Exploring options to test a ‘halt the decline’ nature target using a multi-species abundance indicator

RSPB Research Report 83

June 2025

Fiona Burns, Tom Bradfer-Lawrence, Richard Gregory

RSPB Centre for Conservation Science, The Lodge, Sandy, Bedfordshire, SG19 2DL, UK

Contents

Recommended citation: Burns, F., Bradfer-Lawrence, T. & Gregory, R.D. (2025) Exploring options to test a 'halt the decline' nature target using a multi-species abundance indicator. RSPB Research Report 83. RSPB Centre for Conservation Science, RSPB, The Lodge, Sandy, Bedfordshire, SG19 2DL. ISBN number: 978-1-905601-80-6.

Summary

Introduction and methods

This report explores options to use a multi-species abundance indicator to test a 'halt the decline' biodiversity target. This relates to the Westminster Environment Act 2021, which sets a legally binding target to halt the decline in species abundance in England by 2030. The England Species Abundance Indicator (ESAI) will be used to test whether the target has been met using a methodology set out in secondary legislation. Here, we consider a suite of assessment methods that could be considered complementary or alternative to the legislative framework.

Using prior knowledge and a rapid literature review, we collated a suite of potential assessment options and grouped them to explore similarities and differences. They are broadly based on assessing the rate of change in an abundance indicator or species time-series over one, or more time periods. We ranked each option using a set of descriptive 'properties', including timeliness, ease of communication, error propagation and risk of false positives or negatives. False positives here, represent the likelihood that the assessment will conclude the target has been met, when the true change is continued decline, and false negatives the opposite. We then refined our review based on a series of online workshops with analytical ecologists. Finally, we conducted a preliminary exploration by simulation of the implications of the secondary target legislation, which states the 2030 species abundance target will have been met if the ESAI estimate in 2030 is at or above the estimate in 2029. The simulation estimated the likelihood of the test correctly concluding whether species decline had halted under a range of different hypothetical scenarios of change in species abundance.

Option Exploration

Options for testing the species abundance target have two components. The first describes the 'simplification' model used to summarise how an indicator is changing over time, from no simplification, meaning the test is conducted on the unsmoothed index values, to a non-linear or linear model with varying number of knots or breakpoints. The second describes the type of 'question' or hypothesis used on the model output. We collated 16 relevant examples of time-series analysis from the literature. These represent a variety of combinations of 'simplification' model and 'question' posed. Six assessment options ask whether a single rate of change is greater than or equal to zero. These include the assessment method described in the secondary target legislation and related methods focussed on change over the final time step, which additionally account for uncertainty around the rate of change. Others compare a linear model with a continuous year term, to an intercept only or null model, and consider a longer assessment period. Three options ask whether a single rate of change is both greater than a pre-set declining threshold and lower than an increasing one and could also be implemented solely considering the declining threshold. Many of remaining options compare the rate of change over the recent past with one or more previous periods, several being variations on breakpoint or broken-stick linear regression.

When comparing the properties of the assessment options, the duration of the assessment period varies among assessment options, with those assessing the rate of change between the final and penultimate year ranking highest. However, these options all use a non-linear simplification model so additional years of data contribute to the estimates for the final two years. The options that only consider the rate of change in the final time step require no changes, or only the addition of a statistical test, to be compatible with the secondary target legislation. We ranked the likely rate of false positives and negatives where the true rate of change was zero, the minimum case where the target would be met. Given that target has been met in this situation, there would be no false positives for the assessment methods asking whether a single rate of change is greater than or equal to zero. The level of false negatives would vary, amongst other things, due to the different statistical thresholds chosen. Choosing the standard statistical threshold of 95% would lead to a high level of false negatives, meaning falsely concluding that decline in species abundance has not halted when it has. The rate of false negatives would be lower for the secondary target legislation assessment method, where considering the observed point estimate (assuming normality) can be thought of as using a threshold of as 50%. False negatives are lowest where the observed estimate of the rate of change is considered to represent zero change if it is above a declining threshold and below an increasing one.

Review and refine

Reviewer feedback supported further consideration of the assessment method for the 2030 target. Reviewers considered that comparing the point values of the ESAI in 2030 and 2029 estimates was unlikely to be sufficiently rigorous and provide a definitive assessment. The most common suggestion was to incorporate uncertainty by using a probabilistic approach, based on the likelihood that the estimated recent rate of change was at or above zero change. There was also support to complement the test by repeating it in subsequent years, repeating it for each taxonomic group within the indicator or by additionally comparing recent change to a previous time-period. Feedback also highlighted the elevated uncertainty and volatility of the final estimate in a smoothed time-series. The secondary target legislation mandates testing the species abundance target as soon as 2030 data become available, in 2032. This means that in 2032, the indicator estimate for 2030 and the estimated change between 2029 and 2030 will be less certain than values for earlier years and subject to greater change once data for subsequent years become available. Therefore, the 2030 target could be judged met in 2032, but with refinement in later years it could become evident that the target had actually been missed or vice versa.

Exploration of the secondary target legislation

We used simple simulations to explore our ability to detect change under a range of 'true' trends and uncertainties. These indicated that if the true average biological change is zero (so the distribution of potential observed change is symmetrical around zero change), then using the assessment method set out in the secondary target legislation, we have a 50% chance of correctly concluding that we have met the target, and a 50% chance of incorrectly concluding that we have failed to meet the target (50% rate of false negatives). As the true trend moves away from zero then the likelihood of a correct test outcome increases i.e., the rate of false positives and false negatives declines. However, away from zero change, the error rate increases with uncertainty in the trend estimate. When uncertainty around the true change is low, we can reliably differentiate between 1% per annum decline and zero change. When uncertainty around the true change is high, we may reach the incorrect conclusion in over a third of occasions when the true rate of change is ±0.5% per annum and the incorrect test outcome may be observed (6% likelihood) even when the rate of 'true' change is substantial, at ±2% per annum.

Recommendations

This report explored assessment options that could be used to test to the 2030 species abundance target and their properties. It does not recommend one option over another as the properties will vary depending on a range of implementation choices. Here we outline potential work to build the supporting evidence needed to judge whether the current secondary target legislation is sufficient, whether supplementary information is required, or whether changes are recommended to the secondary target legislation itself and alternative or additional tests developed. We recommend first repeating the current simulation using observed estimates of trend uncertainty in the ESAI and extending the simulation to consider a range of statistical threshold and test metrics. This could be followed by more complex simulations considering full time-series data rather than estimates of a single year change. These would explore how a set of candidate tests behave when using the observed ESAI dataset or time-series data simulated with similar levels of uncertainty. Finally, we recommend exploring the sensitivity of the test outcomes over time and to the inclusion of particular species or species groups and how applicable the findings from the 2030 simulations are to testing the 2042 target. Given that some decisions related to meeting these nature targets are to a degree value-based rather than purely analytical, we additionally recommend a workshop bringing together researchers, policy and decision makers.

