

England Biodiversity Indicators technical background document

Assessing change in the England Butterfly Indicators

Updated: November 2023

Contents

England Biodiversity Indicators technical background document	1
Assessing change in the England Butterfly Indicators.....	1
Background information	3
Data collection	3
Data capture	5
Standardisation and harmonisation of the UKBMS data set.....	5
Data verification	5
Data analysis	6
1. Calculation of a total abundance estimate for each species, at each site within each year, to account for missing data	6
2. Combining separate site level abundance into a single time series for each species	7
3. Calculation of multi-species (composite) indices and trends	7
4. Methodological changes to the butterfly composite indicators in 2020	9
References	10

Background information

This is the technical background document for:

- Indicator 5a - Butterflies in the wider countryside: farmland
- Indicator 6a - Butterflies in the wider countryside: woodland

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 the [England Biodiversity indicators](#) page on GOV.UK.

This document was written by:

- Ian Middlebrook and Emily Dennis, Butterfly Conservation
- David Roy, UK Centre for Ecology & Hydrology

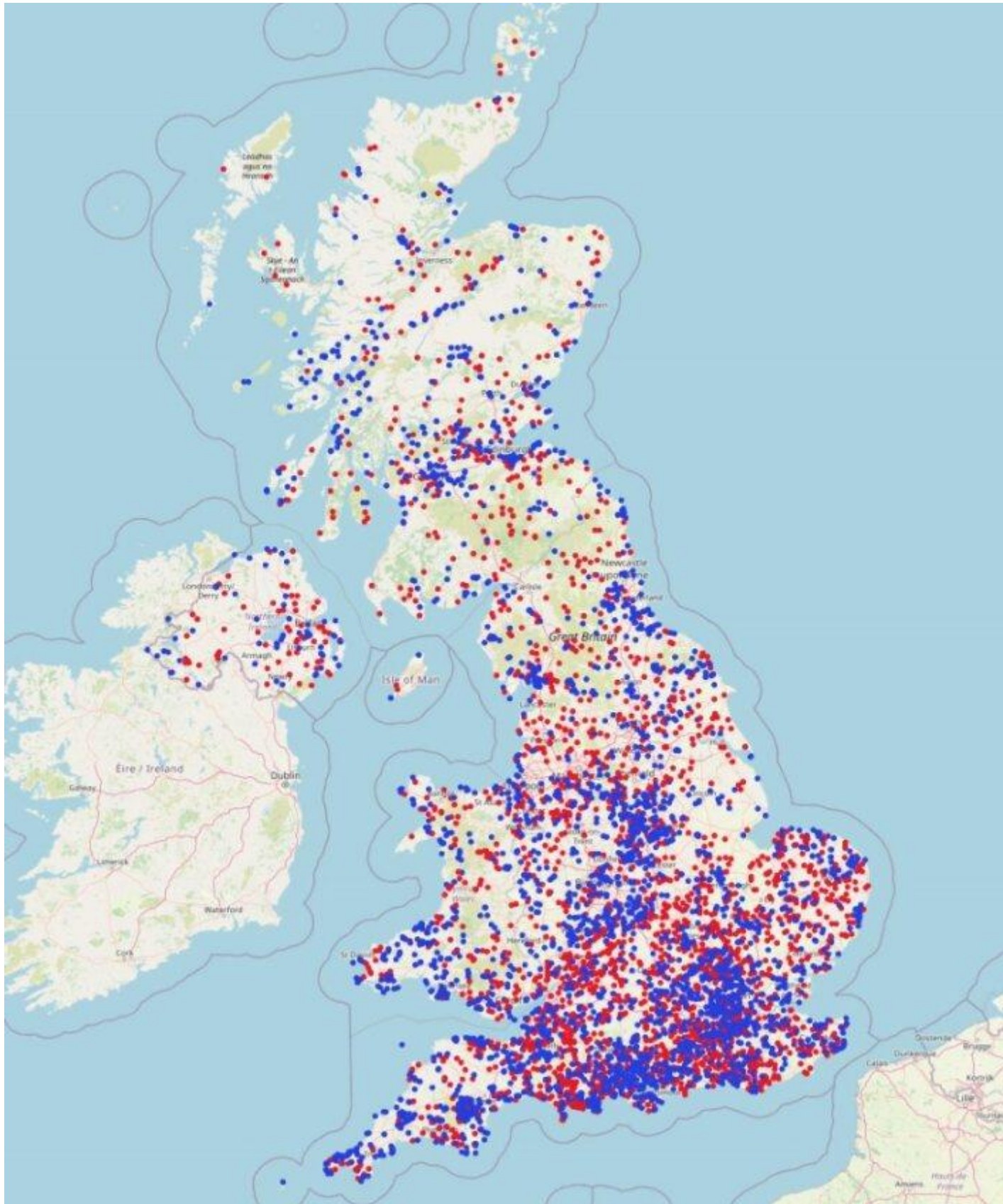
Data collection

The core indicator is comprised of 2 multi-species indices compiled by Butterfly Conservation and the UK Centre for Ecology and Hydrology (UKCEH) from data collated through the UK Butterfly Monitoring Scheme (UKBMS) including the Wider Countryside Butterfly Survey (WCBS). Through the UKBMS, data are collected from around 3,000 sample locations distributed across the UK by around 3,000 skilled volunteers each year. The scheme comprises three survey components:

- traditional butterfly transects (Pollard Walks);
- Wider Countryside Butterfly Survey;
- targeted surveys in which non-transect methods are used to survey specific priority species.

The sampling locations are largely chosen by the recorder, and so are not evenly distributed across the UK. Sites are recorded repeatedly throughout any given year. Volunteer recorders are well supported and receive training and guidance on collecting and submitting data by Butterfly Conservation.

Figure 1: Map showing locations of UKBMS transects that produced a site index (red dots) and WCBS squares (blue dots) that were monitored in 2022



Source: UKBMS.

Data capture

