Ecological Indicators 94 (2018) 348–356

Contents lists available at ScienceDirect

Ecological Indicators

journal homepage: www.elsevier.com/locate/ecolind

Original Articles

Uncertainty in site classification and its sensitivity to sample size and indicator quality – Bayesian misclassification rate in ecological risk assessment

Richard Ottermanns[a,*], Erhard Cramer[c], Benjamin Daniels[a], René Lehmann[b], Martina Roß-Nickoll[a]

[a] Institute for Environmental Research, RWTH Aachen University, Worringerweg 1, 52074 Aachen, Germany
[b] Institute of Empiricism and Statistics, FOM Hochschule für Oekonomie und Management University of Applied Science, Berliner Allee 7, 30175 Hannover, Germany
[c] Institute of Statistics, RWTH Aachen University, Wüllnerstraße 3, 52056 Aachen, Germany

ARTICLE INFO

Keywords:
Disturbance
Precautionary principle
Positive indicator
Reference condition
Statistical power

ABSTRACT

The aim of this study was to quantify uncertainty when assigning field investigation sites according to their species community composition to either undisturbed or disturbed reference sites by use of ecological indicators. In ecological risk assessment this problem arises when selecting control investigation sites or defining reference species communities. Uncertainty is quantified using a Type II error or misclassification rate. A probabilistic Bayesian model is used to integrate a priori domain knowledge, assess the error rate and come to recommendations about an adequate sample size. Application is demonstrated using data from a case study investigating off-crop arthropod communities in German grassy field margins and consequences for impact assessment of pesticides on terrestrial ecosystems. The model allows calculating statistical power when using such a classification system. By means of stochastic simulations, recommendations about experimental design and indicator size are derived. The study shows that to develop a classification system to typify newly observed sites a well-balanced ratio of undisturbed and disturbed sites as well as a high relevance of reference sites are needed. For the given data set, a much larger number of reference sites as well as increased relevance of selected reference sites would be needed to achieve a good classification result. An optimal number of indicators is calculated allowing for a compromise between sampling error and indicator quality. Uncertainty for correct assignment of an investigation site is compared using indicators for disturbance and reference conditions. Finally, misclassification rate is proposed as a new measure for indicator quality.

1. Introduction

The European Union requires that an ecological risk assessment (ERA) be performed for the authorisation process of plant protection products (PPP) (EC 1107/2009). The aim of ERA is to decide whether there may be a risk of unacceptable adverse effects on the environment, e.g. caused by the chemical substances used in pesticides (www.efsa.europa.eu). Negative effects of pesticides on biodiversity are still a problem in European agricultural landscapes (Geiger et al., 2010). However, to provide important ecosystem functions and services (e.g. pollination, food web support, pest control) it is important that the biodiversity of non-target organisms, like plants and soil arthropods be supported (EFSA PPR Panel, 2014, 2015). This holds for in-field sites, as well as those areas surrounding a field (off-field sites). The latter include field margins and buffer strips that may serve as sources of non-target species, facilitating recovery from impacts in the cropped area (Holland and Luff, 2000). Landscape structures are known to determine properties that to a large extent affect the external recovery of populations (EFSA SC, 2016a,b). The spatial distribution of exposed and non-exposed refuge areas is a particularly important driver for the underlying sink-source dynamics. Thus, to make general protection goals operational, effects of plant protection products on the occupancy of non-target organisms must be quantified at the landscape level.

Potential stressors, such as pesticide exposure, can alter the acceptable range of environmental conditions for populations, communities or ecosystems as normally observed in a reference ecosystem (Normal Operating Range, NOR) (Kersting, 1984; Ravera, 1989). In order to uncover such unacceptable effects of plant protection products, the normal operation range has to be defined using suitable local reference sites (Hughes, 1995; Kilgour et al., 1998; Ottermanns et al., 2010). In

* Corresponding author.
E-mail address: ottermanns@bio5.rwth-aachen.de (R. Ottermanns).

https://doi.org/10.1016/j.ecolind.2018.06.024
Received 2 June 2017; Received in revised form 8 June 2018; Accepted 12 June 2018
1470-160X/ © 2018 Elsevier Ltd. All rights reserved.

ecological risk assessment these reference sites have to be found in off-field areas, where the presence of potential stressors can more or less be excluded.

To define reference sites in an ecological risk assessment, the sites have to be assigned to a class of undisturbed or opposed to a class of disturbed sites (statistically referred to as a discriminant analysis). Bioindicators from the observed species compositions can be used to achieve this (Golden and Rattner, 2003). One advantage of using bioindicators is that they tend to integrate effects over time. Indicators can be calculated from species compositions under undisturbed and disturbed conditions using indicator analysis (De Cáceres et al., 2012; Dufrêne and Legendre, 1997). Such an assessment is often focused on detecting a disturbance within the site under investigation. A species that is positively associated with disturbance is called a 'negative indicator' (Carignan and Villard, 2002), and finding such a species can lead to the conclusion that the investigated site is disturbed, and the presence of a potential stressor can be assumed. Not finding the indicator can result in discounting any disturbance or the presence of a potential stressor (statistically referred to as a two-class prediction problem). A species that is positively associated with undisturbed conditions is called a 'positive indicator'. Finding such a species can lead to the conclusion that the investigated site is undisturbed and thus the absence of a potential stressor can be assumed.

When detecting disturbances in such a way, one can make two types of error (Table 1) corresponding to a set of hypotheses. The first hypothesis states that there is no disturbance at the site under investigation. The second hypothesis states that there is a disturbance at the site under investigation. In Type I errors, a disturbance at the site is stated due to finding the indicator for disturbance, despite the absence of a disturbance (α-error, false positive assignment). In Type II errors, a disturbance is neglected due to not finding the indicator for disturbance, despite there being a disturbance (β-error, false negative assignment). A Type II error can be interpreted in different but consistent ways. First, it means that a site is assigned to the undisturbed class although it belongs to the disturbed class. In this case β can be called a misclassification rate. Second, it is the probability of a site belonging to the disturbed class although the indicator for disturbance has not been observed. The same applies to the detections of undisturbed conditions.

It has been pointed out that, in accordance with precautionary principles, β should be minimized in environmental risk assessment and decision-making based on negative indicators. This results in more powerful statistical testing (Power = 1 − β) (Buhl-Mortensen, 1996; Peterman and M'Gonigle, 1992; Sanderson and Petersen, 2002; Santillo et al., 1998; Underwood and Chapman, 2003). In terms of risk protection, for consumers as well as parts of the ecosystem false negative assignments are much more severe and relevant than false positives. The demand for protectiveness is especially important when looking at non-target organisms (Atlas et al., 1978; Montesinos, 2003; Pereira et al., 2009). Due to their important role in ecosystem functioning and services, as well as their sensitivity, arthropods are suitable bioindicators to detect adverse effects. One example is studying the effects of pesticides in impacted German agricultural landscapes (Frampton, 1997; Holland and Luff, 2000; Huusela-Veistola, 1996; Kremen et al., 1993; Rob-Nickoll et al., 2004; Ottermanns, 2008). Given this context and the probabilistic interpretation for β from above, the Type II error can also be referred to as the probability that a pesticide effect exists at a site, but was overlooked using arthropods as bioindicators.

Table 1
The two types of error that can be made when detecting disturbance or undisturbed conditions.

Observation: Disturbance or undisturbed conditions stated?	yes	Type I error α	✓
	no	✓	Type II error β
		no	yes
		Truth: Disturbance or undisturbed conditions present?	

