Gillespie, Mark A.K.; Baude, Mathilde; Biesmeijer, Jacobus; Boatman, Nigel; Budge, Giles E.; Crowe, Andrew; Memmott, Jane; Morton, R. Daniel; Pietravalle, Stephane; Potts, Simon G.; Seenapathi, Deepa; Smart, Simon M.; Kunin, William E. 2017. **A method for the objective selection of landscape-scale study regions and sites at the national level**. *Methods in Ecology and Evolution*, 8 (11). 1468-1476, which has been published in final form at https://doi.org/10.1111/2041-210X.12779

DR. MARK ANDREW KUSK GILLESPIE (Orcid ID : 0000-0001-9808-4836)

DR. DEEPA SENAPATHI (Orcid ID : 0000-0002-8883-1583)

Article type : Research Article

Editor : Justin Travis

Title: A method for the objective selection of landscape-scale study regions and sites at the national level

Running title: Landscape site selection

Authors: Mark A. K. Gillespie[1,2], Mathilde Baude[3,4], Jacobus Biesmeijer[5], Nigel Boatman[6], Giles E Budge[6,7], Andrew Crowe[6], Jane Memmott[3], R. Daniel Morton[8], Stephane Pietravalle[6], Simon G. Potts[9], Deepa Senapathi[9], Simon M. Smart[8], William E. Kunin[1,10]

Affiliations

[1]School of Biology, University of Leeds, Leeds LS2 9JT, UK

[2]Department of Science & Engineering, Western Norway University of Applied Sciences, PB 133, 6851 Sogndal, Norway

[3]School of Biological Sciences, University of Bristol, Woodland Road, Bristol, BS8 1UG UK

[4] Collegium Sciences et Techniques EA 1207 LBLGC, Université d'Orléans, F-45067, Orléans, France

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1111/2041-210X.12779

[5]Naturalis Biodiversity Center, 2333 CR Leiden, the Netherlands & Institute for Biodiversity and Ecosystem Dynamics, University of Amsterdam, the Netherlands

[6]Fera Science Ltd (previously Food and Environment Research Agency), Sand Hutton, York, YO41 1LZ UK

[7]Institute for Agri-Food Research and Innovation, Newcastle University, Newcastle upon Tyne, NE1 7RU, UK.

[8]NERC Centre for Ecology & Hydrology, Library Avenue, Bailrigg, Lancaster LA1 4AP, UK

[9]Centre for Agri-Environmental Research, School of Agriculture, Policy and Development, University of Reading, Reading RG6 6AR, UK

[10] Stellenbosch Institute for Advanced Study (STIAS), Wallenberg Research Centre at Stellenbosch University, Stellenbosch 7600, South Africa.

Corresponding author: Mark Gillespie, Department of Science & Engineering, Western Norway University of Applied Sciences, PB 133, 6851 Sogndal, Norway, markg@hvl.no

Keywords: Experimental design, Site selection, Remote sensing, Insecticides, Floral resources, Habitat diversity, Honey bees, Pollinators, accidental experiments, natural experiments

Abstract

1) Ecological processes operating on large spatio-temporal scales are difficult to disentangle with traditional empirical approaches. Alternatively, researchers can take advantage of "natural" experiments, where experimental control is exercised by careful site selection. Recent advances in developing protocols for designing these "pseudo-experiments" commonly do not consider the selection of the focal region and predictor variables are usually restricted to two. Here we advance this type of site selection protocol to study the impact of multiple landscape scale factors on pollinator abundance and diversity across multiple regions.
2) Using datasets of geographic and ecological variables with national coverage, we applied a novel hierarchical computation approach to select study sites that contrast as much as possible in four key variables, while attempting to maintain regional comparability and national representativeness. There were three main steps to the protocol: i) selection of six 100 km x 100 km regions that collectively provided land cover representative of the national land average, ii) mapping of potential sites into a multivariate space with axes representing four key factors potentially influencing insect pollinator abundance, and iii) applying a selection algorithm which maximised differences between the four key variables, while controlling for a set of external constraints.
3) Validation data for the site selection metrics were recorded alongside the collection of data on pollinator populations during two field campaigns. While the accuracy of the metric estimates varied, the site selection succeeded in objectively identifying field sites that differed significantly in values for each of the four key variables. Between variable correlations were also reduced or eliminated, thus facilitating analysis of their separate effects.

4) This study has shown that national datasets can be used to select randomised and replicated field sites objectively within multiple regions and along multiple interacting gradients. Similar protocols could be used for studying a range of alternative research questions related to land use or other spatially explicit environmental variables, and to identify networks of field sites for other countries, regions, drivers, and response taxa in a wide range of scenarios.

Introduction

A major challenge facing researchers of large-scale ecological processes is to find appropriate methods to characterise relationships between land use and biodiversity patterns (Diamond 1983; Hargrove & Pickering 1992; Dilts, Yang & Weisberg 2010; Smart *et al.* 2012; HilleRisLambers *et al.* 2013). At the landscape scale, it is extremely difficult and expensive to apply a classical experimental approach involving establishing controls, manipulating "treatments", assigning large-scale experimental units to treatments randomly or achieving true replication (Hargrove & Pickering 1992; Rundlof *et al.* 2015). In response to these issues, landscape ecology as a discipline has developed a number of tools to study large-scale natural phenomena (Diamond 1983; Hargrove & Pickering 1992; Sagarin & Pauchard 2010; HilleRisLambers *et al.* 2013). Many landscape-scale observational studies take place within "natural" or "accidental experiments", making use of existing environmental variation occurring due to some sudden event or the gradual change brought about by humans, nature or both. When the goal of the study is to make statistical inferences about a broader population of landscapes, control of confounding factors can be applied through the careful, non-random selection of sites in so-called "pseudo-experiments" (Diamond 1983; Fahrig *et al*. 2011). This kind of selection is important to avoid common statistical design flaws such as spatial dependence of sites, the use of a only a portion of the range of landscape variables and collinearity between variables (Eigenbrod *et al*. 2011; Pasher *et al*. 2013)

The recent development of this form of site selection methodology appears to perpetuate two common drawbacks (Table 1): a) the region(s) within which the study sites are selected are not explicitly considered, and b) the number of predictor variables is restricted to two (although see Watts *et al*. 2016). In this study, we argue that some research questions require that the broader study regions are representative of some larger area to enhance generalisability of results. Such regions should also be free from the potential biases and problems of repeatability introduced by only studying well-known landscapes close to the study base or research institution (Dilts, Yang & Weisberg 2010). In addition, while there is a suitable method to select study sites that differ as much as possible in values of two variables (Fahrig *et al*. 2011), future studies seeking to disentangle multiple interacting drivers at large-scales will require a more advanced protocol. Watts *et al*. (2016) present the most promising of approaches to this need, developing a protocol that selects study sites that differ between three variables simultaneously. However, their protocol was not designed for hypothesis testing, was not applied to standardised sites and selected sites within subjectively chosen regions.

