



Kent Academic Repository

Dennis, Emily B., Morgan, Byron J. T., Fox, Richard, Roy, David B. and Brereton, Tom (2019) *Functional data analysis of multi-species abundance and occupancy data sets*. *Ecological Indicators*, 104 . pp. 156-165. ISSN 1470-160X.

Downloaded from

<https://kar.kent.ac.uk/71360/> The University of Kent's Academic Repository KAR

The version of record is available from

<https://doi.org/10.1016/j.ecolind.2019.04.070>

This document version

Author's Accepted Manuscript

DOI for this version

Licence for this version

CC BY-NC-ND (Attribution-NonCommercial-NoDerivatives)

Additional information

Versions of research works

Versions of Record

If this version is the version of record, it is the same as the published version available on the publisher's web site. Cite as the published version.

Author Accepted Manuscripts

If this document is identified as the Author Accepted Manuscript it is the version after peer review but before type setting, copy editing or publisher branding. Cite as Surname, Initial. (Year) 'Title of article'. To be published in *Title of Journal* , Volume and issue numbers [peer-reviewed accepted version]. Available at: DOI or URL (Accessed: date).

Enquiries

If you have questions about this document contact ResearchSupport@kent.ac.uk. Please include the URL of the record in KAR. If you believe that your, or a third party's rights have been compromised through this document please see our [Take Down policy](https://www.kent.ac.uk/guides/kar-the-kent-academic-repository#policies) (available from <https://www.kent.ac.uk/guides/kar-the-kent-academic-repository#policies>).

Functional data analysis of multi-species abundance and occupancy data sets

Emily B. Dennis^{1,2,*}, Byron J.T. Morgan², Richard Fox¹,
David B. Roy³ & Tom M. Brereton¹

April 2, 2019

¹Butterfly Conservation, Manor Yard, East Lulworth, Wareham, Dorset, U.K.

²School of Mathematics, Statistics and Actuarial Science, University of Kent, Canterbury, Kent, U.K.

³Centre for Ecology & Hydrology, Benson Lane, Crowmarsh Gifford, Wallingford, Oxfordshire, U.K.

* Corresponding author

Abstract

Multi-species indicators are widely used to condense large, complex amounts of information on multiple separate species by forming a single index to inform research, policy and management. Much detail is typically lost when such indices are constructed. Here we investigate the potential of Functional Data Analysis, focussing upon Functional Principal Component Analysis (FPCA), which can be easily carried out using standard R programs, as a tool for displaying features of the underlying information. Illustrations are provided using data from the UK Butterflies for the New Millennium and UK Butterfly Monitoring Scheme databases. The FPCAs conducted result in a huge simplification in terms of dimensional reduction, allowing species occupancy and abundance to be reduced to two and three dimensions, respectively. We show that a functional principal component arises for both occupancy and

22 abundance analyses that distinguishes between species that increase or decrease over time,
23 and that it differs from percentage trend, which is a simplification of complex temporal
24 changes. We find differences in species patterns of occupancy and abundance, providing
25 a warning against routinely combining both types of index within multi-species indicators,
26 for example when using occupancy as a proxy for abundance when insufficient abundance
27 data are available. By identifying the differences between species, figures displaying func-
28 tional principal component scores are much more informative than the simple bar plots of
29 percentages of significant trends that often accompany multi-species indicators. Informed
30 by the outcomes of the FPCA, we make recommendations for accompanying visualisations
31 for multi-species indicators, and discuss how these are likely to be context and audience
32 specific. We show that, in the absence of FPCA, using mean species occupancy and total
33 abundance can provide additional, accessible information to complement species-level trends.
34 At the simplest level, we suggest using jitter plots to display variation in species-level trends.
35 We encourage further application to other taxa, and recommend the routine augmentation
36 of multi-species indicators in the future with additional statistical procedures and figures,
37 to serve as an aid to improve communication and understanding of biodiversity metrics, as
38 well as reveal potentially hidden patterns of behaviour and guide additional directions for
39 investigation.

40 Key words: Biodiversity indicators; Butterfly; Citizen science data; Functional principal
41 component analysis; Multi-species indices; Procrustes analysis

42 **1 Introduction**

43 Multi-species indicators are used to combine indices from a set of species and present a simple
44 summary of the species-level information. Indicators provide important metrics for evaluating
45 progress towards reducing the rate of biodiversity loss at a range of scales, including global
46 (Tittensor et al., 2014) and national (Eaton et al., 2015; Burns et al., 2018), as well as taxon-
47 specific assessments, such as for butterflies (Brereton et al., 2011b) and birds (Gregory et al.,
48 2005).

49 The geometric mean of component species indices is widely used to calculate multi-species

50 indicators (Gregory et al., 2005; Buckland et al., 2011; van Strien et al., 2012). However there
51 remains variation among different indicators, for example with regard to if and how uncer-
52 tainty in the estimated species-level indices is incorporated (Soldaat et al., 2017), and in the
53 presentation of both indicators and associated trends. Multi-species indicators are produced
54 for all species within a taxonomic group, or subsets based on classifying the component
55 species. For example, UK butterfly indicators are produced separately for habitat specialist
56 versus wider countryside species (Fox et al., 2015), and separate UK indicators are typically
57 produced for farmland, woodland and wetland bird species (Hayhow et al., 2017). Indica-
58 tors are generally produced from combining species-level indices for either annual estimates
59 of occupancy or an annual index of abundance, for which the underlying methods used to
60 estimate the indices can also vary among taxa.

61 Despite the advantages of providing simple summaries of biodiversity change, much in-
62 formation is necessarily lost when multi-species indicators are formed. One option to address
63 this, which is adopted by UK government biodiversity indicators, presents multi-species in-
64 dicators with adjacent bar charts which define the percentages of species declining versus
65 increasing (Defra, 2018), based on species-level trends. However the classification of such bar
66 charts can vary among taxa, for example by only separating increases from decreases, or by
67 also considering the significance of species trends. Similar visualisations of species trends are
68 also presented in the State of Nature assessment (Hayhow et al., 2016).

69 Given the increasing use and relevance of biodiversity indicators, of interest in this paper
70 is whether it is possible to use relatively simple tools to gain further insights into the ecologi-
71 cal patterns of species' changes in abundance and distribution. In doing so we aim to provide
72 recommendations for improved visualisations that may be used to support multi-species indi-
73 cators, to serve as an aid to improve communication and understanding of biodiversity metrics
74 and the underlying changes in species populations. Specifically, we investigate the potential
75 of Functional Principal Component Analysis (FPCA), which is one of several Functional
76 Data Analysis (FDA) techniques, in order to present simple informative graphical displays
77 (Ramsay et al., 2005), that can display far more of the lost information when multi-species
78 indicators are formed, than just providing indications of trend.

79 The goals of FDA include the following, taken from Ramsay et al. (2005, p.9):

- 80 • to represent the data in ways that aid further analysis,
- 81 • to display the data so as to highlight various characteristics,
- 82 • to study important sources of pattern and variation among the data.

83 These goals are relevant to the aims of this paper, but with novel application to summarising
84 biodiversity indices.

85 **2 Materials and methods**

86 **2.1 Functional Principal Component Analysis**

87 The main technique used in the paper is FPCA. It has similarities with Principal Components
88 Analysis (PCA), which is more familiar, and is described in outline in Appendix A. FPCA
89 performs much like PCA but FPCA operates on curves. In the applications in this paper,
90 species correspond to individuals and smoothed annual estimates for each species correspond
91 to the measurements on the individuals.

92 Interpretation of functional principal components can be made with the aid of harmonics
93 plots, however the primary objective of FPCA, as with PCA, is to reduce the dimensionality
94 of a problem, and if possible to provide plots of species, in our case, which may be inspected,
95 with species which have similar indices appearing close to each other. Importantly, PCA and
96 FPCA are objective techniques, so that derived components are data driven. In addition
97 to FPCA, we also apply Procrustes matching, for which the results can be found in the
98 Supplementary material, as well as axis rotation for functional principal components when
99 appropriate.

100 **2.2 Application to biodiversity indices**

101 The techniques used in this paper may be applied to abundance or occupancy indices for
102 multiple species of any taxon (or combination of multiple taxa). For demonstration we analyse
103 data from the Butterflies for the New Millennium (BNM) database and the UK Butterfly
104 Monitoring Scheme (UKBMS). Prior to the application of FDA, appropriate annual indices

105 of occupancy and abundance were produced from the two data sets. We consider data from
106 the BNM and UKBMS from 1980 onwards because most species have a full run of UKBMS
107 data from 1980. Based on the data available, we consider 1980-2014 for BNM and 1980-2016
108 for UKBMS. This resulted in occupancy and abundance data sets for 47 UK butterfly species
109 (out of a total of 59, of which 50 typically contribute to UK biodiversity indicators), which
110 are listed in the Supplementary material along with the species codes using in the paper.