Uncertainty arises at different stages of the risk assessment due to a lack of knowledge and to natural variability (EFSA SC, in press, 2016b). In risk assessment in the field, a crucial source of uncertainty comes from the selection of potentially unaffected references sites. It is especially difficult to find suitable reference systems in heavily modified agricultural landscapes (EFSA SC, 2016b). Nevertheless, to understand the impact of uncertainty on the final assessment outcome, ecological risk assessment must (1) clearly identify the sources of uncertainty, (2) reliably find the range of possible outcomes and (3) exactly quantify the probability of their occurrence (EFSA SC, in press).

The aim of this study was to quantify uncertainty when assigning field investigation sites according to their species community composition by use of ecological indicators to one of two classes, either undisturbed or disturbed sites. In ecological risk assessment this problem arises, for example, when selecting control investigation sites or defining reference species communities. Uncertainty is quantified using the Type II error or misclassification rate. Both classes are characterized by specific indicators, consisting of one or more indicator species i.e., a multiple indicator set. A simple probabilistic Bayesian model was used to integrate a priori domain knowledge, assess the error rate and come to recommendations about an adequate sample size when developing indicators for assessment. Misclassification rate is proposed as a new measure for indicator quality. This is demonstrated using a data set of vegetation and arthropods in grassy field margins from three German macrochores belonging to a class of undisturbed off-field sites not affected by adjacent land use (called references) or a class of off-field sites potentially affected by adjacent land use (spray-drift, called non-target sites). Finally the uncertainty in the correct assignment of an investigation site to the class of undisturbed references was compared using indicators for disturbance (negative indicators) and indicators for reference conditions (positive indicators).

2. Material and methods

2.1. Notation

Throughout this study the following notation is used:

$G = 1$: site belongs to disturbed class = group 1
$G = 0$: site belongs to undisturbed class = group 0
$IG1 = 1$: characteristic indicator (set) for disturbed conditions has been found
$IG1 = 0$: characteristic indicator (set) for disturbed conditions has not been found
$IG0 = 1$: characteristic indicator (set) for undisturbed conditions has been found
$IG0 = 0$: characteristic indicator (set) for undisturbed conditions has not been found
$occ(i,G = k)$: Occurrence of species i over all sites belonging to class k $(k=\{0,1\})$
$abu(i,G = k)$: Abundance of species i over all sites belonging to class k $(k=\{0,1\})$
$P(IG1 = 1|G = 1)$: Probability of finding the indicator for disturbed conditions at a site given the site belongs to the disturbed class ($B(G = 1)$) (resp. $IG1 = 0$, $IG0 = 1$, $IG0 = 0$ and $G = 0$), B means the sensitivity from indicator species analysis
$P(G = 1|IG1 = 1)$: Probability of a site belonging to the disturbed class given the indicator for disturbed conditions has been found ($A(G = 1)$), A means the positive predictive value from the indicator species analysis
$P(G = 1|IG1 = 0)$: Probability of a site belonging to the disturbed class given the indicator for disturbed conditions has not been found → false-negative assignment

2.2. Data set and context

The application of the approach is demonstrated using data from a case study, investigating off crop arthropod communities in grassy field margins and the consequences for impact assessment of pesticides on terrestrial ecosystems. The data set used in this application came from Rob-Nickoll et al. (2004), a study that investigated the consequences for impact assessment of pesticides on arthropods of grassy field margins (off crop) in terrestrial ecosystems. This study analyzed the epigeic arthropod and vegetation communities in grassy field margins in three macrochores, which are typical representatives for intensively used agricultural landscapes in Germany: the Jülicher Börde (J), the north-eastern edge of the Leipzig lowlands (L), and the Würzburg area in Franconia (W). In each macrochore 24 sites were sampled. Four of these sites were a priori classified as unaffected by adjacent land use and further denoted as undisturbed reference sites (controls, G = 0). Twenty sites were located in the direct neighbourhood of arable areas, hereafter referred to as disturbed non-target sites (treatments, G = 1). In this case study these a priori assignments were necessary to develop indicator species, which afterwards could be applied to classify new sites as either reference or undisturbed.

The analysis included a variety of abiotic parameters, vegetation composition as well as arthropods from different trophic levels, different mobility and structural niches: beetles, spiders, springtails, hymenoptera, hoverflies and lady birds. These species are also relevant in the context of ecotoxicological test methods and sufficient ecological background knowledge on the ecology of the species was available.

All sites were a priori classified into one of two system states (states of G) in each macrochore: undisturbed (reference, control) sites referred to as JR, LR, WR, and disturbed (non-target, treatment) sites referred to as J, L, W.

2.3. Indicator analysis and feature selection

Prior to indicator analysis, singletons in abundance were deleted and suitable species were pre-selected using the following criteria. The overall abundance of species i in both site classes had to be larger than zero and the occurrence of species i at both site classes could not be one.

Indicator analysis (feature selection) was performed in different analyses for all macrochores using function `indicators` from the `indicspecies` package (De Cáceres et al., 2012; Dufrêne and Legendre, 1997) in the R statistical environment (R Core Team, 2015). This is an ecologically motivated method to reduce dimensionality to a smaller number of site specific indicators (single species or multiple species combinations). As a type of discriminant approach, it is applied to the two site classes defined a priori and provides indicators for each of the two site classes (McCune and Mefford, 2011). Indicator analysis calculates positive predictive values (A), sensitivities (B) and indicator values (Indval) from the relative abundances, mean abundances and the relative degree of presence of a species in a group of sites.

A positive predictive value (A) reflects the probability of a site belonging to a certain class G from a set of possible classes (for example {G = 0,G = 1}) given an indicator for one of those classes has been found (for example {IG0 = 1,IG1 = 1}): P(G = 0|IG0 = 1), P(G = 1|IG0 = 1). Sensitivity (B) means the probability of finding an indicator for a class at a site (IG1 = 1) if the site belongs to the given class (G = 1): P(IG1 = 1|G = 1). Statistical significance was determined by a permutation test. As the number of sites in the two classes are not equal, group size equalized indicator values were calculated to ensure independence from group size. This approach supplies those species which can be recognized as statistically significant indicators for the local conditions in reference and non-target sites in Jülich, Leipzig and Würzburg.

2.4. Selection of indicators for stochastic simulations

Based on the results of the indicator analysis, feasible indicators were selected for stochastic simulation studies from the list of indicators for disturbed conditions using the following procedure. First, we checked sampling efforts using Jacknife estimator from species accumulation curves (Ottermanns, 2008). Sample sizes proved to be adequate as the species numbers in this data set were found to be between 75 and 86% of the Jacknife estimator. Second, all indicators with A(i,G = 1) = 1 were removed from the list. This condition was imposed to rule out methodological sampling errors, because A(i,G = 1) = 1 means that species i was only observed at disturbed sites but not at undisturbed sites. Third, the indicators were sorted in descending order according to sqrt(Indval), and the indicator with maximum value was selected. The indicator was kept for simulation if the indicator value (Indval) was > 0.7 (sqrt(Indval) > 0.837). This threshold value proved suitable for the indication of local conditions in past studies (Ottermanns, 2008). Additionally, significance from non-parametric bootstrapping had to be < 0.05 and B(i,G = 1) had to be > B(i,G = 0). This additional condition was imposed because we required the sensitivity for disturbed sites needed to be larger than the sensitivity for reference sites. Otherwise the next indicator from the ranked list was selected.

2.5. Bayesian misclassification rate

Usually, newly observed sites are assigned to the class k for which they have the largest positive predictive value A(i,k) given an indicator species i. As these predictive values are always < 1, the assignment of a site to one of the classes will inevitably lead to a misclassification. A common frequentistic misclassification rate quantifies the inaccurate assignment of an investigation site to the undisturbed class (G = 0), although it belongs to the disturbed class (G = 1). It can be calculated from the ratio of the number of false positive assignments and the number of false positive + number of true positive assignments. Due to the ecotoxicological context of this application, such false negative assignments for G = 1 are of special interest.

