England Biodiversity Indicators 2020

This documents supports

5a. Butterflies in the wider countryside: farmland
6a. Butterflies in the wider countryside: woodland

Technical background document – Assessing change in the England Butterfly Indicators

Tom Brereton, Butterfly Conservation
David Roy, UK Centre for Ecology and Hydrology

For further information on the England Biodiversity Indicators; the Species in the wider countryside, farmland indicator (5); and the Species in the wider countryside, woodland indicator (6) visit:
https://www.gov.uk/government/statistics/england-biodiversity-indicators

Technical background document – Assessing change in the England Butterfly Indicators

Tom Brereton, Butterfly Conservation
David Roy, UK Centre for Ecology and Hydrology

August 2020 update

Data capture

The primary method for capturing UK Butterfly Monitoring Scheme (UKBMS) data, including the Wider Countryside Butterfly Survey (WCBS), is through the online capture system available at www.ukbms.org/mydata. This includes site details (e.g. location, habitat and management information), species counts through transect walks and other survey methods (e.g. timed counts and egg/larval counts).

A proportion of data are also captured via the Transect Walker software package or via spreadsheets.

Data are processed on an annual basis. The majority of data are from surveys conducted in the previous summer, but data from previous years are also often collated. All data are processed in the same way.

Standardisation and harmonisation of the UKBMS dataset

All UKBMS data are collated into a single dataset to enable analysis and reporting. As of 2019, the dataset comprises over 7 million butterfly counts. Data are standardised to conform with the UKBMS database structure, including: standardised species nomenclature, data integrity checks to ensure that all mandatory information is captured, valid date and time information and accurate geographic location information.

Data verification

The UKBMS online data capture system is built using the Indicia software tools and links to the iRecord verification system (www.brc.ac.uk/irecord) to enable review of the data by experts approved by Butterfly Conservation or other National Recording Schemes (for records for non-lepidoptera). To support verification, iRecord applies automated data checks against known species distributions (e.g. derived from the Butterflies for the New Millennium recording scheme) and timing of adult flight periods. Experts can use these checks and other information to confirm observations.

The UKBMS online data capture system (www.ukbms.org/mydata) also provides data summaries to enable UKBMS Branch Co-ordinators to review all transect data for their area and make corrections.

Further review and correction is undertaken by staff at Butterfly Conservation and the UK Centre for Ecology and Hydrology at the end of each field season, including the following checks that are discussed with Branch Co-ordinators and/or transect recorder:

- Counts outside of known distribution;
- Counts outside of the standard flight period for a species;
- Species newly recorded on a transect site;
- Species recorded on a transect site after being absent for more than 5 years; and
- Potential data input errors or misidentifications - all counts of specialist butterfly species

are closely scrutinised and summary tables for generalist species are reviewed for anomalies.

Transect visits which are undertaken outside the criteria for butterfly activity (e.g. based on weather conditions and time of day) are flagged and excluded from the main data analyses; data is retained within the database for use in other analysis.

Data analysis

a. Classification of separate generations for bivoltine species

For bivoltine species, separate generations are identified by defining the time of year where there is a gap between generations. Classification of generations is supported by visual inspection of the seasonal pattern of counts through the season at each transect site.

b. Calculation of phenology metrics

Algorithms are applied to butterfly counts throughout the season for each species at each site to estimate phenology metrics for each year (and separately for each generation of bivoltine species). The following metrics are calculated for each site, year and species (generation):

- Number of generations;
- Date of gap between generations;
- Date of first positive count (for each generation);
- Date of last positive count (for each generation);
- Date of highest positive count (for each generation);
- Count at date of highest positive count (for each generation);
- Mean date of flight period (for each generation), as defined as the weighted date of counts (Brakefield, 1987); and
- Length of flight period (for each generation), as defined as the standard deviation of counts (Brakefield, 1987).