111 **2.2.1 Producing species-level indices**

112 The BNM data consist of opportunistic records of species' presence gathered by volunteers
113 from any location in the UK and on any date. Over 7.5 million presence records were
114 collated for 1980-2014 for the 47 species considered in this paper. For each species and year
115 we estimate the occupancy probability for the UK for that species, using the occupancy
116 model approach of Dennis et al. (2017). For each species the set of these estimates over
117 time forms an occupancy index (see Figure 1a for examples and Supplementary Figure 1 for
118 indices for all 47 species). Covariates included in the fitted occupancy models followed those
119 used in Dennis et al. (2017), since species-specific model selection would be time-consuming.
120 Some species-level indices (Supplementary Figure 1) show irregular estimates for a small
121 number of years which could be due to the start values used, or as a result of over-fitting.
122 Preliminary comparisons were made with occupancy indices produced using a simpler set of
123 covariates (easting and northing and associated quadratics), but did not influence the overall
124 conclusions of this study.

125 The UKBMS consists of a long-running network of transects which began in 1976 with
126 34 sites, but has grown to nearly 1500 transects monitored each year (Brereton et al., 2017).
127 Since 2009 this additionally includes reduced-effort data from the Wider Countryside But-
128 terfly Survey (Brereton et al., 2011a). Under standardised weather conditions, counts are
129 made weekly from the beginning of April until the end of September (Pollard and Yates,
130 1993). Indices of relative abundance are estimated from the UKBMS for each species using a
131 Generalised Abundance Index approach (Dennis et al., 2016). Species-level indices are given
132 for four illustrative species in Figure 1b, and for all 47 species in Supplementary Figure 2.
133 UKBMS indices are typically presented on the \log_{10} scale where they either start at 2 or have

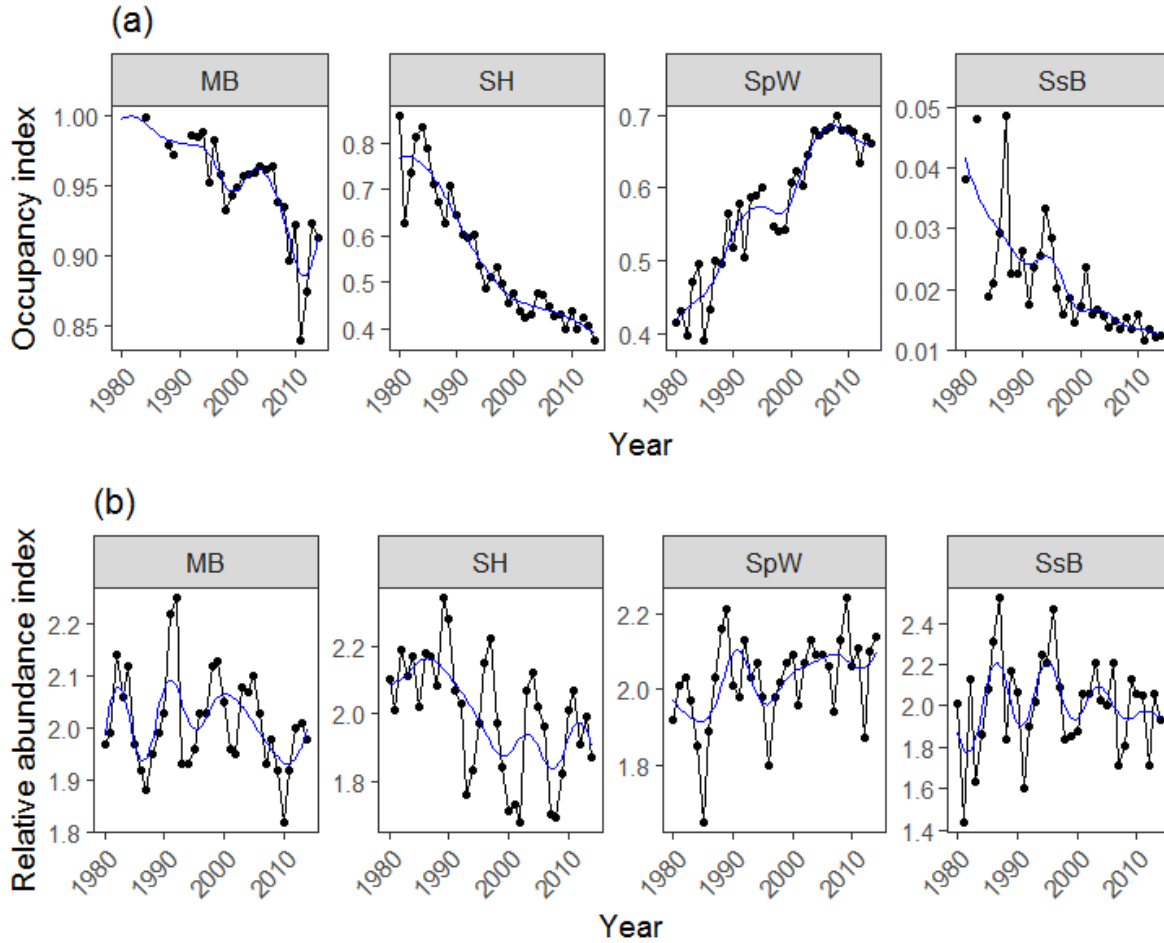


Figure 1: Occupancy (a) and relative abundance (b) indices for four illustrative butterfly species. Smoothed indices (blue) were produced using B-splines. Plots for all 47 species are given in the Supplementary Material. A small number of occupancy estimates are missing, particularly in early years, when the model-fitting fails due to insufficient data.

134 a mean of 2. It will be seen that there is therefore a fundamental difference between these
 135 indices and those relating to occupancy, when the entire probability range was possible.

136 2.2.2 Calculating species-level trends

137 For each species, a weighted logistic regression was fitted to the occupancy index, where
 138 the inverses of the index standard errors were used as weights. The standard errors were
 139 calculated using the Delta method (Morgan, 2009, p129), rather than the bootstrapping
 140 approach in Dennis et al. (2017), which can under perform in cases with limited data. Per-

141 centage changes for 1980-2014 were then estimated from the predicted values of the regression.
142 Whereas weighted regression was used for occupancy, to account for the variable standard er-
143 rors over time, percentage changes in relative abundance were estimated in the standard way
144 by fitting simple linear regressions to the species' indices of relative abundance for 1980-2016.

145 **2.2.3 Calculating multi-species indicators**

146 Multi-species indicators were produced separately for abundance and occupancy using by cal-
147 culating the geometric mean of the species-level indices. For both abundance and occupancy
148 the indices were scaled so that each species' index starts at 100, and the geometric average
149 then taken. We used the `BRCindicators` package (August et al., 2017), which accounts for
150 cases where a species-level index contains some missing year values. In brief, where a species
151 enters the indicator after the first year, the first year of that species' index is set to the
152 geometric mean of the series for species that are already in the indicator for that year.

153 **2.2.4 Applying FPCA**

154 We apply FPCA to occupancy and abundance indices from the BNM and UKBMS, respec-
155 tively. All analyses were performed using the `fda` package (Ramsay et al., 2009, 2017), in R
156 (R Core Team, 2018).

157 The input to the FPCAs is a set of smoothed curves of the species indices, with one
158 per species, separately for each of occupancy or relative abundance. These are displayed
159 for all 47 species in Supplementary Figures 1 and 2 for both occupancy and abundance.
160 Prior to smoothing, small numbers of missing year index values were interpolated (only
161 for Duke of Burgundy for abundance, and for 31 species for the occupancy indices). The
162 smoothed estimates were produced using the `fda` package using B-splines with 10 basis
163 functions and order 3. Alternative spline smooths were considered and there was a striking
164 stability in the results and conclusions with regard to how much smoothing was adopted.
165 It is possible, however, that for detailed scientific application to small numbers of species
166 it would be interesting to explore the use of cross-validation for choice of the amount of
167 smoothing, for each species separately. The smoothing used in these analyses does not take
168 account of relative precision of the species-level indices, where more recent estimates and

169 better recorded/monitored species are typically more precise.

170 For each survey separately, because the index values for any species at each time have
171 similar ranges, FPCA operates on the covariance matrices. In addition, for each species each
172 smoothed set of indices is centered by removing the mean over time before analysis.

173 We first review the associated harmonics plots, which display the principal component
174 functions, and then the corresponding functional principal component scores. The scores are
175 formed in an analogous way to how principal component scores are obtained for standard
176 PCA, though it is more complicated due to the use of curves rather than measurements
177 (Ramsay et al., 2005, p. 149). We distinguish between habitat specialists, migrants and wider
178 countryside species, based on the classification in Asher et al. (2001). We draw comparisons
179 with species-level abundance and occupancy trends estimated from the associated indices. A
180 three-dimensional plot for the first three principal components for the UKBMS analysis was
181 created using the `plotly` package (Sievert et al., 2017).

182 Necessarily, results obtained from a FPCA depend upon the time periods analysed, and
183 it is sometimes informative to consider how trends and indices change for different time inter-
184 vals. We compare results from different time periods in Sections 3 and 4 of the Supplementary
185 material. In particular we use Procrustes analysis (Gower, 1975) to match component plots
186 from different time periods. Further comparisons of abundance and occupancy using FDA
187 techniques are also given in Section 5 of the Supplementary Material.

188 **3 Results and discussion**

189 **3.1 Indicators for occupancy data**

190 Multi-species occupancy indicators, formed using the geometric mean, are shown in Figure
191 2, where habitat specialists display a greater decline in occupancy since 1980 compared
192 to wider countryside species. The associated species-level occupancy indices are given in
193 Supplementary Figure 1. For illustration, a bar chart displaying the percentages of species
194 increasing and decreasing (including significance) is given in Figure 2, which is produced
195 separately for subsets of species in biodiversity indicators.