Our site selection protocol brings together the best aspects of its predecessors, enhances the objectivity and control of site selection, improves the description and testing of the protocol and allows application of the method to a broader array of situations. The method was originally developed to study the links between land use / management variables and insect pollinator populations and communities, but the approach is generic and could be used at a range of spatial scales and applied to almost any taxa or system. The objectives of the site selection methodology were to improve on previous landscape-scale pseudo-experimental designs by: i) enhancing objectivity of region selection (i.e., using a systematic approach with a transparent methodology which could be readily reproduced by other researchers), ii) enabling the study of several key factors simultaneously, and interactions between them, by selecting sites contrasting along multiple axes, and iii) enhancing the generality of results by selecting sites from areas that are representative of an entire country. To do this, national datasets were used first to select a set of focal regions that would be representative of Britain, and then to characterise each potential field site within those regions in terms of four key landscape-scale metrics that are thought to affect insect pollinator populations (habitat diversity, floral resource availability, insecticide loadings, managed honey bee density). Field sites were chosen to contrast as much as possible in each of the four key metrics while attempting to maintain regional comparability and representativeness. Verification of the protocol was conducted by validating the values of the four metrics through *in situ* surveys. The data demonstrate that landscape scale variation can be estimated using available national datasets, and thus suggest that similar approaches may be effective in addressing other large-scale issues.

Methods

The site selection protocol consists of three parts: 1) focal region selection, 2) assigning values of key variables to potential sites within each region, and 3) a site selection algorithm. This is followed by validation of the variable estimates used in site selection. These aspects are outlined briefly below with full details given in the Supporting information.

Focal Regions

To simplify field logistics and costs by limiting the amount of travel between sites, it was decided to first select six representative "focal regions" of 100 x 100 km, and then choose study landscapes within them. The regions were selected to be as representative as possible of the British landscape across vegetation and environmental gradients and the number of regions was chosen as the minimum number to allow sufficient statistical power for paired contrasts. However, the protocol could easily be applied to a different number of regions. The selection of focal regions began with two 100 km resolution grids: the standard UK Ordnance Survey grid at 100 km resolution and a second grid diagonally offset by 50 km to the east and north. The second grid was used to double the pool of regions from which to choose. All possible six-region combinations that did not include adjacent or overlapping cells were examined. For each six-region combination, the area of each broad habitat (from the 2007 Land Cover Map (LCM2007); Morton *et al.* 2011) was summed and the proportional contribution to the overall area calculated. A national proportional contribution for each habitat type was also calculated. For each habitat type, the Euclidian distance between the six-region proportion and the national proportion was calculated, and then a mean distance for all habitat types was taken. This distance then corresponds to how well the six-region combination represents Britain in terms of land cover categories. This process was also completed for ITE Land Classes (Bunce *et al.* 1996) which represent topography, climate and human infrastructure. The combination of six regions that had the shortest mean distance for both classification schemes was considered most representative of Britain, and was chosen as the set of focal regions to be studied.

Survey sites

The aim of the site selection protocol was to identify sites that contrasted as much as possible in four landscape-scale metrics: 1) habitat diversity, 2) floral resource availability, 3) insecticide loadings and 4) managed honey bee density. These four metrics were chosen because previous studies have demonstrated that they may be important drivers of local pollinator population decline in the UK. Strong links have been made between pollinator populations and the complexity of the landscape (Shackelford *et al.* 2013), the diversity and density of floral resources in agricultural settings (Potts *et al.* 2003; Gabriel & Tscharntke 2007) and increased insecticide usage (Rortais *et al.* 2005; Brittain *et al.* 2010). There is also evidence that managed stocks of honey bees can affect the condition of wild pollinator stocks either through spill-over of parasites (e.g., Evison *et al*. 2012) or through competitive interactions (Goulson & Sparrow 2009; Elbgami *et al.* 2014), although the landscape-scale population impact of honey bees on wild pollinators remains untested. In order to study the effects of these four factors individually and in combination, 16 sites in each study region were sought. We wanted these 16 sites to represent every possible combination of "high" and "low" values of each metric (i.e., site 1 = relatively "high" values for all four metrics, site 2 = "high" for three metrics and low for one metric, and so on) in a similar fashion to a full-factorial experiment. To this end, we used a computer algorithm technique to select sites with extreme values of each metric, as outlined below and in more detail in Appendix S1 (Supporting information).

Data sources and manipulation

Datasets were compiled using the UK Ordnance Survey National Grid reference system, the system of geographic grid references in the UK. The finest scale at which most agricultural and biodiversity datasets are available is the "tetrad" scale (2 x 2 km). Given the relatively high mobility of many pollinating insects (Westphal, Steffan-Dewenter & Tscharntke 2006), we opted to define our sites at this scale. For each of the 2,500 potential sites or tetrads within a 100 x 100 km region, a value for each of the metrics was calculated from national datasets. Full details of the calculations are given in Appendix S1 (Supporting information), but they are briefly outlined here:

1) **Habitat diversity** was calculated as a Shannon diversity index of broad habitats present, with each weighted by the area covered within each candidate tetrad. Habitat areas were derived from the LCM2007 (Morton et al. 2011).
2) **Floral resource availability** was calculated from nectar data only, as pollen data are less well recorded for British plants. This variable is expressed in terms of kilograms of sugar per hectare per year, and was derived by a) estimating flowering plant species cover per unit area of each habitat type in each site by combining finely-resolved regional vegetation quadrat data from Countryside Survey 2007 (CS2007; Carey *et al.* 2008) with the satellite-derived LCM 2007, b) modelling nectar sugar values for the 220 commonest insect-pollinated species based on published values for 124 species at the time of the study (see Table S2 for details and references), c) accounting for additional floral resources in mass-flowering crops, agri-environment schemes and in organic arable fields.
3) **Insecticide loadings**, a score of the hazard to bees of different insecticide types and application rates, were calculated by multiplying the area under cultivation of each of 36 crop groups within the sites estimated from national agricultural statistics, by a regional hazard score for agrichemicals used on that crop group, derived from Pesticide Usage Survey data for each crop combined with honey bee toxicity data for each insecticide applied.

