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

Emily Dennis[1,2,*], Byron Morgan[2], Richard Fox[1], David Roy[3] & Tom Brereton[1]

September 21, 2018

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

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

[3]Centre 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 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 temporal changes. We find differences in species patterns of occupancy and abundance, providing a warning against routinely combining both types of index within multi-species indicators, for example when using occupancy as a proxy for abundance when sufficient abundance data are not available. By identifying the differences between species, figures displaying functional principal component scores are much more informative than the simple bar plots of percentages of significant trends that often accompany multi-species indicators. Informed by the outcomes of the FPCA, we make recommendations for accompanying visualisations for multi-species indicators, and discuss how these are likely to be context and audience specific. We show that, in the absence of FPCA, using mean species occupancy and total abundance can provide additional, accessible information to complement species-level trends. At the simplest level, we suggest using jitter plots to display variation in species-level trends. We recommend the routine augmentation of multi-species indicators in the future with additional statistical procedures and figures, to serve as an aid to improve communication and understanding of biodiversity metrics, as well as reveal potentially hidden patterns of behaviour and guide additional directions for investigation.

Key words: Biodiversity indicators; BNM; Citizen science data; Functional principal component analysis; Multi-species indices; Outlier detection; Procrustes analysis; UKBMS;

1 Introduction

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

The geometric mean of component species indices is widely used to calculate multi-species indicators (Gregory et al., 2005; Buckland et al., 2011; Van Strien et al., 2012). However there remains variation among different indicators, for example with regard to if and how uncertainty in the estimated species-level indices is incorporated (Soldaat et al., 2017), and in the presentation of both indicators and associated trends. Multi-species indicators are produced for all species within a taxonomic group, or subsets based on classifying the component species. For example, UK butterfly indicators are typically produced separately for habitat specialist versus wider countryside species (Fox et al., 2015), and separate UK indicators are typically produced for farmland, woodland and wetland bird species (Hayhow et al., 2017). Indicators are typically produced from combining species-level indices for either annual estimates of occupancy or an annual index of abundance, for which the underlying methods used to estimate the indices can also vary among taxa.

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

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

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

- to represent the data in ways that aid further analysis,

- to display the data so as to highlight various characteristics,
- to study important sources of pattern and variation among the data.

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

2 Materials and methods

2.1 Functional Principal Component Analysis

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

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

2.2 Application to biodiversity indices

The techniques used in this paper may be applied to abundance or occupancy indices for multiple species of any taxon (or combination of multiple taxa). For demonstration we analyse data from the Butterflies for the New Millennium (BNM) database and the UK Butterfly Monitoring Scheme (UKBMS). Prior to the application of FDA, appropriate annual indices of occupancy and abundance were produced from the two data sets. We consider data from the BNM and UKBMS from 1980 onwards because most species have a full run of UKBMS data from 1980. Based on the data available, we consider 1980-2014 for BNM and 1980-2016 for UKBMS. This resulted in occupancy and abundance data sets for 47 UK butterfly species (out of a total of 59, of which 50 typically contribute to UK biodiversity indicators), which are listed in the Supplementary material along with the species codes using in the paper.

2.2.1 Producing species-level indices

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

The UKBMS consists of a long-running network of transects which began in 1976 with 34 sites, but has grown to nearly 1500 transects monitored each year (Brereton et al., 2017). Since 2009 this additionally includes reduced-effort data from the Wider Countryside Butterfly Survey (Brereton et al., 2011a). Under standardised weather conditions, counts are made weekly from the beginning of April until the end of September (Pollard and Yates, 1993). Indices of relative abundance are estimated from the UKBMS for each species using a Generalised Abundance Index approach (Dennis et al., 2016). Species-level indices are given for four illustrative species in Figure 1b, and for all 47 species in Supplementary Figure 2. UKBMS indices are typically presented on the $\log_{10}$ scale where they either start at 2 or have a mean of 2. It will be seen that there is therefore a fundamental difference between these indices and those relating to occupancy, when the entire probability range was possible.