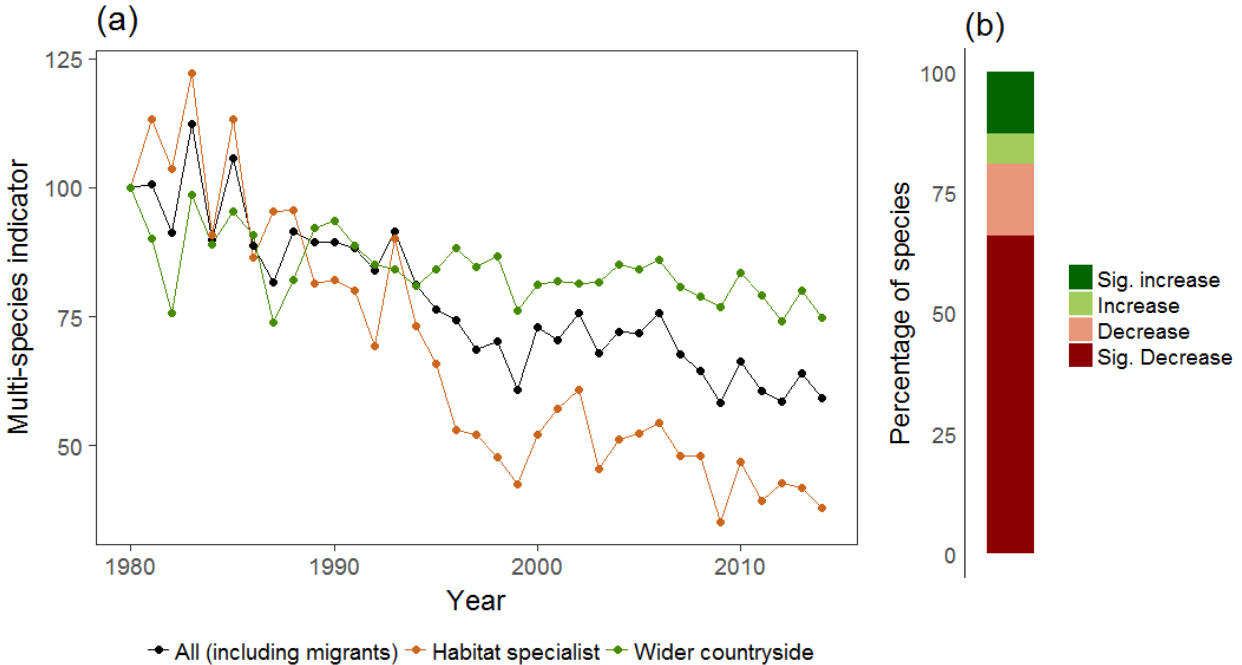


Figure 2: (a) Multi-species occupancy indicators calculated by the geometric mean of the occupancy indices for 47 UK butterflies for 1980–2014. (b) Bar plot giving percentage increases/decreases of individual butterfly species, where significance refers to the 5% level.

3.2 FPCA of occupancy data

3.2.1 Harmonics plots

Figure 3 provides us with a potential means of interpreting the first two principal components of FPCA applied to the BNM occupancy indices by showing a harmonic plot for each functional principal component. The first principal component orders species according to whether they have high or low occupancy, essentially corresponding to an average occupancy over time: at one end of the scale are species with near constant high occupancy, while at the other end are species with near constant low occupancy. This first component describes 97.4% of the total variance. The second component contrasts species that are declining over the time period with species that are increasing, although in both cases the harmonics level out for the most recent few years. Thus although it does not explain much of the total variance, just 1.9%, this component has a clear interpretation.

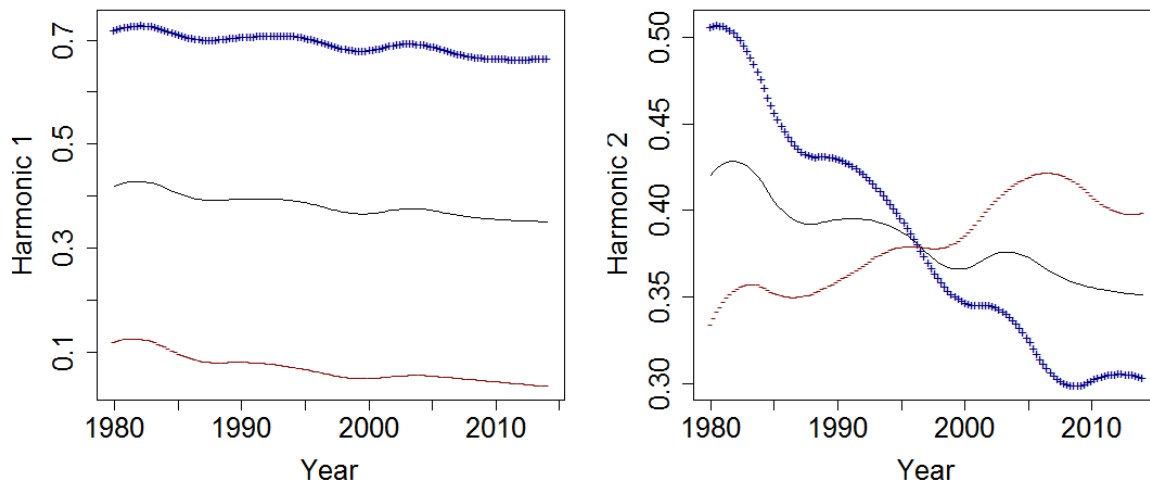


Figure 3: Harmonics plots of the first two functional principal components for the BNM data. The arithmetic average of the species indices is shown by a solid line, with the end of each component plotted using + (blue) and - (red). The percentages of variance for the first two components are 97.4% and 1.9%.

208 Both plots in Figure 3 show the arithmetic mean of all smoothed indices for all years,
 209 and this is the same in each case. It therefore plays a similar role to the geometric mean
 210 for all 47 species (Figure 2a). The first two functional principal components describe most
 211 of the total variance, so that we have reduced the information in the species-level curves
 212 (Supplementary Figure 1), and can represent the species as points in two-dimensional space
 213 (see Figure 5a, with discussion to follow), with coordinates given by the first two functional
 214 principal component scores. This is a great simplification compared to having 35 (annual)
 215 data points for each species.

216 With minor differences, we have found the general patterns of the harmonics plots of
 217 Figure 3 to appear in other occupancy analyses, for example of Scottish moths (Dennis
 218 et al., 2019), when occupancy data on 225 moth species were analysed (Section 2 of the
 219 Supplementary Material). The same is also true if we divide the data into the first half
 220 and second half time periods and analyse the two halves separately (see Section 3 of the
 221 Supplementary Material).

222 3.2.2 Comparison with species-level occupancy trends

223 Figure 4a shows the estimated percentage trend for each species, plotted against the corre-
224 sponding second functional principal component score, denoted by X_2 . Note that all principal
225 component scores are centered on zero due to the mean centering at each individual time
226 point. As we might expect from the interpretation of the second component provided above
227 by Figure 3, there is a relationship between the trend and the second functional component
228 score, however it is not a linear one. The association is approximately linear for wider coun-
229 tryside species, however habitat specialists, with generally lower occupancy, necessarily have
230 smaller absolute changes, resulting in relatively small values for X_2 .

231 Figure 4b distinguishes between values that are significantly changing (increasing or de-
232 creasing), each at the 5% level. While there is a correlation between the X_2 and trend values,
233 the X_2 axis is reflecting shapes of the individual species indices in a more complex way than
234 simply ordering the species according to their estimated trend value. It is instructive to
235 relate the points back to the index plots for the species that they represent. Rug plots are
236 displayed along the axes in Figure 4, which indicate the values taken by species along those
237 axes, and this feature recurs in similar plots in the paper.

238 In Figure 5a each species is plotted according to the scores of its first two functional prin-
239 cipal components, X_1 measuring average occupancy, and X_2 indicating whether the species
240 is increasing or decreasing over time. Figure 5a identifies two main clusters of species, driven
241 by the size of occupancy, suggesting that it might be of interest to analyse these two clusters
242 separately. This is in fact what is essentially done when multi-species indicators are produced
243 separately for habitat specialists and wider countryside species (Figure 2a). However this
244 distinction is not clear cut in that a small number of the wider countryside species appear
245 similarly placed to the habitat specialists. These are the wider countryside species with rela-
246 tively low estimates of occupancy probability. The second component corresponds to species
247 that are increasing/declining over the entire time period, and therefore provides the relevant
248 information in the individual species occupancy indices in Supplementary Figure 1. Thus
249 here the X_2 values alone, on the y-axis, illustrate important information that is hidden when
250 the geometric mean indicator is formed.