4) **Managed honey bee population density** was estimated from data held by the national "Beebase" database (www.nationalbeeunit.com). The number of adult bees present in mid-summer for an average colony was estimated and this was combined with the typical number of colonies present in each of three apiary classes. Honey bee density in surrounding landscapes was modelled by using published honey bee foraging data (Waddington *et al.* 1994; Beekman & Ratnieks 2000). The apiary location was used as a centroid and the estimated number of honey bee foragers grouped into concentric 200 m bins (see Supporting information).

Site selection algorithm

Once assigned, the metric values were standardised by a Box-Cox transformation and converted to z scores (zero-centred), so that a score below zero for a metric corresponded to a "low" value relative to regional norms, and a score above zero represented a "high" value. The objective of the algorithm was to select a combination of 16 sites within a 100 x 100 km focal region to maximise the width of each of the four gradients sampled as well as the orthogonality between them. The number of ways of drawing unique sets of 16 sites from the 2,500 options in a focal region is enormous ($1.06055 * 10^{41}$ combinations). It was therefore essential to reduce computing time by constraining the site combinations using a series of design criteria. These criteria included removing the sites closest to the mean value for any of the four variables, restricting the maximum distance between sites within a cluster to 50 km (for logistical reasons), restricting the amount of urban and water cover allowed per site, and ensuring topographic comparability between sites (e.g., to avoid comparing sites on mountain tops vs valley floors). See Appendix S1 (Supporting information) for full details of the selection criteria. Once a feasible combination of field sites had been selected, landowners were identified and contacted for access permission. If access permission was refused to more than 30% of the site, the next feasible combination of field sites was chosen.

Site selection: validation

As the four metrics were all assessed indirectly with varying degrees of reliability, their values were validated during a two-year field campaign. This aim of this fieldwork was both to validate the metrics and to sample the field sites for wild pollinators. The full details of the validation processes are given in Appendix S1 (Supporting information) but are outlined briefly here:

1) **Habitat diversity** values were validated during field surveys by confirming or correcting the habitat types as mapped in the LCM2007. Corrected habitat areas were then used in new diversity index calculations.
2) **Floral resource availability.** Validation for this metric required several stages: a) actual floral reward production per flower per day was sampled for 175 species, and remodelled for a further 62 (2012) and 86 (2013) species (Baude *et al.* 2016), b) transect surveys were conducted to assess actual floral cover of each species for each broad habitat within each site, c) data from (a) and (b) were combined with corrected habitat areas to calculate the total floral resource per site.
3) **Insecticide loadings** were collated by conducting questionnaire surveys of all land managers for land within the field sites. The response rate to these questionnaires was approximately 50%, corresponding to an area of approximately 30% of the field sites. It was not possible therefore to validate the entire metric. Instead, direct comparison was made between the estimated and measured values for the fields covered by the questionnaire responses. Field values were summed for each tetrad.

4) **Managed honey bee density** was assessed by surveying each site using field observations along the predetermined transects used for floral resource validation, and using pan-trapping. Pan traps were set out on good weather days primarily to sample the wild pollinator community and any caught honey bees were added to the density count.

Results

Region and site selection

The six focal regions and 96 survey sites chosen by the protocol are shown in Fig. 1. From southeast to northwest, the focal regions covered parts of 1) Cambridgeshire, Suffolk and Norfolk, 2) Wiltshire and Gloucestershire, 3) Staffordshire, Cheshire, Shropshire and North East Wales, 4) North Yorkshire and Cumbria, 5) Ayrshire, Lanarkshire and East Renfrewshire, and 6) Inverness-shire.

Survey sites (listed in Table S6, Supporting Information with metric estimates) were generally well selected in line with the criteria of the protocol, with some exceptions. Fig. 2 illustrates the contrasting values of the four estimated metrics for the Cambridgeshire/Suffolk region as an example. The goal of this part of the selection protocol was to effectively ensure that the bars were as high as possible for the "high" values (positive values in Fig. 2) and as low as possible for the "low" values (negative values in Fig. 2). In practice, we appreciated that the indirect assessment of focal variables (and regression towards the mean) would tend to narrow or erase the gap between high and low categories, such that each axis should be treated as continuous rather than categorical. Our protocol, however, helps ensure that as wide a range of variation as possible is sampled. Furthermore, although it was not a site selection criterion, the site selection protocol removed the inherent correlation between the estimated values of the four metrics both for all regions (Table 2), and within individual regions (Fig. S4 – S6).

Validation

In order to validate the site selection protocol, the observed values of each of the four metrics were tested against the predictions derived from national datasets using simple Spearman's rank correlation tests (R base package; R Core Team 2014). These correlations are shown graphically in Fig. 3 and the coefficients are given in Table 3, together with results from linear mixed effects models using measured values as response variable, predicted values as explanatory variable, and region as random effect. Mixed models were performed using the package *nlme* in R 3.1.1 (R Core Team 2014), and were considered valid following inspection of residuals for normal distribution, heteroscedasticity and influential values (Zuur et al. 2009). All four metrics showed significant positive relationships between the observed and predicted values. According to the correlation coefficients, the best predicted metric was habitat diversity, followed by insecticide loadings, floral resources, and honey bee density. However, it should be noted that the insecticide loading comparison omits tetrads for which questionnaire responses were not received, and tetrads for which measured insecticide could be assumed to be zero due to the absence of arable fields. If the latter are included, the Spearman's rank correlation coefficient is 0.57 ($p < 0.001$) but the slope of the regression is only 0.25 ($p<0.01$).

