ORIGINAL PAPER

Efficiency of spatial sampling designs in estimating abundance and species richness of carabids at the landscape level

Jan Thiele · **Gunda Schulte auf'm Erley** ·
Michael Glemnitz · **Doreen Gabriel**

Abstract

Context Declining biodiversity in agricultural landscapes has increased the need for research and monitoring of insect abundance and diversity at the landscape level.

Objectives We investigated the accuracy of different spatial sampling designs in estimating landscape-level abundance and species richness of carabids in agricultural landscapes and, further, which sample size per landscape section was required and whether dominating land use or landscape subdivision affected accuracy and required sample size.

Methods We developed a simulation model that created raster maps of agricultural landscapes, compiled local carabid communities (species composition and abundances) within raster cells and simulated the sampling of carabids with pitfall traps using different spatial sampling designs and sample sizes between 4 and 49 sampling points per landscape section. Spatial sampling designs included random, systematic grid-based, stratified and clustered schemes.

Results To estimate landscape-level abundance, area-proportional stratified random sampling was most accurate followed by systematic grid-based designs. A sample size of 25 appeared to be the best trade-off between accuracy and sampling cost. Accuracy was not affected substantially by landscape characteristics in most cases. With respect to species richness, all designs except for clustered sampling had comparable accuracies, but even 49 samples were not sufficient to detect 80% of the species.

Conclusion Systematic grid-based designs are generally recommendable for sampling of carabids in agricultural landscapes and, in case a carabid-specific habitat classification is available, area-proportional stratified random sampling provides optimal accuracy for estimating landscape-level abundance. For assessment of species richness, large sample size is more important than spatial sampling design.

J. Thiele (✉)
Thünen Institute, Institute of Biodiversity, Bundesallee 65, 38116 Braunschweig, Germany
e-mail: jan.thiele@thuenen.de

G. Schulte auf'm Erley · D. Gabriel
Julius Kühn Institute, Institute of Crop and Soil Science, Bundesallee 58, 38116 Braunschweig, Germany

M. Glemnitz
Leibniz Centre for Agricultural Landscape Research (ZALF), Eberswalder Straße 84, 15374 Müncheberg, Germany

Introduction

The abundance of insects declined in agricultural landscapes over the last decades (Hallmann et al. 2017; Seibold et al. 2019) leading to a loss of ecosystem services and cascading effects on food webs (Cardoso et al. 2020). This has increased the need for research on the effectiveness of biodiversity measures in supporting insects and for monitoring of the development of insect communities in agricultural landscapes (Brooks et al. 2012; Kunin 2019; Scherber et al. 2021). Particularly, an increasing number of research and monitoring projects are seeking to assess the abundance and species richness of insects at landscape level (Scherber et al. 2019). In this regard, the questions arise of how to place sampling points in a landscape so that estimates of abundance and species richness are representative of the whole landscape, and how many sampling points are required?

Agricultural landscapes are spatially heterogeneous as they comprise different crops and semi-natural habitats. In heterogeneous landscapes with several land-use/land-cover (LULC) types that differ in species composition and richness, the placement of sampling points may have substantial influence on landscape-level estimates of abundance and species richness (Miller and Ambrose 2000). This may be particularly severe with ground-dwelling insects who have a limited radius of action and whose abundance and species richness varies among LULC types, such as carabid beetles (Knapp and Řezáč 2015; Knapp et al. 2019).

Several spatial sampling designs are available for placing sampling points on a section of the earth's surface in order to estimate the expected value of some spatial variable (Wang et al. 2012). Here we will focus on sampling designs that do not require prior measurements of the spatial variable of interest because such data are rarely available for insects in agricultural landscapes. That is, we will not consider sampling designs based on estimates of spatial-autocorrelation and model-based sampling methods (Wang et al. 2012; Benedetti et al. 2015). Instead, we will look at sampling designs that we deem generally applicable to ground-dwelling insects in agricultural landscapes and that require either no prior knowledge of landscape heterogeneity or use a spatial stratification based on some sort of land-cover or habitat classification. These include simple spatial random

sampling, systematic sampling designs based on spatial grids, stratified random sampling, area-proportional stratified random sampling and clustered sampling (Wang et al. 2012).

Previous theoretical and empirical studies have shown that systematic and stratified sampling designs can be more efficient than simple spatial random sampling when the variable of interest shows spatial structure, such as spatial auto-correlation or patchy distribution of different mean values (Miller and Ambrose 2000; Wang et al. 2010; McGarvey et al. 2016). However, rankings of efficiencies of spatial sampling designs may vary in the presence of anisotropy or non-stationarity of the variable of interest (Wang et al. 2012). Regarding ground-dwelling insects in agricultural landscapes, species richness and abundances vary among different LULC types and, thus, are spatially auto-correlated (Knapp and Řezáč 2015; Knapp et al. 2019). Further, crop rotation will presumably lead to non-stationary distributions of abundance and species richness (Beduschi et al. 2015). Therefore, it is uncertain which type of spatial sampling design—simple random, systematic, stratified—would most reliably estimate abundances and species richness of ground-dwelling insects at landscape level.

Beyond general anisotropy and non-stationarity, the appropriateness of spatial sampling designs may be affected by landscape composition and landscape structure, i.e., the areal proportions of LULC types and the spatial configuration of fields and landscape elements, respectively (Plotkin and Muller-Landau 2002). Regarding landscape composition, the uneven distribution of areal proportions of LULC types may be a challenge for the efficiency of sampling designs (Kivinen et al. 2006). Normally, agricultural landscapes are strongly dominated by either arable fields or grassland while subordinate land-use types and, particularly, linear landscape elements, such as hedges and field margins, have lower areal proportions and may be underrepresented by random or systematic sampling designs. In this regard, the magnitude of the difference in abundance and species richness between the dominating and subordinate LULC types may affect the efficiency of sampling designs. With respect to landscape structure, the efficiency of systematic sampling designs may depend on the scale of spatial auto-correlation, i.e., field size, in relation to the size of the grid used for drawing

sampling points (Ripley 2005). Comparing different agricultural landscapes, field sizes may vary strongly between less than a hectare to hundreds of hectares (Lesiv et al. 2019). Thus, it is questionable whether systematic sampling design will perform equally well in different types of agricultural landscapes.

In the present study, we assessed the efficiency of different spatial sampling designs in estimating landscape-level abundance and species richness of carabids in agricultural landscapes using a simulation approach. For this purpose, we created artificial landscape maps representing 2×2 km in nature that comprised LULC types typical of agricultural landscapes in temperate Europe, including linear elements. The landscapes were either dominated by arable land or by grassland. Further, we varied the degree of subdivision of the landscapes by controlling the number of fields (including both arable and grassland parcels) between 100, 500 and 1000 so that average field size varied between approximately 0.5 and 5 ha.