The primary method for capturing UK Butterfly Monitoring Scheme (UKBMS) data, including the Wider Countryside Butterfly Survey (WCBS), is through the [UKBMS online data capture system](#). This includes site details (for example, location, habitat and management information), species counts through transect walks and other survey methods (for example, 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 data set

All UKBMS data are collated into a single data set to enable analysis and reporting. As of 2021, the data set comprises over 8.5 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](#) 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 (for example, 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 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 recorders:

- 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 (for example, based on weather conditions and time of day) are flagged and excluded from the main data analyses; data are retained within the database for use in other analysis.

Data analysis

The calculation of species trends from UKBMS data is not a straightforward calculation because not all transect sites in the UKBMS data set have been recorded each year and the number of weeks with transect counts varies markedly between sites and year. The analytical steps taken to produce the estimates of butterfly populations in England are as follows:

1. Calculation of a total abundance estimate for each species, at each site within each year, to account for missing data
2. Combining separate site level abundance into a single time series for each species
3. Calculation of multi-species (composite) indices and trends

1. Calculation of a total abundance estimate for each species, at each site within each year, to account for missing data

Not all transect sites in the UKBMS data set 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. A Generalized Abundance Index (GAI) method is used which is designed for seasonal invertebrates and is applied to the UKBMS data to calculate annual indices of abundance and assess trends. This method can account for missing patchy data and so 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 (that is, flight curve) of butterfly counts for each species and year, using generalised additive models (GAMs) applied to weekly summarised data. This stage relies heavily on the traditional UKBMS transect data with good coverage throughout the season. For a given species and year, a site index, which represents an estimate of the expected total number of butterflies had a site been surveyed fully that season, is calculated

by scaling the total observed count by the proportion of the species flight curve that was surveyed.