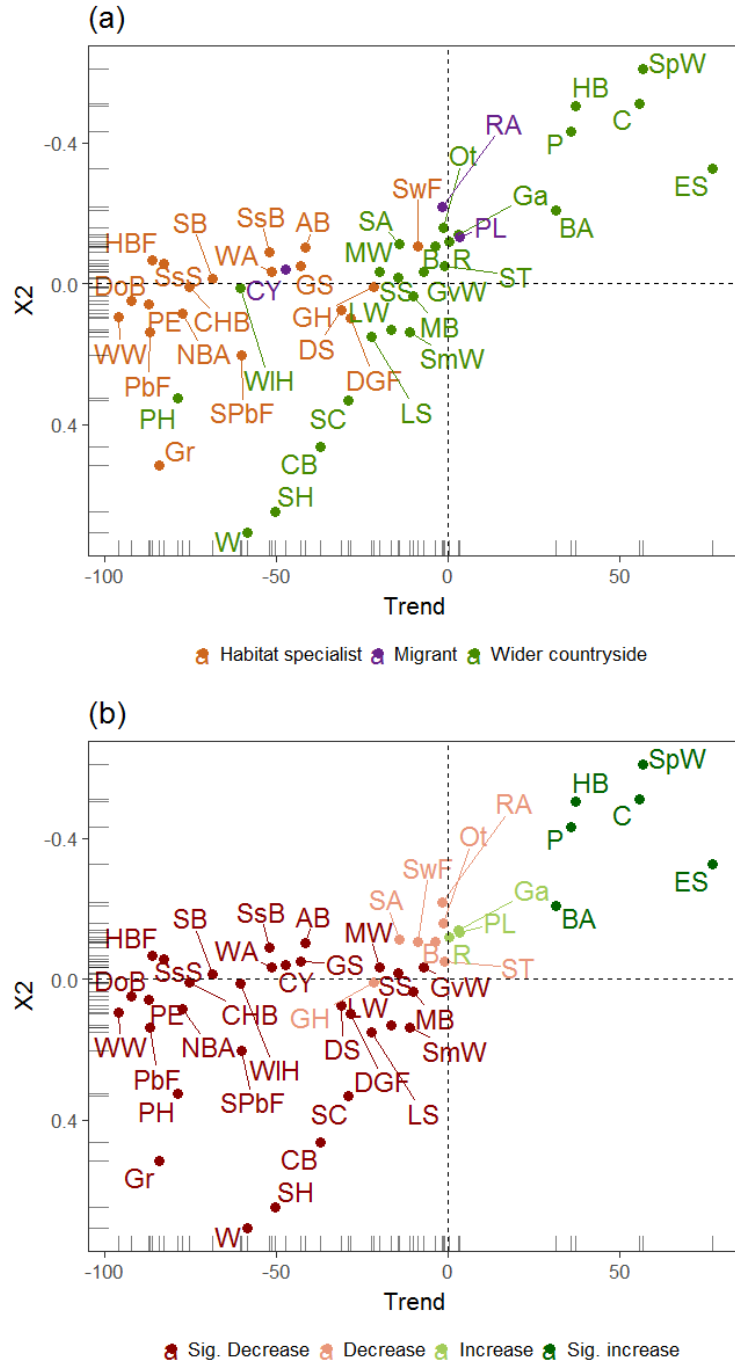


Figure 4: Estimated species occupancy trends (percentage changes) versus the corresponding scores that result on the second axis (X_2) from the FPCA analysis of the BNM data; the locations of points are the same in both plots. (a) Colours indicate species classification: habitat specialists, migrants and wider countryside; (b) colours indicate category of trend, as summarised in Figure 2b. The vertical and horizontal dashed lines indicate no change in occupancy and X_2 values of zero, respectively.

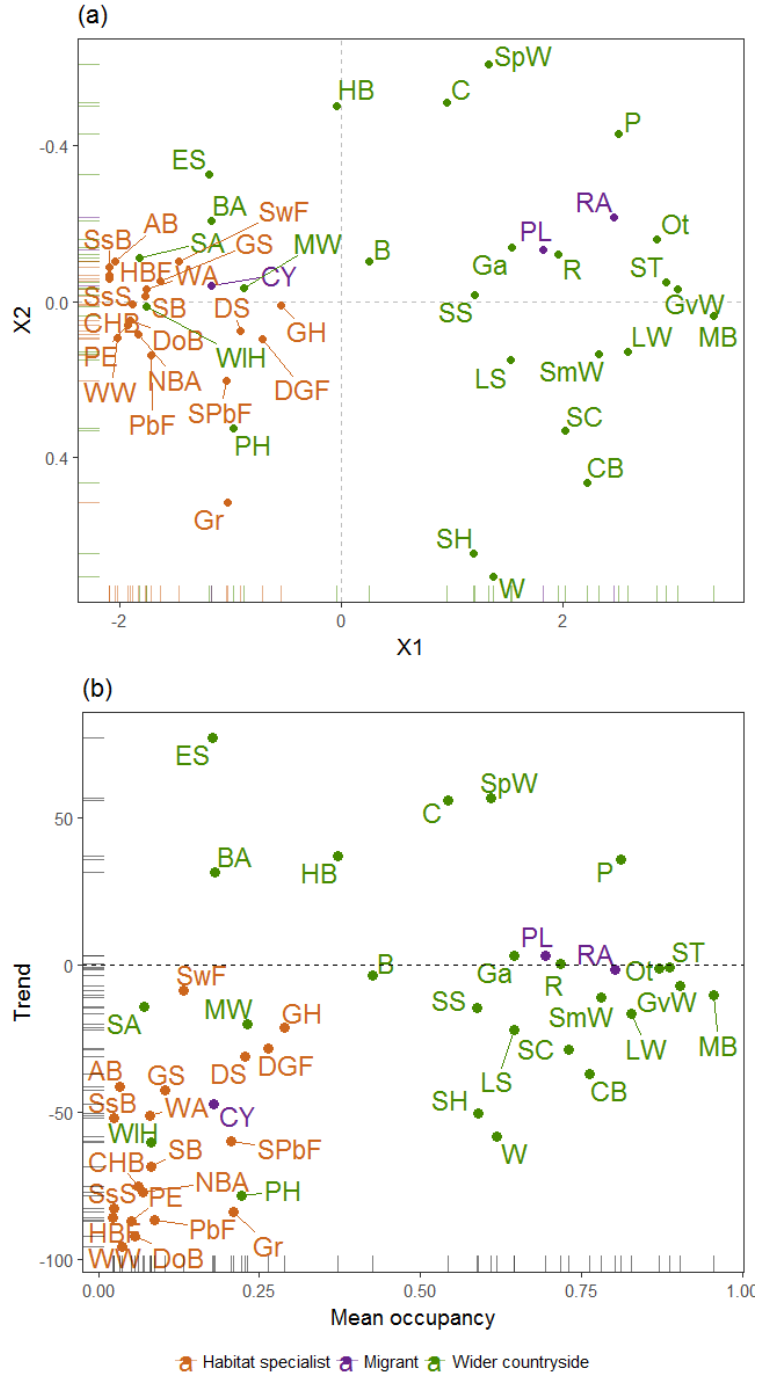


Figure 5: (a) Plot of the two functional principal component scores, X_1 , measuring average occupancy, and X_2 , measuring increase or decrease, for all 48 species for the full time period. The axis for X_2 has been reversed. The dashed lines indicate score values of zero. (b) For comparison we replace X_1 by the average occupancy index value and X_2 by the estimated species occupancy trend. The horizontal dashed line indicates no change in occupancy.

251 Species to the right of X_1 have high occupancy, and those to the left have low occupancy.
252 Species at the top of X_2 are increasing, and those at the bottom are decreasing. It is
253 easy to verify this: see for example the positions of Meadow Brown (MB, high occupancy
254 and minimal change over time), Speckled Wood (SpW, medium occupancy and increasing
255 over time), and Grayling (Gr, relatively low occupancy and much temporal decline) for which
256 species-level occupancy indices are shown in Figure 1a and Supplementary Figure 1. Wall and
257 Small Heath stand out as showing the lowest values of X_2 , representing the largest absolute
258 declines in occupancy, and despite being wider countryside species they are considered to be
259 priority species for conservation.

260 FPCA has demonstrated a great economy in description of occupancy of 47 butterfly
261 species over the time period. It provides a huge improvement over a single bar plot, at the
262 cost of just introducing one extra dimension of plotting (2 dimensions, rather than 1), and
263 does not have to replace a bar plot, but can be considered in association with it.

264 Figure 5b is motivated by Figure 5a, and provides an alternative display of potentially
265 similar information. Given that FPCA is objective, it is interesting that there are some
266 similarities between the two figures. Figure 5b has the advantage that it might be easier to
267 understand than Figure 5a, since FPCA is not needed and percentage change information
268 is included. However in this case the two variables are now correlated, as they have not
269 resulted from a FPCA. It is useful to combine mean occupancy with percentage trend in
270 a single plot, as we can see that the species with the largest percentage declines have the
271 smallest occupancy. This information is lacking in a standard bar chart summarising species
272 trends (see Figure 2a). Figure 5b is suggested by Figure 5a, and it is only for Figure 5a that
273 we know that most variance is described. Thus we can with confidence consider the spatial
274 location of species in relation to others, as close points in Figure 5a indicate species which
275 exhibit similar species indices.

276 **3.3 Indicators for abundance data**

277 Multi-species indicators for the relative abundance of butterflies, formed using the geometric
278 mean, are shown in Figure 6a, for all species and also for habitat specialists and wider
279 countryside species separately. The patterns of behaviour shown here are somewhat different

280 from those in Figure 2a, and we note also that there is a degree of apparent cycling for the
 281 indicators. The relevant species indices of abundance are given in Supplementary Figure 2.