Long-term and decadal phenology trends are calculated for each species (and generation) at each site, where sufficient data are available, using linear regression models on the timing and duration phenology metrics.

c. Calculation of abundance indices for each species, site and year

Algorithms are applied to butterfly counts throughout the season for each species at each site to estimate a total abundance for the year (and separately for each generation of bivoltine species). This can be interpreted as the area under the flight period distribution curve. The following metrics are calculated for each site, year and species (generation):

- Number of observations, including zero counts;
- Number of positive counts;
- Sum of observed counts; and

The following metrics based on the methods described in Rothery and Roy (2001):

- Index of abundance calculated by Trapazoidal rule fitted to counts;
- The smoothing parameter used for the Generalized Additive Model (GAM) fitted to counts;
- Sum of fitted counts from GAM;
- Sum of Imputed counts (observed or GAM fitted counts);

- Sum of Imputed counts (observed or Trapazoidal estimate);
- Highest seasonal count is a GAM estimate (yes/no);
- Index of abundance estimated via a GAM (GAM Index); and
- Proportion of GAM index contributed by estimated counts versus observed counts.

Long-term and decadal abundance trends are calculated for each species at each site, where sufficient data are available, using linear regression models on the site indices.

d. Estimation of zero index for species, site, year

Zero indices are not produced by the GAM models as it only deals with counts data. Where a species is not recorded at a site in a given year there is no count (no data). This may mean that the species was not seen, but could simply be because the site was not walked enough during the flight period of that species. A series of automated and manual checks were run to determine where site indices of zero are considered likely.

e. Calculation of collated indices (regional index of abundance for each year) and trends

Although alternative methods are used for specific applications, the main methods used to calculate collated indices and trends in status for individual species are as follows:

The calculation of species trends from UKBMS data is not a straightforward calculation because not all transect sites in the UKBMS dataset have been recorded each year and the number of weeks with transect counts varies markedly between sites and year. A statistical model is therefore needed to produce a regional or national index of how butterfly populations have changed each year. Since 2017, a Generalized Abundance Index (GAI) method that is designed for seasonal invertebrates has been applied to the UKBMS data to calculate annual indices of abundance and assess trends. This method combines all UKBMS data including timed counts and data from the WCBS. Briefly, the method (Dennis *et al.*, 2016) adopts a two-stage approach. Firstly, all butterfly counts in a season from both traditional UKBMS transects and WCBS are used to estimate the seasonal pattern of butterfly counts for that year, either via a GAM model or other statistical model of the flight period pattern. This stage relies heavily on the traditional UKBMS transect data with good coverage throughout the season. A second stage of the model is then applied to the full set of annual counts, accounting for where the counts occur within the flight season, to then calculate annual population indices using a statistical model to account for sites and years in a comparable way described above. In common with most butterfly and bird monitoring schemes in Europe (ter Braak *et al.*, 1994), the statistical model uses log-linear Poisson regression. The national collated index is the mean (on a log scale) of the imputed and recorded site indices for each year. Long-term and decadal trends are calculated for each species at UK and country level where sufficient data are available, using linear regression models on the collated indices.

f. Calculation of multi-species (composite) indices and trends

The England Biodiversity Indicators use multi-species (composite) indices of abundance for butterflies in different habitats e.g. farmland and woodland. Composite indices are calculated following methods developed for UK birds, derived by calculating the geometric mean index across each species assemblage.

Long time series of species abundance data such as those collected through the UKBMS and used to compile UK Butterfly Indicators cannot always be summarised adequately by linear trend lines. These long time series may show alternating periods of increase and decrease, and it can be difficult to separate patterns of genuine change from annual fluctuations. Consequently, methods that model smoothed trend lines through abundance data are becoming increasingly popular. An extension of the linear trend approach is the application of a smoothing technique that describes the pattern by assigning a trend level (= modelled abundance) to *each year* in the time series (similar to a moving average). There are several smoothing methods available such as polynomial regression, splines and Loess estimators. These models may be summarised as 'flexible trend models'. The most popular flexible trend models for the analysis of wildlife populations are GAMs and these, for example, are used to produce the UK Bird Indicators. GAMs do not however present the complete time series and do not account for serial correlation which limits their applicability to butterfly data.

