improve ecological analyses of vegetation structure in open, low-stature habitats such as grasslands and wetlands (Alexander et al., 2015; Zlinszky et al., 2014). Additionally, full-waveform LiDAR could also improve the ability to quantify understorey vegetation in forests and woody habitats (Anderson et al., 2016). The availability of such LiDAR data over broad spatial extents would greatly enhance analyses of species distributions, ecological niches and habitat preferences of invertebrates and other taxa. Moreover, synergies with other remote sensing data such as spectral imagery or information obtained from synthetic aperture radar (SAR) could improve the quantification of habitat aspects in grasslands that are not captured by LiDAR, for example the identification of specific grasses or host plant species (Marcinkowska-Ochtyra et al., 2018), the quantification of seasonal growth dynamics in grasslands (Metz et al., 2014) or the extent of bare ground patches. Furthermore, SDMs of host plant species could be used to improve butterfly SDMs through an assessment of host plant distribution and availability, a crucial habitat factor for most butterfly species.

## 5 | CONCLUSIONS

Our study shows how the distribution of grassland and woodland butterflies depends on different aspects of vegetation structure, including the vertical variability of vegetation and the horizontal heterogeneity of vegetation and microtopography at the landscape scale. As vegetation structure is a key habitat determinant for many invertebrate species and their host plants (Dennis et al., 2003, 2006), the ability of LiDAR to quantify vegetation structure offers promising new ways to gain insights into invertebrate-habitat relationships (Davies & Asner, 2014; Moeslund et al., 2019; Zellweger et al., 2013). Such information can be beneficial for improving the management and conservation of threatened species by identifying and quantifying habitat preferences and specific habitat thresholds. Our study demonstrates that LiDAR metrics are not only informative for species inhabiting woody habitats, but also for invertebrates occurring in low-stature habitats. The efficient, scalable and distributed processing of large,

multi-terabyte LiDAR datasets and the development of dedicated algorithms and software will facilitate and enhance applications of ALS data in ecology and biodiversity science (Meijer et al., 2020; Roussel et al., 2020). This offers ample new opportunities for developing ecosystem structure EBVs (Valbuena et al., 2020) and for applying LiDAR-based habitat analyses to invertebrates, not only in grasslands but also in dunes, heathlands and wetlands, and other non-forest habitats that are threatened throughout Europe and other parts of the world.

## ACKNOWLEDGEMENTS

This work is part of the project "eScience infrastructure for Ecological applications of LiDAR point clouds" (eEcoLiDAR; Kissling et al., 2017), funded by the Netherlands eScience Center (<https://www.esciencecenter.nl>), grant number ASDI.2016.014. The Dutch Butterfly Monitoring Scheme is a joint scheme of Dutch Butterfly Conservation and Statistics Netherlands and is financed by the Ministry of Agriculture, Nature and Food Quality in the framework of the Dutch Network Ecological Monitoring programme. We are greatly indebted to the many volunteers that contributed to the data collection. We thank Chris van Swaay for discussion and providing photographs of butterflies.

## PEER REVIEW

The peer review history for this article is available at <https://publons.com/publon/10.1111/ddi.13272>.

## DATA AVAILABILITY STATEMENT

All scripts for processing the raw LiDAR data (point clouds) into LiDAR metrics and all data and R scripts to reproduce the statistical analyses are available from the Dryad Digital Repository (de Vries et al., 2021; <https://doi.org/10.5061/dryad.g79cnp5pf>).