In contrast to a frequentistic misclassification rate a priori information (i.e. domain knowledge) was integrated into a Bayesian misclassification rate in the study. This refers to the fact that undisturbed and disturbed sites are not distributed uniformly in the field. From our experience, a ratio of about 1:5 can be assumed in German agricultural landscapes (p(G = 0) = 1/6; p(G = 1) = 5/6).

If a site is disturbed then it belongs to class G = 1. Overlooking (= not observing) the indicator for disturbed conditions (IG1 = 0) results in assigning this site to the wrong class of undisturbed sites (G = 0). So, the Type II error is given by P(G = 1|IG1 = 0). Minimizing Type II errors results in a higher certainty that the indication derived from the positive predictive value A(i,k) is correct.

2.6. Probabilistic model

Selected indicators were used in a simple probabilistic Bayesian model to assess the rate of misclassification. The network topology models the two-class problem in which a binary parent variable (state G) determines the values of a binary child variable (indicator IG1) (see Fig. 1). It reflects a case-control study with observed disturbed state conditions observed (G = 1) or not observed disturbed state conditions not observed (G = 0, equals undisturbed conditions observed in this two class problem, G = 1 and G = 0 are mutually exclusive and complementary events), disturbance specific indicator set observed (IG1 = 1) or disturbance specific indicator set not observed (IG1 = 0, be aware that this does not mean that the undisturbed state specific

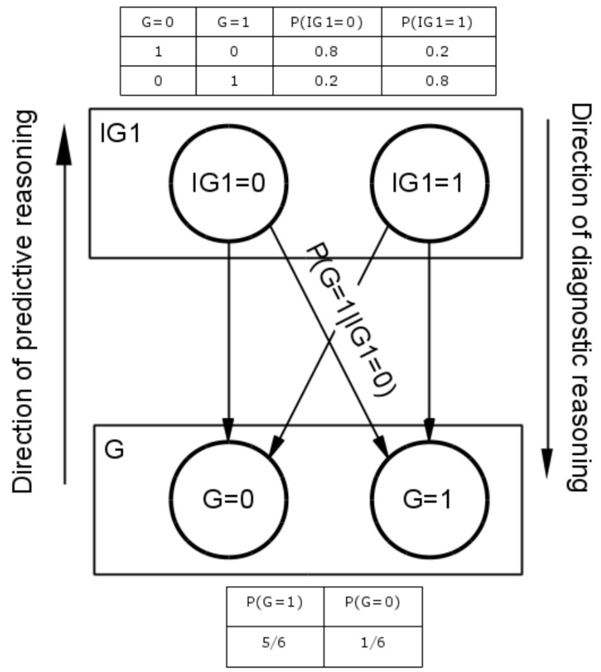

Fig. 1. Topology of the Bayesian network with discrete state for observation of the disturbance indicator (IG1 = {0,1}) and resulting system state (G = {0,1}), including marginal probability table for system state, conditional probability table for indicator state, and misclassification rate P(G = 1|IG1 = 0).

indicator set has been observed). G and IG1 are random variables taking values of {0,1}. The probability of classifying an observation as disturbed due to the observation of the indicator set specific for disturbed state is called P(G = 1|IG1 = 1), whereas the probability of an observation belonging to the disturbed state although the indicator set specific for disturbed state was not observed is denoted as P (G = 1|IG1 = 0). The latter is defined as the misclassification rate that is relevant in environmental monitoring for investigating ecological risk (Type II error). The definitions of the prior and posterior probabilities used in this model are given in the Supplementary material.

2.7. Simulation studies to optimize misclassification rate

The probabilistic network was used in different simulation studies to derive recommendations about how to reduce the misclassification rate. In scenario 1 the number of reference sites was increased (i.e. an increase in the prior probability of references sites P(G = 0)). In scenario 2 the probability of finding indicators for undisturbed conditions given an undisturbed site P(IG1 = 0|G = 0) was increased (i.e. an increase in the indicator value or the quality of the multiple indicator). In scenario 3 the probability of finding an indicator for undisturbed conditions given a disturbed site P(IG = 0|G = 1) was reduced (i.e. a reduction of false-positive indication of the multiple indicator). All other variables were kept constant within the simulations. The results were finally used to come to recommendations about a useful number of species within the multiple indicator in the attempt to reduce Type II errors (misclassification rate) under an optimal model complexity.

Table 2

Ten best significant indicators for disturbed conditions (G = 1) from the indicator analysis (negative indicators) in all macrochores (J/L/W). A = positive predictive value, B = sensitivity, sqrt = square root of indicator value Indval. Selected indicators in bold (for abbreviation of species see Table 4).

	G = 1			G = 0		
Jülich (J)	A	B	sqrt(Indval)	A	B	sqrt(Indval)
Nebrbrev	**0.911**	**0.80**	**0.854**	**0.089**	**0.50**	**0.211**
Nebrbrev + Trecquad	0.947	0.70	0.814	0.053	0.25	0.115
Harptard + Nebrbrev	0.910	0.60	0.739	0.090	0.25	0.150
Trecquad	0.677	0.80	0.736	0.323	0.75	0.493
Harptard + Nebrbrev + Trecquad	0.886	0.50	0.666	0.114	0.25	0.169
Sminaure	0.726	0.60	0.660	0.274	0.50	0.370
Harptard	0.591	0.70	0.643	0.409	0.75	0.554
Melamell	0.811	0.45	0.604	0.189	0.25	0.217
Nebrbrev + Sminaure	0.730	0.45	0.573	0.270	0.5	0.260
Sphapumi	0.538	0.60	0.568	0.462	0.75	0.589
Leipzig (L)	A	B	sqrt(Indval)	A	B	sqrt(Indval)
Amarsimi	**0.935**	**0.90**	**0.918**	**0.065**	**0.25**	**0.127**
Amarcomm	0.876	0.95	0.912	0.124	0.25	0.176
Synuviva	0.897	0.85	0.873	0.103	0.25	0.160
Panabipu	0.951	0.80	0.872	0.049	0.25	0.111
Lepptenu	0.845	0.90	0.872	0.155	0.50	0.279
Melamell + Amarsimi	0.910	0.80	0.853	0.090	0.25	0.150
Amarcomm + Erigatra	0.909	0.80	0.853	0.091	0.25	0.151
Amarsimi + Lepptenu	0.891	0.80	0.844	0.109	0.25	0.165
Amarsimi + Synuviva	0.888	0.80	0.843	0.112	0.25	0.168
Amarsimi + Panabipu	0.925	0.70	0.833	0.075	0.25	0.137
Würzburg (W)	A	B	sqrt(Indval)	A	B	sqrt(Indval)
Melamell	**0.968**	**0.95**	**0.959**	**0.032**	**0.25**	**0.090**
Ptermela	0.978	0.90	0.938	0.022	0.25	0.075
Pardpull	0.896	0.95	0.923	0.104	0.75	0.279
Pardoull-Melamell	0.938	0.90	0.919	0.063	0.25	0.125
Pardpull + Ptermela	0.974	0.85	0.910	0.026	0.25	0.800
Lepttenu	0.836	0.95	0.891	0.164	0.75	0.351
Pardpull + Lepttenu	0.851	0.90	0.875	0.149	0.50	0.273
Ptermela + Lepttenu	0.894	0.85	0.872	0.106	0.25	0.163
Pardpull + Draspusi	0.926	0.80	0.861	0.074	0.25	0.136
Pardpull + Ptermela + Lepttenu	0.887	0.80	0.842	0.113	0.25	0.168

Table 3
Ten best significant indicators for undisturbed conditions (G = 0) from the indicator analysis (positive indicators) in all macrochores (J/L/W). A = positive predictive value, B = sensitivity, sqrt = square root of indicator value Indval. Selected indicators in bold (for abbreviation of species see Table 4).