Another flexible trend method from the class of structural time series analysis has been developed (Visser, 2005) and applied to European birds (Gregory *et al.*, 2007) and European butterflies (Brereton *et al.*, 2011) using TrendSpotter software (Visser, 2004). This is the approach used to describe and assess changes in the UK Butterfly Indicator updates published and updated annually from 2008 onwards. Unlike the GAM approach, the confidence interval of the trend line is not calculated by a bootstrapping method but by application of a time series analysis and the Kalman filter (Visser, 2004). This approach uses one observation per time point (e.g. year or month) and therefore the uncertainty in the estimate of yearly index values (e.g. confidence intervals around each year index) is modelled indirectly in the annual fluctuations. The main advantage of the TrendSpotter analysis however is the calculation of confidence intervals for the *differences* between the trend level of the last year and each of the preceding years, taking into account serial correlation which is unique for flexible trend methods. This allows short-term trends to be usefully assessed.

A statistical test is performed in TrendSpotter to compare the difference in the index in the latest year versus other years in the series. Yearly change rates and confidence intervals produced in the TrendSpotter output are used to classify the trends per year (see Appendix). The trend classification applied to the composite index (see Soldaat *et al*., 2007) is given in Table 1. This classification is not the same as that used for the individual species trends presented in the dataset (increased, decreased and no change).

Table 1: Classification of composite trends on the basis of the 95% confidence intervals of the yearly change rate in TrendSpotter smoothed indices. CL = confidence limit; CI = confidence interval (see Soldaat *et al*., 2007 for explanation).

Trend class	Criteria	Description
Strong Increase	Lower CL > 1.05	*> 5% increase / year (≈ doubling in 15 years)*
Moderate Increase	1.00 < lower CL ≤ 1.05	*Increase, but unsure whether > 5% / year*
Stable	Confidence interval contains 1.00 AND lower CL ≥ 0.95 AND upper CL ≤ 1.05	*Population changes less than 5% / year*
Moderate Decrease	0.95 ≤ upper CL < 1.00	*Decrease, but unsure whether > 5% / year*
Steep Decrease	Upper CL < 0.95	*> 5% decrease / year (≈ halving in 15 years)*
Uncertain	Confidence Interval contains 1.00 AND (lower CL < 0.95 OR upper CL > 1.05)	*CI too large for reliable classification*

In summary, structural time series models are essentially regression models in which the explanatory variables are functions of time and the parameters are time-varying. The Kalman filter is an efficient recursive filter that estimates the state of a dynamic system from a series of incomplete and noisy measurements. For mathematical details about structural time-series analysis and the Kalman filter please refer to Harvey (1989).

TrendSpotter is currently considered the best-available technique in the assessment of Butterfly Indicators. Regular reviews of methods to assess changes in butterfly indicators are needed however; techniques to model trends are an active area of statistical development.

g. Methodological changes to UK and England butterfly composite indicators in 2020

Improvements were made to the analytical techniques in 2020 to better account for the colonisation of sites. The change has been to add pre-colonisation zero abundance counts for species at sites they have colonised, where the site was monitoring prior to colonisation. These improvements have greatest effect where sites have been monitored for a number of years prior to the arrival of species and/or where species are notably expanding their range. In general, the effect of these changes has been most notable for expanding species whereby there has been a slight reduction in their population indices for the earlier years, relative to the latter years. An example of a species where the effect of these improvements was noticeable is Silver-washed Fritillary which has spread considerably during recent decades.

This analysis improvement has coincided with relatively favourable recent years for butterflies. The combination of the relative reductions in the indices of earlier years for colonising species with the relatively high indices in recent years have resulted in the current indicator assessment differing from previous assessments to a greater extent than in previous updates. The difference is most noticeable for the UK & England Farmland indicators. These indicators are over a relatively short time period and include relatively few species and are therefore sensitive to changes in estimated population indices for component species.

The previous Farmland indicator assessments for UK & England were both categorised as “moderate decline” showing steady long-term declines, albeit with a noticeable levelling off in the latter part of the series in the England Farmland indicator. The latest UK & England Farmland indicators are categorised as “stable”. The current farmland indicators still show steady declines but now this is limited to the first half of the series, with the latter half showing stabilisation or even slight recovery in the case of the England Farmland indicator. Although the changes in indicator have been emphasised by the methodological improvements they are not dramatic alterations as the indicators were already showing signs of stabilisation and the addition of another relatively good year would have increased this further.