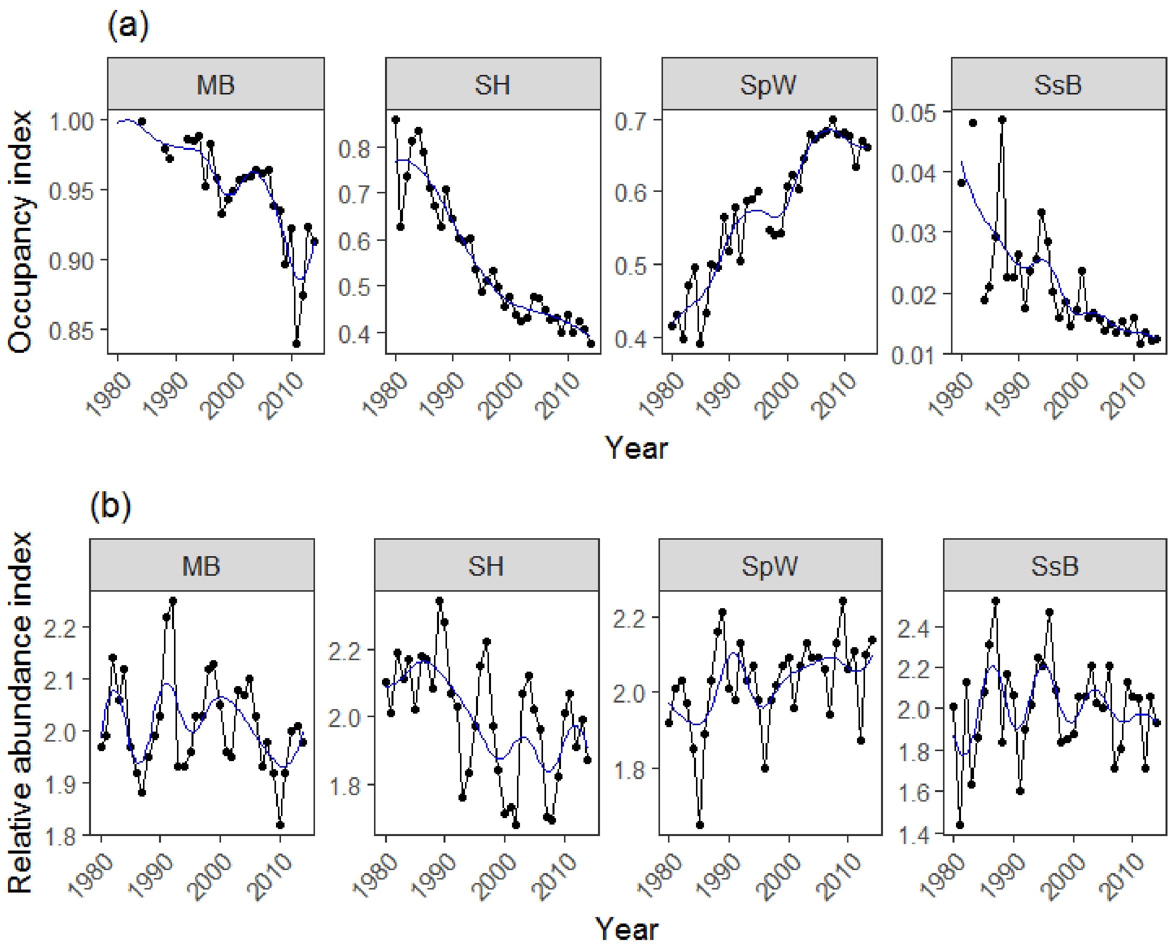

Figure 1: Occupancy (a) and relative abundance (b) indices for four illustriative butterflu species. Smoothed indices (blue) were produced using B-splines. Plots for all 47 species are given in the Supplementary Material.

2.2.2 Calculating species-level trends

For each species, a weighted logistic regression was fitted to the occupancy index, where the inverse of the index standard errors were used as weights. The standard errors were calculated using the Delta method, rather than the bootstrapping approach in Dennis et al. (2017), which can under perform in cases with limited data. Percentage changes for 1980-2014 were then estimated from the predicted values of the regression. Percentage changes in relative abundance were estimated by fitting simple linear regressions to the species' indices of relative abundance for 1980-2016.

2.2.3 Calculating multi-species indicators

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

2.2.4 Applying FPCA

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

The input to the FPCAs is a set of smoothed curves of the species indices, with one per species, separately for each of occupancy or relative abundance. These are displayed for all 47 species in Supplementary Figures 1 and 2 for both occupancy and abundance. Prior to smoothing, small numbers of missing year index values were interpolated (only for Duke of Burgundy for abundance, and for 31 species for the occupancy indices). The smoothed estimates were produced using the `fda` package using B-splines with 10 basis functions and order 3. Alternative spline smooths were considered and there was a striking stability in the results and conclusions with regard to how much smoothing was adopted. The smoothing used in these analyses does not take account of relative precision of the species-level indices, where more recent estimates and better recorded/monitored species are typically more precise.

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

We first review the associated harmonics plots, which display the principal component functions, and then the corresponding functional principal component scores. The scores are formed in an analogous way to how principal component scores are obtained for standard

PCA, though it is more complicated due to the use of curves rather than measurements (Ramsay et al., 2005, p. 149). We distinguish between habitat specialists, migrants and wider countryside species, based on the classification in Asher et al. (2001). We draw comparisons with species-level abundance and occupancy trends estimated from the associated indices. A three-dimensional plot for the first three principal components for the UKBMS analysis was created using the `plotly` package (Sievert et al., 2017).

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

3 Results and discussion

3.1 Indicators for occupancy data

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

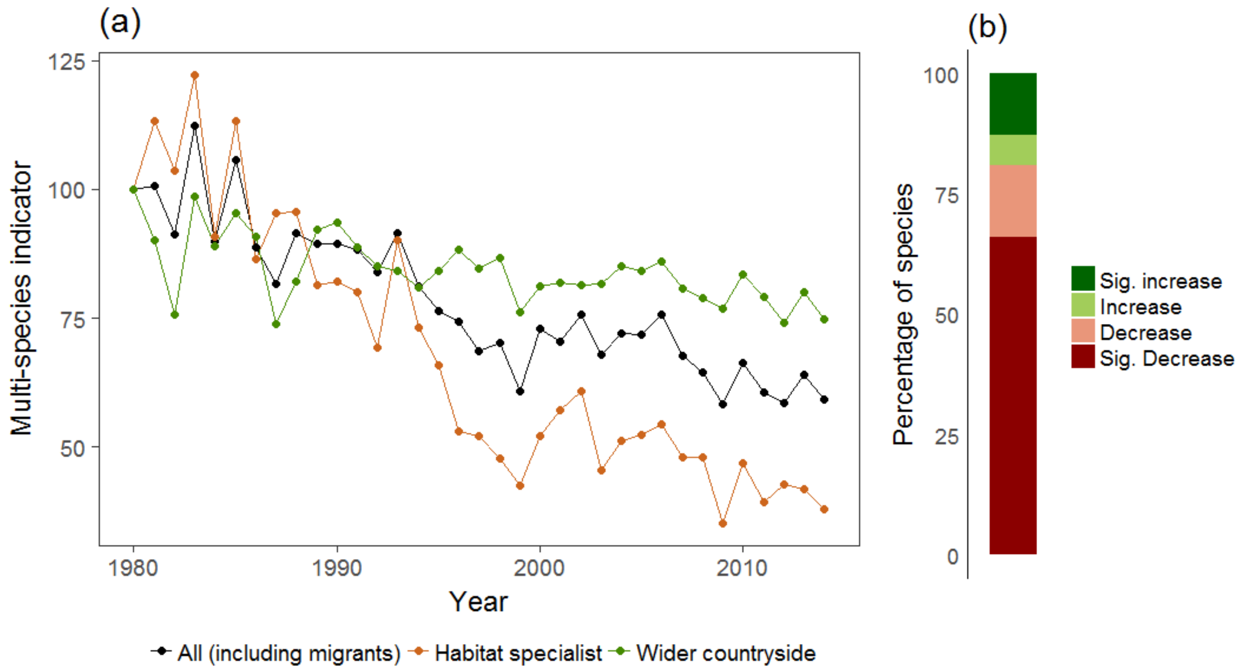

Figure 2: (a) Multi-species occupancy indicators calculated by the geometric mean of the occupancy indices for 47 UK butterflies for 1980–2014.(b) Bar plot giving percentage increases/decreases of individual butterfly species, where significance refers to the 5% level.