	G = 0			G = 1		
Jülich (J)	A	B	sqrt(Indval)	A	B	sqrt(Indval)
Harptard + Lepipara	0.968	0.75	0.852	0.032	0.05	0.040
Harptard + Trecquad + Lepipara	**0.968**	**0.75**	**0.852**	**0.032**	**0.05**	**0.040**
Panabipu + Lepipara	0.968	0.75	0.852	0.032	0.05	0.040
Trecquad + Panabipu + Lepipara	0.968	0.75	0.852	0.032	0.05	0.040
Trecquad + Sphapumi + Panabipu	0.956	0.75	0.847	0.044	0.05	0.047
Trecquad + Panabipu	0.950	0.75	0.844	0.050	0.05	0.050
Harptard + Panabipu	0.949	0.75	0.844	0.051	0.10	0.071
Dicynigr	0.948	0.75	0.843	0.052	0.05	0.051
Harptard + Sphapumi + Lepipara	0.946	0.75	0.842	0.054	0.05	0.052
Pardnigr	0.946	0.75	0.842	0.054	0.05	0.052
Leipzig (L)	A	B	sqrt(Indval)	A	B	sqrt(Indval)
Lepicyan	0.964	0.75	0.850	0.036	0.05	0.043
Melamell + Lepicyan	0.959	0.75	0.848	0.041	0.05	0.045
Amarpleb + Erigatra + Isotviri	**0.946**	**0.75**	**0.842**	**0.054**	**0.05**	**0.052**
Alopprat + Coccsept + Lepicyan	0.938	0.75	0.839	0.063	0.05	0.056
Alopprat + Isotviri + Achimill	0.938	0.75	0.839	0.063	0.05	0.056
Alopprat + Lepicyan	0.938	0.75	0.839	0.063	0.05	0.056
Amarpleb + Achimill	0.938	0.75	0.839	0.063	0.05	0.056
Amarpleb + Alopprat + Achimill	0.938	0.75	0.839	0.063	0.05	0.056
Amarpleb + Isotviri + Achimill	0.938	0.75	0.839	0.063	0.05	0.056
Amarpleb + Melamell + Achimill	0.938	0.75	0.839	0.063	0.05	0.056
Würzburg (W)	A	B	sqrt(Indval)	A	B	sqrt(Indval)
Pardlugu + Micapuli	0.952	0.75	0.845	0.048	0.05	0.049
Pardpull + Pardlugu + Micapuli	**0.952**	**0.75**	**0.845**	**0.048**	**0.05**	**0.049**
Dysderyt	0.915	0.75	0.829	0.085	0.01	0.092
Draslute + Micapuli	0.909	0.75	0.826	0.091	0.05	0.067
Haplumbr + Draslute + Micapuli	0.909	0.75	0.826	0.091	0.05	0.067
Pardpull + Draslute + Micapuli	0.909	0.75	0.826	0.091	0.05	0.067
Pardlugu + Draslute	0.897	0.75	0.820	0.103	0.05	0.072
Pardpull + Pardlugu + Draslute	0.897	0.75	0.820	0.103	0.05	0.072
Haplumbr + Micapuli	0.889	0.75	0.816	0.111	0.01	0.105
Pardpull + Haplumbr + Micapuli	0.889	0.75	0.816	0.111	0.01	0.105

3. Results

3.1. Indicator analysis and feature selection

Indicators are compiled in Tabs. 2 and 3. If an indicator is found in a newly observed site, this site can be assigned to one of the two classes with a probability equal to A(i,j,G = k), k = {0,1} (De Cáceres and Legendre, 2009) (e.g. typifying using a crisp two-class discriminative function). In Table 2 the ten best significant indicators for disturbed conditions (G = 1) from the indicator analysis are compiled, whereas in Table 3 the ten best significant indicators for undisturbed conditions (G = 0) can be found.

3.2. Selection of indicators for stochastic simulations and probabilistic model

The total number of sites per macrochore was 24, the number of undisturbed sites (G = 0) and disturbed sites (G = 1) per macrochore were 4 and 20 respectively. From these ratios the a priori probabilities for the system states of G (priors) are defined as $P_k(G = 0) = 4/24 = 1/6$ and $P_k(G = 1) = 20/24 = 5/6$, respectively (Bernoulli distributed with p = 1/6, k = {J,L,W}).

The resulting a posteriori probabilities for the events IG1 = {0,1} (likelihoods) undisturbed sites (P(IG1 = 1|G = 0) as well as disturbed sites (P(IG1 = 1|G = 1)) for the ten best significant indicators are shown in Table 2. In Jülich *Nebria brevicollis*, in Leipzig *Amara similata* and in Würzburg *Melanostoma mellinum* were selected as the best indicators given the proposed selection criteria (for species information see Table 4).

Table 4
Species abbreviations.

Abbreviation	Species	Group
Achimill	*Achillea millefolium*	Plants
Alopprat	*Alopecurus pratensis*	Plants
Amarcomm	*Amara communis*	Carabidae
Amarpleb	*Amara plebeja*	Carabidae
Amarsimi	*Amara similata*	Carabidae
Coccsept	*Coccinella septempunctata*	Coccinellidae
Dicynigr	*Dicymbium nigrum*	Araneae
Draslute	*Drassyllus lutetianus*	Araneae
Draspusi	*Drassyllus pusillus*	Araneae
Dysderyt	*Dysdera erythrina*	Araneae
Erigatra	*Erigone atra*	Araneae
Haplumbr	*Haplodrassus umbratilis*	Araneae
Harptard	*Harpalus tardus*	Carabidae
Isotviri	*Isotoma viridis*	Collembola
Lepicyan	*Lepidocyrtus cyaneus*	Collembola
Lepipara	*Lepidocyrtus paradoxus*	Collembola
Lepttenu	*Lepthyphantes tenuis*	Araneae
Melamell	*Melanostoma mellinum*	Syrphidae
Micapuli	*Micaria pulicaria*	Araneae
Nebrbrev	*Nebria brevicollis*	Carabidae
Panabipu	*Panagaeus bipustulatus*	Carabidae
Pardlugu	*Pardosa lugubris*	Araneae
Pardnigr	*Pardosa nigriceps*	Araneae
Pardpull	*Pardosa pullata*	Araneae
Ptermela	*Pterostichus melanarius*	Carabidae
Sminaure	*Sminthurinus aureus*	Collembola
Sphapumi	*Sphaeridia pumilis*	Collembola
Synuviva	*Synuchus vivalis*	Carabidae
Trecquad	*Trechus quadristriatus*	Carabidae