In terms of the correlations between validated metrics, there were significant relationships between the metrics for three out of the six pair-wise comparisons overall (Table 4), although the correlation coefficients were all below the commonly used threshold of 0.7 for including variables in the same analysis. Measured floral resources was significantly correlated with measured honey bee density (Spearman's $\rho = 0.31$, $p = 0.002$) and with measured insecticide loadings (Spearman's $\rho = -0.47$, $p < 0.05$). In addition, measured honey bee density was strongly linked to measured insecticide loadings (Spearman's $\rho = 0.54$, $p < 0.05$). However, for the individual regions (Fig. S7 – S9) the only significant correlations were for measured habitat diversity vs measured honey bee density in Inverness (Spearman's $\rho = 0.54$, $p = 0.03$; Fig. S7), measured insecticide loadings vs measured habitat diversity in Wiltshire (Spearman's $\rho = -0.92$, $p < 0.01$; Fig S9) and for measured honey bee density vs measured insecticide loadings in Cambridgeshire (Spearman's $\rho = -0.65$, $p = 0.04$; Fig. S9).

Discussion

The methodology described here aimed to build on previous site selection protocols to select sites that varied in four main gradients, while at the same time ensuring comparability between sites and representation of Britain more widely. Although estimations of the four metrics were made with some uncertainty, the low level of correlation between verified metrics at the regional and national scales suggest that the site selection method provides a suitable sample of sites for investigating links between land management and pollinator biodiversity.

Region selection

One of the main differences between previous approaches and our protocol is in the objective selection of study regions, chosen here to represent Britain in terms of land class and land cover variables. Regions are often chosen in landscape studies because they are well known and have been used several times before in previous work. This manner of selecting focal regions is sufficient for studies that aim to understand basic or local mechanisms or processes. For example, Watts *et al.* (2016) chose two regions of the UK due to previous knowledge of the areas and of the variation in woodland habitats. Such a selection approach was expedient and suitable for the authors' study question, which focused on landscape conservation and links between woodland biodiversity and gradients of woodland characteristics. Furthermore, the inferential scope of this study is likely restricted to British lowland woodlands within these two regions. By contrast, our research project sought to link the regional variation in land management drivers across a broad range of habitat types to the regional variation in pollinator diversity, thereby supporting inference about Britain as a whole. With this target of broader generality of results, the location of regions should ideally be more objectively selected (Dilts, Yang & Weisberg 2010) and subject to the same levels of control as site selection. The addition of this regional selection protocol is therefore recommended for studies seeking broad statistical inference and a replicated pseudo-experimental design (Table 1).

Site selection

The second main difference in our approach was in the number of focal variables used simultaneously to select sites. Previous approaches have selected sites for different variables in a similarly hierarchical fashion, simultaneously selecting sites based on two variables (Holzschuh, Steffan-Dewenter & Tscharntke 2010; Hopfenmueller, Steffan-Dewenter & Holzschuh 2014; Steckel *et al.* 2014). Some such studies also detail selecting sites in the four quadrants of a 2-dimensional bivariate plot to remove the correlation between variables in the selected sites (Fahrig *et al*. 2011; Pasher *et al*., 2013). Pasher *et al.* (2013) further suggested the extension of this selection system to *n* dimensions, and Watts *et al*. (2016) attempted it with three dimensions. However, each additional selection variable greatly increases the number of possible combinatorial possibilities, which can soon become unmanageable. Here, we have presented the first attempt to use four dimensions and provide detailed instructions for manageable repetition of the method.

While there was some uncertainty in estimating our four metrics, the set of sites selected was sufficiently dispersed in variable space to allow analysis using continuous variables with values across the full ranges of each (Pasher *et al*. 2013). Randomly selected focal sites tend to cluster around mean values, providing relatively low resolving power for discerning the effects of landscape-scale drivers. Our original choice of what were modelled to be extreme values might be criticised for missing out these typical parameter values, but in practice the imprecise models combined with the inevitable regression towards the mean resulted in a wide exploration of parameter space of variables individually and in combination. An additional benefit of the protocol is that it greatly reduces the degree of correlation between focal variables, allowing valid inferences to be drawn about their separate and interacting impacts (Eigenbord *et al*. 2011; Pasher *et al*. 2013). Furthermore, studies of this kind do not normally assess correlations based on validated data, but we have demonstrated here that some caution is required if the calculation of focal variables is subject to high levels of uncertainty. Improvements to our metric estimates are likely to lead to further decoupling of metrics at the national scale.

Site validation

The estimates of the four metrics varied in their accuracy quite widely. The most accurate was the habitat diversity metric, which was based on the proportion of habitat covers calculated from remote sensing data. The high accuracy of this metric is not surprising as the estimates required the fewest steps in making the calculations, and verification was relatively straightforward. Even where the precise nature of land cover was misclassified on LCM2007, the spatial configuration of habitats as determined on the ground, and thus the Shannon index value, was generally quite close to our estimates from the LCM data. The level of accuracy is also similar to previous verification efforts (Morton *et al.* 2011).

The insecticide metric was also relatively well predicted when only considering those fields for which questionnaire responses were received. However, this result masks the large number of tetrads (especially in the North) for which large positive insecticide loadings were predicted when no arable fields were found on the ground. Although insecticides are applied on non-arable fields, the extent of application is unlikely to warrant a “high” insecticide loading value. These inappropriate values were probably caused in part by the satellite classification of reseeded pastures as arable fields and partly by changes in the crop areas between the 2010 census and 2012/13 survey years due to normal crop rotation.

The floral resource metric proved to have relatively low accuracy for a number of reasons related to the data available for making estimates: 1) some habitat cover estimates were incorrect due to misclassification in LCM2007 as described above, 2) actual floral reward data were only available for relatively few species at the time of site selection, 3) estimates of species cover per habitat were based on regional averages per broad habitat and so were not sensitive to within-region variation, and 4) mean nectar availability reported in databases does not capture the high variability observed in the field due to site differences in climate, soil and nectar consumption. Validation of these factors inevitably led to some widely differing values of site-level floral resource availability.

The honey bee density metric was the least well verified of the four drivers partly because the methods used to count the number of honey bees visiting sites proved to be unsuitable. As honey bees are social foragers, using scouts to alert workers to rich floral resource patches, the use of pan trapping to sample them is extremely inefficient (Westphal *et al.* 2008). Further, attempts to observe honey bees on the wing or foraging along transects suffered from a lack of available survey time: only 3 full days per season per site were used, often in poor weather conditions. Where data are available, they show a good relationship with the estimated density. However, such is the noise in the data and the high presence of zeros that subsequent analysis will need to use the original estimated values as an explanatory variable. Better estimates of honey bee numbers would require either greater investment in survey time or an alternative method such as the use of baited traps or estimating the number of hives present through, for example, surveys of farmers and beekeepers. As a result of these problems, we are not able to verify the accuracy of the honey bee density estimation technique.