We aimed at answering the following research questions:

1. Which spatial sampling design is most accurate in estimating abundance/ species richness of carabids at landscape level?
2. Which sample size is required to get an accurate estimate of landscape-level abundance/species richness and which sample size is optimal with respect to the trade-off between sampling effort and accuracy?

3. Does the estimation accuracy and the required/ optimal sample size depend on landscape composition or subdivision?

As criteria for accurate estimates, we considered an acceptable deviation of $\pm 10\%$ for abundance and a detection rate of 80% for species richness.

Methods

Spatial sampling designs

We tested seven spatial sampling designs (Table 1) that appeared generally suitable for sampling biodiversity in agricultural landscapes and do not require prior knowledge of the spatial distribution of the variable of interest apart from land-surface classifications, such as LULC or habitat types, that are used in stratified designs (Wang et al. 2012). Examples of the spatial distributions of sampling points for the tested sampling design are shown in Fig. 1. We excluded sampling designs that take into account spatial auto-correlation because statistics of auto-correlation of occurrence and abundance of carabid species will usually not be available.

Empirical data of carabid occurrence and abundance

We aimed at compiling realistic local communities of carabids with respect to species composition, frequency and abundance on the landscape rasters

Table 1 Description and implemented parameter values of spatial sampling designs. Abbreviations: LULC: land use/ land cover

| Sampling design	Description	Parameter value
Random	Sampling points are placed randomly within the landscape	
Stratified random	Each LULC type receives the same number of sampling points, which are placed randomly within the area of the respective type	
Area-proportional stratified random	Each LULC type receives a number of sampling points proportional to its area in the landscape, which are placed randomly within the area of the type	
Systematic random	The landscape is superimposed by a regular grid and each grid cell receives N sampling points that are placed randomly	$N = 1$
Systematic regular	Grid-based as above, but all sampling points have the same X and Y offset from the centre of the grid cell	$N = 1$
Systematic unaligned	Grid-based as above, but the X offset from the centre of the grid cells is constant within each row while the Y offset is constant within each column of the grid	$N = 1$
Clustered	N clusters of sampling points are scattered around N random locations in the landscape	$N = 2$
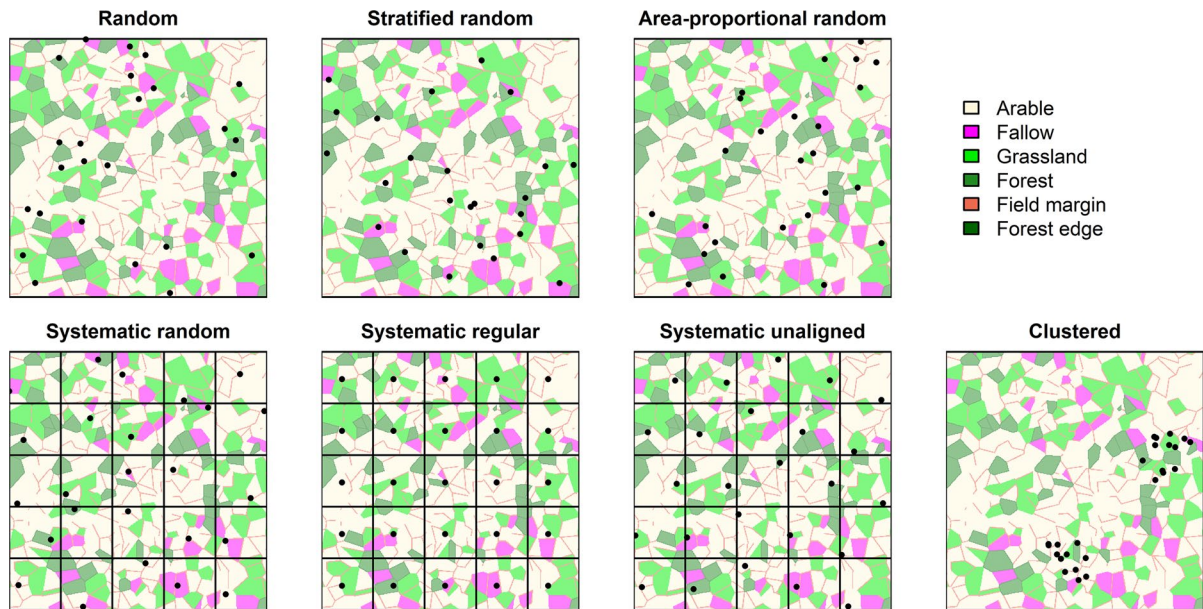
Fig. 1 Examples of spatial distributions of sampling points in landscapes of 2×2 km for the seven spatial sampling designs tested in the simulation study

used for the simulations of spatial sampling. For this purpose, we used field data from a research study performed by the Leibniz Center for Agricultural Landscape Research (ZALF) in 2011 and 2012 in Scheyern, Bavaria, Germany (Glemnitz et al. 2013) where carabids were sampled in all major LULC types in a section of an agricultural landscape of approximately 500×500 m. Pitfall traps were placed on neighbouring LULC patches comprising two arable fields with different crop types, one fallow field, one meadow, one field margin, one forest edge and one forest interior site (Fig. S1). In each habitat, five traps were installed. The distance between sampling sites was minimum 100 m. The trap replicates within single sites had a regular distance of 10 m.

The traps were sampled every two weeks from beginning of April until end of October in both years, summing up to 29 dates and 890 pitfall-trap samples in total. Carabid beetles were caught and preserved with a 4% formaldehyde solution with a drop of detergent. Pitfall content was sorted out and stored in 70% ethanol until identification to species level using Müller-Motzfeld (2004). The nomenclature followed Köhler and Klausnitzer (1998).

In addition to the field data described above, we compiled a hybrid dataset based on data from multiple projects that conducted samplings of carabids with pitfall traps in various regions of Germany. However different studies used different exposure times and numbers of pitfall traps per site and, further, the sampling periods varied considerably. Therefore, we could only use a subset of available data with similar sampling periods. We scaled the data to a standard exposure time and one trap per site. This second dataset was used to check if results were consistent for different sources of carabid data.

Pitfall traps introduce several biases into occurrence and abundance data of arthropods (Zaller et al. 2015). The number of trapped individuals represents the activity density around the trap, rather than real abundance, and depends on species' traits so that some species are overrepresented while others may be missing in the traps (Knapp et al. 2020). Further, the number of catches depends on temperature (Saska et al. 2013; Engel et al. 2017), and the efficiency of pitfall traps varies with choice of collecting fluid and trap size (Koivula et al. 2003; Schmidt et al. 2006). We chose to use pitfall traps in awareness of these limitations, because it is the most widely used method in practice and other sampling methods are associated with other, partly even greater limitations (Zaller et al. 2015).