2. Combining separate site level abundance into a single time series for each species

The second stage of the approach is then applied to the corrected total annual counts, accounting for where the counts occur within the flight season, to then calculate annual population indices (or time series) for each species using a statistical model to account for sites and years. Data from non-transect surveys are also incorporated at this modelling stage. In common with most butterfly and bird monitoring schemes in Europe (ter Braak *et al.*, 1994), the statistical model uses log-linear Poisson regression to account for the fact that not all sites are sampled in every year. 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, applying linear regression models to the collated indices.

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

The England Biodiversity Indicators use multi-species (composite) indices of abundance for butterflies in different habitats, for example, farmland and woodland. Composite indices are 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 England 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 (that is, a 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.

TrendSpotter software (Visser, 2004) is used to identify periods of significant change in butterfly abundance. Under this approach, confidence intervals around the trend line is calculated time series analysis and the Kalman filter (Visser, 2004). This approach uses one

observation per time point (for example, year or month) and therefore the uncertainty in the estimate of yearly index values (for example, 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.

Periods of significant change are identified by comparing the difference in the index for the first and last year of any given time period. Thresholds for determining change are given in Table 1 (see Soldaat *et al.*, 2007). This classification is not the same as that used for the individual species trends presented in the data set (increased, decreased and no change).

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

Trend class	Criteria	Description
Strong increase	Lower confidence limit greater than 1.05	Increase greater than 5% per year (approximately equal to doubling in 15 years)
Moderate increase	Lower confidence limit greater than 1.00 and less than or equal to 1.05	Increase, but unsure whether greater than 5% per year
Stable	Confidence interval contains 1.00 AND lower confidence limit greater than or equal to 0.95 AND upper confidence limit less than or equal to 1.05	Population changes less than 5% per year
Moderate decrease	Upper confidence limit greater than or equal to 0.95 and less than 1.00	Decrease, but unsure whether greater than 5% per year
Steep decrease	Upper confidence limit less than 0.95	Decrease greater than 5% per year (approximately equal to halving in 15 years)
Uncertain	Confidence interval contains 1.00 AND lower confidence limit less than 0.95 OR upper confidence limit greater than 1.05	Confidence interval 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.

4. Methodological changes to the butterfly composite indicators in 2020

Improvements were made to the analytical techniques in 2020 to better account for the colonisation of new sites (UKBMS transects and WCBS squares). The change was to add pre-colonisation zero abundance counts for species at sites they colonised, where the site was being monitored prior to colonisation. These improvements had the greatest effect where sites had been monitored for a number of years prior to the arrival of species and/or where species were notably expanding their range. In general, the effect of these changes was most notable for expanding species whereby there was 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 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 has resulted in the indicator assessments presented from 2020 onwards differing from those presented prior to 2020 to a greater extent than would have otherwise been expected. The difference is most noticeable for the Farmland indicator. This indicator is over a relatively short time period (since 1990) and includes relatively few species and is therefore sensitive to changes in estimated population indices for component species.

Prior to the methodological changes in 2020, the Farmland indicator assessment for England was categorised as 'moderate decline' showing steady long-term declines, albeit with a noticeable levelling off in the latter part of the series. Since 2020, the indicator has been categorised as 'stable'. The current farmland indicator still shows a steady decline but now this is limited to the first half of the series, with the latter half showing a slight recovery. Although the changes in indicator have been emphasised by the methodological improvements, they are not dramatic alterations as the indicator was already showing signs of stabilisation and the addition of another relatively good year in 2020 would have increased this further.

These indicators are updated and published annually and can be viewed at: [Butterflies in England: species of the wider countryside on farmland and in woodland - GOV.UK \(www.gov.uk\)](https://www.gov.uk/government/collections/butterflies-in-england-species-of-the-wider-countryside-on-farmland-and-in-woodland).

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. & 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. & 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. & 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. & 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. & 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., & 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. Hagemeyer & 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 55000200.