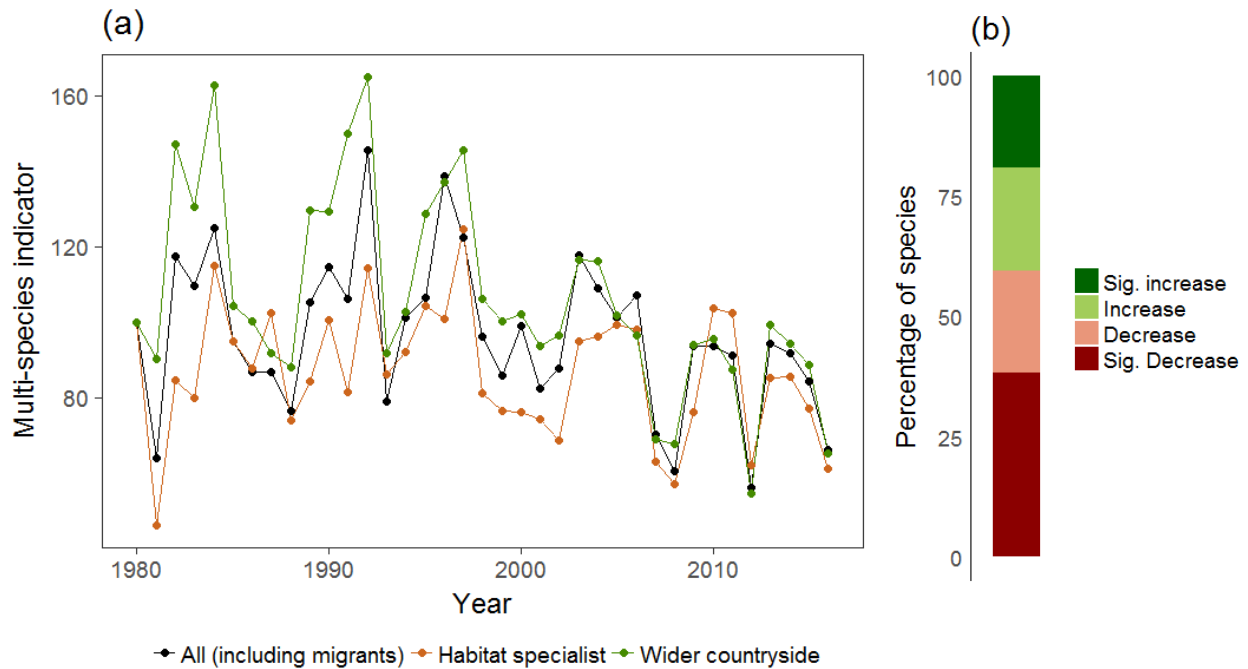


Figure 6: (a) Multi-species abundance indicators calculated by the geometric mean of the relative abundance indices for 47 UK butterflies for 1980–2016.(b) Bar plot giving percentage increases/decreases of individual butterfly species, where significance refers to the 5% level.

282 3.4 FPCA of abundance data

283 3.4.1 Harmonics plots

284 The harmonics plots resulting from the FPCA applied to the relative abundance indices
 285 (Figure 7) show differences compared to those obtained for occupancy indices (Figure 3),
 286 partly due to the differences in scale of the two types of indices. Since the relative abundance
 287 indices are all standardised (on the \log_{10} scale with a mean of 2), the dominant first component
 288 for the occupancy case is no longer present, and instead we have as the first component one
 289 that resembles the second component for the occupancy FPCA, in this case indicative of an
 290 increase or decline in abundance.

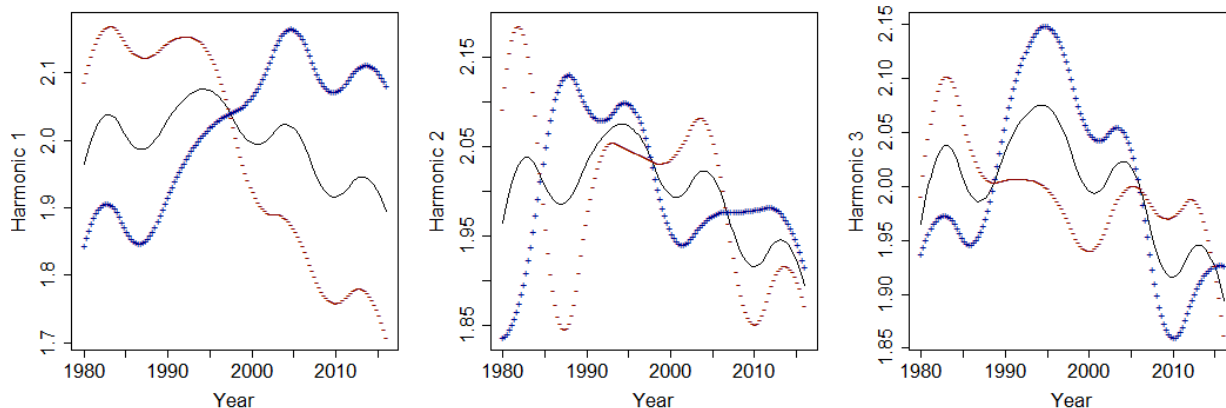


Figure 7: Harmonics plots of the first three functional principal components for the UKBMS data. The arithmetic average of the species indices is shown by a solid line, with the end of each component plotted using + (blue) and - (red). The percentages of variance for the first three components are 59.2%, 18.5% and 8.8%.

291 Both the second and third components are more difficult to interpret. For example, the
 292 second component distinguishes at one end of the range species that increase from a low
 293 abundance before declining again, and at the other end of the range species which behave
 294 similarly, but after an initial decrease from an initial high abundance. Thus one might regard
 295 the latter type of species as behaving in a similar way to the former type of species, but later
 296 in the time period, and this can be checked by reference to the species' index plots.

297 3.4.2 Comparison with species-level abundance trends

298 Plotting the first abundance functional principal component scores vs the estimated trends,
 299 as was done for the occupancy study, gives the near-linear plot of Figure 8 when a logarithmic
 300 transformation is used for the trend, which is an interesting and unexpected feature.
 301 This is due in part to the fact that what is measured is relative abundance, so that similar
 302 denominators feature when percentage trends are formed, in contrast to the situation with
 303 occupancy data.

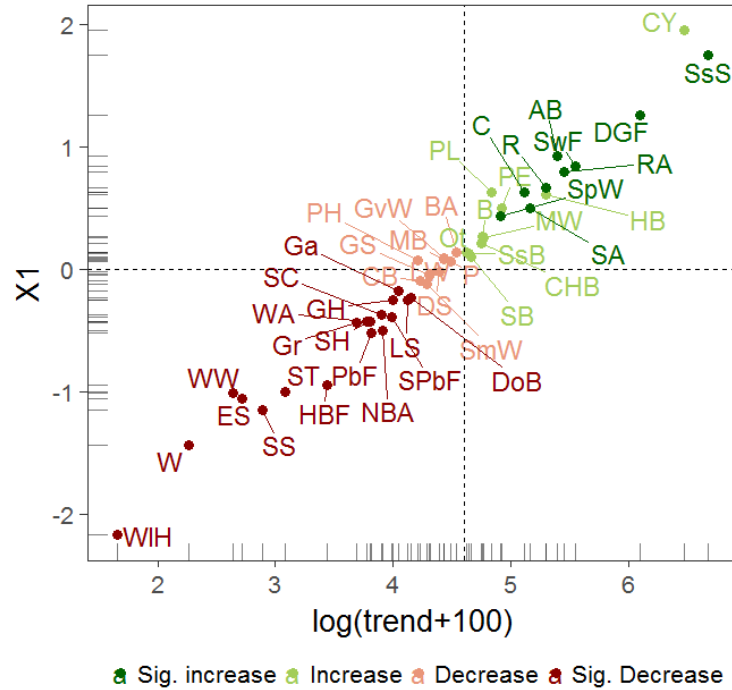


Figure 8: Plot of the first functional principal component score for the FPCA of UKBMS data plotted vs a logarithmic transformation of the estimated trend for each species. The vertical and horizontal dashed lines indicate no change in abundance and score values of zero, respectively.

304 The plot of species according to the first two functional principal components is shown
 305 in Figure 9a. The first component now measures abundance trend, and the second compo-
 306 nent distinguishes different patterns to the changes, as explained above. Note that these
 307 two components explain 77.7% of the total variance. If we include the third component
 308 then the percentage explained increases to 86.5%. A particular three-dimensional plot is
 309 given in Supplementary Figure 4 and the three-dimensional configuration can be accessed at
 310 <https://plot.ly/~EBDennis/1>. This allows the three-dimensional plots to be rotated, and the
 311 identity of individual points to be revealed.