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

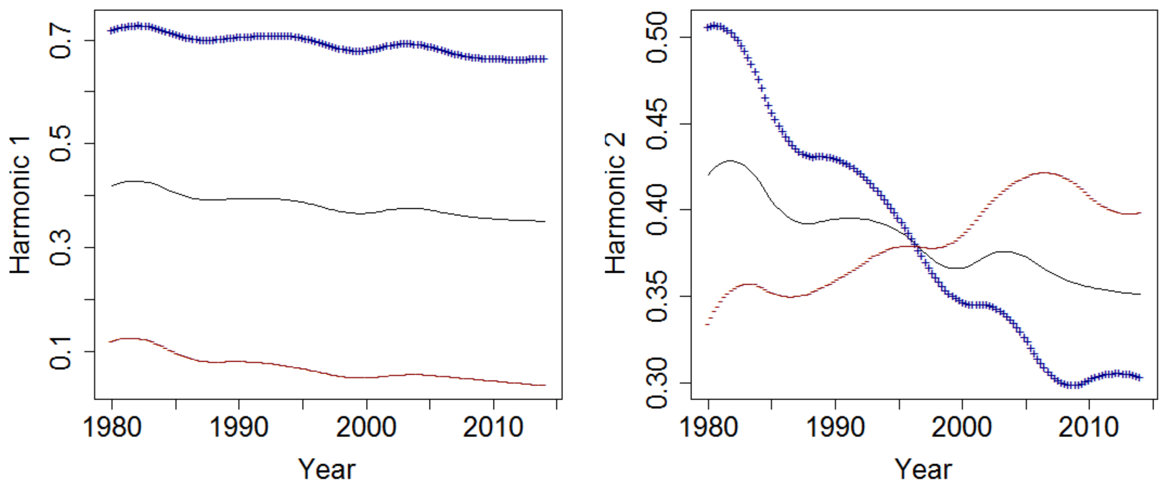

Figure 3: Harmonics plots of the first two functional principal components for the BNM data. The arithmetic average of the species indices is shown by a solid line, with the end of each component plotted using + (blue) and - (red). The percentages of variance for the first two components are 97.4% and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

variance, just 1.9%, this component has a clear interpretation.

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

With minor differences, we have found the general patterns of the harmonics plots of Figure 3 to appear in other occupancy analyses, for example of Scottish moths (Dennis and Brereton, 2018), when occupancy data on 225 moth species were analysed (Figure 3 of the Supplementary Material). The same is also true if we divide the data into the first half and second half time periods and analyse the two halves separately (see Section 3 of the Supplementary Material).

3.2.2 Comparison with species-level occupancy trends

Figure 4a shows the estimated percentage trend for each species, plotted against the corresponding second functional principal component score, denoted by X2. Note that all principal component scores are centered on zero due to the mean centering at each individual time point. As we might expect from the interpretation of the second component provided above by Figure 3, there is a relationship between the trend and the second functional component score, however it is not a linear one. The association is approximately linear for wider countryside species, however habitat specialists, with generally lower occupancy, necessarily have smaller absolute changes, resulting in relatively small values for $X2$.

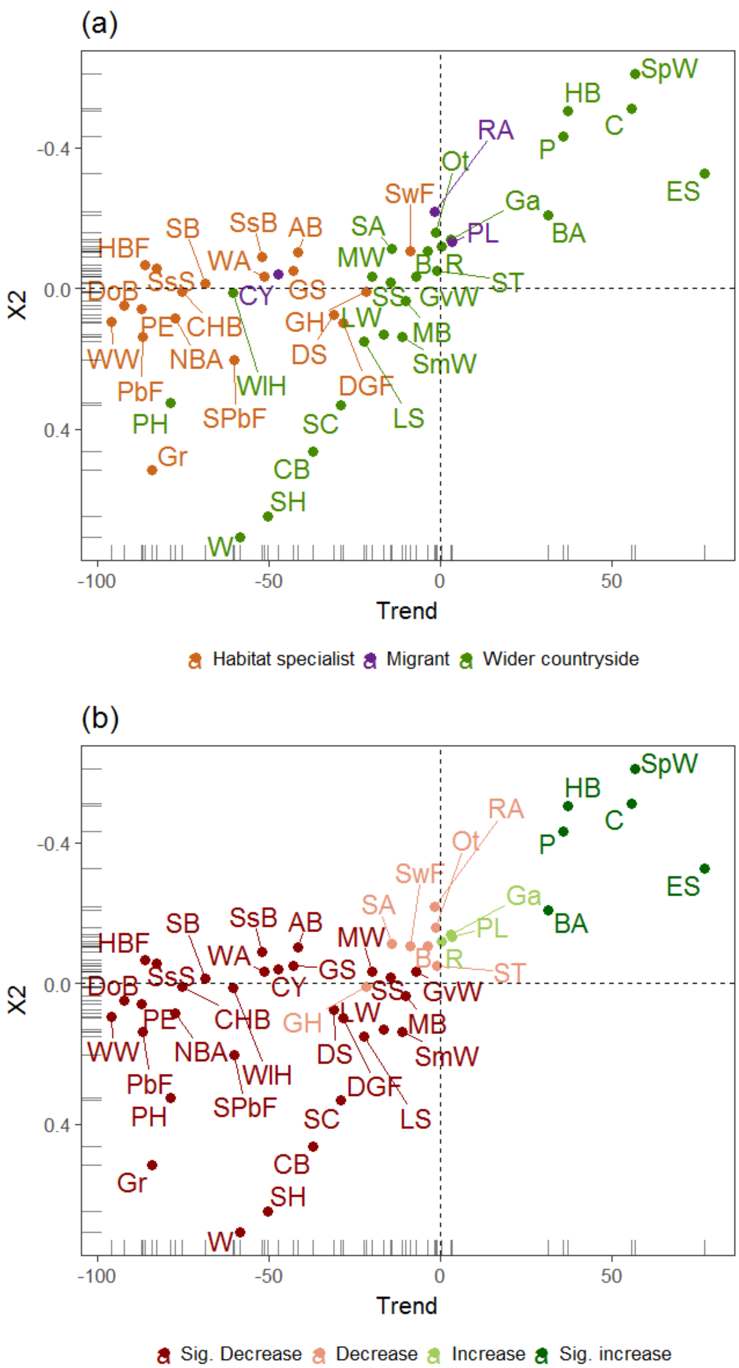

Figure 4: Estimated species occupancy trends (percentage changes) versus the corresponding scores that result on the second axis ($X2$) from the FPCA analysis of the BNM data; the locations of points are the same in both plots. (a) Colours indicate species classification: habitat specialists, migrants and wider countryside; (b) colours indicate category of trend, as summarised in Figure 2b. The vertical and horiztonal dashed lines indicate no change in occupancy and X2 values of zero, respectively.

Figure 4b distinguishes between values that are significantly changing (increasing or decreasing), each at the 5% level. While there is a correlation between the $X2$ and trend values, the $X2$ axis is reflecting shapes of the individual species indices in a more complex way than simply ordering the species according to their estimated trend value. It is instructive to relate the points back to the index plots for the species that they represent. Rug plots are displayed along the axes in Figure 4, which indicate the values taken by species along those axes, and this feature recurs in similar plots in the paper.