1. Introduction

Given substantial ecosystem degradation and biodiversity loss [1, 2] , nature targets are being set at a variety of geographic scales to engender progress towards recovery [3, 4]. A range of biodiversity indicators are available or are being developed to help assess progress towards these targets. These include species abundance indicators, which chart the average change in the relative abundance of a range of species over time [5, 6].

One category of nature targets aims to 'halt the loss' of nature, and therefore arrest declines in biodiversity. It is difficult to assess whether 'halt the decline' targets have been met, because these can be satisfied with no trend - as well as an improvement - in the species abundance indicator. Determining whether a declining trend has switched to no trend is not straightforward using standard statistical approaches. Such approaches rely on determining whether the null hypothesis can be rejected. In the case of 'halt the decline' targets, the null hypothesis is that there is no significant negative change in the indicator trend. Failure to reject this null hypothesis does not mean the null hypothesis can be accepted. A non-significant result may therefore disguise situations where the target has been met. A further consideration is how to differentiate between no trend and an uncertain trend (due either to scant data or highly variable trends in individual species).

This report explores options to use a multi-species abundance indicator to test a 'halt the decline' target. This relates specifically to the Environment Act 2021, which sets a legally binding target to halt the decline in species abundance in England by 2030 and will use the England Species Abundance Indicator (ESAI) to test whether the target has been met. A methodology for assessing the target is set out in secondary legislation linked to the Environment Act 2021. In this work we consider a wide suite of assessment methods, that could be considered complementary, additional or alternative to the current legislative framework.

1.1 The Legislation

The Environment Act 2021[3] compels the "Secretary of State to set a target in respect of a matter relating to the abundance of species, the date of which must be 31 December 2030". The secretary of state must be satisfied that meeting the target would halt the decline in the abundance of species. This was the first legally binding target of its kind globally and so there is little precedence on how it should be assessed.

The Environment Act was followed by more detailed secondary legislation, in the form of a statutory instrument (SI)[7]. This sets out the target and when and how it will be assessed. It also sets out in Schedule 2 the species that must contribute to the assessment as part of the ESAI. The SI states:

> *"1) The 2030 species abundance target is to be measured by calculating the difference between the overall relative species abundance index for the years 2029 and 2030 in order to establish whether the overall relative species abundance index for the year 2030 is the same as, or higher than, the overall relative species abundance index for the year 2029.*
>
> *2) The overall relative species abundance index for a year is derived from the calculation of the geometric mean of the relative species abundance indices for every species listed in Schedule 2 for that year, which is smoothed to reduce the impact of between-year fluctuations in data collected over time.*
>
> *3) The reporting date for the target in regulation 11 is 15th April 2032 (when data running to 2030 are expected to be available)."*

Throughout this report we use 'the 2030 species abundance target' to represent the general target set out in the primary legislation without reference to how and when it is intended to be tested, and the ‘secondary target legislation’, when we are referring to the details of the assessment method and timing set out in the Environment targets statutory instrument.

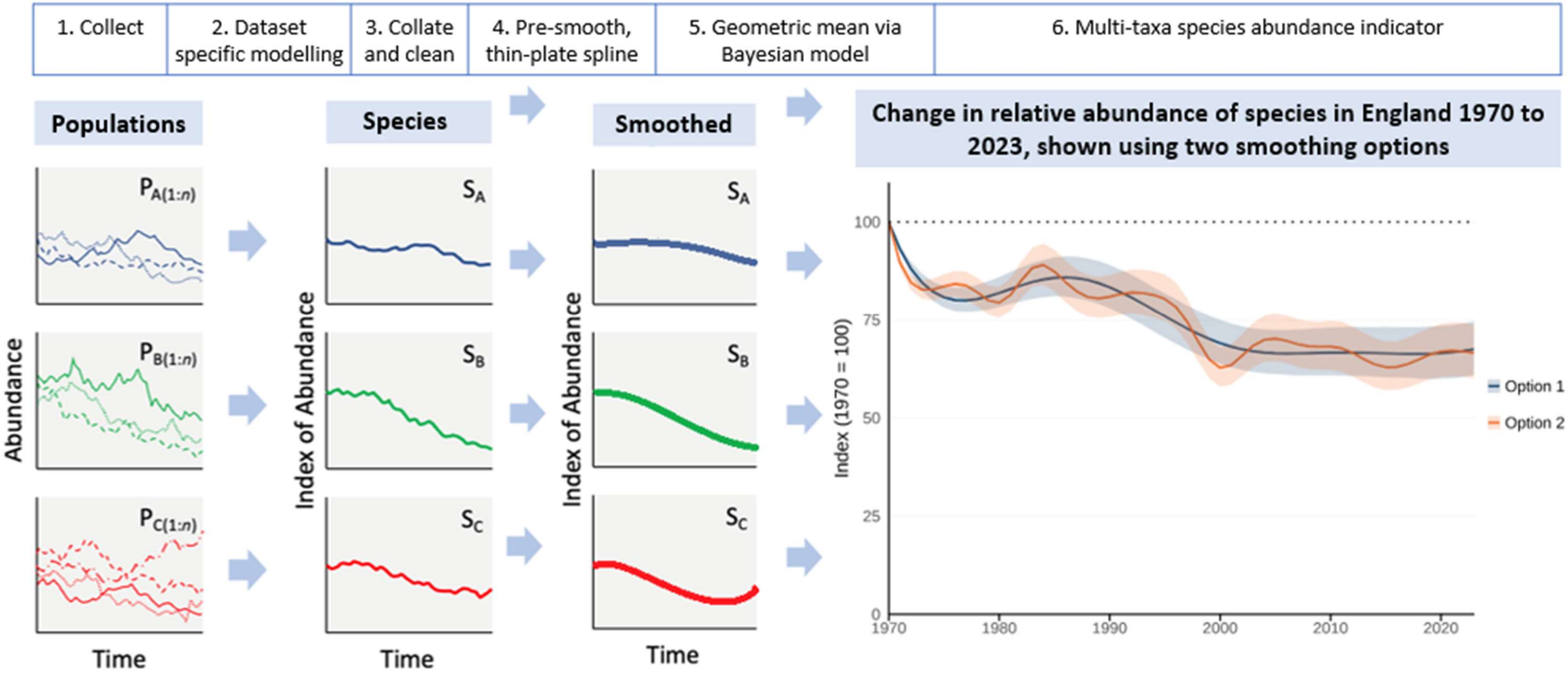
1.2 The England Species Abundance Indicator (ESAI)

The ESAI was first published by Defra in 2024[8]. It is an official ‘statistic in development’ and as such the methods continue to be revised. Like other similar species abundance indicators, the metric illustrates the average proportional change in relative abundance across a range of species over time, often termed Multi-Species Indicators. The overall trend therefore shows the balance across all the species included in the indicator. The individual constituent species may be increasing or decreasing in abundance. All species are weighted equally in the indicator regardless of their total population size or any other characteristic (such as rarity or importance for a particular ecosystem service).