Overall evaluation and implications

The aims of this site selection methodology were to improve on previous landscape-scale natural experimental designs by i) increasing objectivity of region selection to enhance the ability to generalise results to the wider landscape, and ii) to improve the selection of sites based on the values of multiple focal variables. This has been achieved by developing a hierarchical region selection protocol and by explicitly testing previously conceived ideas of site selection using multiple variables simultaneously. The additional complexities we have introduced to landscape-scale site selection will not be necessary for every research question, but provide a basis for increasing the inferential scope and complexity of landscape-scale pseudo-experiments.

We have also shown that it is possible to use national datasets to derive credible and objective sets of study sites that cover multiple environmental gradients, without bias from researcher's personal knowledge of landscapes in the site selection. The implications of this methodological development are important for landscape ecology and national scale monitoring programmes in any region or country with sufficient data, with a network of well-chosen sampling sites being a vital tenet of a well-designed national monitoring scheme.

Acknowledgements

We are especially grateful to our network of farmers, landowners and land managers who provided access to their land to conduct validation and our field assistants for conducting the field work: Nicole Dunn, Hayley Wiswell, Ewan Munro, Katy Donald, Jessica Heikkinen, Katherine White, Clare Pemberton, Paul Webb, Mark Tilzey, Sara Iversen, Robin Curtis, Paul Hill, Sam Bacon, Paul Wilson, Bex Cartwright, John Fitzgerald, Patrick Hancock, Mel Stone, Robert Day, James McGill, Tracie Evans. Kate Somerwill and Mette Kusk Gillespie are thanked for assistance with the protocol design. This research was supported by the UK Insect Pollinator Initiative project "AgriLand: Linking agriculture and land use change to pollinator populations", funded under the Living with Environmental change programme, a collaboration between Biotechnology and Biological Sciences Research Council (BBSRC), the Wellcome Trust, Scottish Government, Department of Environment, Food and Rural Affairs (DEFRA) and Natural Environment Research Council (NERC): grant BB/H014934/1 (www.agriland.leeds.ac.uk).

Data Accessibility: LCM2007, CS2007 and ITE Land class data are available from the NERC Environmental Information Data Centre catalogue (LCM2007: https://doi.org/10.5285/a1f88807-4826-44bc-994d-a902da5119c2; CS2007: https://doi.org/10.5285/57f97915-8ff1-473b-8c77-2564cbd747bc; ITE: https://doi.org/10.5285/5f0605e4-aa2a-48ab-b47c-bf5510823e8f). Detailed Defra datasets are available on request; for the June Agricultural Survey see https://data.gov.uk/dataset/june_survey_of_agriculture_and_horticulture_uk and for the Pesticide Usage Survey see https://secure.fera.defra.gov.uk/pusstats/. BeeBase summary data can be obtained under the government open data initiative (see https://data.gov.uk). Sugar production per flower per day for the 175 sampled flower species can be found here: (http://dx.doi.org/10.5285/69402002-1676-4de9-a04e-d17e827db93c and http://dx.doi.org/10.5285/6c6d3844-e95a-4f84-a12e-65be4731e934). Two further datasets (modelled sugar production and study site estimated and measured metrics) are provided in the Supporting information and are available to download from the NERC Environmental Information Data Centre catalogue https://catalogue.ceh.ac.uk/eidc/documents.

Author contributions: WK conceived and led the project in collaboration with JB, NB, JM, RDM and SGP, and all authors were involved in its development and completion. RDM designed the region and site selection algorithm in collaboration with WK and SS, selected the focal regions and processed the LCM2007 data to calculate habitat diversity. SS provided and processed CS2007 data and contributed to floral resource value estimations together with WK, NB, MB, JM and MG. NB and AC provided Pesticide Usage and Hazard Index data and related calculations. GB, AC and SPi estimated honey bee density values. MG performed the study site selection and validation with assistance from field staff and all co-authors. MG performed analyses and wrote the initial draft of the manuscript. All authors contributed to revisions of the manuscript.

Supporting Information:

Appendix S1: Full details of site selection methodology, with further information about data sources and processing, site selection algorithm, validation data collection and processing and supporting figures (S3-S10) cited in the main text.

Table S2: List of flower species and the traits used in sugar production modelling, including published values, associated references, and fitted values of sugar production from two linear models.

Table S6: List of sites selected for the project together with estimated and measured values of the four focal predictor variables.

References

Baude, M., Kunin, W.E., Boatman, N.D., Conyers, S., Davies, N., Gillespie, M.A.K., Morton, R.D., Smart, S.M. & Memmott, J. (2016) Historical nectar assessment reveals the fall and rise of floral resources in Britain. *Nature,* **530,** 85-88.

Beekman, M. & Ratnieks, F.L.W. (2000) Long-range foraging by the honey-bee, Apis mellifera L. *Functional Ecology,* **14,** 490-496.

Brittain, C.A., Vighi, M., Bommarco, R., Settele, J. & Potts, S.G. (2010) Impacts of a pesticide on pollinator species richness at different spatial scales. *Basic and Applied Ecology,* **11,** 106-115.

Bunce, R.G.H., Barr, C.J., Clarke, R.T., Howard, D.C. & Lane, A.M.J. (1996) ITE Merlewood Land Classification of Great Britain. *Journal of Biogeography,* **23,** 625-634.

Carey, P.D., Wallis, S., Chamberlain, P.M., Cooper, A., Emmett, B.A., Maskell, L.C., McCann, T., Murphy, J., Norton, L.R., Reynolds, B., Scott, W.A., Simpson, I.C., Smart, S.M. & Ullyett, J.M. (2008) Countryside Survey: UK Results from 2007. Centre for Ecology & Hydrology.

Diamond, J.M. (1983) Ecology - laboratory, field and natural experiments. *Nature,* **304,** 586-587.

Dilts, T.E., Yang, J. & Weisberg, P.J. (2010) The Landscape Similarity Toolbox: new tools for optimizing the location of control sites in experimental studies. *Ecography,* **33,** 1097-1101.

Eigenbrod, F., Hecnar, S.J. & Fahrig, L. (2011) Sub-optimal study design has major impacts on landscape-scale inference. Biological Conservation, 144, 298-305.