These indicators are updated and published annually and can be viewed at:
https://www.gov.uk/government/statistics/england-biodiversity-indicators

References

Brakefield, P. M., (1987). Geographical variability in, and temperature effects on, the phenology of Maniola jurtina and Pyronia tithonus (Lepidoptera, Satyrinae) in England and Wales. Ecological entomology, 12(2), pp.139-148.

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

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

Gregory, R. D., Vorisek, P., van Strien, A. J., Gmelig Meyling, A. W., Jiguet, F., Fornasari, L., Jiri, R., Chylarecki, P. and Burfield, I. J., (2007). Population trends of widespread woodland birds in Europe. Ibis: 149 (Suppl. 2), 78–97.

Harvey, A. C., (1989). Forecasting structural time series models and the Kalman filter. Cambridge University Press, London.

Rothery, P. and Roy, D. B., (2001). Application of generalized additive models to butterfly transect count data. Journal of Applied Statistics, 28(7), pp.897-909.

Soldaat, L. L., Visser, P., van Roomen, M. and van Strien, A. (2007). Smoothing and trend detection in waterbird monitoring data using structural time-series analysis and the Kalman filter. Journal of Ornithology. Vol. 148 suppl. 2. Dec. 2007.

ter Braak, C. J. F., van Strien, A. J., Meijer, R., and Verstrael, T. J., (1994). Analysis of monitoring data with many missing values: which method? In Bird Numbers 1992: Distribution, monitoring and ecological aspects. (eds W. Hagemeijer & T. Verstrael), pp. 663-673. SOVON, Beek-Ubbergen, Netherlands.

Visser, H., (2004). Estimation and detection of flexible trends. Atm Environment 38: 4135-4145.

Visser, H., (2005). The significance of climate change in the Netherlands. An analysis of historical and future trends (1901-2020). MNP report 550002007.

Appendix. Trend results for England butterflies of the wider countryside indicators, 1990 to 2019.