In Figure 5a each species is plotted according to the scores of its first two functional principal components, $X1$, measuring average occupancy, and $X2$ indicating whether the species is increasing or decreasing over time. Figure 5a identifies two main clusters of species, driven by the size of occupancy, suggesting that it might be of interest to analyse these two clusters separately. This is in fact what is essentially done when multi-species indicators are produced separately for habitat specialists and wider countryside species (Figure 2a). However this distinction is not clear cut in that a small number of the wider countryside species appear similarly placed to the habitat specialists. These are the wider countryside species with relatively low estimates of occupancy probability. The second component corresponds to species that are increasing/declining over the entire time period, and therefore provides much of the information in the individual species occupancy indices in Supplementary Figure 1. Thus here the $X2$ values alone, on the y-axis, illustrate much of the information that is hidden when the geometric mean indicator is formed.

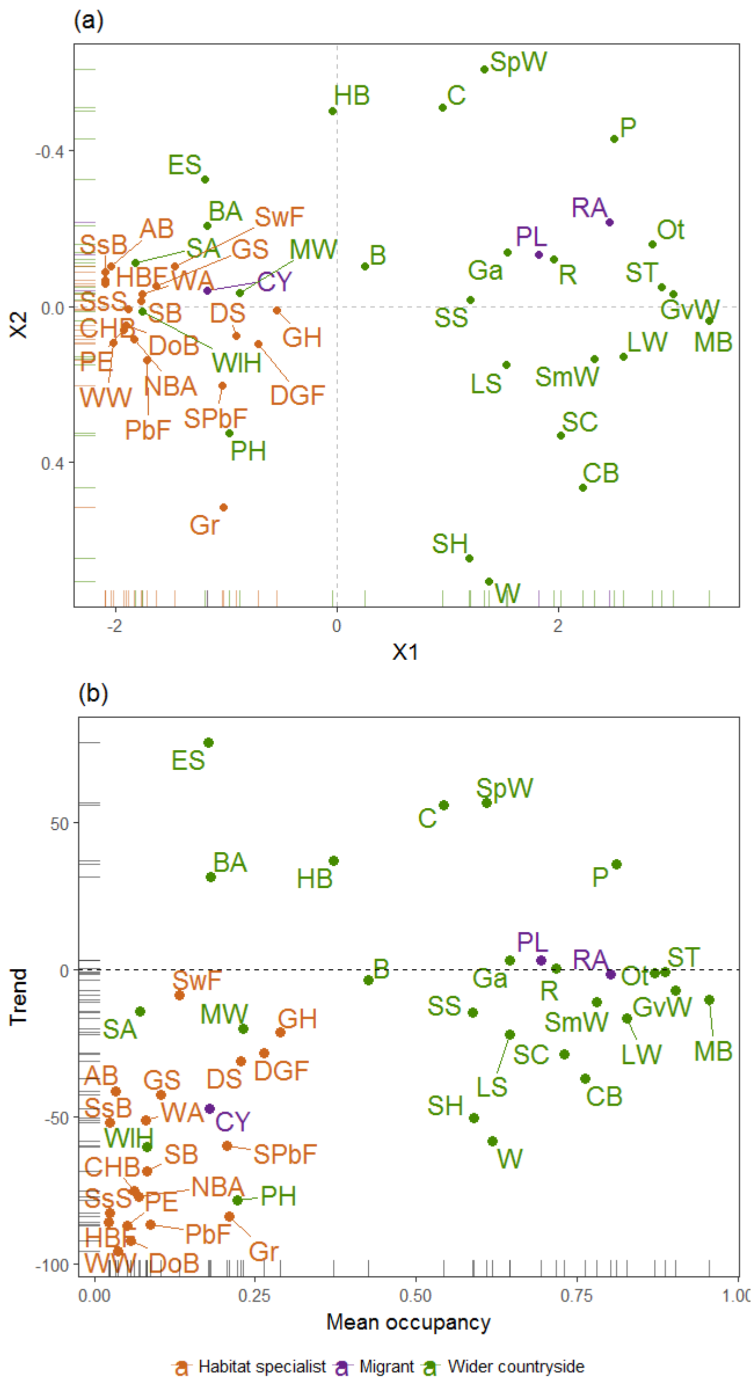

Figure 5: (a) Plot of the two functional principal component scores, $X1$, measuring average occupancy, and $X2$, measuring increase or decrease, for all 48 species for the full time period. The axis for $X2$ has been reversed. The dashed lines indicate score values of zero.(b) For comparison we replace $X1$ by the average occupancy index value and $X2$ by the estimated species occupancy trend. The vertical and horiztonal dashed lines indicate no change in abundance and score values of zero, respectively. The horiztonal dashed line indicates no change in occupancy.

Species to the right of $X1$ have high occupancy, and those to the left have low occupancy. Species at the top of $X2$ are increasing, and those at the bottom are decreasing. It is easy to verify this: see for example the positions of Meadow Brown (MB, high occupancy and minimal change over time), Speckled Wood (SpW, medium occupancy and increasing over time), and Grayling (Gr, relatively low occupancy and much temporal decline) for which species-level occupancy indices are shown in Figure 1a. Wall and Small Heath stand out as showing the lowest values of $X2$, representing the largest absolute declines in occupancy, and despite being wider countryside species they are considered to be priority species for conservation.

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

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

3.3 Indicators for abundance data

Multi-species indicators for the relative abundance of butterflies, formed using the geometric mean, are shown in Figure 6a, for all species and also for habitat specialists and wider countryside species separately. The patterns of behaviour shown here are somewhat different from those in Figure 2a, and we note also that there is a degree of apparent cycling for the indicators. The relevant species indices of abundance are given in Supplementary Figure 2.