Species occurrence was calculated by dividing the number of each individual pitfall sample containing the respective species by the total number of samples for each LULC type. The species' abundances were calculated as the mean number of individuals in pitfall-trap samples where they were present per LULC type. Accordingly, the standard deviations of species' abundances were also calculated based on samples containing the species.

Highest total carabid abundances were found in arable fields, followed by forest edges (Fig. 2). Medium abundances were found on fallow land, grassland and in the forest interior, and the lowest on the field margins. Species numbers showed the same ranking among habitats, but with lower variation.

Simulation model of spatial sampling

The simulation of spatial sampling comprised, first, the creation of raster-based artificial landscapes with different landscape compositions and subdivisions, then, the simulation of local carabid communities in landscape raster cells and, finally, the sampling of carabid communities with different spatial sampling designs and different sample sizes. The simulations were repeated 100 times. We implemented the simulation model in R version 3.6.0 using the packages NLMR (Sciaini et al. 2018), plyr (Wickham 2011), raster (Hijmans 2021), rgdal (Bivand et al. 2021), rgeos (Bivand and Rundel 2021), sf (Pebesma 2018), sp (Bivand et al. 2013), spatialEco (Evans 2021) and

spatstat (Baddeley et al. 2015). The program code is contained in the file 'SpatSam_1.0.R' that can be accessed through the GitLab project SpatSam (https://gitlab.com/jan.thiele/spatsam). A detailed description of the simulation model is provided in the documentation file in the GitLab project.

The landscapes represented 2.2×2.2 km in nature and the size of raster cells was 5×5 m. We used two types of landscape compositions: (a) arable-dominated and (b) grassland-dominated. The subdivision of landscapes was controlled by the number of fields, which was either 100, 500 or 1000. We first created the respective number of fields using Voronoi polygons. Then, we assigned areal Land-use/ land-cover (LULC) types to the polygons based on predefined proportions that are given in Table 2. Thereafter, all boundaries of forest polygons were converted to spatial lines of class 'forest edge', while 50% of the remaining polygon boundaries were assigned to the class 'field margin'. Finally, we created landscape rasters by first rasterising the LULC polygons and, thereafter, inserting the also rasterised spatial lines of forest edges and field margins. For an example of a simulated landscape see Fig. S2. The proportions of LULC types varied to some degree among the simulated landscapes due to randomness in the creation and classification of Voronoi polygons. The mean proportional areas of areal LULC types were lower than the predefined values because parts of their polygon margins were reclassified as linear landscape elements. Field margins covered between 1.1 and 8.5%

Fig. 2 Abundance sums per pitfall trap and number of species (species pools) of carabids by habitat types: arable fields (arable), fallow land (fallow), agricultural grasslands (grassl), field margin (margin), forest edge (f. edge) and forest
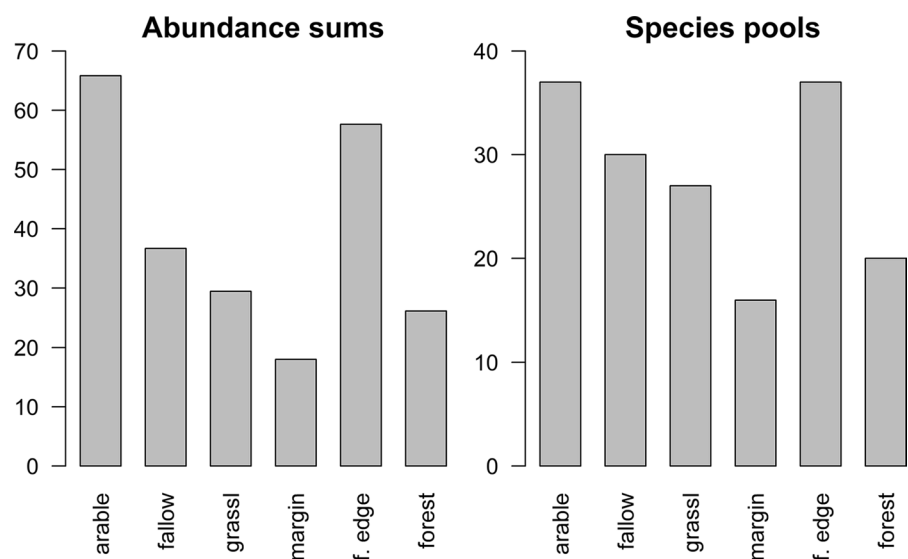
Table 2 Proportions (%) of land-use/ land-cover (LULC) types for creating artificial landscapes

Landscape composition	Land-use/land-cover type					
	Arable fields	Fallow fields	Grassland	Forest	Field margin	Forest edge
Arable-dominated	60	10	20	10	50	100
Grassland-dominated	18	9	55	18	50	100

For areal LULC types, proportions relate to the number of polygons in the landscape, while for linear landscape elements, i.e. field margins and forest edges, proportions relate to the number of line segments of the boundaries of field and forest polygons, respectively

and forest edges between 0.3 and 5.9% of the area (Fig. S3), with proportions increasing with the number of fields in the landscape.

The local carabid communities were simulated based on LULC-type specific frequencies and abundances. For each of the 57 carabid species, we first defined the presence cells by drawing a random sample of raster cells proportional to the species' frequency stratified by LULC type. For the simulation of abundances, we transformed the empirical mean abundances in pitfall traps to abundances per raster cell. For this purpose, we estimated the effective sampling area of a pitfall trap using a Gaussian kernel of 5×5 raster cells with a standard deviation of 0.6. The kernel size of 625 m^2 was chosen according to estimated catchment areas of pitfall traps for carabids of 620–640 m^2 (Bergeron 2019). This Gaussian kernel was later also used for sampling of carabids assuming a catch rate of 40% of carabid individuals on the central raster cell (Table S1). The effective area of the sampling kernel was 22.7 m^2 meaning that the number of individuals caught in the pitfall trap represented abundance per 22.7 m^2. This abundance was then scaled to the size of raster cells (25 m^2) and divided by catch rate so that it represented total abundances of the species including individuals that were not caught in the trap. Then, we drew abundances of the respective species for each presence cells from either a Poisson distribution to avoid negative values when the mean abundance was smaller than five or, otherwise, a Gaussian distribution parameterised with the transformed mean abundance and transformed standard deviation. In this way, we created random distribution patterns of presences and abundances of species per habitat type (see Fig. S2 for an example).