The key steps taken to produce the ESAI are as follows (see Figure 1) (amended from Defra, 2025[8]):

1. Collection of observations in the field. Data are derived from national monitoring schemes, where each scheme follows standardised protocols to collect abundance data for a particular group of species. In the case of vascular plants, abundance is measured on a categorical cover scale, rather than the number of individuals.
2. Calculation of an England-level time-series of abundance for each species in each year. For most datasets, this step is performed by the individual monitoring schemes. Time-series are of variable length and the analytical methods by each scheme used are tailored to the properties of each dataset.
3. Data collation and consolidation of individual species trends.
4. Pre-smoothing individual species trends to remove short-term fluctuations and reveal long term trends.
5. Calculation of geometric mean smooth multispecies (composite) indicator in a Bayesian framework. This estimates multi-species growth rates, describing the change between each pair of years (also termed the first derivative), each of which are assumed to conform to a lognormal distribution [9]. Two smoothing options are currently estimated: a ten-year timescale and a three-year timescale.

There is a range of types of error that influence how certain we can be in how individual species and the overall indicator are changing. Propagating uncertainty through from data collection to a multi-species indicator is challenging and the subject of current research [10]. Uncertainty in the indicator will be influenced by constituent species’ life history, data collection method, population metric and sample size, amongst a range of other factors. The level of uncertainty is one important factor that will influence the outcome of any statistical test of whether, or not, the 2030 species abundance target has been met.



Group:	Moths	Birds	Butterflies	Mammals	Fish	Freshwater invertebrates	Bumblebees	Flowering plants	TOTAL
Species:	444	168	55	17	37	235	11	209	1166
Data from:		1970s		1990s	2000s		2010s		

Figure 1: Steps involved in constructing the England Species Abundance Indicator (extended from figure 1 in Bane et al 2021[11]), with Figure 1 from the latest version of the ESAI in the right-most panel [8] available at: [Indicators of species abundance in England - GOV.UK](). The ESAI shows the two options for the smoothed trend (solid line) with their 95% credible intervals (shaded area). Index values represent change from the baseline value in 1970, the credible interval widens as the index gets further from the 1970 value and confidence in the estimate of change relative to the baseline falls. The table shows the taxonomic and temporal coverage of the indicator, all icons are reproduced under a CC0 1.0 Universal Public Domain Dedication license and were downloaded from phylopic.org.

2. Methods

This report explores and contrasts assessment methods that could be employed, either in isolation or in concert, to test the 2030 species abundance target using a multispecies abundance indicator. Three strands of work informed this report: a rapid literature search, workshop discussions, and simple simulations.

2.1. Option Exploration

2.1.1 Option collation and framing

We used a combination of our existing knowledge, a rapid review of the literature and recent work in this area to collate an initial suite of options. The options considered are broadly based on assessing the rate of change in the indicator over one or more time periods. For instance, testing the likelihood that the rate of change over a single recent time period is greater than or equal to zero, or whether the rate of change over a recent period differs from the rate in one or more previous periods. We sought examples of assessments that had been conducted on existing multispecies indicators or on time-series of population trends more broadly. We did not assume that the multispecies indicator would have been constructed using the same indicator model as the ESAI. We considered frequentist, Bayesian and machine learning approaches and both linear and non-linear models. Once we had a suite of assessment examples, we summarised and grouped them in order to start to explore similarities and differences.

2.1.2 Properties of assessment options

We identified a set of putative 'properties', for example, timeliness or ease of communication, against which we could score the different assessment options. Each property was scored on a three-point scale. This allowed us to begin to identify which aspects of each option facilitated its use in testing the 2030 species abundance target and to highlight suitable attributes from across the range of examples that could potentially be combined into a future assessment. These scores are provisional and do not reflect the overall strengths of each assessment option, only their applicability to testing the 2030 target.

As well as the specific statistical test undertaken, there are a number of things that could influence the outcomes of any assessment method, such as:

Length of time series: Key aspects are the number of years over which the assessment is made and the number of years of data contributing to the assessment. These would be the same where a linear model was used to estimate an average trend across a range of years. However, they may differ if a non-linear model (like a running average) was used to smooth the indicator and the trend was assessed over a subset of years. In such cases, the estimate for each year would be influenced by the estimates for the surrounding years. Where non-linear models are used, the **level of smoothing** chosen would dictate which previous data would influence the values within the assessment period. A final consideration, where smoothing is used, is the availability of **data subsequent to the assessment period.** This is because within smoothed time-series, the final estimate is less certain than those within the time-series and is subject to greater change once additional years of data are added.

Underlying biological change: The direction and magnitude of change both at the point of assessment and in the preceding years will have a bearing on the test outcome and interact with the level of smoothing.

Accuracy: Despite the structured nature of each underpinning monitoring scheme, the data are subject to a range of biases, such as uneven uptake of survey squares or observer turnover, which are controlled for to varying extents in the modelling methods used to generate the species time-series. These biases will influence the multispecies indicator and any assessment of whether, or not, a declining trend has halted.

Level of uncertainty/precision: As described in section 1.2, the level of uncertainty in the estimate(s) of change will influence the assessment outcome and could influence which assessment method was most suitable.

Level of risk accepted: For each assessment method there will be some risk of incorrectly concluding the target has been met when it has not (false positives – concluding species abundance decline has halted when it has not), incorrectly concluding the target has not been met when it has (false negatives – concluding species abundance continues to decline when declines have halted), or both. The rate of each type of error can be estimated under different potential situations for different assessment methods, however, it is a societal decision as to what balance of these risks is acceptable.

Many of the aspects listed above varied amongst the assessment options we collated. Whilst it was possible therefore to make a comparison of the specific implementations used within each assessment option, it was challenging to draw broader generalisations that held true when, for example, the period assessed, or the level of uncertainty varied.

2.1.3 Potential test metrics describing a single rate of change

To start to describe more general analytical approaches, rather than specific assessment examples, we visualised the range of test metrics or questions that could be asked of a single rate of change estimate and associated uncertainty, irrespective of the period over which the rate of change was estimated or the statistical model used.

2.2 Review and refine

We shared the results of section 2.1 with a range of analytical ecologists for review and arranged a series of online workshops to share and discuss their views. Prior to reading our exploration of options, we asked each reviewer to share their initial thoughts on how they would assess whether species abundance decline has halted in England using a multispecies abundance indicator.

We then asked them to review our work and provide feedback on:

- The range of options presented;
- How we had grouped and framed them;
- Our approach to drawing out the properties of each assessment option.

As well as comparing between assessments methods, it would be possible to use either;

(1) Two or more different tests on the same time-period and indicator
(2) The same test in more than one year
(3) The same test on a different subset of the indicator, for example, each species or taxonomic group, or on the full dataset omitting one species or taxonomic group at a time.

We were not able to capture the implications of conducting multiple tests within our scoring framework (2.1.2), but these were covered during the workshop discussions.

Each aspect of our work was revised following these discussion sessions.

2.3 Exploration of secondary target legislation and potential future work

Given that we already have an assessment method set out in the secondary target legislation, we considered that it was important to understand how this test may behave under different possible scenarios of future change in species abundance.