312 Figure 9 suggests that there is no indication of clustering of species, and we have a main
 313 core of species, together with a number of outlying species. Here, and also in the case of
 314 occupancy analysis, such results are useful in suggesting how one might group indices for
 315 presentation, as well as for categorisations for indicators. Outliers may be detected formally

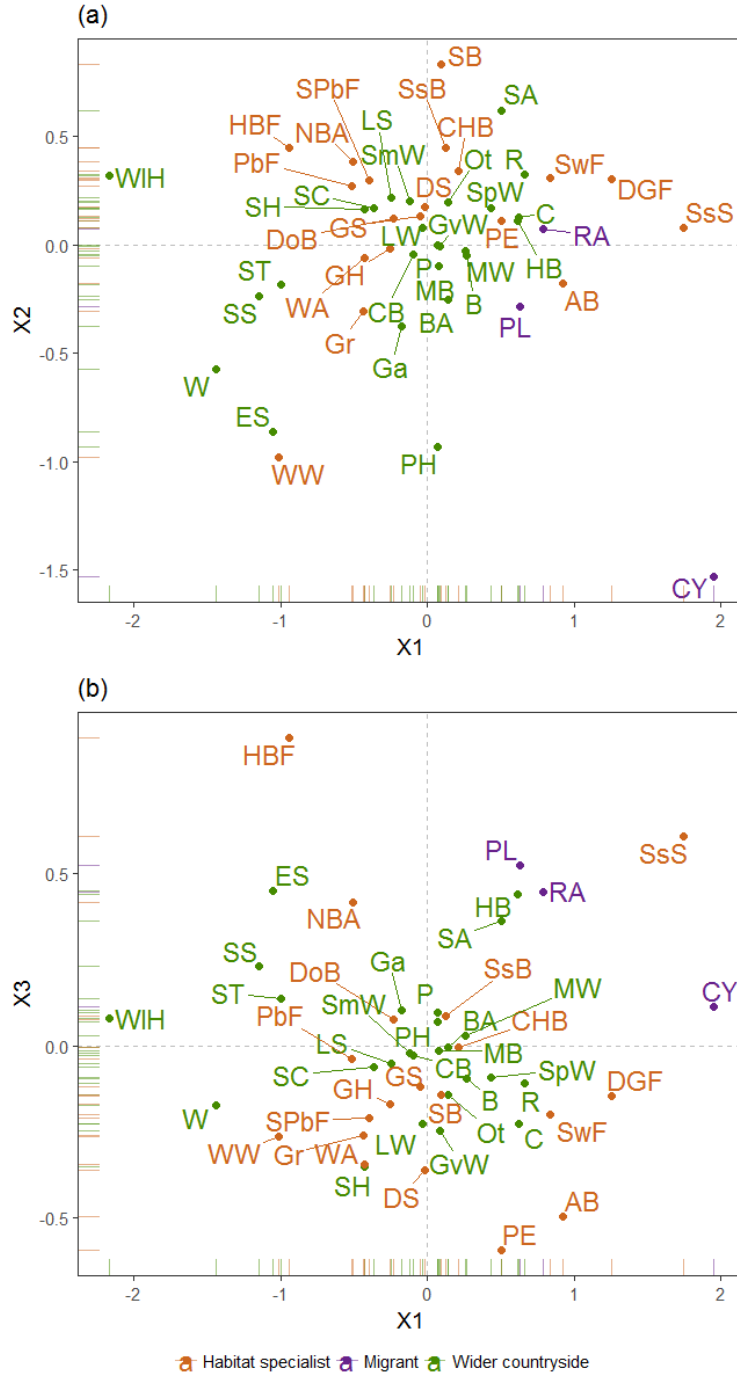


Figure 9: Plot of the functional principal component scores following a FPCA of the abundance indices: (a) components 1 and 2 and (b) components 1 and 3. The dashed lines indicate score values of zero.

316 in a variety of ways; see eg., the formal peeling approach of Barnett (1976). We note here in
 317 particular the species CY, HBF, WIH, W, WW, SsS and PE.

318 It is interesting to note the increases in abundance in all three migrants. WIH and CY
319 are at opposite ends of dimension $X1$, and their indices correspond to the extremities of that
320 axis suggested in Figure 7. The same is true of the indices of PE and HBF, at opposite ends
321 of dimension $X3$. In addition to considering the interpretation of dimensions, as here, we
322 can also use the plots in this abundance case in three dimensions in order to identify which
323 species are close to which, and therefore show similar abundance indices.

324 The three different categories of butterfly species are not as separate as for the BNM
325 case, which is in part a consequence of the standardisation of indices in the UKBMS case
326 (as seen from Supplementary Figure 2). This ties in well with the relative agreement of the
327 multi-species indicators of Figure 6.

328 **3.5 Comparison of abundance and occupancy trends**

329 In Figure 10 abundance and occupancy trends are compared, where in Figure 10a log trends
330 are shown in order to improve the presentation. There was a slight difference in the time
331 periods considered (1980-2014 and 1980-2016). We note from Figure 8 that in Figure 10b
332 the abundance axis, $X1$, is similar to $\log(\text{trend}+100)$, where “trend” refers to the abundance
333 trend, and this contributes to similarities between the two plots in Figure 10. There is
334 a greater correlation in panel (b) ($\rho = 0.36$, $p < 0.05$) than in panel (a) ($\rho = 0.20$, not
335 significant at the 5% level). Differences arise because the occupancy trends (Figure 10a) are
336 relative to the scale of the occupancy index, whereas $X2$ (Figure 10b), represents overall
337 change on the occupancy scale, since $X1$ and $X2$ are uncorrelated.

338 The positions of migrant species provide an interesting comparison and verification. In
339 terms of occupancy, all three are increasing, though not dramatically so. There is no stan-
340 dardisation in this case and CY has a smaller estimated occupancy probability than the
341 other two migrant species, in line with common observation. However in terms of abundance,
342 where there is standardisation, the three species appear to have more in common, including
343 increases in relative abundances, which might possibly be related to climate change (Sparks
344 et al., 2005).

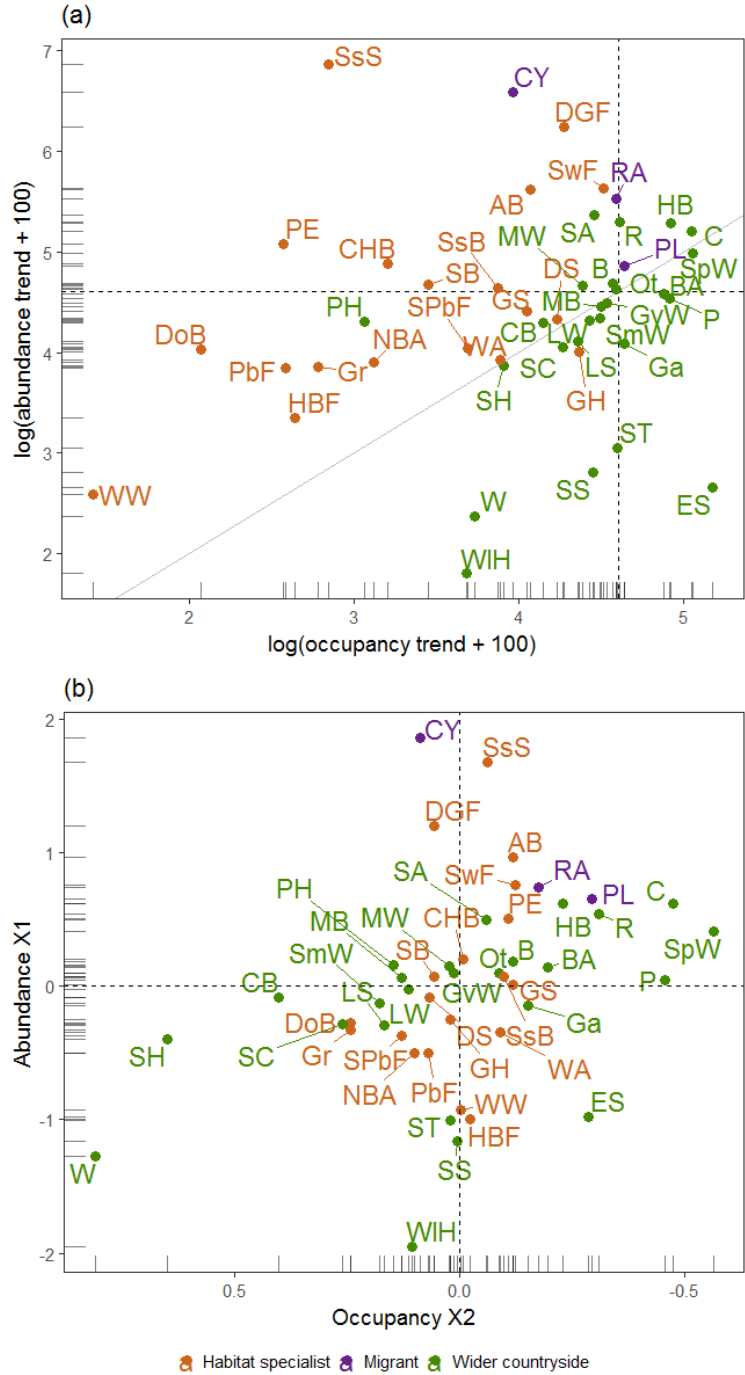


Figure 10: (a) $\log(\text{occupancy trend})$ vs $\log(\text{abundance trend})$. The grey line represents the 1-1 line and the dashed lines indicate no change. (b) Plot of the scores of the second axis ($X2$) from the FPCA of BNM vs the first axis ($X1$) from the FPCA of UKBMS. The dashed lines indicate score values of zero. The axis for occupancy $X2$ has been reversed.