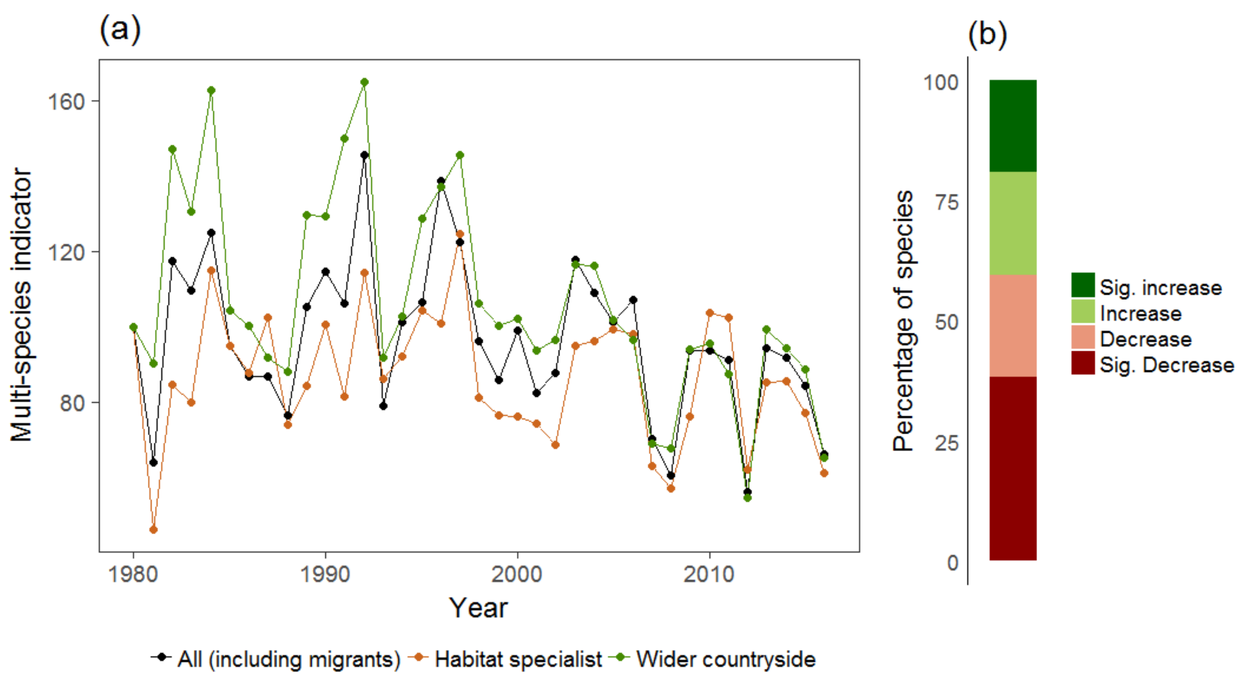

Figure 6: (a) Multi-species abundance indicators calculated by the geometric mean of the relative abundance indices for 47 UK butterflies for 1980–2016.(b) Bar plot giving percentage increases/decreases of individual butterfly species, where significance refers to the 5% level.

3.4 FPCA of abundance data

3.4.1 Harmonics plots

The harmonics plots resulting from the FPCA applied to the relative abundance indices (Figure 7) show differences compared to those obtained for occupancy indices (Figure 3), partly due to the differences in scale of the two types of indices. Since the relative abundance indices are all normalised in the same way, the dominant first component for the occupancy case is no longer present, and instead we have as the first component one that resembles the second component for the occupancy FPCA, in this case indicative of an increase or decline in abundance.

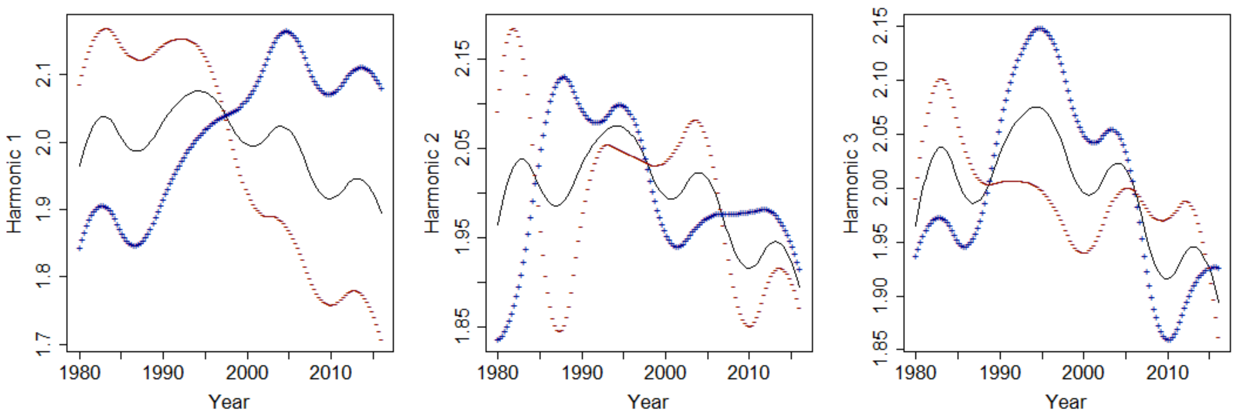

Figure 7: Harmonics plots of the first three functional principal components for the UKBMS data. The arithmetic average of the species indices is shown by a solid line, with the end of each component plotted using + (blue) and - (red). The percentages of variance for the first three components are 59.2%,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

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

3.4.2 Comparison with species-level abundance trends

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

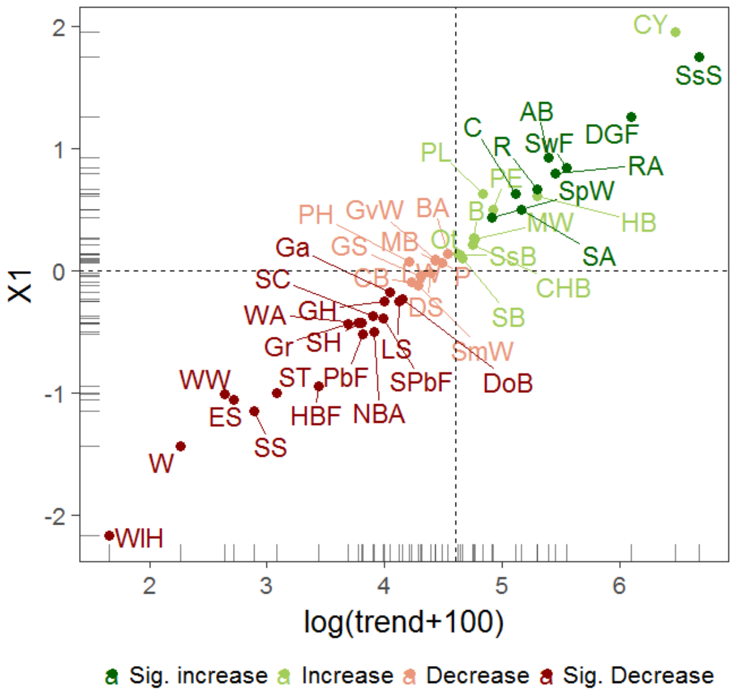

Figure 8: Plot of the first functional principal component score for the FPCA of UKBMS data plotted vs a logarithmic transformation of the estimated trend for each species. The vertical and horizontal dashed lines indicate no change in abundance and score values of zero, respectively.