## ORCID

W. Daniel Kissling  <https://orcid.org/0000-0002-7274-6755>

## REFERENCES

- Aguirre-Gutiérrez, J., WallisDeVries, M. F., Marshall, L., van't Zelfde, M., Villalobos-Arámbula, A. R., Boekelo, B., Bartholomeus, H., Franzén, M., & Biesmeijer, J. C. (2017). Butterflies show different functional and species diversity in relationship to vegetation structure and land use. *Global Ecology and Biogeography*, 26(10), 1126–1137. <https://doi.org/10.1111/geb.12622>
- Aiello-Lammens, M. E., Boria, R. A., Radosavljevic, A., Vilela, B., & Anderson, R. P. (2015). spThin: An R package for spatial thinning of species occurrence records for use in ecological niche models. *Ecography*, 38(5), 541–545. <https://doi.org/10.1111/ecog.01132>
- Alexander, C., Deák, B., Kania, A., Mücke, W., & Heilmeyer, H. (2015). Classification of vegetation in an open landscape using full-waveform airborne laser scanner data. *International Journal of Applied Earth Observation and Geoinformation*, 41, 76–87. <https://doi.org/10.1016/j.jag.2015.04.014>
- Allouche, O., Tsoar, A., & Kadmon, R. (2006). Assessing the accuracy of species distribution models: Prevalence, kappa and the true skill statistic (TSS). *Journal of Applied Ecology*, 43(6), 1223–1232. <https://doi.org/10.1111/j.1365-2664.2006.01214.x>
- Anderson, K., Hancock, S., Disney, M., & Gaston, K. J. (2016). Is wave-form worth it? A comparison of LiDAR approaches for vegetation and landscape characterization. *Remote Sensing in Ecology and Conservation*, 2(1), 5–15.
- Bakx, T. R., Koma, Z., Seijmonsbergen, A. C., & Kissling, W. D. (2019). Use and categorization of light detection and ranging vegetation metrics in avian diversity and species distribution research. *Diversity and Distributions*, 25(7), 1045–1059. <https://doi.org/10.1111/ddi.12915>
- Bergman, K. O., Ask, L., Asklund, J., Ignell, H., Wahlman, H., & Milberg, P. (2008). Importance of boreal grasslands in Sweden for butterfly diversity and effects of local and landscape habitat factors. *Biodiversity and Conservation*, 17(1), 139–153. <https://doi.org/10.1007/s10551-007-9235-x>
- Bos, F., Bosveld, M., Groenendijk, D., van Swaay, C. A., Wynhoff, I., & De Vlinderstichting. (2006). De dagvlinders van Nederland, verspreiding en bescherming (Lepidoptera: Hesperioidea, Papilionidea). *Nederlandse Fauna 7*. Leiden. Nationaal Natuurhistorisch Museum Naturalis, KNNV Uitgeverij & European Invertebrate Survey – Netherlands.
- Breiman, L. (2001). Random forests. *Machine Learning*, 45(1), 5–32.
- Brotons, L., Thuiller, W., Araújo, M. B., & Hirzel, A. H. (2004). Presence-absence versus presence-only modelling methods for predicting bird habitat suitability. *Ecography*, 27(4), 437–448. <https://doi.org/10.1111/j.0906-7590.2004.03764.x>
- Carvalho, L. G., Kunin, W. E., Keil, P., Aguirre-Gutiérrez, J., Ellis, W. N., Fox, R., & van de Meutter, F. (2013). Species richness declines and biotic homogenisation have slowed down for NW-European pollinators and plants. *Ecology Letters*, 16(7), 870–878.
- Cozzi, G., Müller, C. B., & Krauss, J. (2008). How do local habitat management and landscape structure at different spatial scales affect fritillary butterfly distribution on fragmented wetlands? *Landscape Ecology*, 23(3), 269–283. <https://doi.org/10.1007/s10980-007-9178-3>
- Davies, A. B., & Asner, G. P. (2014). Advances in animal ecology from 3D-LiDAR ecosystem mapping. *Trends in Ecology & Evolution*, 29(12), 681–691. <https://doi.org/10.1016/j.tree.2014.10.005>
- de Vries, J. P. R., Koma, Z., WallisDeVries, M., & Kissling, W. D. (2021). Data and code for: Identifying fine-scale habitat preferences of threatened butterflies using airborne laser scanning. Dryad, <https://doi.org/10.5061/dryad.g79cnp5pf>
- Dennis, R. L. H., Shreeve, T. G., & van Dyck, H. (2003). Towards a functional resource-based concept for habitat: A butterfly biology viewpoint. *Oikos*, 102, 417–426.
- Dennis, R. L. H., Shreeve, T. G., & van Dyck, H. (2006). Habitats and resources: The need for a resource-based definition to conserve butterflies. *Biodiversity & Conservation*, 15(6), 1943–1966. <https://doi.org/10.1007/s10531-005-4314-3>
- Elith, J., Ferrier, S., Huettmann, F., & Leathwick, J. (2005). The evaluation strip: A new and robust method for plotting predicted responses from species distribution models. *Ecological Modelling*, 186(3), 280–289.
- Eskildsen, A., Carvalho, L. G., Kissling, W. D., Biesmeijer, J. C., Schweiger, O., & Høye, T. T. (2015). Ecological specialization matters: Long-term trends in butterfly species richness and assemblage composition depend on multiple functional traits. *Diversity and Distributions*, 21(7), 792–802.
- Essens, T., van Langevelde, F., Vos, R. A., van Swaay, C. A., & WallisDeVries, M. F. (2017). Ecological determinants of butterfly vulnerability across the European continent. *Journal of Insect Conservation*, 21(3), 439–450. <https://doi.org/10.1007/s10841-017-9972-4>
- Fox, R., Brereton, T. M., Asher, J., August, T. A., Botham, M. S., Bourn, N. A. D., Cruickshanks, K. L., Bulman, C. R., Ellis, S., Harrower, C. A., Middlebrook, I., Noble, D. G., Powney, G. D., Randle, Z., Warren, M. S., & Roy, D. B. (2015). *The State of the UK's Butterflies 2015*. Butterfly Conservation and the Centre for Ecology & Hydrology.