After that, we simulated spillover of carabid individuals from grasslands, field margins und forest edges into arable fields using a Gaussian kernel and multiplying the kernel values with the species' abundances in habitat cells adjacent to the arable fields. Spillover accounted for approximately 1–4% of total carabid abundance in the landscape depending on landscape composition and subdivision (Fig. S4).

The spatial sampling was simulated with sample sizes of 4, 9, 16, 25, 36 and 49 pitfall traps by drawing the respective number of sampling points from the central 2×2 km of the landscape rasters. We omitted the outer 100 m of the landscape to avoid boundary effects. For stratified random sampling, sample sizes were 6, 12, 18, 24, 36 and 48 because, by design, all six LULC types had to receive the same number of sampling points. The proportions of LULC types among the sampling points varied in the random sampling designs, but areal proportions were generally well represented in samples of size 16 and above (Fig. S5).

We used the R function rSSI (package spatstat) with a minimum distance of zero for simple random sampling. Further, we used sampleStratified (raster) for stratified random sampling, rsyst (spatstat) for systematic regular, rstrat (spatstat) for systematic random, spsample (sp) for systematic unaligned and clustered, and own source code for area-proportional random sampling.

For each sampling point, the presence and abundance of all species was recorded and, for each sample, the total species richness and projected total abundance of carabids at landscape-level was calculated. We projected total abundance by calculating the sample mean, dividing by the effective sampling area and multiplying by the total landscape area. For stratified random sampling, we additionally calculated weighted estimates of landscape-level abundance by area of the LULC types.

We conducted sensitivity analyses for those model parameters that were uncertain, i.e., the standard

deviation of the Gaussian sampling kernel and the catch rate. For this purpose, we varied the standard deviation between 0.1 and 1.5 (values: 0.1, 0.3, 0.6, 0.9, 1.2, 1.5) and the catch rate between 0.1 and 1.0 (values: 0.1, 0.2, 0.4, 0.6, 0.8, 1.0). Further, we tested model variants with clustered spatial distribution of carabids, with a uniform sampling kernel instead of Gaussian, and without spillover.

Analysis of simulation results

We assessed the estimation accuracy of landscape-level abundance and species richness of carabids with Root Mean Squared Error (RMSE) calculated with estimated and true abundance/ species richness over the 100 repetitions of the simulations. RMSE was calculated separately for each combination of landscape composition and subdivision, sampling design and sample size. Additionally, we calculated Mean Percentage Error (MPE) to assess if abundances were over- or underestimated.

We calculated 95% confidence intervals of RMSE based on 1000 bootstrap resamples of estimated and true abundance/ species richness in the 100 repetitions. We further used the RMSE values from the bootstrap resamples to conduct pairwise Wilcoxon tests (function 'pairwise.wilcox.test') comparing all sampling methods and sample sizes within each landscape type, i.e., combination of landscape composition and subdivision. To correct for multiple testing, we applied Benjamini-Hochberg's adjustment of p-values in the Wilcoxon tests.

Further, we tested the estimation accuracy with equivalence tests against the true values of abundance/ species richness. Regarding abundances, we applied paired Two One-Sided Tests (TOST) using the function 'tost' (package 'equivalence'; Robinson 2016) with a region of similarity of $\pm 10\%$. For species richness, we conducted one-sided equivalence tests (non-inferiority tests) using the function 'equiv. test' (package 'eqivUMP') against 80% detection of species.

Results

Abundance

At a sample size of 25 pitfall traps per 2×2 km landscape, area-proportional random sampling was most accurate in estimating landscape-level carabid abundance, followed by systematic random and systematic regular sampling according to RMSE (Fig. 3). Differences in accuracy between sampling designs were significant, except for systematic random and regular sampling in grassland-dominated landscapes (Table 3). All sampling designs but stratified random sampling provided estimates of abundance equivalent to the true values according to Equivalence Tests with a region of similarity of $\pm 10\%$ (p < 0.05). When using area-weighted estimates of landscape-level abundance, stratified random sampling also yielded results equivalent to the true values and performed particularly well in grassland-dominated landscapes, but moderately in arable-dominated ones (Fig. S6). The rankings of sampling designs did not change consistently with increasing sample sizes. Swapping of ranks

Table 3 Rankings of spatial sampling designs by estimation accuracy of landscape-level abundance and species richness of carabids in arable- and grassland-dominated landscapes of 2×2 km with 500 fields at sample size 25

Target value	Landscape type	Sampling designs						
		Random	Stratified random	Area-prop. strat-ified random	System. random	System. regular	System. unaligned	Clustered
Abundance	arable	<u>4</u>	7	<u>1</u>	<u>2</u>	<u>3</u>	<u>5</u>	<u>6</u>
	grassl	<u>5</u>	6	<u>1</u>	2	2	4	<u>7</u>
Species richness	arable	3	1	2	4	5	5	7
	grassl	3	1	2	5	4	7	6