We conducted a preliminary exploration by simulation of the implications of working within the current wording of the secondary target legislation, which states we will have met the 2030 species abundance target if the ESAI estimate in 2030 is at or above the estimate in 2029, without consideration of uncertainty around either estimate or consideration of the rate of change in previous time periods. We used a simple simulation to estimate the likelihood of correctly concluding whether the target had been met under a range of different possible scenarios. In each case, we consider the rate of change between two time points.

In the first simulation, we consider three possible patterns of future biodiversity change: slight decline (0.5% decline per year), zero change (0% change per year) and slight increase (0.5% increase per year), all rates on a log-scale. To give these rates of change some context, if the indicator increased at 0.5% per year, then after 25 years it would have risen by 13%. Equally if the indicator declined by 0.5% per year, after 25 years it would have declined by 12%. These patterns represent possible average change across species, within which individual species are likely to show a variety of trends. We included zero change here, as this represents the lowest true pattern of change where the target would be met, and helps us understand the confidence we may have in differentiating between a true shallow ongoing decline (where the target would not be met) and zero change (where the target would be met).

As well as understanding the test behaviour under different patterns of biodiversity change, it is also important to understand how the test behaves under different levels of uncertainty introduced by data collection and analytical processes, in other words, how far the rate of change we observe may be from the true biological change. We considered three levels of uncertainty around our measures of change: low (0.3%), medium (0.8%) and high (1.3%).

If the true biological change in the environment was zero, then we may observe a point estimate of that change that is anything from slightly negative to slightly positive. In all cases we would have met the target, however, given the wording of the test, we would only conclude that we had met it in a subset of cases, depending on whether the observed change estimate was at or above zero. Here we quantify the likelihood of correctly concluding whether or not the target has been met, as indicated by the true rate of change, across the potential distribution of observed point values.

For each of the nine combinations of true rate of change and trend uncertainty, we described a lognormal distribution using the rate of change as the mean and the estimate of uncertainty as the standard deviation. We then asked the probability of the test of the 2030 species abundance target providing the correct outcome. The correct outcome being that the target is met where the true change is zero or slightly increasing, represented by the proportion of the normal distribution at or above zero, and not met where the true change is slightly declining, represented by the proportion of the normal distribution less than zero.

In our second simulation, we extend the values we used to describe the true pattern of biodiversity change from the three values representing slight decline, zero change and slight increase, to a sequence of values ranging from a 3% decline per year to a 3% increase. A 3% per annum decline if maintained would represent a decline of 53% over 25 years, whilst a 3% per annum increase would result in an increase of 112% over the same time period. Using the same three levels of trend uncertainty, (0.3%, 0.8% and 1.3%) we again estimate the likelihood of the correct test outcome.

Finally, in our Recommendations (section 4), we set out a range of further analytical work and activities that would further our understanding of which assessment method(s) may be most appropriate to test the 2030 species abundance target under a range of potential future scenarios.

3. Results

3.1 Option Exploration

3.1.1 Options collation and framing

Options for testing the species abundance target vary at two key stages (Figure 2). The first stage describes the ‘simplification’ model used to summarise how an indicator or time-series is changing over time. Choices range from no simplification, meaning the test is conducted on the unsmoothed index values, to a linear model with varying number of breakpoints or a non-linear model with varying levels of smoothing. The second stage describes the type of ‘question’ used on the model output. For example, asking whether or not a rate of change is at or above zero. This essentially describes the hypothesis to be tested in each case. The ‘questions’ in the second stage can be divided into two broad categories. Question groups A and B are concerned with a single estimate of the change in the indicator between the final time point and some previous time point, including but not restricted to the penultimate time point. Question groups C and D put recent change into a wider context by contrasting the recent rate of change with one or more historical rates of change. In many cases, this means looking at longer-term patterns than those considered by Question groups A and B. We collated 16 examples of time-series analysis relevant to this report from the literature (Figure 3). These represent a variety of combinations of ‘simplification’ model and ‘question’. Fuller details for each example are given in Appendix 1.

Six of the assessment options fall into Question group A, which seeks to determine whether the rate of change is greater than or equal to zero (Figure 3). These include the assessment method described in the secondary target legislation (Figure 3, Option 1) and related methods which also consider rate of change between the final and penultimate indicator estimates, but account for the uncertainty around the rate of change as well as the point estimate (Options 2, 3). The first four options all use a non-linear model to estimate the rate of change. The final two options in Question group A compare a linear model with a continuous year term, to an intercept only or null model and consider a longer assessment period.

Three tests in Question group B are similar to those in A, but rather than a single threshold of zero change, these tests seek to determine whether the rate of change falls within a predefined range where it is both greater than a declining threshold and lower than an increasing one. This assessment can be done based on the magnitude of change alone (Option 7), or additionally incorporate uncertainty around the point estimate (Options 8, 9). These approaches could also be implemented using only the lower threshold and asking whether the rate of change is above a declining threshold.

Tests in Question groups C and D compare the rate of change over the recent past with either longer-term change, or with similar lengths of time from earlier in the time-series. Several of the options collated are variations on breakpoint or broken-stick linear regression, which look at the strength of evidence for a difference in the rate of change (Options 11-13). In contrast to this, one option uses progressive partial regression to identify periods over which the indicator could be considered to have a trend of zero (Option 14).

The options set out in Figure 3 relate to specific examples in the literature. It would be possible to use any other combination of simplification model and question. For example, you could ask whether the rate of change over the final time step is at or above zero (Options 1-3) based on an unsmoothed, (Simplifying process i)) rather than a smoothed multispecies indicator, or you could describe the difference in the rate of change between two time-periods (Question group C) categorising it based on the magnitude of the point estimate (Option 7).

Figure 2: Conceptual framing exploring options to test a 'halt decline' target. We start with an Indicator and consider two separate processes that could be combined in a variety of ways to conduct an assessment. Firstly, a simplifying process (Numerals i) – iii)), involving applying a statistical model to the indicator to summarise change over time to varying degrees. A variety of questions (Letters A – D) can then be asked of the simplified indicator. There are a variety of statistical approaches to each question. Examples of the different combinations of simplifying process and question are given in Figure 3.