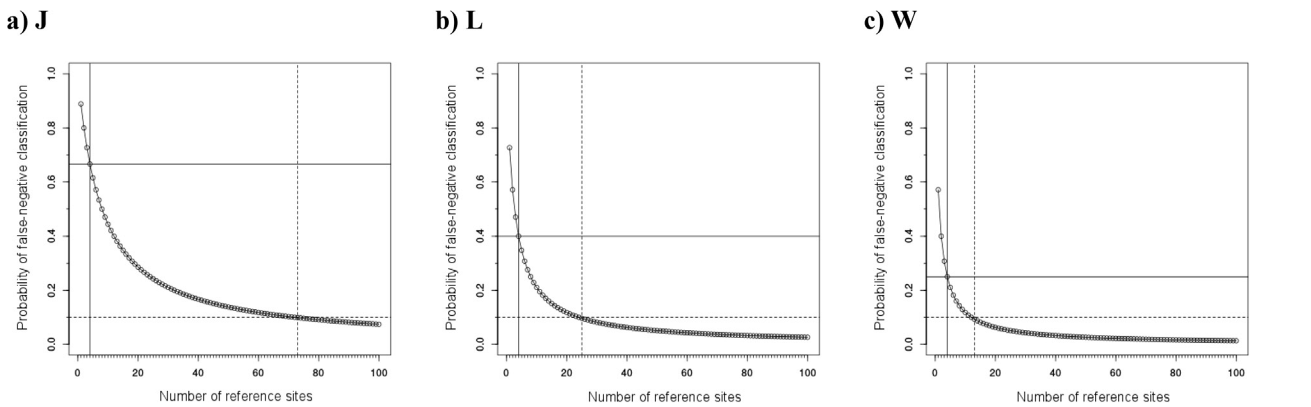

Fig. 2. Simulations for reducing the misclassification rate $P(G = 1|IG1 = 0)$ for Scenario 1: Increasing the number of undisturbed sites (= increasing $P(G = 0)$) (solid lines: given number of undisturbed sites = 4 and resulting misclassification rate, dashed lines: acceptable misclassification rate of 10% and needed number of undisturbed sites).

The evidence (a priori probabilities for the events $IG1 = \{0,1\}$) can be calculated using the rule of total probability. A priori probabilities (e.g. for Jülich) for the system states of G (priors) were calculated from the total number of sites and the number of undisturbed sites ($P_J(G = 0) = 1/6$) as well as disturbed sites ($P_J(G = 1) = 5/6$). From Table 2 it can be concluded that the probability of observing the top negative indicator *Nebria brevicollis* in a disturbed site in Jülich is $P_J(IG1 = 1|G = 1) = 0.8$ (B), and the probability that this indicator will be observed in an undisturbed site is ($P_J(IG1 = 1|G = 0) = 0.5$. The evidence for Jülich (a priori probability for the event $\{IG1 = 0\}$ = non-occurrence of the indicator) can thus be calculated according to (2) as $P_J(IG1 = 0) = (1–0.8) * 5/6 + (1–0.5) * 1/6 = 0.250$. The evidence for the top negative indicators in Leipzig (*Amara similata*) and Würzburg (*Melanostoma mellinum*) can be calculated as $P_L(IG1 = 0) = 0.208$ and $P_W(IG1 = 0) = 0.167$, respectively.

Using Bayesian identity the posterior probability for a site belonging to the disturbed class ($G = 1$) although the characteristic indicator for disturbance has not been observed ($IG1 = 0$), that is the misclassification rate under the given conditions, can be calculated for Jülich as $P_J(G = 1|IG1 = 0) = 1 - P(IG1 = 1|G = 1) * P(G = 1)/P(IG1 = 0) = (1–0.8) * 5/6/0.250 = 0.667$ and $P_L(G = 1|IG1 = 0) = 0.401$ and $P_W(G = 1|IG1 = 0) = 0.250$, respectively.

3.3. Simulation studies to optimize misclassification rate

In scenario 1, the number of undisturbed sites (= the prior probability of undisturbed sites $P(G = 0)$) was increased (see Fig. 2). To reduce the misclassification rate to less than 10%, the number of undisturbed site would have to be increased to at least 73 in Jülich, 25 in Leipzig and 13 in Würzburg (under the otherwise constant condition of 20 disturbed sites per macrochore). In scenario 2, the sensitivity B of indicators for disturbed conditions in disturbed sites (= the posterior probability $P(IG1 = 1|G = 1)$) was increased (see Fig. 3). Given the frame conditions (priors) at hand (4 undisturbed sites and 20 disturbed sites), the sensitivity of the indicator for disturbed conditions in disturbed sites would have to be increase to at least 99% in Jülich, Leipzig and Würzburg in order to reduce the misclassification rate to less than 10%. Finally, in scenario 3 the sensitivity B of indicators for disturbed conditions in undisturbed sites (= the posterior probability $P(IG1 = 1|G = 0)$) was reduced (see Fig. 4). Even when reducing the sensitivity of indicators of disturbed conditions in undisturbed sites to zero, a desirable misclassification rate less than 10% cannot be achieved in Jülich, Leipzig and Würzburg.

4. Discussion

From the calculation of Bayesian evidence, it is obvious that the probability of not observing the selected indicators for disturbance is low at all three macrochores ($P_J(IG1 = 0) = 0.250$, $P_L(IG1 = 0) = 0.208$, $P_W(IG1 = 0) = 0.167$). Ecologically this is not a surprise as it is known that ecological disturbance (landscape-scale simplification and in-field intensification) promotes biotic homogenization and the occurrence of more generalist species (Ekroos et al., 2010; Gámez-Virués et al., 2015; Vellend et al., 2007).

For the undisturbed sites it is assumed that a disturbance (e.g. pesticide impact) can be ruled out, thus the indicators for disturbance

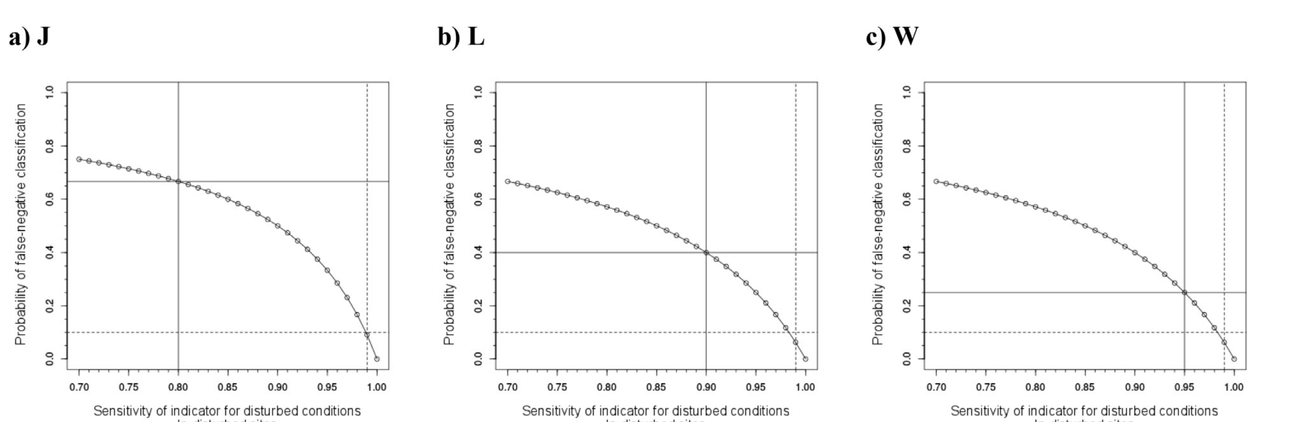

Fig. 3. Simulations for reducing the misclassification rate $P(G = 1|IG1 = 0)$ for Scenario 2: Increasing the sensitivity of the indicators for disturbed conditions in disturbed sites (= increase of the posterior probability $P(IG1 = 1|G = 1)$) (solid lines: given sensitivity and resulting misclassification rate, dashed lines: acceptable misclassification rate of 10% and needed sensitivity).