Designs with different ranks have significantly different mean Root Mean Squared Error (RMSE) in Wilcoxon tests calculated on 1000 bootstrap resamplings. Rank numbers with underlining indicate that abundance/ richness estimates were equivalent to the true values in Equivalence Tests (p < 0.05) with a region of similarity of $\pm 10\%$ of mean true abundance or > 80% of true species richness
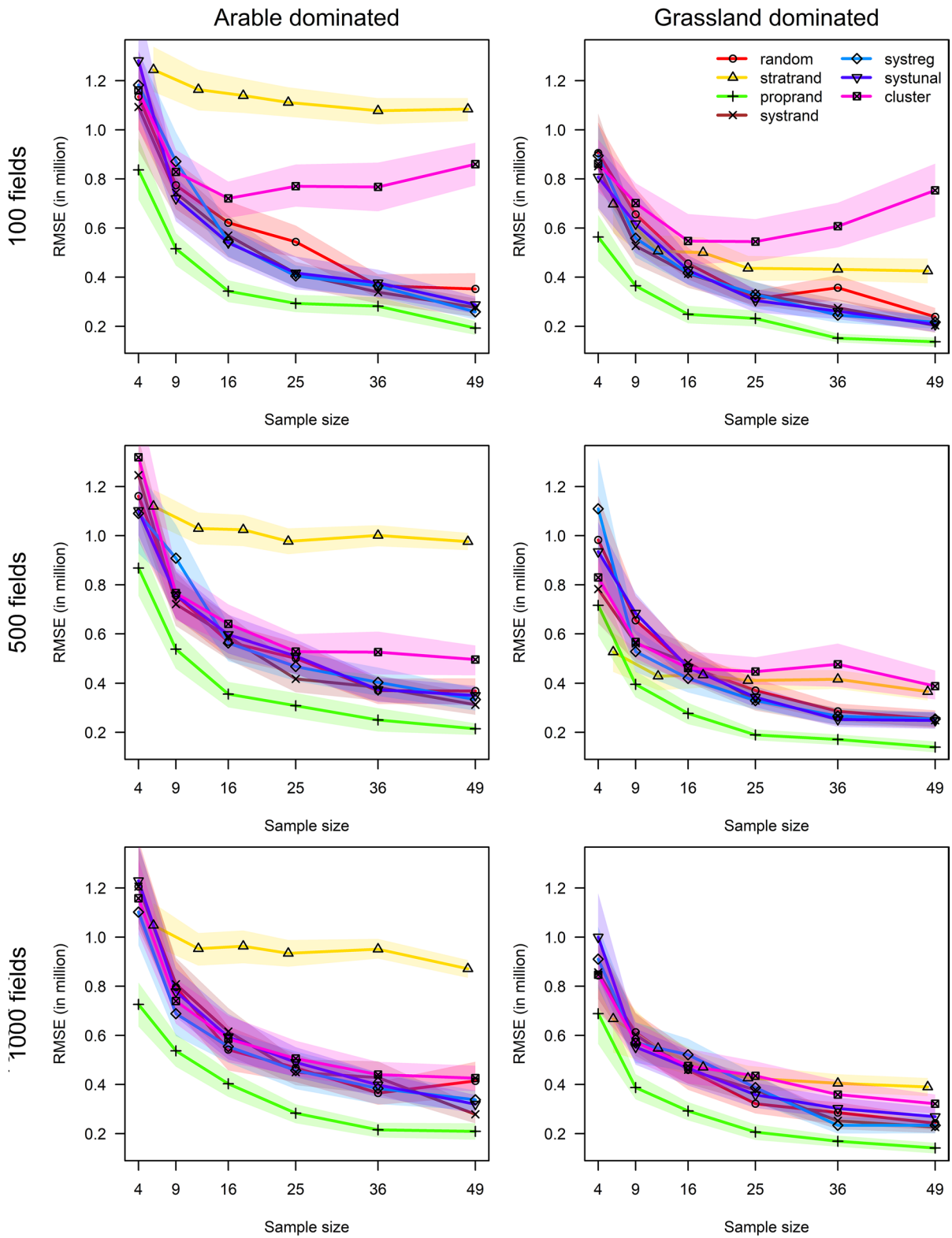
◄**Fig. 3** Accuracy of estimated total carabid abundance at landscape level expressed as root mean squared error (RMSE) of 100 simulation runs versus sample size for seven spatial sampling designs. Simulations were run for six landscape types with either arable fields or grasslands as the dominating landuse and different degrees of subdivision (100, 500 or 1000 fields per landscape). Shaded areas indicate 95% confidence intervals based on 1000 bootstrap resamples

occurred only among designs that showed moderate differences in RMSE (Fig. 3).

Even though most sampling designs provided abundance estimates equivalent to the true values at a sample size of nine, estimation accuracy further increased, i.e. RMSE decreased, with increasing sample size with the exception of clustered sampling in coarse-grained landscapes (100 fields) and stratified random sampling in arable-dominated landscapes (Fig. 3). The curves of RMSE levelled off at sample sizes between 25 and 36. Further increases in estimation accuracy from 36 to 49 sampling points were significant (Wilcoxon tests on 1000 resamples $p < 0.001$; exceptions: random sampling in arable-dominated landscape with 500 fields: $p = 0.460$; systematic regular sampling in grassland-dominated landscape with 1000 fields: $p = 0.572$), but moderate except for area-proportional random sampling in arable-dominated landscapes with 100 fields.

In general, area-proportional random sampling was the most accurate design for estimating abundances followed by systematic random and regular sampling being the next best designs with moderate differences in performance. This main pattern did not change with landscape composition and subdivision. However, stratified random sampling performed much worse in arable-dominated landscape compared to grassland-dominated ones. Further, clustered sampling showed improving accuracy with increasing subdivision of the landscapes albeit never reaching top ranks.

Deviations of estimates of landscape-level carabid abundance from true values did not show consistent patterns of over- or underestimation for most of the sampling designs, but stratified random sampling substantially underestimated abundances in arable-dominated landscapes, whereas it overestimated abundances in grassland-dominated ones (Figs. S7–S9).

In the sensitivity analysis, increasing the standard deviation of the Gaussian sampling kernel and the catch rate decreased estimation errors, but otherwise showed results comparable to the main model. That is area-proportional random sampling was confirmed to be most accurate, while stratified random and clustered sampling performed poorly. Only the intermediate sampling designs showed frequent changes of ranks among model variants. Also model variants with clustered instead of random distribution of species in the landscape, or without simulating spillover or with uniform sampling kernel yielded similar results as the main model. Finally, we also found virtually the same results when using the hybrid dataset of pitfall samples compiled from several studies conducted in different regions of Germany (Fig. S10).

Species richness

Stratified random sampling was the best design for estimating species richness of carabids at landscape level at a sample size of 25 (Table 3; N.B.: for stratified random sampling the sample size was 24 in fact) and at almost all other sample sizes (Fig. 4). Differences in accuracy among the remaining sampling designs were moderate, apart from clustered sampling, even though they were statistically significant in Wilcoxon tests ($p < 0.001$, with few exceptions). The rankings of sampling designs did not change substantially with sample size. However, all sampling designs failed to detect 80% of landscape-level species richness. Stratified random sampling detected on average 56.3% of the species at a sample size of 25, while the maximum detection rate was 66.8% at 49 sampling points. The second best sampling design, area-proportional random sampling detected on average 55.8% of species at a sample size of 25 and up to 66.3% at sample size of 49 (cf. Fig. S11).