Elbgami, T., Kunin, W.E., Hughes, W.O.H. & Biesmeijer, J.C. (2014) The effect of proximity to a honeybee apiary on bumblebee colony fitness, development, and performance. *Apidologie,* **45,** 504-513.

Evison, S.E.F., Roberts, K.E., Laurenson, L., Pietravalle, S., Hui, J., Biesmeijer, J.C., Smith, J.E., Budge, G. & Hughes, W.O.H. (2012) Pervasiveness of Parasites in Pollinators. *Plos One,* **7**.

Fischer, C., Thies, C. & Tscharntke, T. (2011) Mixed effects of landscape complexity and farming practice on weed seed removal. Perspectives in Plant Ecology Evolution and Systematics, 13, 297-303.

Gabriel, D. & Tscharntke, T. (2007) Insect pollinated plants benefit from organic farming. *Agriculture Ecosystems & Environment,* **118,** 43-48.

Gabriel, D., Sait, S.M., Hodgson, J.A., Schmutz, U., Kunin, W.E. & Benton, T.G. (2010) Scale matters: the impact of organic farming on biodiversity at different spatial scales. Ecology Letters, 13, 858-869.

Goulson, D. & Sparrow, K. (2009) Evidence for competition between honeybees and bumblebees; effects on bumblebee worker size. *Journal of Insect Conservation,* **13,** 177-181.

Hargrove, W.W. & Pickering, J. (1992) Pseudoreplication - a sine-qua-non for regional ecology. *Landscape Ecology,* **6,** 251-258.

HilleRisLambers, J., Ettinger, A.K., Ford, K.R., Haak, D.C., Horwith, M., Miner, B.E., Rogers, H.S., Sheldon, K.S., Tewksbury, J.J., Waters, S.M. & Yang, S. (2013) Accidental experiments: ecological and evolutionary insights and opportunities derived from global change. *Oikos,* **122,** 1649-1661.

Holzschuh, A., Steffan-Dewenter, I. & Tscharntke, T. (2010) How do landscape composition and configuration, organic farming and fallow strips affect the diversity of bees, wasps and their parasitoids? *Journal of Animal Ecology,* **79,** 491-500.

Hopfenmueller, S., Steffan-Dewenter, I. & Holzschuh, A. (2014) Trait-Specific Responses of Wild Bee Communities to Landscape Composition, Configuration and Local Factors. *Plos One,* **9**.

Morton, D., Rowland, C., Wood, C., Meek, L., Marston, C., Smith, G., Wadsworth, R. & Simpson, I. (2011) Final Report for LCM2007-the new UK land cover map. *Countryside Survey*. Centre for Ecology & Hydrology.

Pasher, J., Mitchell, S.W., King, D.J., Fahrig, L., Smith, A.C. & Lindsay, K.E. (2013) Optimizing landscape selection for estimating relative effects of landscape variables on ecological responses. Landscape Ecology, 28, 371-383.

Potts, S.G., Vulliamy, B., Dafni, A., Ne'eman, G. & Willmer, P. (2003) Linking bees and flowers: How do floral communities structure pollinator communities? *Ecology,* **84,** 2628-2642.

R Core Team (2014) R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.

Rortais, A., Arnold, G., Halm, M.P. & Touffet-Briens, F. (2005) Modes of honeybees exposure to systemic insecticides: estimated amounts of contaminated pollen and nectar consumed by different categories of bees. *Apidologie,* **36,** 71-83.

Rundlof, M., Andersson, G.K.S., Bommarco, R., Fries, I., Hederstrom, V., Herbertsson, L., Jonsson, O., Klatt, B.K., Pedersen, T.R., Yourstone, J. & Smith, H.G. (2015) Seed coating with a neonicotinoid insecticide negatively affects wild bees. *Nature,* **521,** 77-U162.

Sagarin, R. & Pauchard, A. (2010) Observational approaches in ecology open new ground in a changing world. *Frontiers in Ecology and the Environment,* **8,** 379-386.

Shackelford, G., Steward, P.R., Benton, T.G., Kunin, W.E., Potts, S.G., Biesmeijer, J.C. & Sait, S.M. (2013) Comparison of pollinators and natural enemies: a meta-analysis of landscape and local effects on abundance and richness in crops. *Biological Reviews,* **88,** 1002-1021.

Smart, S.M., Henrys, P.A., Purse, B.V., Murphy, J.M., Bailey, M.J. & Marrs, R.H. (2012) Clarity or confusion? - Problems in attributing large-scale ecological changes to anthropogenic drivers. *Ecological Indicators,* **20,** 51-56.

Smart, S.M., Ellison, A.M., Bunce, R.G.H, Marrs, R.H., Kirby, K.J., Kimberley, A., Scott, W.A. & Foster, D.R. (2014) Quantifying the impact of an extreme climate event on species diversity in fragmented temperate forests: the effect of the October 1987 storms on British broadleaved woodlands. Journal of Ecology, 102, 1273-1287.

Steckel, J., Westphal, C., Peters, M.K., Bellach, M., Rothenwoehrer, C., Erasmi, S., Scherber, C., Tscharntke, T. & Steffan-Dewenter, I. (2014) Landscape composition and configuration differently affect trap-nesting bees, wasps and their antagonists. *Biological Conservation,* **172,** 56-64.

Waddington, K.D., Visscher, P.K., Herbert, T.J. & Richter, M.R. (1994) Comparisons of forager distributions from matched honey-bee colonies in suburban environments. *Behavioral Ecology and Sociobiology,* **35,** 423-429.

Watts, K., Fuentes-Montemayor, E., Macgregor, N.A., Peredo-Alvarez, V., Ferryman, M., Bellamy, C., Brown, N. & Park, K.J. (2016) Using historical woodland creation to construct a long-term, large-scale natural experiment: the WrEN project. Ecology and Evolution, 6, 3012-3025

Westphal, C., Bommarco, R., Carre, G., Lamborn, E., Morison, N., Petanidou, T., Potts, S.G., Roberts, S.P.M., Szentgyorgyi, H., Tscheulin, T., Vaissiere, B.E., Woyciechowski, M., Biesmeijer, J.C., Kunin, W.E., Settele, J. & Steffan-Dewenter, I. (2008) Measuring bee diversity in different european habitats and biogeographical regions. *Ecological Monographs,* **78,** 653-671.

Westphal, C., Steffan-Dewenter, I. & Tscharntke, T. (2006) Bumblebees experience landscapes at different spatial scales: possible implications for coexistence. *Oecologia,* **149,** 289-300.