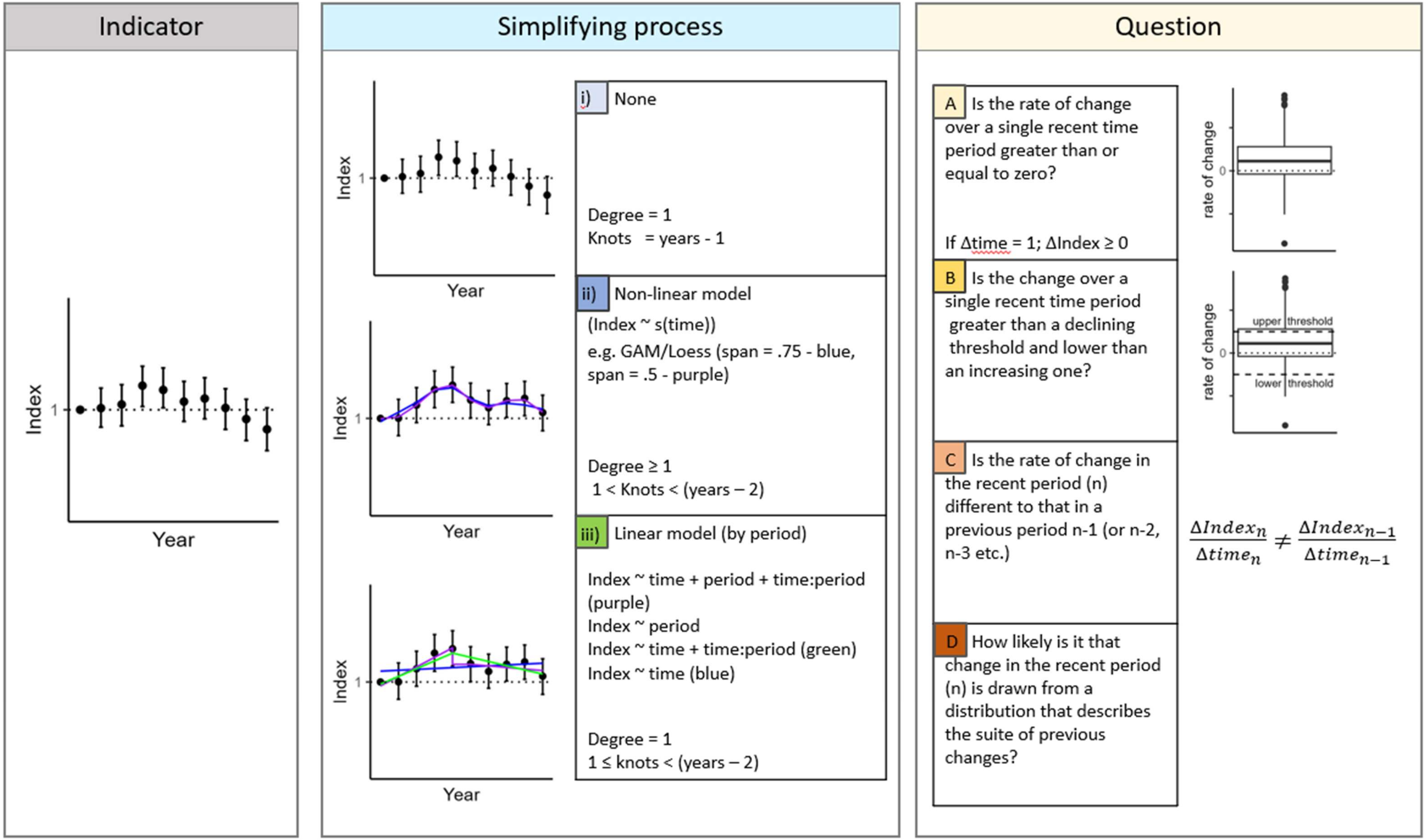
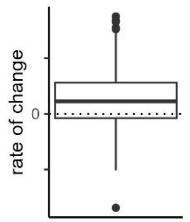
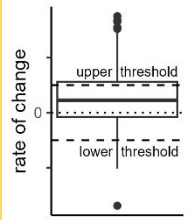
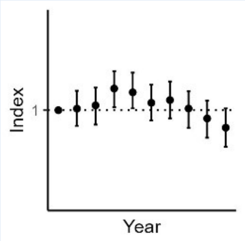
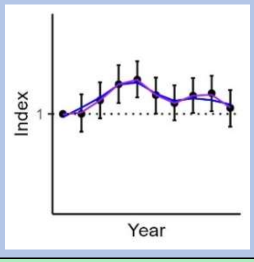
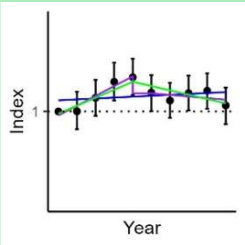


Figure 3: Examples of indicator/time-series analysis using different combinations of simplifying models and questions. We identify 16 test options numbered 1-16. The numbers in square brackets refer to specific published examples of those tests: see References below. Throughout UI indicates Uncertainty Interval, m = slope.

Question / **Simplify**	**A. Is the rate of change over a single time period greater than or equal to zero** 	**B. Is the rate of change over a single time period > than a declining threshold & < than an increasing one?** 	**C. Is the rate of change in the recent period (n) different to that in a previous period n-1 (or n-2, n-3 etc.)**	**D. How likely is it that change in the recent period (n) is drawn from a distribution that describes the suits of previous changes?**
i) None 				**Anomalies versus change points** 15. Machine learning approach to differentiate between anomalies and change points[12]
ii) Non-linear model 	**Final rate of change (Δyear = 1):** 1. Without uncertainty (2° legislation) [7] 2. Bayesian UI, proportion of posterior distribution at or above zero [13] 3. Frequentist UI, 95% threshold [14] **Final to baseline comparison (Δyear > 1):** 4. Final index UI ≥ baseline (e.g. WBI[15], LPI).	**Categorise final to baseline cf.** 7. Species level average growth rate estimated from final to baseline comparison and categorised, e.g. WBI [15]	**Change point analysis** 10. Use 2nd derivatives and associated UI to identify change points in the indicator [16]	
iii) Linear model 	**Distribution of species level growth rates** 5. Compare distribution of average species growth rates from linear model (Index ~ time) to one centred around zero with equal levels of variation [17] **Linear model cf. to intercept only** 6. Does linear model (Index ~ time) have greater AIC weight than the null model (Index ~ 1)? Plus, do index values predict the rate of change - density dependence test ($Index_{t+1}/Index_t \sim Index_t$) [18].	**Categorise linear model trend** 8. Equivalence test of two one-sided t-tests [19] 9. Categorise slope of bootstrapped linear model (Index ~ time) using its magnitude and significance [6]	**Breakpoint regression** 11. Index ~ time * period [20] 12. Index ~ time + time:period [21] 13. $Index_{period} \sim m.time_{period}$; Δm ≥ 0 [6] **Progressive partial regression** 14. Identify year ranges where there is zero change in the index [22]	**Piecewise regression** 16. Does recent change (final decade estimate of: Index ~ time * decade) differ from global average [1]

3.1.2 Properties of assessment options

We scored each assessment example against a range of 'properties', for example, timeliness, ease of communication or propagation of error (Table 1), in each case on a three-point categorical scale. We split options addressing Question groups A and B (Table 2) from those addressing Question groups C and D (Table 3) given that these consider different things. The former ask how consistent a single rate of change is with zero, and the latter consider the difference between two or more rates of change. In each case, we scored the specific implementation used in each assessment option in Figure 3, rather than, for example, describing more broadly the strengths and weaknesses of a linear simplification model compared to a non-linear one. We use a categorical scale as the scores are not additive. The scores consider only a subset of the properties or characteristics that could influence an assessment method's suitability and some properties may variably be considered more or less important than others.

Table 1: Description of properties, and the levels of each, used to describe assessment options in Table 2 and 3.