- Glad, A., Reineking, B., Montadert, M., Depraz, A., & Monnet, J.-M. (2020). Assessing the performance of object-oriented LiDAR predictors for forest bird habitat suitability modelling. *Remote Sensing in Ecology and Conservation*, 6(1), 5–19.
- Graham, L. J., Spake, R., Gillings, S., Watts, K., & Eigenbrod, F. (2019). Incorporating fine-scale environmental heterogeneity into broad-extent models. *Methods in Ecology and Evolution*, 10(6), 767–778. <https://doi.org/10.1111/2041-210X.13177>
- Guisan, A., Thuiller, W., & Zimmermann, N. E. (2017). *Habitat suitability and distribution models*. Cambridge University Press.
- Hallmann, C. A., Sorg, M., Jongejans, E., Siepel, H., Hofland, N., Schwan, H., Stenmans, W., Müller, A., Sumser, H., Hörrén, T., Goulson, D., & de Kroon, H. (2017). More than 75 percent decline over 27 years in total flying insect biomass in protected areas. *PLoS One*, 12(10), e0185809. <https://doi.org/10.1371/journal.pone.0185809>
- Hess, A. N., Falkowski, M. J., Webster, C. R., Storer, A. J., Pocewicz, A., & Martinuzzi, S. (2013). Employing lidar data to identify butterfly habitat characteristics of four contrasting butterfly species across a diverse landscape. *Remote Sensing Letters*, 4(4), 354–363. <https://doi.org/10.1080/2150704X.2012.734932>
- Hesselbarth, M. H. K., Sciaini, M., With, K. A., Wiegand, K., & Nowosad, J. (2019). Landscapemetrics: An open-source R tool to calculate landscape metrics. *Ecography*, 42(10), 1648–1657.
- Karlsson, B., & Wiklund, C. (2005). Butterfly life history and temperature adaptations; dry open habitats select for increased fecundity and longevity. *Journal of Animal Ecology*, 74(1), 99–104. <https://doi.org/10.1111/j.1365-2656.2004.00902.x>
- Kissling, W. D., Seijmonsbergen, A., Foppen, R., & Bouten, W. (2017). eEcoLiDAR, eScience infrastructure for ecological applications of LiDAR point clouds: Reconstructing the 3D ecosystem structure for animals at regional to continental scales. *Research Ideas and Outcomes*, 3, e14939. <https://doi.org/10.3897/rio.3.e14939>
- Koma, Z., Seijmonsbergen, A. C., & Kissling, W. D. (2020). Classifying wetland-related land cover types and habitats using fine-scale lidar metrics derived from country-wide Airborne Laser Scanning. *Remote Sensing in Ecology and Conservation*, 7, 80–96. <https://doi.org/10.1002/rse2.170>
- Lederer, G. (1960). Verhaltensweisen der Imagines und der Entwicklungsstadien von *Limnitis camilla camilla* L. (Lep. Nymphalidae). *Zeitschrift Für Tierpsychologie*, 17(5), 521–546. <https://doi.org/10.1111/j.1439-0310.1960.tb00404.x>
- Lucas, C., Bouten, W., Koma, Z., Kissling, W. D., & Seijmonsbergen, A. C. (2019). Identification of linear vegetation elements in a rural landscape using LiDAR point clouds. *Remote Sensing*, 11, 292. <https://doi.org/10.3390/rs11030292>
- Maes, D., Ghesquiere, A., Logie, M., & Bonte, D. (2006). Habitat use and mobility of two threatened coastal dune insects: Implications for conservation. *Journal of Insect Conservation*, 10(2), 105–115. <https://doi.org/10.1007/s10841-006-6287-2>
- Maes, D., & van Dyck, H. (2001). Butterfly diversity loss in Flanders (north Belgium): Europe's worst case scenario? *Biological Conservation*, 99(3), 263–276. [https://doi.org/10.1016/S0006-3207\(00\)00182-8](https://doi.org/10.1016/S0006-3207(00)00182-8)
- Marcinkowska-Ochtyra, A., Jarocińska, A., Bzdęga, K., & Tokarska-Guzik, B. (2018). Classification of expansive grassland species in different growth stages based on hyperspectral and LiDAR data. *Remote Sensing*, 10(12), 2019. <https://doi.org/10.3390/rs10122019>
- Meijer, C., Grootes, M. W., Koma, Z., Dzigan, Y., Gonçalves, R., Andela, B., van den Oord, G., Rangelova, E., Renaud, N., & Kissling, W. D. (2020). Laserchicken—A tool for distributed feature calculation from massive LiDAR point cloud datasets. *SoftwareX*, 12, 100626. <https://doi.org/10.1016/j.softx.2020.100626>
- Metz, A., Marconcini, M., Esch, T., Reinartz, P., & Ehlers, M. (2014). Classification of grassland types by means of multi-seasonal TerraSAR-X and RADARSAT-2 imagery. In *2014 IEEE Geoscience and Remote Sensing Symposium* (pp. 1202–1205). IEEE, Quebec City, QC, Canada.