Zuur, A.F., Ieno, E.N., Walker, N.J., Saveliev, A.A., & Smith, G.M. (2009) Mixed Effects Models and Extensions in Ecology with R. Springer, New York.

Tables

Table 1: Comparison of previous and current site selection protocols of studies incorporating a landscape scale pseudo-experimental approach

Study	Number of simultaneous focal selection variables	Number of regions (size)	Number of study sites/ landscapes (size)	True population	Method useful for:	Limitations of method
Gabriel *et al.* (2010)	1	2 (not given)	16 (10x10km)	The two regions studied	Nested or multi-scale designs, paired landscapes, ensuring non-target environmental conditions remain similar	Regions selected subjectively, one categorical focal selection variable
Fischer, Thies and Tscharntke (2011)	2*	3 (not given)	100* (forests: 100 x 100m; grassland: 50 x 50m)	The three regions studied; Central European grassland and forest areas?	Selecting sites along variable gradients, multi-criteria selection, focus on particular habitat types	Regions selected subjectively, restricted to two selection variables, limited control of external factors
Pasher *et al.* (2013)	2	1 (~15,500km^2)	100 (100ha)	The study region	Avoiding correlations between landscape variables, maximizing variability in variables	Region chosen subjectively, restricted to two selection variables
Smart et al. (2014)	1	2 (~60,000km2)	26 (5-100ha)	The study region; temperate lowland	Avoiding correlations between landscape variables, maximizing contrast between treatment of interest	Difficult to ensure equivalence of numerous other factors across treatment groups
Watts *et al.* (2016)	3	2 (~7335 km^2 & ~8570 km^2)	106 (0.5-32ha)	The two regions studied; temperate lowland agricultural landscapes ?	Selecting sites along variable gradients, multi-criteria selection, focus on particular habitat types, "natural experiments", analysing relative effects of variables, landscape conservation studies	Regions chosen subjectively, focus on woodland only, variable site sizes, not designed for hypothesis testing
This study	4	6 (100 x 100km)	96 (2 x 2km)	The six regions, the British countryside	Replicated pseudo-experimental designs, broad generality of results, hypothesis testing	Time consuming, data intensive

* corresponds to "experimental plots"

Table 2: Spearman correlation coefficients for the four **estimated** metrics (i.e., before validating; Box-Cox transformed Z-scores) for all six study regions. Coefficients are calculated for all possible sites within all regions (n = 12,718 sites) and the sites selected for study (n = 96). Asterisks denote significant correlations (p<0.001). Partial correlation coefficients were calculated controlling for Region, but are not shown as they were not different from the coefficients below.

	Habitat diversity		**Floral resources**		**Insecticide loadings**	
	All possible sites	Selected sites	All possible sites	Selected sites	All possible sites	Selected sites
Floral resources	0.14*	0.11	-	-	-	-
Insecticide loadings	-0.28*	-0.16	-0.20*	-0.16	-	-
Honey bee density	0.10*	0.10	-0.15*	-0.08	0.24*	0.11

Table 3: Spearman's rank correlation and partial correlation coefficients (controlling for Region), and parameters of linear mixed models (Region as random effect) for the estimated versus measured metrics in all regions. The data are Z-scores: box-cox transformed and zero centred. "Mean floral resources" is the total amount of floral resources averaged over the two years of field sampling. Asterisks indicate significant correlations: *** = p<0.001, ** = p<0.01, * = p<0.05

	Overall correlation	**Partial correlation**	**Slope**	**Intercept**	**P**
Habitat diversity	0.77***	0.77***	0.56	-0.05	<0.001
Mean floral resources	0.28**	0.29**	0.20	-0.03	0.005
Insecticide loadings	0.67**	0.60**	0.67	-0.01	0.001
Honey bee density	0.22*	0.21*	0.16	0.03	0.002

Table 4: Spearman's rank correlation and partial correlation (controlling for region) coefficients for the four **measured** metrics (i.e., corrected metrics after validation; Box-Cox transformed Z-scores) for all six study regions. Asterisks indicate significant correlations (* = p<0.05, ** = p<0.01).

	Habitat diversity	**Floral resources**	**Insecticide loadings**
All regions			
Floral resources	0.18		
Insecticide loadings	-0.47*	0.10	
Honey bee density	-0.04	0.31**	-0.54*
All regions (partial correlation)			
Floral resources	0.16		
Insecticide loadings	NA	NA	
Honey bee density	-0.05	0.29**	NA