3.6 Recommendations

By displaying the underlying differences among species, figures presenting functional principal component scores are much more informative than simple bar plots of percentages of significant trends, and could be considered as alternatives. We have seen that a functional principal component arises for both occupancy and abundance analyses that distinguishes between species that increase or decrease over time, and that it differs from percentage trend, which is a simplification of complex indices. Percentage trends provide simple summaries, but have been seen to be crude representations of complex temporal change.

How results of FPCA might be used in practice would depend upon the particular application, and the results obtained. In the context of occupancy, bar plots that supplement multi-species indicators could be replaced, or augmented by a plot comparing species average occupancy versus species trends (for example Figure 5b). Each species could be colour-coded appropriately, for example by the significance of the trends, by a species categorisation, or by taxon in multi-taxon applications. In combination with the multi-species indicators one would then see at a glance which species have different levels of occupancy and changes, and any clear outliers might be visible. Even in scenarios where the indicator is more species rich than the examples shown here, it would be possible to more easily interpret the variation among species, although individual species might not be decipherable. An alternative would be to use a corresponding plot showing principal component scores (for example Figure 5a), however a potential disadvantage would be that the figure may be more difficult to interpret and/or communicate to varied audiences who may use multi-species indicators.

Recommendations for accompanying visualisations for multi-species abundance indicators are more context-specific, given the less readily interpretable X_2 dimension from the FPCA, as well as the desirability of a three-dimensional representation in that case. In the absence of an absolute measure of mean abundance, suggestions similar to those made for occupancy above may be possible, for example by plotting the total species count, as a proxy for representing how abundant a species is, versus the species trends. We compare species' total counts with trends in Supplementary Figure 3, which shows interesting similarities with Figure 5b, although it should be noted that the total count provides only a crude simplification of absolute abundance, for example since missing data have not been accounted for. Alter-

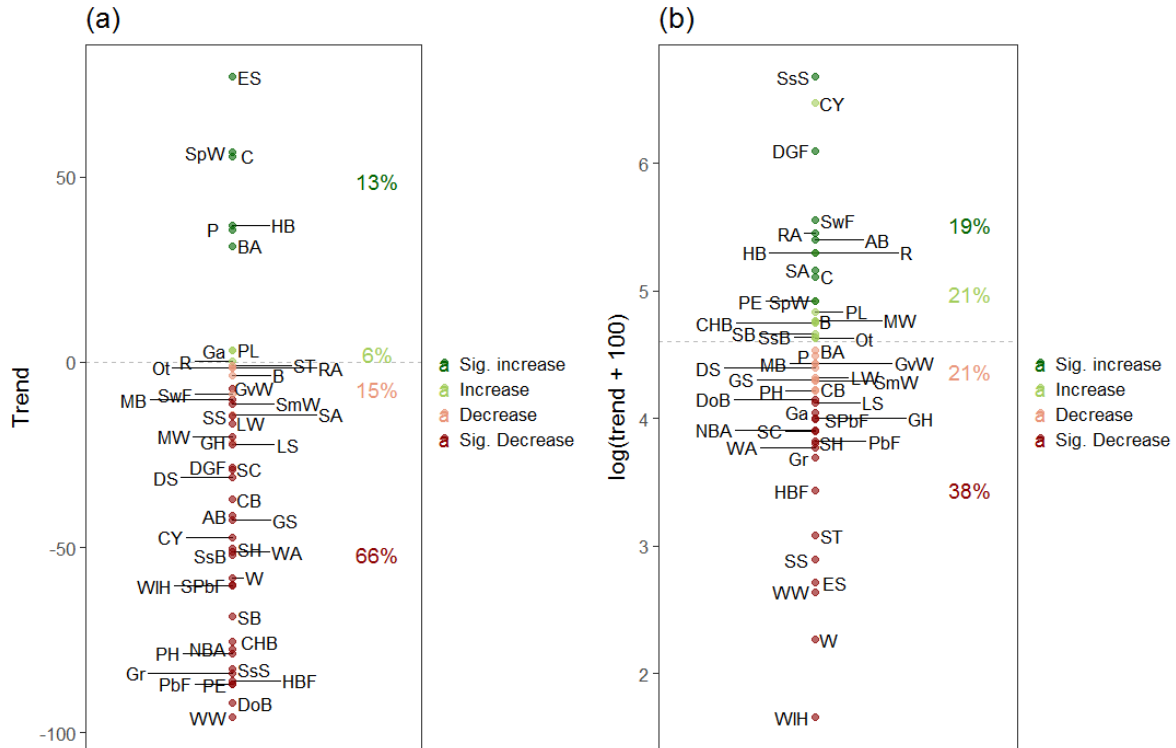


Figure 11: Summary of percentage trends for (a) occupancy and (b) relative abundance for 1980-2014. For abundance logged trends are shown. Points are coloured by significance of the trends, based on a 5% level, and the percentage of species for each category is also displayed. The dashed line indicates no change.

375 natively, where occupancy data are also available, estimates of mean occupancy could also
 376 be used as above to provide additional information when considering changes in abundance.

377 A final suggestion, which would still provide additional information over bar plots of the
 378 species trends and could be used for both abundance and occupancy indicators, would be to
 379 provide a single jitter plot of points representing species trends, or logged species trends, such
 380 as those shown for butterflies in Figure 11. The points in Figure 11 are in fact akin to the
 381 relevant rug plots in Figures 4 and 8. Points can again be categorised in various ways using
 382 colour and could also be readily shown for multiple time periods and/or subsets of species.
 383 Furthermore, the information displayed in bar plots is still displayed via the percentages,
 384 which are displayed in addition to the points in Figure 11.

4 Conclusions

We have demonstrated the potential of FPCA as a powerful new tool for the study and interpretation of species occupancy and abundance indices. Much is already known regarding the changes of UK butterfly populations (Fox et al., 2015), so that the results obtained using FPCA are in part a validation of the usefulness of the approach. As might be anticipated, we have shown that the dominant first component for occupancy, which described the high/low species' occupancy, is not present when considering relative abundance indices which were on a standardised scale. If available, applying FPCA to estimates of absolute abundance would be of interest.

For the two butterfly data sets illustrated in the paper, the analysis of occupancy data by FPCA appears to be more stable and readily interpretable than that of abundance data. This may reflect in part the fact that the abundance of species may respond more rapidly to environmental changes than their distribution (Gaston et al., 2000; van Strien et al., 2016). There is a warning here that one should not routinely combine both types of index, as individually they may exhibit different patterns of behaviour. Abundance and occupancy have been combined in some multi-species indicators, for example when occupancy has been used as a proxy for abundance when insufficient abundance data were available for certain species (Hayhow et al., 2016; van Strien et al., 2016).

Multi-species indicators and accompanying bar plots of trend provide accessible summaries of biodiversity change for reports and in advice to governments and policy-makers. The accompanying bar plots have the potential to be strengthened and/or supplemented based on the suggestions and recommendations made in this paper. The end result would then involve no more plots than existing analyses, but with far more information being displayed. Augmentation could be in terms of providing more information on which species is doing what, in terms of sizes of individual species trends, and how trends for abundance and occupancy relate to each other. This could be done via the output from FPCA analyses, primarily for a research/scientific audience, or more simply, as suggested above, for public consumption, without performing a FPCA analysis.

The recommendations made in this paper are based on analyses of data for butterflies

414 and moths, and we would encourage further investigation of the approaches of this paper
415 via applications to other taxa, and also to the context of multi-species indicators that are
416 constructed for several taxa, as with the Living Planet Index (van Strien et al., 2016). In
417 the case of multiple taxa one might expect FPCA to identify clusters of species from the
418 same taxa, and also possibly to indicate whether indicators are unduly influenced by certain
419 taxa (Buckland and Johnston, 2017), which may assist in the choice of any weightings to be
420 used (Burns et al., 2018). We can expect different features to arise from the analysis of data
421 from different taxa. Importantly the techniques used here are simple to apply using freely
422 available computer programs.

423 **Appendix A: Principal components analysis**

424 The aim of PCA (Jolliffe, 2002) is to analyse a multivariate data set in which p observations
425 are each taken on a number, n , of individuals. Typically these observations are correlated,
426 and PCA produces a set of uncorrelated derived variables known as principal components,
427 each of which is a linear combination of the original variables. PCA is the result of an axis
428 rotation, resulting from an eigen analysis of the correlation matrix of the original variables;
429 in some cases a covariance matrix is used.

430 We can think of each individual as a point in space, the dimensionality of which is the
431 number of variables measured on each individual. The derived principal components will
432 be the same in number, p . Thus in PCA the original set of $n \times p$ variables is replaced by
433 a new set of $n \times p$ variables; for each individual the variables are known as the principal
434 component scores. Principal components are typically ordered in terms of their variance,
435 and the desire is that only a small number will be needed in order to capture a high fraction
436 of the sum of the variances of the original measures. In such a case it is then possible to plot
437 individuals according to their principal component scores in the corresponding far smaller
438 dimensional space. Such plots can then be inspected for interesting features, such as outliers,
439 clusters of individuals and so forth. We shall see examples of this later for functional principal
440 components.