Estimation accuracy increased markedly with sampling size and did not level off at sample size 49. Landscape composition and subdivision did not affect the accuracy and ranking of sampling designs, except for clustered sampling that fell behind in coarse landscapes (100 fields). Model variants with different parameterisations of sampling kernels and different spatial distributions of species showed virtually the same results as the main model.
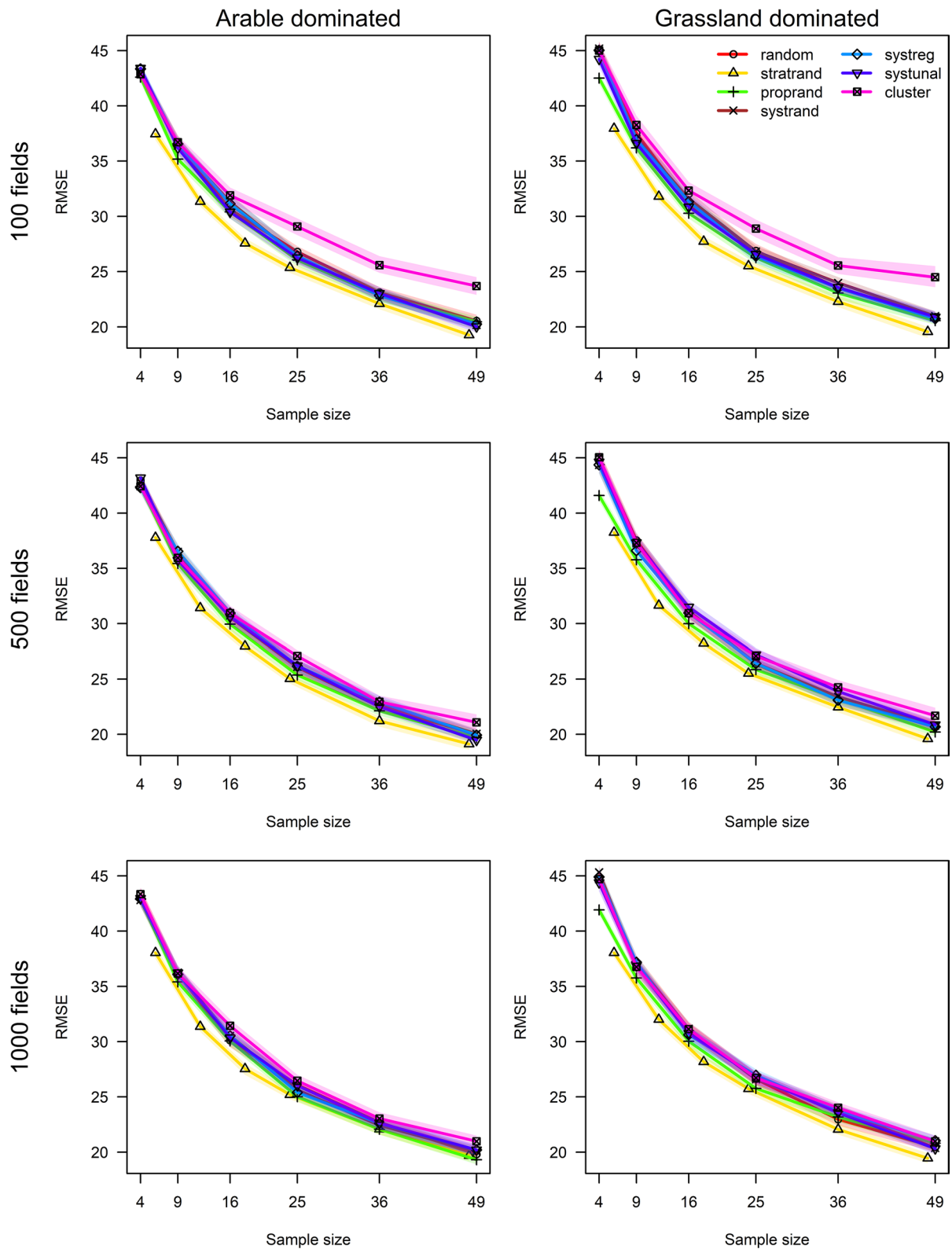
◄**Fig. 4** Accuracy of assessment of total carabid species richness at landscape level expressed as root mean squared error (RMSE) over 100 simulation runs versus sample size for seven spatial sampling designs. Simulations were run for six landscape types with either arable fields or grasslands as the dominating land-use and different degrees of subdivision (100, 500 or 1000 fields per landscape). Shaded areas indicate 95% confidence intervals based on 1000 bootstrap resamples

Discussion

Which spatial sampling design is most accurate in estimating abundance/species richness of carabids at landscape level?

The present simulations suggest that area-proportional stratified random sampling appears to be most accurate in case that a sound habitat classification is available. Further, systematic sampling designs perform better than simple random sampling in estimating landscape-level abundance of carabids in agricultural landscapes although the difference is mostly small. In contrast, stratified random and clustered sampling seem to be inappropriate for estimating carabid abundances.

The suitability of stratification depends on our knowledge of habitat types and their species composition and abundance of carabids. If the habitat classes comprised different carabid communities with varying composition and abundances that covered unknown proportions of the habitat-class areas, then area-proportional sampling might lead to wrong results. Hence, it should be assessed beforehand, if available habitat classifications are compatible with ecological requirements of carabid communities. Since carabids are one of the best-studied groups of insects and often used as bioindicators, such an assessment should be possible where databases with regionalised habitat preferences are available. However, if the knowledge about the habitats and their carabid communities is incomplete, systematic random or systematic regular sampling designs seem to be most suitable (Wang et al. 2012). A systematic design was also proposed for bees in agricultural landscapes (Scherber et al. 2019). Further, projects that investigate multiple species groups might face the problem that there is no single uniform habitat classification suitable for all taxa. Also in this case, systematic designs might be a robust choice.

Regarding species richness, the choice of sampling design appears to be less important, with the exception that clustered sampling is only suitable in fine-grained landscapes. Stratified random sampling is somewhat more efficient in detecting species, which is plausible because rare habitats that are potentially species-rich are sampled with the same intensity as the dominating LULC types. If, however, a research or monitoring project seeks to investigate abundances, in addition to species richness, the choice of sampling design should be based on the suitability for estimating landscape-level abundances of ground-dwelling insects.

In line with the present results, area-proportional stratified random sampling was more accurate than simple stratified random sampling in assessing abundance of marine benthos at regional scale (van Hoey et al. 2019) and also better than simple random sampling in estimating cover of intertidal benthic communities at landscape scale (Miller and Ambrose 2000). Further, another study on monitoring of amphibians and reptiles at national scale found that stratification by environment and protection status improved detection of species substantially (Carvalho et al. 2016). In contrast to our study, random sampling was found to be more efficient than systematic sampling in estimating frequencies of aquatic species in lakes and rivers in a simulation study (Marta et al. 2019).

Which sample size is recommendable for estimating landscape-level abundance and species richness?

According to equivalence tests with a region of similarity of ±10%, nine sampling points would be sufficient to accurately assess the abundance of carabids at landscape level in a 2×2 km agricultural landscape. However, estimation accuracy increases substantially up to 25 or 36 sampling points. Bearing in mind the trade-off between accuracy and sampling effort, we would suggest that a sample size of 25 could be a good choice for scientific studies. For wild bees, 25 samples on 1×1 km were found suitable in recent research (Scherber et al. 2019).

Comprehensive inventories of species richness of carabids at landscape level do not seem to be possible with the tested sample sizes (up to 49) in a single sampling period of two weeks. In addition, carabid species have different periods of peak abundance and activity throughout the year (Wang et al. 2014) and,

therefore, it is unlikely to detect all present species at one point of time. Hence, it seems recommendable to repeat sampling several times per year rather than to increase the sample size on a single time point in order to increase detection rates. In our simulations, detection of species depended on their frequencies in the landscape-raster cells which we took from empirical datasets of pitfall samples. If in reality species had higher frequencies, then a higher percentage of species could be detected in field samplings compared to our simulation.