- Moeslund, J. E., Zlinszky, A., Ejrnæs, R., Brunbjerg, A. K., Bøcher, P. K., Svenning, J. C., & Normand, S. (2019). Light detection and ranging explains diversity of plants, fungi, lichens and bryophytes across multiple habitats and large geographic extent. *Ecological Applications*, 29(5), e01907.
- Müller, J., Bae, S., Röder, J., Chao, A., & Didham, R. K. (2014). Airborne LiDAR reveals context dependence in the effects of canopy architecture on arthropod diversity. *Forest Ecology and Management*, 312, 129–137. <https://doi.org/10.1016/j.foreco.2013.10.014>
- Müller, J., & Brandl, R. (2009). Assessing biodiversity by remote sensing in mountainous terrain: The potential of LiDAR to predict forest beetle assemblages. *Journal of Applied Ecology*, 46(4), 897–905. <https://doi.org/10.1111/j.1365-2664.2009.01677.x>
- Naimi, B., & Araújo, M. B. (2016). sdm: A reproducible and extensible R platform for species distribution modelling. *Ecography*, 39(4), 368–375. <https://doi.org/10.1111/ecog.01881>
- Naimi, B., Hamm, N. A., Groen, T. A., Skidmore, A. K., & Toxopeus, A. G. (2014). Where is positional uncertainty a problem for species distribution modelling? *Ecography*, 37(2), 191–203. <https://doi.org/10.1111/j.1600-0587.2013.00205.x>
- Pearce, J., & Ferrier, S. (2000). Evaluating the predictive performance of habitat models developed using logistic regression. *Ecological Modelling*, 133(3), 225–245. [https://doi.org/10.1016/S0304-3800\(00\)00322-7](https://doi.org/10.1016/S0304-3800(00)00322-7)
- Phillips, S. J., Anderson, R. P., & Schapire, R. E. (2006). Maximum entropy modelling of species geographic distributions. *Ecological Modelling*, 190(3–4), 231–259.
- QGIS Development Team. (2009). *QGIS Geographic Information System. Open Source Geospatial Foundation Project*. <https://qgis.osgeo.org>
- R Core Team. (2019). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. <https://www.R-project.org>
- Reutebuch, S. E., Andersen, H. E., & McGaughey, R. J. (2005). Light detection and ranging (LiDAR): An emerging tool for multiple resource inventory. *Journal of Forestry*, 103(6), 286–292.
- Roussel, J. R., & Auty, D. (2019). *lidR: Airborne LiDAR data manipulation and visualization for forestry applications. R package version 2.1.1*. <https://CRAN.R-project.org/package=lidR>
- Roussel, J.-R., Auty, D., Coops, N. C., Tompalski, P., Goodbody, T. R. H., Meador, A. S., Bourdon, J.-F., de Boissieu, F., & Achim, A. (2020). *lidR: An R package for analysis of Airborne Laser Scanning (ALS) data. Remote Sensing of Environment*, 251, 112061. <https://doi.org/10.1016/j.rse.2020.112061>
- Settele, J., Kudrna, O., Harpke, A., Kühn, I., van Swaay, C., Verovnik, R., Warren, M., Wiemers, M., Hanspach, J., Hickler, T., Kühn, E., van Halder, I., Veling, K., Vliegnerhart, A., Wynhoff, I., & Schweiger, O. (2008). *Climatic risk atlas of European butterflies*. Pensoft, Sofia.
- Simonson, W. D., Allen, H. D., & Coomes, D. A. (2014). Applications of airborne lidar for the assessment of animal species diversity. *Methods in Ecology and Evolution*, 5(8), 719–729. <https://doi.org/10.1111/2041-210X.12219>
- Thomas, J. A. (1995). Why small cold-blooded insects pose different conservation problems to birds in modern landscapes. *Ibis*, 137, 112–119. <https://doi.org/10.1111/j.1474-919X.1995.tb08431.x>
- Thomas, J. A. (2005). Monitoring change in the abundance and distribution of insects using butterflies and other indicator groups. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 360(1454), 339–357.
- Thomas, J. A., Telfer, M. G., Roy, D. B., Preston, C. D., Greenwood, J. J. D., Asher, J., & Lawton, J. H. (2004). Comparative losses of British butterflies, birds, and plants and the global extinction crisis. *Science*, 303(5665), 1879–1881.
- Valbuena, R., O'Connor, B., Zellweger, F., Simonson, W., Vihervaara, P., Maltamo, M., Silva, C. A., Almeida, D., Danks, F., Morsdorf, F., Chirici,