Property	Description	Levels
Timeliness - assessment	Period over which the assessment is undertaken	A: < five years, B: Five to ten years, C: 10+ years
Timeliness - contributing	Period over which data are contributing to the assessment (including smoothing)	A: < five years, B: Five to ten years, C: 10+ years
Propagate uncertainty	Does the assessment account for uncertainty in the indicator and in the species time-series?	A: Both and fully; B: Some aspects; C: Little or no consideration of uncertainty
Compatibility with secondary legislation	What level of change would be needed to the legislation to implement the assessment?	A: No change or very little; B: Some changes required; C: Significant changes required
Ease of communication	How easy is it to describe the assessment to a range of audiences?	A: Easy to describe; B: Mostly straightforward; C: Challenging
False positives	Likelihood of the assessment falsely concluding a decline has halted when it has not	A: Unlikely; B: Moderately likely; C: Likely
False negatives	Likelihood of the assessment falsely concluding a decline has not halted when it has	A: Unlikely; B: Moderately likely; C: Likely

We have two measures of '**Timeliness**'; the number of years the assessment is made over, for example, between the final and penultimate year, and the number of years of data that contribute to the indicator values within the assessment period. The options that assess the rate of change between the final and penultimate year have the highest scores for '**Timeliness – assessment'** (Table 2). However, these examples all use a non-linear simplification model so many more years of data contribute to the values estimated for the final two years and they therefore have lower scores for '**Timeliness – contributing'**. The number of years of data contributing to the final two index values will vary with the level of smoothing used. In most forms of non-linear model it is difficult to know precisely how each year of data has contributed to the final two index values as the change points in the smoothed trend, the 'knots', will be unevenly spaced through time.

The options that only consider the rate of change in the final time step are also the closest to the **current wording of the secondary target legislation** and would require no changes (Option 1), or only the addition of a statistical test to describe how the final and penultimate values are to be compared (Options 2,3). In all cases where a smoothed indicator is used (including Options 2,3), then consideration could also be given to delaying the test until at least one year of additional data are available. This would mean waiting until 2031 data are available, which would help to anchor the 2030 value and reduce uncertainty around the estimate of change between 2029 and 2030. Using other options would require more substantial changes to the wording of the secondary target legislation.

Most of the assessment options account for the uncertainty around the rate of change estimate, however, several do not fully **propagate uncertainty** in the species level time-series through to the assessment (e.g. Options 2, 5, 6).

It is not possible to give an overall score for the rate of **false positives** or **false negatives** for each option (Table 1), as these rates vary depending on the rate of change. Instead, we scored false positives and negatives for the case where the rate of change was the minimum required for the target to have been met, that is where the true rate of change is zero. Given we are estimating this change with uncertainty, then the observed difference may be negative or positive. The correct test outcome in this situation is that the target has been met, so there will be low or no false positives for the assessment methods addressing Question group A. The level of false negatives will vary, amongst other things, due to the different statistical thresholds used in each case. Where the standard statistical threshold of 95% is used, then the level of false negatives is high (e.g. Options 3, 4). This means there is a good chance of falsely concluding that the target has been missed when it has been met. The rate of false negatives is lower for the assessment method based on the secondary target legislation (Option 1), where the threshold can be thought of as 50%, which is equivalent to considering the point estimate assuming the uncertainty interval is lognormally distributed. False negatives are lowest where the point estimate of the rate of change is considered to represent zero change if it is above a declining threshold and below an increasing one (Option 5). Looking at the rate of false positives and negatives where the rate of change is zero only gives us a partial picture and ideally this thought exercise would be repeated for a range of different patterns of change.

The assessment options that consider Question groups C and D (Table 3) tend to cover longer time-periods and therefore have low scores for both aspects of **Timeliness**. Several of the options are complex or have multiple components and are therefore likely to be more **difficult to communicate** with a range of audiences (e.g. Options 15, 16). Many of the options addressing Question group C use a 95% statistical level of confidence, meaning that there is a high rate of **false negatives** when the difference between the rate of change over two time periods is around zero (Options 10:13,16). Finally, all the options in Table 3 are further from the current **wording of the secondary target legislation**, with more substantial changes needed to allow them to be used to test to the 2030 species abundance target.

Table 2: Summary of the properties of assessment options (1:9) outlined in Figure 3 for Question groups A and B. Definitions of properties in Table 1. Throughout 'UI' indicates Uncertainty interval

Option			Properties						
			Timeliness		Propagates uncertainty	Ease of communication	Compatibility with SI	False positives	False negatives
			Assessed over	Contributing				When rate of change ≈ zero	
A.	ii)	1. Final rate of change, no UI	A	B	C	A	A	A	B
		2. Final rate of change, α UI ≥ 0	A	B	B	B	A	A	No threshold set
		3. Final rate of change, 95% threshold	A	B	A	B	A	A	C
		4. Final to baseline comparison	B	C	B	A	B	A	C
	iii)	5. Species growth rate distribution	C	C	B	B	C	A	C
		6. Linear cf. null with DD test	C	C	B	C	C	A	A
B.	ii)	7. Final to baseline categorisation	B	C	C	B	B	A	A
	iii)	8. Equivalence test	B	B	B	B	C	B	B
		9. Categorise linear trend	B	B	A	B	B	B	B

Table 3: Summary of the properties of assessment options (10:16) outlined in Figure 3 for Question groups C and D. Definitions of properties in Table 1. Throughout UI indicates Uncertainty interval

Option			Properties						
			Timeliness		Propagate uncertainty	Ease of communication	Compatibility with SI	False positives	False negatives
			Assessed over	Contributing				When rate of change between periods ≈ zero	
C.	ii)	10. Change point analysis	A	B	A	B	C	A	C
	iii)	11-13. Breakpoint regression	C	C	A	A	C	A	C
		14. Prog. partial regression	C	C	B	C	B	A	A
D.	i)	15. Anomalies versus change points	A	C	B	C	C	B	B
	iii)	16. Piecewise regression	C	C	A	C	C	A	C

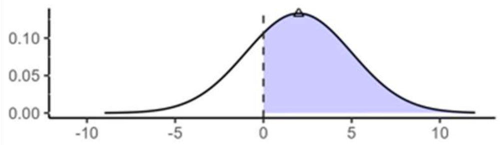
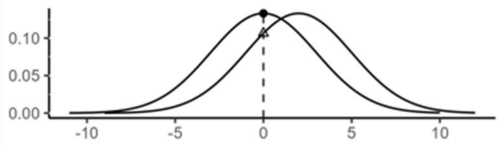
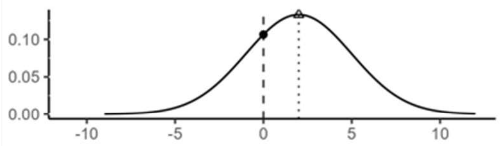
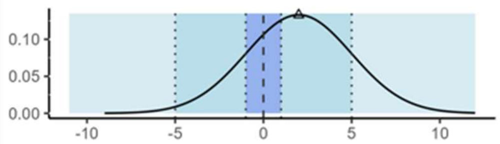
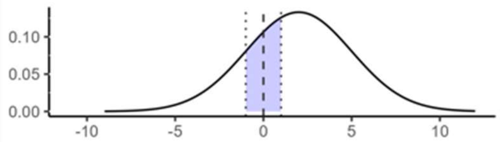
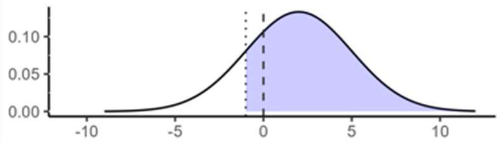
3.1.3 Potential test metrics describing a single rate of change