Does estimation accuracy and optimal sample size depend on landscape composition or subdivision?

The accuracy and required sample sizes of most sampling designs were not affected substantially by landscape composition and subdivision. Marked landscape effects were observable only for stratified random and clustered sampling with respect to the assessment of landscape-level abundances.

Stratified random sampling was less accurate in the arable-dominated landscapes where it substantially underestimated carabid abundances. The reason for this is that the dominating arable-fields had the highest local abundances in our model (Fig. 2), but stratified random sampling placed 5 out of 6 sampling points in subordinate LULC types that all had lower local abundances. In contrast, stratified random sampling performed better in grassland-dominated landscapes, but overestimated abundances, because the local abundance in grasslands was closer to the overall mean, but lower. More generally, underrepresentation of the main LULC type in the samples distorts landscape-level estimates when the local mean deviates from the global mean. In conclusion, stratified random sampling is not suitable for assessing landscape-level abundance in cases where there is a dominating LULC type that has abundances well above or below the average.

Clustered sampling performed most poorly in coarse-grained landscapes, but substantially better in fine-grained ones. The likely reason for this is that most sampling points of a cluster could be located on a single large field in the coarse landscape, so that other LULC types were undersampled or not represented in the samples at all. Regarding shorter travel distances among sampling points, clustered sampling could possibly be a cost-effective alternative to other sampling designs in very fine-grained landscapes, but it proved to be consistently less accurate than other sampling designs over the gradient of subdivision tested in this study.

Generalisability of results

Regarding landscape composition and subdivision, it appears that the main results of the present study are generally valid for agricultural landscapes independent of the proportions of arable fields, grasslands and semi-natural habitats, except for stratified random sampling. Further, the results seem to be valid for all studies using pitfall traps regardless of sampling intervals and regional differences in the composition of arthropod communities, at least for agricultural landscapes in Central Europe, as results based on the hybrid carabid dataset were tantamount to the local dataset presented here. Thus, biases of pitfall samples introduced due to different sampling intervals and temperature effects (Schirmel et al. 2010; Saska et al. 2013; Engel et al. 2017) do not seem to affect the accuracy rankings of spatial sampling methods. However, we suggest to verify our results in regions with markedly different climate conditions or habitat inventories. Still, we would suggest that results are likely applicable to agricultural landscapes in other regions of the world as different landscape settings as well as spillover effects (yes or no) and spatial distributions of species (random or clustered) did not affect the results.

Different sampling methods introduce different biases into carabid community data (Zaller et al. 2015). Therefore, it is not safe to generalise the results to studies using other sampling methods for ground-dwelling carabids, such as emergence traps and suction sampling. The results should first be validated using field data collected with these methods. Further, verification of the results is needed for other groups of arthropods that show markedly different community structure or behaviour. In particular, our simulations are not transferable to flying insects, such as bees, butterflies and hoverflies. Assessments of spatial sampling designs for flying insects would require simulating their foraging behaviour in the landscape.

Conclusion

All systematic grid-based sampling designs are recommendable for sampling abundances of ground-dwelling insects in agricultural landscapes unless a sound habitat classification tailored for the target species group is available, in which case area-proportional stratified random sampling is most accurate. A sample size of 25 traps per 2×2 km landscape provides a good balance between accuracy and sampling cost. Regarding species richness, differences in efficiency between sampling designs are moderate except for clustered sampling which is less accurate. High sample size and/ or repeated samplings is more important than choice of sampling design for detecting species of ground-dwelling insects in agricultural landscapes.

Author contributions All authors contributed to the study conception and design. Collection of field data was performed by MG. Data preparation was conducted by GSaE and DG. Development and coding of the simulation model was performed by JT. The first draft of the manuscript was written by JT and all authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

Data availability The R script of the simulation model and the model documentation as well as auxiliary R scripts for data preparation, processing multiple runs and analysis of results, and the data used and generated during the current study are available in the GitLab repository 'SpatSam', https://gitlab.com/jan.thiele/spatsam.git.

Declarations

References

Baddeley A, Rubak E, Turner R (2015) Spatial point patterns: methodology and applications with R. Chapman and Hall/CRC Press, London

Beduschi T, Tscharntke T, Scherber C (2015) Using multi-level generalized path analysis to understand herbivore and parasitoid dynamics in changing landscapes. Landscape Ecol 30(10):1975–1986

Benedetti R, Piersimoni F, Postiglione P (2015) Sampling spatial units for agricultural surveys. Springer, Berlin

Bergeron C (2019) Estimates of catchment area of pitfall traps for carabids. ARPHA Conf Abstracts 2:e38518

Bivand R, Rundel C (2021) rgeos: Interface to Geometry Engine-Open Source ('GEOS'). R package version 0.5-8. https://CRAN.R-project.org/package=rgeos

Bivand R, Pebesma E, Gomez-Rubio V (2013) Applied spatial data analysis with R, 2nd edn. Springer, New York

Bivand R, Keitt T, Rowlingson B (2021) rgdal: bindings for the 'Geospatial' data abstraction library. R package version 1.5–27. https://CRAN.R-project.org/package=rgdal

Brooks DR, Bater JE, Clark SJ, Monteith DT, Andrews C, Corbett SJ, Beaumont DA, Chapman JW (2012) Large carabid beetle declines in a United Kingdom monitoring network increases evidence for a widespread loss in insect biodiversity. J Appl Ecol 49:1009–1019

Cardoso P, Barton PS, Birkhofer K, Chichorro F, Deacon C, Fartmann T, Fukushima CS, Gaigher R, Habel JC, Hallmann CA, Hill MJ, Hochkirch A, Kwak ML, Mammola S, Noriega JA, Orfinger AB, Pedraza F, Pryke JS, Roque FO, Settele J, Simaika JP, Stork NE, Suhling F, Vorster C, Samways MJ (2020) Scientists' warning to humanity on insect extinctions. Biol Conserv 242:108426

Carvalho SB, Goncalves J, Guisan A, Honrado JP (2016) Systematic site selection for multispecies monitoring networks. J Appl Ecol 53(5):1305–1316

Engel J, Hertzog L, Tiede J, Wagg C, Ebeling A, Briesen H, Weisser WW (2017) Pitfall trap sampling bias depends on body mass, temperature, and trap number: insights from an individual-based model. Ecosphere 8:e01790

Evans JS (2021) spatialEco R package version 1.3-6. https://github.com/jeffreyevans/spatialEco