- G., Lucas, R., Coomes, D. A., & Coops, N. C. (2020). Standardizing ecosystem morphological traits from 3D information sources. *Trends in Ecology & Evolution*, 35(8), 656–667. <https://doi.org/10.1016/j.tree.2020.03.006>
- van Strien, A. J., Meyling, A. W. G., Herder, J. E., Hollander, H., Kalkman, V. J., Poot, M. J. M., Turnhout, S., van der Hoorn, B., van Strien-van Liempt, W. T. F. H., van Swaay, C. A. M., van Turnhout, C. A. M., Verweij, R. J. T., & Oerlemans, N. J. (2016). Modest recovery of biodiversity in a western European country: The Living Planet Index for the Netherlands. *Biological Conservation*, 200, 44–50. <https://doi.org/10.1016/j.biocon.2016.05.031>
- van Strien, A. J., van Swaay, C. A., van Strien-van Liempt, W. T., Poot, M. J., & WallisDeVries, M. F. (2019). Over a century of data reveal more than 80% decline in butterflies in the Netherlands. *Biological Conservation*, 234, 116–122. <https://doi.org/10.1016/j.biocon.2019.03.023>
- van Swaay, C. A. (2019). *Basisrapport Rode Lijst Dagvlinders 2019 volgens Nederlandse en IUCN criteria. Rapport VS2019.001, De Vlinderstichting*. Wageningen, the Netherlands.
- van Swaay, C. A., Nowicki, P., Settele, J., & van Strien, A. J. (2008). Butterfly monitoring in Europe: Methods, applications and perspectives. *Biodiversity and Conservation*, 17(14), 3455–3469. <https://doi.org/10.1007/s10531-008-9491-4>
- van Swaay, C. A., Warren, M., & Loïs, G. (2006). Biotope use and trends of European butterflies. *Journal of Insect Conservation*, 10(2), 189–209. <https://doi.org/10.1007/s10841-006-6293-4>
- Vanreusel, W., Maes, D., & van Dyck, H. (2007). Transferability of species distribution models: A functional habitat approach for two regionally threatened butterflies. *Conservation Biology*, 21(1), 201–212. <https://doi.org/10.1111/j.1523-1739.2006.00577.x>
- Vierling, K. T., Bässler, C., Brandl, R., Vierling, L. A., Weiß, I., & Müller, J. (2011). Spinning a laser web: Predicting spider distributions using LiDAR. *Ecological Applications*, 21(2), 577–588. <https://doi.org/10.1890/09-2155.1>
- Warren, M. S. (1987a). The ecology and conservation of the heath fritillary butterfly, *Mellicta athalia*. I. Host selection and phenology. *Journal of Applied Ecology*, 24(2), 467–482. <https://doi.org/10.2307/2403887>
- Warren, M. S. (1987b). The ecology and conservation of the heath fritillary butterfly, *Mellicta athalia*. II. Adult population structure and mobility. *Journal of Applied Ecology*, 24(2), 483–498. <https://doi.org/10.2307/2403888>
- Wösten, J. H. M., de Vries, F., Denneboom, J., & van Holst, A. F. (1988). *Generalisatie en bodemfysische vertaling van de Bodemkaart van Nederland 1: 250 000 ten behoeve van de PAWN-studie. Rapport 2055*. STIBOKA, Wageningen, the Netherlands.