The plot of species according to the first two functional principal components is shown in Figure 9a. The first component now measures abundance trend, and the second component distinguishes different patterns to the changes, as explained above. Note that these two components explain 77.7% of the total variance. If we include the third component then the percentage explained increases to 86.5%. A particular three-dimensional plot is given in Supplementary Figure 3 and the three-dimensional configuration can be accessed at https://plot.ly/∼EBDennis/1. This allows the three-dimensional plots to be rotated, and the identity of individual points to be revealed.

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

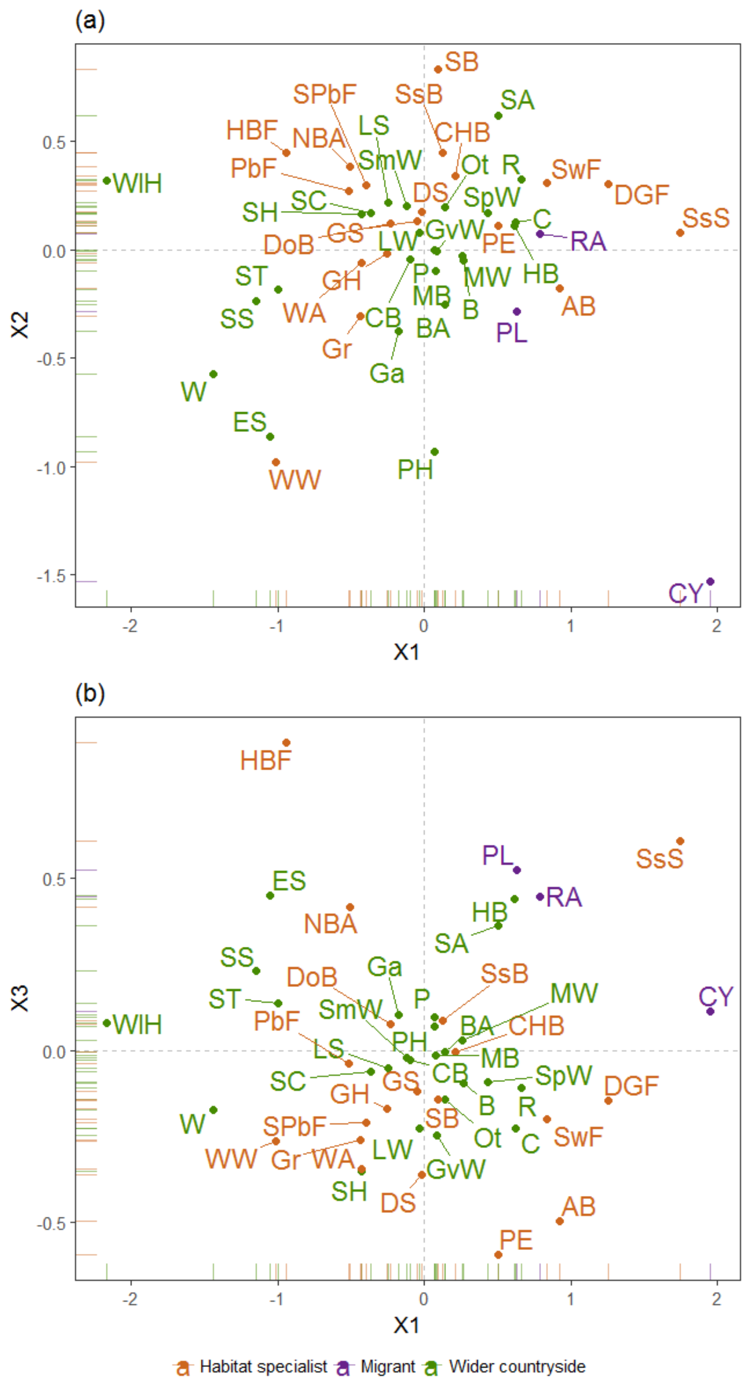

Figure 9: Plot of the functional principal component scores following a FPCA of the abundance indices: components 1 and 2 (a) and components 1 and 3 (b). The dashed lines indicate score values of zero.