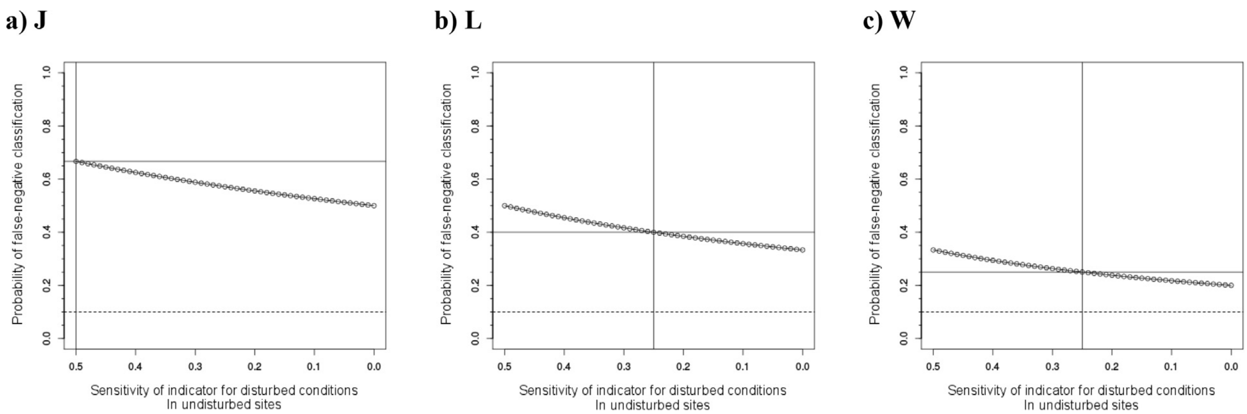

Fig. 4. Simulations for reducing the misclassification rate P(G = 1|IG1 = 0) for Scenario 3: Decreasing the sensitivity of the indicators for disturbed conditions in undisturbed sites (= decrease of the posterior probability P(IG1 = 1|G = 0)) (solid lines: given sensitivity and resulting misclassification rate, dashed line: acceptable misclassification rate of 10%).

should not be observed in this class (event {IG1 = 0}). For the disturbed sites a disturbance (e.g. pesticide impact) can at least not be ruled out, so the indicators should be observed (event {IG1 = 1}). So, from the ecological point of view the question posed to the probabilistic network could be expressed as: "*What is the probability that within the investigated site a potential pesticide effect exists, but was overlooked?*" This probability was calculated to be P(G = 1|IG1 = 0) = 0.667 for Jülich and 0.401 and 0.250 for Leipzig and Würzburg respectively. From these results it can be concluded that the probability that randomly chosen sites belong to the class of disturbed sites, although the indicators for disturbance were not observed (misclassification (P(G = 1|IG1 = 0). This means that a potential pesticide impact was overlooked about 67% of the time at Jülich, 40% at Leipzig and 25% at Würzburg. If the stated indicators for disturbance should be used for environmental monitoring, such a high magnitude of misclassification is not acceptable.

The simulation studies showed that theoretically it is possible to reduce the misclassification rate to less than 10%. Although the prior probability for undisturbed sites reflects a fixed characteristic of the landscape under investigation, which cannot be changed in practice, the simulation study manipulating the numbers of undisturbed sites is helpful from two perspectives. First, finding undisturbed conditions in a field study is a very time-consuming process and takes much longer than finding sites under disturbed conditions. This is particularly the situation for highly fragmented agricultural landscapes. More undisturbed sites may be found, so the ratio of undisturbed to disturbed sites would have to be reconsidered. This is an important cost-benefit aspect in ecological risk studies influencing misclassification rates. Secondly, the availability of undisturbed sites depends on the landscape under investigation. In more intensively used landscapes the chance of finding undisturbed conditions is smaller than in extensively used landscapes. So, prior probabilities should not be taken for granted when moving to another investigation area because this landscape characteristic triggers the misclassification rate via the prior probability. In this study the number of undisturbed sites would have to be increased to over 73 in Jülich, 25 in Leipzig and 13 in Würzburg (see Fig. 2) assuming the indicator values are constant and independent from sample size (as an equalized group size was used). In the attempt to develop indicators such a large number of undisturbed sites would not be manageable. This is not only because financial resources are mostly limited, but also because often it is not possible to find such a large number of undisturbed sites in the field. Nevertheless, the scenario gives a good feeling about how far we are often away from reasonable indication of disturbance using indicators for a disturbance.

Additionally, to the increase of the number of undisturbed sites, the misclassification rate can be reduced by enhancing the sensitivity of the disturbance indicators (see Fig. 3). This goal could be achieved by selecting more representative indicators for disturbance conditions. To do so, would increase the sensitivity of the indicators. The sensitivity would have to be increased to at least 99%, but this improvement is less effective than increasing the number of undisturbed sites. Due to the weighting with the class sizes in the Bayesian evidence, reducing the sensitivity of indicators of disturbed conditions in undisturbed sites has only little effect on the misclassification rate. Additional use of a larger number of undisturbed sites (see scenario 1) would considerably reduce this effort. A larger number of sampling sites would also increase the reliability of the significance test on the indicator values. A ratio of 30 undisturbed sites to 70 disturbed sites would be recommended to get reliable p-values from the non-parametric bootstrap in this study (De Cáceres and Legendre, 2009).

In Würzburg the misclassification rate cannot be optimized by choosing a different indicator (see Table 2). In Leipzig the 2nd best indicator *Amara communis* could be chosen which would result in a misclassification rate that is the same as for Würzburg (0.250 instead of 0.401). This indicator has a slightly larger sensitivity for disturbed conditions (P(IG1 = 1|G = 1)), but a smaller indicator value and positive predictive value for the disturbed class (P(G = 1|IG1 = 1)). Thus the reduction in misclassification rates comes with an increased uncertainty when assigning a newly observed site to the class of disturbed sites. In Jülich the 2nd best indicator could be chosen instead, i.e. *Nebria brevicollis* + *Trechus quadristriatus*. Although this indicator has a lower indicator value (0.814 instead of 0.854) and a lower sensitivity in the disturbed class (P(IG1 = 1|G = 1) = 0.7 instead of 0.8), it has a smaller sensitivity for the undisturbed class (P(IG1 = 1|G = 0) = 0.25 instead of 0.5). This results overall in the same misclassification rate as for *Nebria brevicollis* (0.67). Moreover, this indicator has a larger positive predictive value for the disturbed class P(G = 1|IG1 = 1) = 0.947 instead of 0.941), which reduces the uncertainty when assigning a newly observed site using this indicator. The fact that multiple indicators (species combinations and communities) can have higher positive predictive values than species taken independently has already been pointed out before (De Cáceres et al., 2012; Frampton, 1997).

Nevertheless, increasing the dimensionality of the indicator can increase the estimation error of the statistical model (see Fig. 5 top left). This fact is in concordance with the general modelling principle of Ockham's razor, a parsimony principle that demands the preference of more simple models over complex ones to keep the estimation error low and thus insure the generalization potential of models, at least for Bayesian statistical models (Jefferys and Berger, 1992; Rasmussen and Ghahramani, 2001). Conversely, a lower number of species within the indicator is disadvantageous from the view of ecological methodology. The risk of a false-negative result increases disproportionately with a lower number of integrated species, as overlooking a species during field sampling has a large influence on the sampling error (the deduction error, see Fig. 5 top right). By combining both aspects (estimation

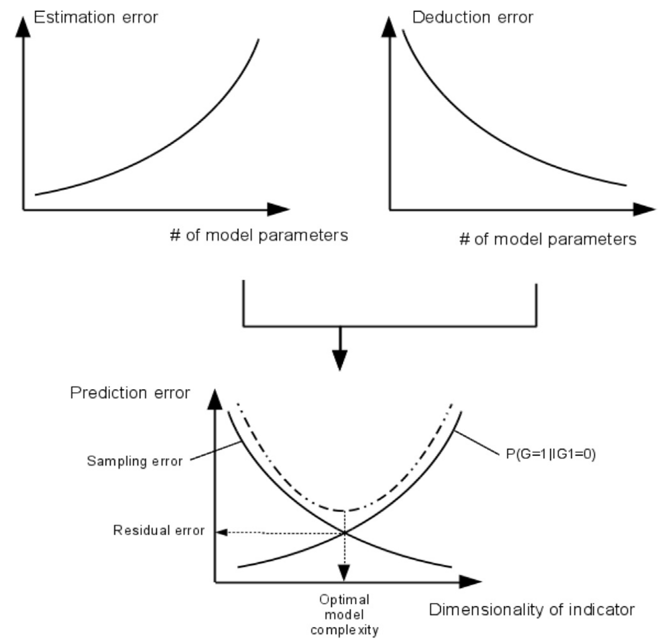

Fig. 5. Concept of optimal indicator dimensionality complexity (= number of species in the indicator) (top left: misclassification rate = estimation error, top right: sampling error = deduction error, bottom: resulting residual error = dashed line).

error and deduction error) it is possible to derive an optimal dimensionality of the indicator to find an optimal model complexity with minimal residual error to reduce the misclassification rate P (G = 1|IG1 = 0) (see Fig. 5 bottom).