Composite Indicator	Year	Index	Smoothed Index (SI)	SI Lower CL.	SI Upper CL.	Yearly Change Rate (YCR)	YCR Lower CL.	YCR Upper CL.	Trend Classification
Farmland	1990	100	109	94.1	124.2	1.00	0.99	1.00	STABLE
Farmland	1991	120	108	95.2	120.9	1.00	0.99	1.00	STABLE
Farmland	1992	125	107	95.8	118.0	1.00	0.99	1.00	STABLE
Farmland	1993	75	106	95.8	115.6	1.00	0.99	1.00	STABLE
Farmland	1994	89	105	95.4	113.7	1.00	0.99	1.00	STABLE
Farmland	1995	105	103	94.8	112.2	1.00	0.99	1.00	STABLE
Farmland	1996	116	102	93.8	110.8	1.00	0.99	1.01	STABLE
Farmland	1997	132	101	92.6	109.4	1.00	0.99	1.01	STABLE
Farmland	1998	100	100	91.2	107.9	1.00	0.99	1.01	STABLE
Farmland	1999	89	98	89.7	106.4	1.00	0.99	1.01	STABLE
Farmland	2000	95	96	88.1	104.9	1.00	0.99	1.01	STABLE
Farmland	2001	87	95	86.6	103.3	1.00	0.99	1.01	STABLE
Farmland	2002	86	93	85.1	101.8	1.00	0.99	1.01	STABLE
Farmland	2003	106	92	83.7	100.4	1.00	0.99	1.01	STABLE
Farmland	2004	104	91	82.3	99.1	1.00	0.99	1.02	STABLE
Farmland	2005	91	89	81.0	97.8	1.01	0.99	1.02	STABLE
Farmland	2006	92	88	79.9	96.7	1.01	0.99	1.02	STABLE
Farmland	2007	68	87	79.0	95.8	1.01	0.99	1.02	STABLE
Farmland	2008	65	87	78.4	95.2	1.01	0.99	1.03	STABLE
Farmland	2009	93	87	78.1	94.9	1.01	0.99	1.03	STABLE
Farmland	2010	95	87	78.2	94.9	1.01	0.99	1.03	STABLE
Farmland	2011	87	87	78.5	95.2	1.01	0.99	1.03	STABLE
Farmland	2012	53	87	79.1	95.9	1.02	0.99	1.04	STABLE
Farmland	2013	104	88	79.9	96.9	1.02	0.99	1.04	STABLE
Farmland	2014	101	90	80.9	98.3	1.02	0.99	1.04	STABLE
Farmland	2015	94	91	81.7	100.1	1.02	0.99	1.05	STABLE
Farmland	2016	70	92	82.5	102.3	1.02	0.99	1.05	STABLE
Farmland	2017	89	94	82.9	105.2	1.02	0.98	1.05	UNCERTAIN
Farmland	2018	111	96	83.0	108.7	1.02	0.98	1.06	UNCERTAIN
Farmland	2019	110	98	82.6	112.8	0.00	0.00	0.00	
Woodland	1990	100	94	81.8	106.7	0.98	0.97	0.99	MODERATE DECLINE
Woodland	1991	90	91	80.2	101.2	0.98	0.97	0.99	MODERATE DECLINE
Woodland	1992	110	87	78.1	96.3	0.98	0.97	0.99	MODERATE DECLINE
Woodland	1993	56	84	75.7	91.9	0.98	0.97	0.99	MODERATE DECLINE
Woodland	1994	65	80	73.0	87.9	0.98	0.97	0.99	MODERATE DECLINE
Woodland	1995	87	77	70.1	84.5	0.98	0.97	0.99	MODERATE DECLINE
Woodland	1996	86	74	67.2	81.2	0.99	0.97	1.00	MODERATE DECLINE
Woodland	1997	95	71	64.2	78.2	0.99	0.97	1.00	MODERATE DECLINE

Woodland	1998	59	68	61.3	75.2	0.99	0.97	1.00	MODERATE DECLINE
Woodland	1999	51	66	58.6	72.5	0.99	0.97	1.00	STABLE
Woodland	2000	51	63	56.1	70.1	0.99	0.97	1.00	STABLE
Woodland	2001	49	61	54.0	67.9	0.99	0.97	1.01	STABLE
Woodland	2002	56	59	52.1	66.0	0.99	0.97	1.01	STABLE
Woodland	2003	71	57	50.4	64.3	0.99	0.97	1.01	STABLE
Woodland	2004	74	56	48.9	62.8	1.00	0.97	1.01	STABLE
Woodland	2005	56	54	47.4	61.3	1.00	0.97	1.02	STABLE
Woodland	2006	57	53	46.1	60.0	1.00	0.97	1.02	STABLE
Woodland	2007	37	52	44.9	58.9	1.00	0.97	1.02	STABLE
Woodland	2008	36	51	44.0	58.0	1.00	0.97	1.03	STABLE
Woodland	2009	56	50	43.3	57.3	1.00	0.97	1.03	STABLE
Woodland	2010	63	50	42.8	56.8	1.01	0.97	1.03	STABLE
Woodland	2011	56	49	42.5	56.4	1.01	0.97	1.04	STABLE
Woodland	2012	29	49	42.3	56.3	1.01	0.97	1.04	STABLE
Woodland	2013	57	49	42.3	56.3	1.01	0.97	1.05	STABLE
Woodland	2014	51	50	42.4	56.7	1.01	0.97	1.05	UNCERTAIN
Woodland	2015	49	50	42.4	57.4	1.01	0.96	1.06	UNCERTAIN
Woodland	2016	35	50	42.3	58.5	1.01	0.96	1.06	UNCERTAIN
Woodland	2017	47	51	42.0	60.2	1.02	0.96	1.07	UNCERTAIN
Woodland	2018	67	52	41.3	62.5	1.02	0.96	1.07	UNCERTAIN
Woodland	2019	58	53	40.2	65.2	0.00	0.00	0.00	