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

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

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

3.5 Comparison of abundance and occupancy trends

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

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

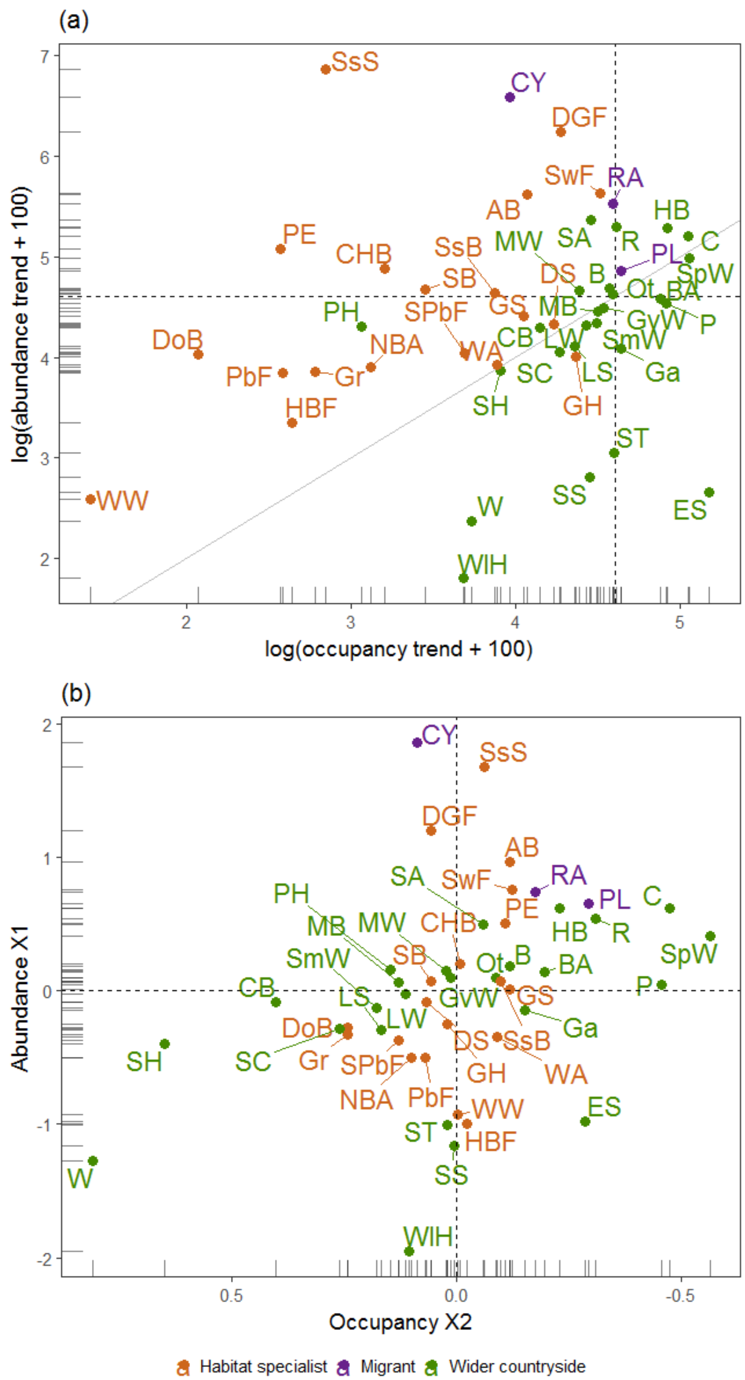

Figure 10: (a) Log(occupancy trend) vs log(abundance trend). The grey line represents the 1-1 line and the dashed lines indicate no change. (b) Plot of the scores of the second axis ($X2$) from the FPCA of BNM vs the first axis ($X1$) from the FPCA of UKBMS. The dashed lines indicate score values of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

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. It has been applied to the two main butterfly databases in the UK. 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. We have demonstrated the differences that can arise between using normalised and non-normalised indices, as well as between relative abundance and occupancy.

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. In the context of multi-species indicators, abundance and occupancy have been combined where for some species data are insufficient to produce an abundance index, therefore a species occupancy index is instead used as a proxy, see for example the UK State of Nature assessment (Hayhow et al., 2016; Burns et al., 2018) and the Living Planet Index for the Netherlands (Van Strien et al., 2016).

By displaying the underlying differences among species, figures displaying 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.

The use of splines for the FDA showed a robustness of the results regarding using different amounts of smoothing. It is possible, however, that for detailed scientific application to small numbers of species that it would be interesting to explore the use of cross-validation for choice of the amount of smoothing, for each species separately.

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. 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 $X2$ 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. Alternatively, where occupancy data are also available, estimates of mean occupancy could also be used as above to provide additional information when considering changes in abundance. A final suggestion, which would still provide additional information over bar plots of the species trends and could be used for both abundance and occupancy indicators, would be to provide a single jitter plot of points representing species trends, or logged species trends, such as those shown for butterflies in Figure 11. The points in Figure 11 are in fact akin to the relevant rug plots in Figures 4 and 8. Points can again be categorised in various ways using colour and could also be readily shown for multiple time periods and/or subsets of species.

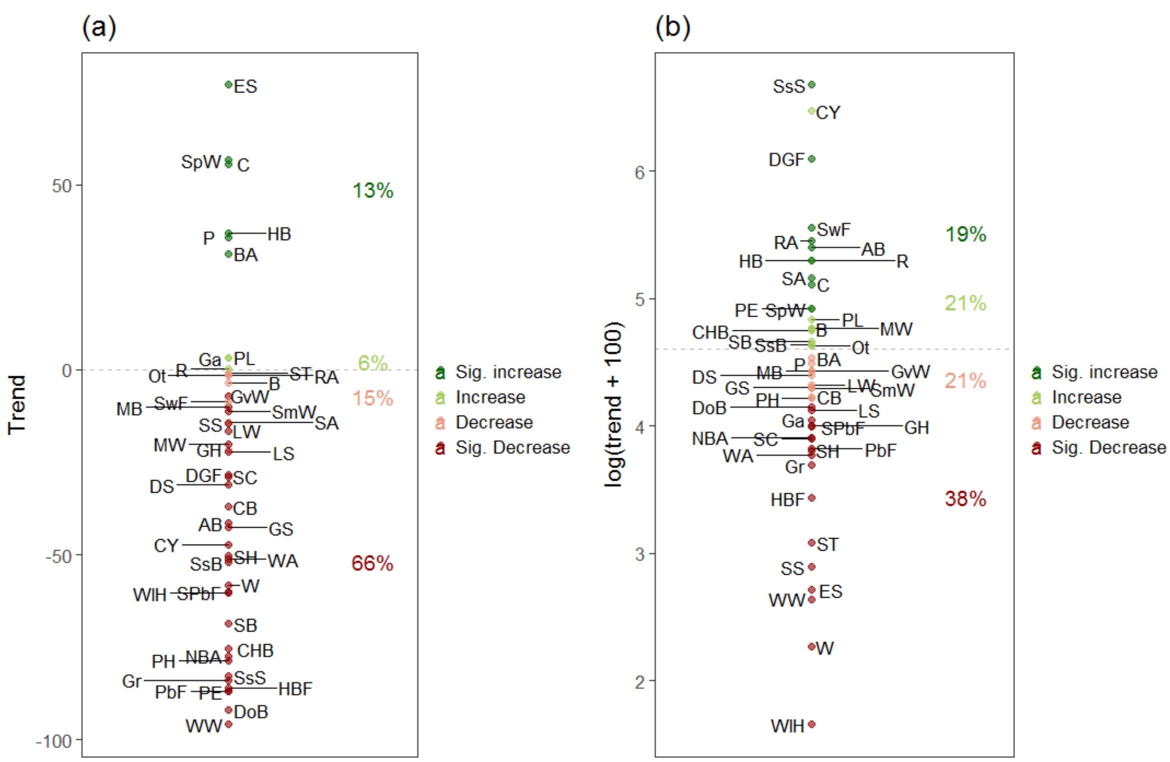

Figure 11: Summary of percentage trends for (a) occupancy and (b) relative abundance for 1980-2014. For abundance logged trends are shown. Points are coloured by significance of the trends, based on a 5% level, and the percentage of species for each category is also displayed. The dashed line indicates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

Furthermore, the information displayed in bar plots is still displayed via the percentages, which are displayed in addition to the points in Figure 11.

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 above. 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 approaches of this paper are applicable to other taxa, and also to when multi-species indicators are constructed for several taxa, as with the Living Planet Index (Van Strien et al., 2016). In the case of multiple taxa one might expect FPCA to identify clusters of species from the same taxa, and also possibly to indicate whether multi-species indicators are unduly influenced by certain taxa (Buckland and Johnston, 2017), to potentially assist in the choice of taxonomic level taken when weightings are used (Burns et al., 2018). We can expect different features to arise from the analysis of data from different taxa. Importantly the techniques used here are simple to apply using freely available computer programs.