- Wösten, J. H. M., de Vries, F., Hoogland, T., Massop, H. T. L., Veldhuizen, A. A., Vroon, H. R. J., Bolman, A. (2013). *BOFEK2012, de nieuwe bodemfysische schematisatie van Nederland. Rapport 2387*. Alterra, Wageningen, The Netherlands.
- Zellweger, F., Braunisch, V., Baltensweiler, A., & Bollmann, K. (2013). Remotely sensed forest structural complexity predicts multi species occurrence at the landscape scale. *Forest Ecology and Management*, 307, 303–312. <https://doi.org/10.1016/j.foreco.2013.07.023>
- Zellweger, F., De Frenne, P., Lenoir, J., Rocchini, D., & Coomes, D. (2019). Advances in microclimate ecology arising from remote sensing. *Trends in Ecology & Evolution*, 34, 327–341. <https://doi.org/10.1016/j.tree.2018.12.012>
- Zielewska-Büttner, K., Heurich, M., Müller, J., & Braunisch, V. (2018). Remotely sensed single tree data enable the determination of habitat thresholds for the three-toed woodpecker (*Picoides tridactylus*). *Remote Sensing*, 10(12), 1972. <https://doi.org/10.3390/rs10121972>
- Zlinszky, A., Schroiff, A., Kania, A., Deák, B., Mücke, W., Vári, Á., Székely, B., & Pfeifer, N. (2014). Categorizing grassland vegetation with full-waveform airborne laser scanning: A feasibility study for detecting Natura 2000 habitat types. *Remote Sensing*, 6(9), 8056–8087. <https://doi.org/10.3390/rs6098056>
- Zurell, D., Franklin, J., König, C., Bouchet, P. J., Dormann, C. F., Elith, J., Fandos, G., Feng, X., Guillera-Arroita, G., Guisan, A., Lahoz-Monfort, J. J., Leitão, P. J., Park, D. S., Peterson, A. T., Rapacciuolo, G., Schmatz, D. R., Schröder, B., Serra-Diaz, J. M., Thuiller, W., ... Merow, C. (2020). A standard protocol for reporting species distribution models. *Ecography*, 43(9), 1261–1277. <https://doi.org/10.1111/ecog.04960>

#### BIOSKETCH

The research team is interested in understanding the spatial distribution of biodiversity and the habitat preferences and ecological niches of species, using a combination of in situ (species) observations and remote sensing. The extraction of vegetation structural parameters from 3D point clouds derived from Light Detection And Ranging (LiDAR) over broad spatial extents is a particular research focus of the team, as is the application of ecological knowledge and theory to the conservation of species and habitats.

#### SUPPORTING INFORMATION

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

**How to cite this article:** de Vries JPR, Koma Z, WallisDeVries MF, Kissling WD. Identifying fine-scale habitat preferences of threatened butterflies using airborne laser scanning. *Divers Distrib*. 2021;00:1–14. <https://doi.org/10.1111/ddi.13272>