Tables 2 and 3 are one way of comparing the different assessment methods we consider. However, they do not allow us to draw strong conclusions around which approach(es) may be more suitable to test the 2030 target. One challenge in comparing the examples is that they represent a predefined combination of analytical choices, each of which may influence the performance of an assessment method. These include, for example, the duration of the assessment period, the level of smoothing, or the level of uncertainty in the species level time-series or multispecies indicator. Without further analytical interrogation we cannot readily tease apart and examine the implications of these choices individually.

As a way of disaggregating the potential questions we can ask of one or more estimated rates of change, we visualised the range of questions you could ask of any single measure of the rate of change in the multispecies indicator and its associated uncertainty (Figure 4). Figure 4 shows a hypothetical posterior distribution of a rate of change from a Bayesian model (but this could equally be a distribution of bootstraps). This distribution could describe the change over one or multiple time-steps and could be generated using a range of modelling approaches. With this single distribution we show six test statistics that could be used to describe the consistency of the posterior distribution with zero change. We link each assessment option addressing Question groups A and B (Options 1:9, Figure 3) to the most closely aligned test statistic. Some of these statistics, such as the Bayes Factor, can be used as both a measure of the support for a difference from zero change, or support for consistency with zero change.

The assessment options addressing Question group C (Options 10:14) could be described in similar fashion, by treating each distribution as the difference in the rate of change in the recent period compared to a previous time periods.

Figure 4: An alternative way of framing the options 1-9 listed under Question groups A and B in Figure 3. In each case a hypothetical posterior distribution of a rate of change is shown, with different questions asked of the data and the options that align to each metric shown (Based on Figure 1 of Makowski et al 2019 [23]).

	Posterior distribution of change	Question	Aligned options
A		What is the likelihood that the rate of change is at or above zero? With or without a binary threshold.	**1**: Threshold of 50%, equivalent of asking whether the maximum density estimate (Δ) is ≥ 0. **2**: No threshold **3,4**: Threshold of 95%, i.e. 95% of posterior is ≥ 0.
		How far do the observed values move from a null model. Bayes Factor: null density at zero (•) divided by the observed density at zero (Δ).	**5** **6**: Step 1
		What are the odds that the rate of change differs to the null hypothesis. MAP-based p-value ; the estimate at zero(•) divided by the maximum *a posteriori* (MAP) (Δ).	No options
B		Does the maximum density (Δ) estimate fall within a pre-defined 'stable' category (mid blue)?	**7**
		What is the likelihood that the rate of change is within a pre-defined 'region of equivalence' (mid-blue).	**8** **9**
		What is the likelihood that the rate of change is at or above a specified declining threshold?	No options

3.2 Review and refine

As described above, we conducted a series of discussions with nine experienced analysts. A summary of the key points raised by the reviewers is outlined below.

Most reviewers felt that simply comparing the absolute values of two index estimates (as outlined in the secondary target legislation) was insufficient as a test the 2030 species abundance target, with the most common preference being to take a probabilistic approach, based on the likelihood that an estimate of the recent rate of change was at or above zero.

Some questions were raised around the wording of the legislation. For example, does the halt need to be maintained, or is meeting the target in a single year sufficient? A number of reviewers thought it would be important to meet the target in each subsequent year after 2030. Related to this, several reviewers commented on the elevated uncertainty and volatility of the final estimate in a smoothed time-series. The secondary target legislation mandates testing the species abundance target as soon as 2030 data become available, in 2032. This means that in 2032 the indicator estimate for 2030, and the estimated change between 2029 and 2030 used to assess the target, will be less certain than values for earlier years and subject to greater change once data for subsequent years become available. In theory, the 2030 target could be judged met in 2032, but with refinement in later years it could become evident that the target had actually been missed.

There were variable views on whether we should focus solely on the recent period, or first compare whether recent change is more positive than historic change and subsequently if recent change is zero or positive.

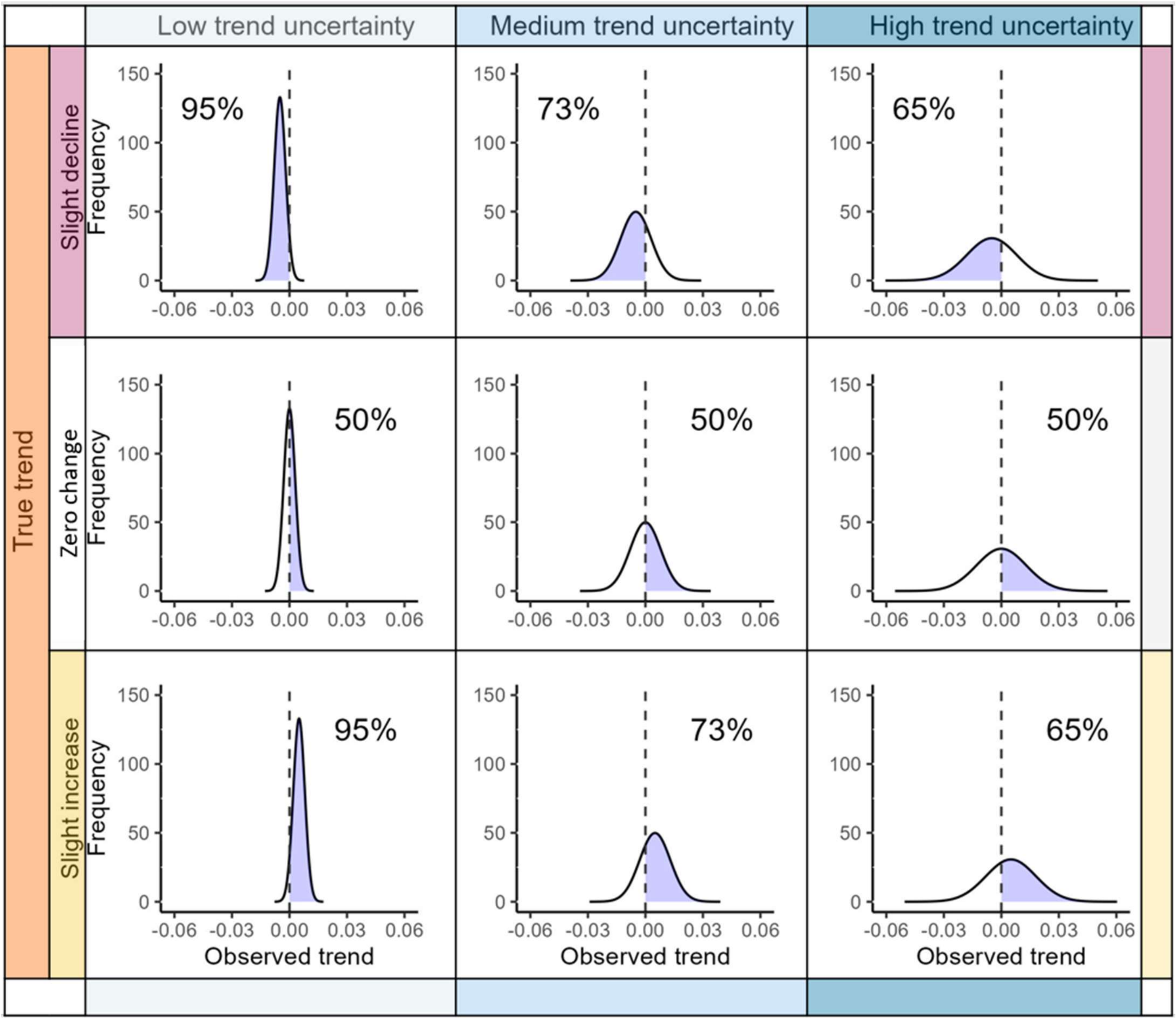
Several suggestions were put forward for analytical work. Most reviewers felt that it would be useful to analytically explore the impact of the duration of the time-period to include in any test and/or the level of smoothing on the assessment outcome. There was support for including all levels of smoothing in any simulation study, from an unsmoothed indicator to a linear model. There was also support to investigate whether subsets of the species met the target, for example individual taxonomic groups, or to investigate the sensitivity of each test to which taxa were included, akin to a jackknife.