Glemnitz M, Platen R, Krechel R, Konrad J, Wagener F (2013) Can short-rotation coppice strips compensate structural deficits in agrarian landscapes. Aspects Appl Biol 118:153–162

Hallmann CA, Sorg M, Jongejans E, Siepel H, Hofland N, Schwan H, Stenmans W, Müller A, Sumser H, Hörren T (2017) More than 75% decline over 27 years in total flying insect biomass in protected areas. PLoS ONE 12(10):e0185809

Hijmans RJ (2021) raster: geographic data analysis and modeling. R package version 3.5-2. https://CRAN.R-project.org/package=raster

Kivinen S, Luoto M, Kuussaari M, Helenius J (2006) Multispecies richness of boreal agricultural landscapes: effects of climate, biotope, soil and geographical location. J Biogeogr 33:862–875

Knapp M, Řezáč M (2015) Even the smallest non-crop habitat islands could be beneficial: distribution of carabid beetles and spiders in agricultural landscape. PLoS ONE 10(4):e0123052

Knapp M, Seidl M, Knappová J, Macek M, Saska P (2019) Temporal changes in the spatial distribution of carabid beetles around arable field-woodlot boundaries. Sci Rep 9:8967

Knapp M, Knappová J, Jakubec P, Vonicka P, Moravec P (2020) Incomplete species lists produced by pitfall trapping: how many carabid species and which functional traits are missing? Biol Conserv 245:108545

Köhler F, Klausnitzer B (1998) Verzeichnis der Käfer Deutschlands. Entomologische Nachrichten und Berichte Beiheft 4:1–185

Koivula M, Kotze J, Hiisivuori L, Rita H (2003) Pitfall trap efficiency: do trap size, collecting fluid and vegetation structure matter? Entomol Fennica 14:1–14

Kunin WE (2019) Robust evidence of insect declines. Nature 574:641–642

Lesiv M, Laso Bayas JC, See L, Duerauer M, Dahlia D, Durando D, Hazarika R, Sahariah PK, Vakolyuk M, Blyshchyk V, Bilous A, Perez-Hoyos A, Gengler S, Prestele R, Bilous S, ul Hassan Akhtar I, Singha K, Choudhury SB, Chetri T, Malek Ž, Bungnamei K, Saikia A, Sahariah D, Narzary W, Danylo O, Sturn T, Karner M, McCallum I, Schepaschenko D, Moltchanova E, Fraisl D, Moorthy I, Fritz S (2019) Estimating the global distribution of field size using crowdsourcing. Glob Change Biol 25:174–186

Marta S, Lacasella F, Romano A, Ficetola GF (2019) Cost-effective spatial sampling designs for field surveys of species distribution. Biodivers Conserv 28(11):2891–2908

McGarvey R, Burch P, Matthews JM (2016) Precision of systematic and random sampling in clustered populations: habitat patches and aggregating organisms. Ecol Appl 26(1):233–248

Miller AW, Ambrose RF (2000) Sampling patchy distributions: comparison of sampling designs in rocky intertidal habitats. Mar Ecol Prog Ser 196:1–14

Müller-Motzfeld G (2004) Die Käfer Mitteleuropas, Vol. 2: Adephaga 1: Carabidae, 2nd edn. Spektrum, Heidelberg

Pebesma E (2018) Simple features for R: standardized support for spatial Vector Data. R J 10(1):439–446

Plotkin JB, Muller-Landau HC (2002) Sampling the species composition of a landscape. Ecology 83:3344–3356

Ripley BD (2005) Spatial statistics. Wiley, New Jersey

Robinson A (2016) Equivalence: provides tests and graphics for assessing tests of equivalence. R package version 0.7.2. https://CRAN.R-project.org/package=equivalence

Saska P, van der Werf W, Hemerik L, Luff ML, Hatten TD, Honek A (2013) Temperature effects on pitfall catches of epigeal arthropods: a model and method for bias correction. J Appl Ecol 50:181–189

Scherber, Beduschi T, Tscharntke T (2019) Novel approaches to sampling pollinators in whole landscapes: a lesson for landscape-wide biodiversity monitoring. Landsc Ecol 34:1057–1067

Scherber C, Beduschi T, Tscharntke T (2021) A grid-based sampling approach to insect biodiversity monitoring in agricultural landscapes. In: Mueller L (ed) Exploring and optimizing agricultural landscapes. Springer, New York, pp 415–424

Schirmel J, Lenze S, Katzmann D, Buchholz S (2010) Capture efficiency of pitfall traps is highly affected by sampling interval. Entomol Exp Appl 136:206–210

Schmidt MH, Clough Y, Schulz W, Westphalen A, Tscharntke T (2006) Capture efficiency and preservation attributes of different fluids in pitfall traps. J Arachnol 34:159–162

Sciaini M, Fritsch M, Scherer C, Simpkins CE (2018) NLMR and landscapetools: an integrated environment for simulating and modifying neutral landscape models in R. Methods Ecol Evol 9:2240–2248

Seibold S, Gossner MM, Simons NK, Bluethgen N, Mueller J, Ambarli D, Ammer C, Bauhus J, Fischer M, Habel JC, Linsenmair KE, Nauss T, Penone C, Prati D, Schall P, Schulze ED, Vogt J, Woellauer S, Weisser WW (2019) Arthropod decline in grasslands and forests is associated with landscape-level drivers. Nature 574:671–674

van Hoey G, Wischnewski J, Craeymeersch J, Dannheim J, Enserink L, Guerin L, Marco-Rius F, O'Connor J, Reiss H, Sell AF, Vanden Berghe M, Zettler ML, Degraer S, Birchenough SNR (2019) Methodological elements for optimising the spatial monitoring design to support regional benthic ecosystem assessments. Environ Monit Assess 191(7):423

Wang J, Haining R, Cao Z (2010) Sample surveying to estimate the mean of a heterogeneous surface: reducing the error variance through zoning. Int J Geogr Inf Sci 24(4):523–543

Wang JF, Stein A, Gao BB, G Y (2012) A review of spatial sampling. Spat Stat 2:1–14

Wang X, Müller J, An L, Ji L, Liu Y, Wang X, Hao Z (2014) Intra-annual variations in abundance and species composition of carabid beetles in a temperate forest in Northeast China. J Insect Conserv 18(1):85–98

Wickham H (2011) The split-apply-combine strategy for data analysis. J Stat Softw 40(1):1–29

Zaller JG, Kerschbaumer G, Rizzoli R, Tiefenbacher A, Gruber E, Schedl H (2015) Monitoring arthropods in protected grasslands: comparing pitfall trapping, quadrat sampling and video monitoring. Web Ecol 15:15–23