441 Illustrative examples of PCA include when the observations are characteristics of human

442 patients, for example, and also when there are morphometric measurements on individuals
443 (Pack et al., 1988). As each principal component is a linear function of the original variables,
444 then by considering the coefficients associated with each variable in a principal component it
445 may be possible to interpret the component. For example when the correlation matrix is used,
446 the first principal component, the one with the largest variance, is typically a measurement
447 of size; we would realise this because the coefficients would all be roughly the same size with
448 the same sign. Potentially the more interesting components are those with smaller variances,
449 and in terms of shape measurements on human beings this can be a contrast between the
450 size of the head and the size of the rest of the body; this would manifest itself if the sign of
451 the head coefficient was different from those of the other shape measurements.

452 **Acknowledgements**

453 The BNM is run by Butterfly Conservation with support from Natural England. The UKBMS
454 is run by Butterfly Conservation, the Centre for Ecology & Hydrology, and the British Trust
455 for Ornithology, in partnership with a consortium of government agencies. The UKBMS is
456 indebted to all volunteers who contribute data to the scheme and thanks all recorders who
457 contribute to the BNM and NMRS. BJTM was supported by a Leverhulme fellowship.

458 **References**

- 459 Asher, J., Warren, M., Fox, R., Harding, P., Jeffcoate, G., and Jeffcoate, S. (2001). *The*
460 *Millennium Atlas of Butterflies in Britain and Ireland*. Oxford University Press.
- 461 August, T., Powney, G., Outhwaite, C., and Issac, N. (2017). *BRCindicators: Creating*
462 *biodiversity indicators for species occurrence data*. R package version 1.0.
- 463 Barnett, V. (1976). The ordering of multivariate data. *Journal of Royal Statistical Society*
464 *Series A*, 139:318–354.
- 465 Brereton, T. M., Botham, M. S., Middlebrook, I., Randle, Z., Noble, D., and Roy, D. B.
466 (2017). United Kingdom Butterfly Monitoring Scheme report for 2016. Technical report,
467 Centre for Ecology & Hydrology & Butterfly Conservation.

- 468 Brereton, T. M., Cruickshanks, K. L., Risely, K., Noble, D. G., and Roy, D. B. (2011a).
469 Developing and launching a wider countryside butterfly survey across the United Kingdom.
470 *Journal of Insect Conservation*, 15:279–290.
- 471 Brereton, T. M., Roy, D. B., Middlebrook, I., Botham, M. S., and Warren, M. S. (2011b).
472 The development of butterfly indicators in the United Kingdom and assessments in 2010.
473 *Journal of Insect Conservation*, 15(1–2):139–151.
- 474 Buckland, S. and Johnston, A. (2017). Monitoring the biodiversity of regions: Key principles
475 and possible pitfalls. *Biological Conservation*, 214:23–34.
- 476 Buckland, S. T., Studeny, A. C., Magurran, A. E., Illian, J. B., and Newson, S. E. (2011). The
477 geometric mean of relative abundance indices: a biodiversity measure with a difference.
478 *Ecosphere*, 2:1–15.
- 479 Burns, F., Eaton, M., Hayhow, D., Outhwaite, C., Al Fulaij, N., August, T., Boughey, K.,
480 Brereton, T., Brown, A., Bullock, D., et al. (2018). An assessment of the state of nature
481 in the United Kingdom: A review of findings, methods and impact. *Ecological Indicators*,
482 94:226–236.
- 483 Defra (2018). *UK Biodiversity Indicators 2018*. Department for Environment Food and Rural
484 Affairs, London, UK.
- 485 Dennis, E., Brereton, T., Morgan, B., Fox, R., Shortall, C., Prescott, T., and Foster, S.
486 (2019). Trends and indicators for quantifying moth abundance and occupancy in scotland.
487 *Journal of Insect Conservation*.
- 488 Dennis, E. B., Morgan, B. J. T., Freeman, S. N., Brereton, T., and Roy, D. B. (2016). A
489 generalized abundance index for seasonal invertebrates. *Biometrics*, 72(4):1305–1314.
- 490 Dennis, E. B., Morgan, B. J. T., Freeman, S. N., Ridout, M. S., Brereton, T. M., Fox, R.,
491 Powney, G. D., and Roy, D. B. (2017). Efficient occupancy model-fitting for extensive
492 citizen-science data. *PLoS ONE*, <https://doi.org/10.1371/journal.pone.0174433>.

493 Eaton, M. A., Burns, F., Isaac, N. J., Gregory, R. D., August, T. A., Barlow, K. E., Brereton,
494 T., Brooks, D. R., Al Fulaij, N., Haysom, K. A., et al. (2015). The priority species indicator:
495 measuring the trends in threatened species in the uk. *Biodiversity*, 16(2-3):108–119.

496 Fox, R., Brereton, T. M., Asher, J., August, T. A., Botham, M. S., Bourn, N. A. D.,
497 Cruickshanks, K. L., Bulman, C. R., Ellis, S., Harrower, C. A., Middlebrook, I., Noble,
498 D. G., Powney, G. D., Randle, Z., Warren, M. S., and Roy, D. B. (2015). *The State of the*
499 *UK's Butterflies 2015*. Butterfly Conservation and the Centre for Ecology & Hydrology,
500 Wareham, Dorset.

501 Gaston, K. J., Blackburn, T. M., Greenwood, J. J. D., Gregory, R. D., Quinn, R. M., and
502 Lawton, J. H. (2000). Abundance - occupancy relationships. *J. Appl. Ecol.*, 37(Suppl.
503 1):39–59.

504 Gower, J. C. (1975). Generalized procrustes analysis. *Psychometrika*, 40:33–51.

505 Gregory, R. D., Van Strien, A., Vorisek, P., Meyling, A. W. G., Noble, D. G., Foppen, R.
506 P. B., and Gibbons, D. W. (2005). Developing indicators for European birds. *Philosophical*
507 *Transactions of the Royal Society B: Biological Sciences*, 360:269–288.

508 Hayhow, D. B., Ausden, M. A., Bradbury, R. B., Burnell, D., Copeland, A. I., Crick, H.
509 Q. P., Eaton, M. A., Frost, T., Grice, P. V., Hall, C., Harris, S. J., Morecroft, M. D.,
510 Noble, D. G., Pearce-Higgins, J. W., Watts, O., and Williams, J. M. (2017). *The State*
511 *of the UK's Birds 2017*. RSPB, BTO, WWT, DAERA, JNCC, NE and NRW, Sandy,
512 Bedfordshire.

513 Hayhow, D. B., Burns, F., and Eaton, M. A. e. a. (2016). *State of Nature*. The State of
514 Nature partnership.

515 Jolliffe, I. T. (2002). *Principal Component Analysis, Second Edition*. Springer-Verlag, New
516 York.

517 Morgan, B. J. T. (2009). *Applied Stochastic Modelling*. Texts in Statistical Science. CRC
518 Chapman & Hall, 2nd edition.

- 519 Pack, P., Jolliffe, I., and Morgan, B. (1988). Influential observations in principal component
520 analysis: a case study. *Journal of Applied Statistics*, 15:39–52.
- 521 Pollard, E. and Yates, T. J. (1993). *Monitoring Butterflies for Ecology and Conservation:
522 the British Butterfly Monitoring Scheme*. Chapman & Hall, London.
- 523 R Core Team (2018). *R: A Language and Environment for Statistical Computing*. R Foun-
524 dation for Statistical Computing, Vienna, Austria.
- 525 Ramsay, J., Hooker, G., and Graves, S. (2009). *Functional data analysis with R and MAT-
526 LAB*. Springer Science & Business Media.
- 527 Ramsay, J. O., , and Silverman, B. W. (2005). *Functional Data Analysis, Second edition*.
528 Springer, New York.
- 529 Ramsay, J. O., Wickham, H., Graves, S., and Hooker, G. (2017). *fda: Functional Data
530 Analysis*. R package version 2.4.7.
- 531 Sievert, C., Parmer, C., Hocking, T., Chamberlain, S., Ram, K., Corvellec, M., and Despouy,
532 P. (2017). *plotly: Create Interactive Web Graphics via 'plotly.js'*. R package version 4.7.1.
- 533 Soldaat, L. L., Pannekoeka, J., Verweija, R. J. T., van Turnhout, C. A. M., and van Strien,
534 A. J. (2017). A Monte Carlo method to account for sampling error in multi-species indi-
535 cators. *Ecological Indicators*, 81:340–347.
- 536 Sparks, T. H., Roy, D. B., and Dennis, R. L. H. (2005). The influence of temperature on
537 migration of Lepidoptera into Britain. *Global Change Biology*, 11(3):507–514.
- 538 Tittensor, D. P., Walpole, M., Hill, S. L., Boyce, D. G., Britten, G. L., Burgess, N. D.,
539 Butchart, S. H., Leadley, P. W., Regan, E. C., Alkemade, R., et al. (2014). A mid-term
540 analysis of progress toward international biodiversity targets. *Science*, 346(6206):241–244.
- 541 van Strien, A. J., Gmelig Meyling, A. W., and Herder, J. E. e. a. (2016). Modest recovery of
542 biodiversity in a western European country: the Living Planet Index for the Netherlands.
543 *Biological Conservation*, 200:44–50.

544 van Strien, A. J., Soldaat, L. L., and Gregory, R. D. (2012). Desirable mathematical prop-
545 erties of indicators for biodiversity change. *Ecological Indicators*, 14:202–208.