Appendix A: Principal components analysis

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

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

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

Acknowledgements

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

References

Asher, J., Warren, M., Fox, R., Harding, P., Jeffcoate, G., and Jeffcoate, S. (2001). *The Millennium Atlas of Butterflies in Britain and Ireland*. Oxford University Press.

August, T., Powney, G., Outhwaite, C., and Issac, N. (2017). *BRCindicators: Creating biodiversity indicators for species occurrence data*. R package version 1.0.

Barnett, V. (1976). The ordering of multivariate data. *Journal of Royal Statistical Society Series A*, 139:318–354.

Brereton, T. M., Botham, M. S., Middlebrook, I., Randle, Z., Noble, D., and Roy, D. B. (2017). United Kingdom Butterfly Monitoring Scheme report for 2016. Technical report, Centre for Ecology & Hydrology & Butterfly Conservation.

Brereton, T. M., Cruickshanks, K. L., Risely, K., Noble, D. G., and Roy, D. B. (2011a). Developing and launching a wider countryside butterfly survey across the United Kingdom. *Journal of Insect Conservation*, 15:279–290.

Brereton, T. M., Roy, D. B., Middlebrook, I., Botham, M. S., and Warren, M. S. (2011b). The development of butterfly indicators in the United Kingdom and assessments in 2010. *Journal of Insect Conservation*, 15(1–2):139–151.

Buckland, S. and Johnston, A. (2017). Monitoring the biodiversity of regions: Key principles and possible pitfalls. *Biological Conservation*, 214:23–34.

Buckland, S. T., Studeny, A. C., Magurran, A. E., Illian, J. B., and Newson, S. E. (2011). The geometric mean of relative abundance indices: a biodiversity measure with a difference. *Ecosphere*, 2:1–15.

Burns, F., Eaton, M., Hayhow, D., Outhwaite, C., Al Fulaij, N., August, T., Boughey, K., Brereton, T., Brown, A., Bullock, D., et al. (2018). An assessment of the state of nature in the United Kingdom: A review of findings, methods and impact. *Ecological Indicators*, 94:226–236.

Defra (2018). *UK Biodiversity Indicators 2018*. Department for Environment Food and Rural Affairs, London, UK.

Dennis, E. and Brereton, T. (2018). Scottish moths: data sources, methods and trends. Scottish Natural Heritage Research Report No. 1036.

Dennis, E. B., Morgan, B. J. T., Freeman, S. N., Brereton, T., and Roy, D. B. (2016). A generalized abundance index for seasonal invertebrates. *Biometrics*, 72(4):1305–1314.

Dennis, E. B., Morgan, B. J. T., Freeman, S. N., Ridout, M. S., Brereton, T. M., Fox, R., Powney, G. D., and Roy, D. B. (2017). Efficient occupancy model-fitting for extensive citizen-science data. *PLoS ONE*, https://doi.org/10.1371/journal.pone.0174433.

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

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

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

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

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

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

Hayhow, D. B., Burns, F., Eaton, M. A., Al Fulaij, N., August, T. A., Babey, L., Bacon, L., Bingham, C., Boswell, J., Boughey, K. L., Brereton, T., Brookman, E., Brooks, D. R., Bullock, D. J, ., Burke, O., Collis, M., Corbet, L., Cornish, N., De Massimi, S., Densham, J., Dunn, E., Elliott, S., Gent, T., Godber, J., Hamilton, S., Havery, S., Hawkins, S., Henney, J., Holmes, K., Hutchinson, N., Isaac, N. J. B., Johns, D., Macadam, C. R., Mathews, F., Nicolet, P., Noble, D. G., Outhwaite, C. L., Powney, G. D., Richardson, P., Roy, D. B., Sims, D., Smart, S., Stevenson, K., Stroud, R. A., Walker, K. J., Webb, J. R., Webb, T. J., Wynde, R., and Gregory, R. D. (2016). *State of Nature*. The State of Nature partnership.

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

Pack, P., Joliffe, I., and Morgan, B. (1988). Influential observations in principal component analysis: a case study. *Journal of Applied Statistics*, 15:39–52.

Pollard, E. and Yates, T. J. (1993). *Monitoring Butterflies for Ecology and Conservation: the British Butterfly Monitoring Scheme*. Chapman & Hall, London.

R Core Team (2017). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Ramsay, J., Hooker, G., and Graves, S. (2009). *Functional data analysis with R and MATLAB*. Springer Science & Business Media.

Ramsay, J. O., , and Silverman, B. W. (2005). *Functional Data Analysis, Second edition*. Springer, New York.

Ramsay, J. O., Wickham, H., Graves, S., and Hooker, G. (2017). *fda: Functional Data Analysis*. R package version 2.4.7.

Sievert, C., Parmer, C., Hocking, T., Chamberlain, S., Ram, K., Corvellec, M., and Despouy, P. (2017). *plotly: Create Interactive Web Graphics via 'plotly.js'*. R package version 4.7.1.

Soldaat, L. L., Pannekoeka, J., Verweija, R. J. T., van Turnhout, C. A. M., and van Strien, A. J. (2017). A Monte Carlo method to account for sampling error in multi-species indicators. *Ecological Indicators*, 81:340–347.

Sparks, T. H., Roy, D. B., and Dennis, R. L. H. (2005). The influence of temperature on migration of Lepidoptera into Britain. *Global Change Biology*, 11(3):507–514.

Tittensor, D. P., Walpole, M., Hill, S. L., Boyce, D. G., Britten, G. L., Burgess, N. D., Butchart, S. H., Leadley, P. W., Regan, E. C., Alkemade, R., et al. (2014). A mid-term analysis of progress toward international biodiversity targets. *Science*, 346(6206):241–244.

Van Strien, A. J., Gmelig Meyling, A. W., Herder, J. E., Hollander, H., Kalkman, V. J., Turnhout, S., van der Hoorn, B., van Strien-van Liempt, W. T. F. H., Poot, M. J. M., van Swaay, C. A. M., van Turnhout, C. A. M., Verweij, R. J. T., and Oerlemans, N. J. (2016).

Modest recovery of biodiversity in a western European country: the Living Planet Index for the Netherlands. *Biological Conservation*, 200:44–50.

Van Strien, A. J., Soldaat, L. L., and Gregory, R. D. (2012). Desirable mathematical properties of indicators for biodiversity change. *Ecological Indicators*, 14:202–208.