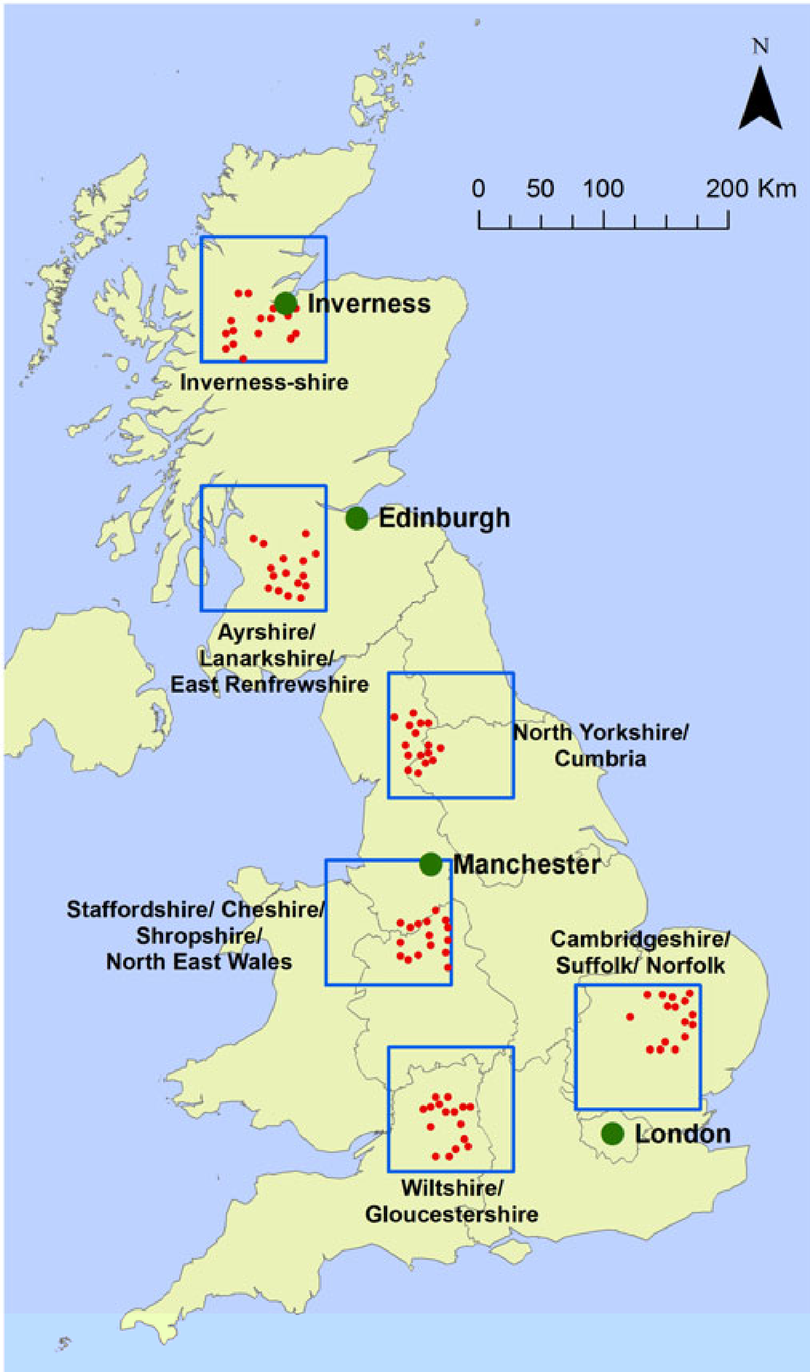
Figure legends

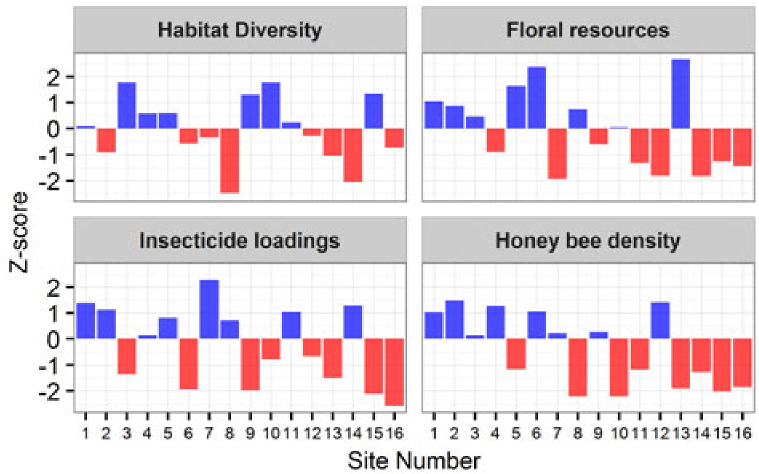
Fig. 1: The extent of the six 100 km^2 regions chosen by the region selection protocol (blue squares), and the 96 field sites (sixteen 2 x 2 km^2 sites per region) chosen by the site selection protocol (red circles). (Service Layer Credit: OS data; Crown copyright and database right 2015)

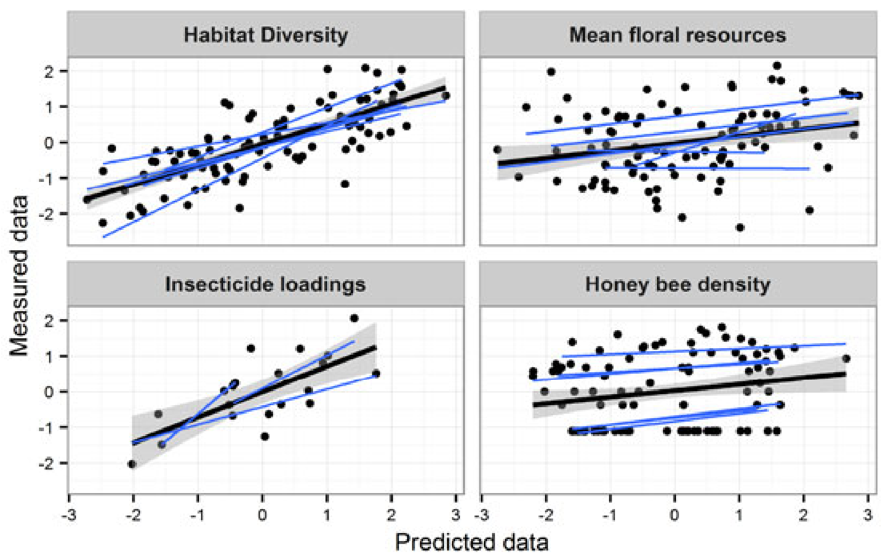
Fig. 2: The estimated Z-scores (Box-Cox transformed and zero centred data) of the four metrics for the final 16 sites of the Cambridgeshire/Suffolk region, shown here as an example. The blue bars are Z-scores above zero, i.e., the site has a "high" score for that metric; the red bars are negative Z-scores, i.e., the site has a "low" score for that metric. The 16 sites represent every combination of high and low values of the four metrics, e.g., site 1

has high values of all four metrics, site 2 has a low value only for habitat diversity, and so on. The data for the remaining regions can be found in Fig. S3.

Fig. 3: Validation of the four key metrics. The data are Z-scores: box-cox transformed and 0 centred, and each point represents a single site. The straight bold line represents the linear regression line for all regions and the shaded area represents 95% confidence intervals. The blue lines are mixed effect regression lines for each of the six regions with "region" as a random effect, displayed here to demonstrate the variation in prediction accuracy between regions. "Mean floral resources" is the total amount of floral resources averaged over the two years of field sampling. Regional graphs are shown in Fig. S10.


N
0 50 100 200 Km
Inverness
Inverness-shire
Edinburgh
Ayrshire/
Lanarkshire/
East Renfrewshire
North Yorkshire/
Cumbria
Manchester
Staffordshire/ Cheshire/
Shropshire/
North East Wales
Cambridgeshire/
Suffolk/ Norfolk
London
Wiltshire/
Gloucestershire


Habitat Diversity
Floral resources
Insecticide loadings
Honey bee density
Z-score
Site Number


Habitat Diversity
Mean floral resources
Insecticide loadings
Honey bee density
Measured data
Predicted data