Finally, one can focus not on the indication of disturbance ('negative indicators', see Table 2), but on the indication of undisturbed conditions ('positive indicators', see Table 3). In this case it can be stated that when indicators for undisturbed conditions are not found a disturbance cannot be ruled out. This approach works much better because ecological indicators for undisturbed conditions can be assumed to be more specific to environmental conditions (specialists) than indicators for disturbance (generalists). In this case the misclassification rate can simply be calculated as the probability of the site belonging to the disturbed class although the indicator for undisturbed conditions (IG0) has been observed (P(G = 1|IG0 = 1) = 1-P(G = 0|IG0 = 1), Type I-error, see Table 1). In indicator analysis this probability is readily given by the positive predictive value for the best indicator of undisturbed conditions in the disturbed sites. In our example *Harpalus tardus* + *Trechus quadristriatus* + *Lepidocyrtus paradoxus* for Jülich, *Amara plebeja* + *Erigone atra* + *Isotoma viridis* for Leipzig and *Pardosa pullata* + *Pardosa lugubris* + *Micaria pulicaria* for Würzburg can be chosen as indicators for undisturbed conditions resulting in misclassification rates of 3.2%, 5.4% and 4.8%, respectively (see Table 3).

5. Conclusions

Given the results, it is concluded that when attempting to detect disturbance or the presence of a potential stressor the development and use of indicators for disturbance ('negative indicators') is not an optimal practice, despite this being used in many ecological assessment approaches. In our application, uncertainty about the question whether a given site belongs to the class of undisturbed references equals the misclassification rate (Type II-error, β-error, false negatives). It is suggested as an additional measure of indicator quality (besides A, B and Indval), but proved to be very large and difficult to control. This resulted from the fact that the indicators for disturbance were developed using a much lower number of undisturbed sites than disturbed sites. To reduce the misclassification rate one could select disturbance indicators with reduced occurrence in undisturbed sites (B(G = 0)), which is generally difficult because indicators of disturbance are often ubiquitous species (generalists) and can show a strong increase under disturbance and landscape management intensification (Cordeiro et al., 2014; Gámez-Virués et al., 2015; Ottermanns, 2008; Rainio and Niemelä, 2003; Rob-Nickoll et al., 2004). Alternatively, the number of undisturbed sites could be increased to very large numbers, which is often impossible because there are not so many undisturbed sites available, especially in the case of agrarian habitats. Not finding an indicator can result from many things, e.g. sampling efficiency or population dynamics. Sampling efficiency can be increased by setting up more relevés or traps in disturbed sites when developing indicators, but this solution may also be limited, e.g. by the much discussed issue of pseudoreplication (Hargrove and Pickering, 1992; Hurlbert, 1984; Schank and Koehnle, 2009).

When focussing on indicators of undisturbed conditions ('positive indicators'), instead of indicators for disturbance, the misclassification rate equals a Type I-error (α-error, false positives) which is much easier to control. Besides the fact that misclassification rates are much lower than in the case of disturbance indicators, these indicators are multiple indicators (like most of the indicators from the list). The chance of overlooking an indicator (sampling error) is thereby reduced. Additionally, these indicators often consist of species from different organism groups, feeding types or other trait groups. This makes them much more robust and relevant from the ecological point of view. Last but not least it can be chosen from a much larger set of indicators, so it is more likely that these indicators will also work on a different set of sites.

Additionally to the well-known ecological arguments, statistical evidence is provided showing that it is better to use indicators for the reference conditions ('positive indicators') than to use indicators for disturbance ('negative indicators') in the attempt to detect disturbance at a site as a deviation from a preferable state. What we desperately need to do this efficiently in practice are good ideas about local reference conditions for focal regions in terms of environmental variables as well as typical species compositions (Carignan and Villard, 2002; Ottermanns et al., 2010; Ratte et al., 2005). Our results show that a carefully balanced study design is needed when using an indicator-based classification system in environmental risk assessment and that suitable reference sites are essential for sound decision making.

References

Atlas, R.M., Pramer, D., Bartha, R., 1978. Assessment of pesticide effects on non-target soil microorganisms. Soil Biol. Biochem. 10, 231–239.
Buhl-Mortensen, L., 1996. Type-II statistical errors in environmental science and the precautionary principle. Mar. Pollut. Bull. 32 (7), 528–531.
De Cáceres, M., Legendre, P., 2009. Associations between species and groups of sites: indices and statistical inference. Ecology 90 (12), 3566–3574.
De Cáceres, M., Legendre, P., Wiser, S.K., Brotons, L., 2012. Using species combinations in indicator value analysis. Methods Ecol. Evol. 3 (6), 973–982.
Carignan, V., Villard, M.-A., 2002. Selecting indicator species to monitor ecological integrity: a review. Environ. Monit. Assess. 78, 45–61.
Cordeiro, E.M., Corrêa, A.S., Guedes, R.N., 2014. Insecticide-mediated shift in ecological dominance between two competing species of grain beetles. PLoS One 9 (6), e100990.
Dufrêne, M., Legendre, P., 1997. Species assemblages and indicator species: the need for a flexible asymmetrical approach. Ecol. Monogr. 67 (3), 345–366.
EC 1107/2009 (2009): Regulation (EC) No 1107/2009 of the European Parliament and of the Council of 21 October 2009 concerning the placing of plant protection products on the market and repealing Council Directives 79/117/EEC and 91/414/EEC.
EFSA PPR Panel (EFSA Panel on Plant Protection Products and their Residues), 2014. Scientific opinion addressing the state of the science on risk assessment of plant protection products for non-target terrestrial plants. EFSA J. 12 (7), 3800. http://dx.doi.org/10.2903/j.efsa.2014.3800.
EFSA PPR Panel (EFSA Panel on Plant Protection Products and their Residues), 2015. Scientific opinion addressing the state of the science on risk assessment of plant protection products for non-target arthropods. EFSA J. 13 (2), 3996. http://dx.doi.org/10.2903/j.efsa.2015.3996.
EFSA SC (EFSA Scientific Committee), 2016a. Guidance to develop specific protection goals options for environmental risk assessment at EFSA, in relation to biodiversity

and ecosystem services. EFSA J. 14 (6), 4499. http://dx.doi.org/10.2903/j.efsa.2016.4499.
EFSA SC (EFSA Scientific Committee), 2016b. Scientific opinion on recovery in environmental risk assessments at EFSA. EFSA J. 14 (2), 4313. http://dx.doi.org/10.2903/j.efsa.2016.4313.
EFSA SC in press (European Food Safety Authority) [or EFSA Acronym Panel (EFSA Panel name)] [or EFSA Scientific Committee], 20YY. [Full title, including output category]. EFSA J. 20YY. volume(issue). NNNN, doi:10.2903/j.efsa.20YY.NNNN.
Ekroos, J., Heliölä, J., Kuussaari, M., 2010. Homogenization of lepidopteran communities in intensively cultivated agricultural landscapes. J. Appl. Ecol. 47, 459–467.
Frampton, G.K., 1997. The potential of Collembola as indicators of pesticide usage: evidence and methods from the UK arable ecosystem. Pedobiologia 41, 179–184.
Gámez-Virués, S., Perović, D.J., Gossner, M.M., Börschig, C., Blüthgen, N., de Jong, H., Simons, N.K., Klein, A.-M., Krauss, J., Maier, G., Scherber, C., Steckel, J., Rothenwöhrer, C., Steffan-Dewenter, I., Weiner, C.N., Weisser, W., Werner, M., Tscharntke, T., Westphal, C., 2015. Landscape simplification filters species traits and drives biotic homogenization. Nat. Commun. 6, 8568.
Geiger, F., Bengtsson, J., Berendse, F., Weisser, W.W., Emmerson, M., Morales, M.B., Ceryngier, P., Liira, J., Tscharntke, T., Winqvist, C., Eggers, S., Bommarco, R., Pärt, T., Bretagnolle, V., Plantegenest, M., Clement, L.W., Dennis, C., Palmer, C., Oñate, J.J., Guerrero, I., Hawro, V., Aavik, T., Thies, C., Flohre, A., Hänke, S., Fischer, C., Goedhart, P.W., Inchausti, P., 2010. Persistent negative effects of pesticides on biodiversity and biological control potential on European farmland. Basic Appl. Ecol. 11, 97–105.
Golden, N.H., Rattner, B.A., 2003. Ranking terrestrial vertebrate species for utility in biomonitoring and vulnerability to environmental contaminants. Rev. Environ. Contam. T. 176, 67–136.
Hargrove, W.W., Pickering, J., 1992. Pseudoreplication: a sine qua non for regional ecology. Landscape Ecol. 6 (4), 251–258.
Holland, J.M., Luff, M.L., 2000. The effects of agricultural practices on Carabidae in temperate agroecosystems. Integrated Pest. Manag. Rev. 5, 109–129.
Hughes, R.M., 1995. Defining acceptable biological status by comparing with reference conditions. In: Davis, W.S., Simon, T.P. (Eds.), Biological Assessment and Criteria: Tools for Water Resource Planning and Decision Making. Lewis Publishers, Boca Raton, Florida, pp. 31–48.
Hurlbert, S.H., 1984. Pseudoreplication and the design of ecological field experiments. Ecol. Monogr. 54 (2), 187–211.
Huusela-Veistola, E., 1996. Effects of pesticide use and cultivation techniques on ground beetles (Col., Carabidae) in cereal fields. Ann. Zool. Fenn. 33, 197–205.
Jefferys, W.H., Berger, J.O., 1992. Ockham's Razor and Bayesian analysis. Am. Sci. 80 (1), 64–72.
Kersting, K., 1984. Normalized ecosystem strain: a system parameter for the analysis of toxic stress in (micro-)ecosystems. Ecol. Bull. 36, 150–153.
Kilgour, B.W., Somers, K.M., Matthews, D.E., 1998. Using the normal range as a criterion for ecological significance in environmental monitoring and assessment. Ecoscience 5 (4), 542–550.
Kremen, C., Colwell, R.K., Erwin, T.L., Murphy, D.D., Noss, R.F., Sanjayan, M.A., 1993. Terrestrial arthropod assemblages: their use in conservation planning. Conserv. Biol. 7 (4), 796–808.
McCune, B., Mefford, M.J., 2011. PC-ORD. Multivariate Analysis of Ecological Data. Version 6. MjM Software, Gleneden Beach, Oregon, U.S.A.
Montesinos, E., 2003. Development, registration and commercialization of microbial pesticides for plantprotection. Int. Microbiol. 6, 245–252.
Ottermanns, R., 2008. Mustererkennung, Dimensionsreduktion und statistische Modellierung in der Ökologie – dargestellt am Beispiel der Lebensgemeinschaften grasiger Feldraine in deutschen Agrarlandschaften (Pattern recognition, dimensionality reduction and statistical modelling in ecology – demonstrated for ecological communities of grassy field margins in German agricultural landscapes). In: Studien zur Mustererkennung. Logos, Berlin, pp. 1–499.
Ottermanns, R., Ratte, H.T., Roß-Nickoll, M., 2010. Darstellung maskierter Nutzungseffekte auf naturraumspezifische Artengemeinschaften grasiger Feldraine mit Hilfe von Restvarianzmustern. UWSF 22 (1), 20–35.
Pereira, J.L., Antunes, S.C., Castro, B.B., Marques, C.R., Gonçalves, A.M.M., Gonçalves, F., Pereira, R., 2009. Toxicity evaluation of three pesticides on non-target aquatic and soil organisms: commercial formulation versus active ingredient. Ecotoxicology 18, 455–463.
Peterman, R.M., M'Gonigle, M., 1992. Statistical Power analysis and the precautionary principle. Mar. Pollut. Bull. 24 (5), 231–234.
Rainio, J., Niemelä, J., 2003. Ground beetles (Coleoptera: Carabidae) as bioindicators. Biodivers. Conserv 12, 487–506.
Rasmussen, C.E., Ghahramani, Z., 2001. Occam's razor. Adv. Neural Inf. Process Syst. 13, 294–300.
R Core Team, 2015. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria URL: https://www.R-project.org/. Assessed 14.01.2017.
Ratte, H.T., Lennartz, F., Roß-Nickoll, M., 2005. Ecosystem dynamics and stability – are the effects of pesticides acceptable? In: Liess (Ed.), Effects of Pesticides in the Field – EPIF. SETAC Press, pp. 88–90.
Ravera, O., 1989. Ecological Assessment of Environmental Degradation, Pollution and Recovery. Elsevier, Amsterdam.
Roß-Nickoll, M., Lennartz, G., Fürste, A., Mause, R., Ottermanns, R., Schäfer, S., Smolis, M., Theißen, B., Toschki, A., Ratte, H.T., 2004. The Arthropods of Grassy Field Margins (Off Crop) and the Consequences for Impact Assessment of Pesticides on Terrestrial Ecosystems. UBA Text Berlin 10/04.
Sanderson, H., Petersen, S., 2002. Power Analysis as a reflexive scientific tool for interpretation and implementation of the precautionary principle in the European Union. Environ. Sci. Pollut. R 9 (4), 221–226.
Santillo, D., Stringer, R.L., Johnston, P.A., Tickner, J., 1998. The precautionary principle: Protecting against failures of scientific method and risk assessment. Mar. Pollut. Bull. 36 (12), 939–950.
Schank, J.C., Koehnle, T.J., 2009. Pseudoreplication is a Pseudoproblem. J. Comp. Psychol. 123 (4), 421–433.
Underwood, A.J., Chapman, M.G., 2003. Power, precaution, Type II error and sampling design in assessment of environmental impacts. J. Exp. Mar. Biol. Ecol. 296, 49–70.
Vellend, M., Verheyen, K., Flinn, K.M., Jacquemyn, H., Kolb, A., van Calster, H., Peterken, G., Graae, B.J., Bellemare, J., Honnay, O., Brunet, J., Wulf, M., Gerhardt, F., Hermy, M., 2007. Homogenization of forest plant communities and weakening of species–environment relationships via agricultural land use. J. Ecol. 95, 565–573.