Finally, the reviewers supported including a measure of confidence in any assessment, even if a binary threshold was used to say whether the target had been met or not. This could be including the magnitude of the test measure as well as the binary outcome, or a categorical scale of confidence, which may include different components.

3.3 Exploring implications of the secondary target legislation

The secondary target legislation requires the species abundance target to be tested by determining if the indicator value in 2030 is at or above that in 2029. Hence, it is useful to understand how likely we would be to conclude that the target had been met given different potential rates of change in species abundances, and how certain we are of those changes.

Figure 5: Visualisation of the posterior distribution of possible observed trends (log scale) given three levels of 'true' trend (rows: slight decline (-0.5% per annum), zero change, slight increase (0.5% per annum)) and three levels of uncertainty (columns: low, medium and high) in the 'true' trend. The area of the probability distribution where the test correctly ascertains whether the target has been met is shaded blue, with the percent likelihood of a correct test outcome shown in each panel. This is the area of the probability distribution greater or equal to zero where the 'true' rate of change is Zero change or Slight Increase and the area of the distribution less than zero where the 'true' change is Slight decline.



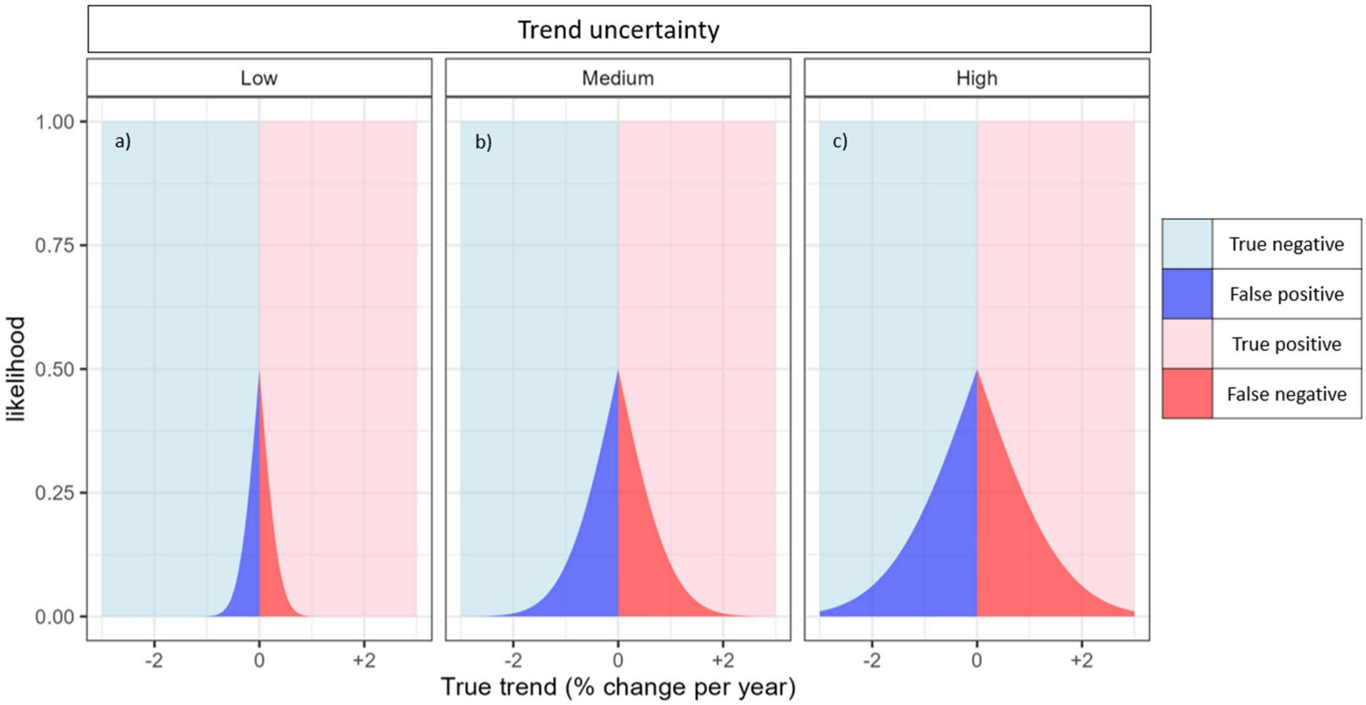
Our simple simulation illustrates that if the true average biological change is zero (so the distribution of observed change is symmetrical around zero change on a log scale, middle row Figure 5), then we have a 50% chance of correctly concluding that we have met the target, and a 50% chance of incorrectly concluding that we have failed to meet the target (50% rate of false negatives). If the true trend is zero, then trend uncertainty does not influence this pattern.

We have greatest chance of reaching an incorrect conclusion when the true rate of change is close to zero. As the true trend moves away from zero then the likelihood of a correct test outcome increases i.e., the rate of false positives and false negatives declines. However, away from zero change, this increased likelihood of a correct conclusion could be counteracted when uncertainty in the trend estimate is greater, which would lead to a higher rate of false positives or false negatives (a higher error rate). For example, in scenarios with 'high' trend uncertainty there is a likelihood of reaching the incorrect conclusion in over one third of occasions (Figure 5).

To illustrate how the error rate varies depending on the 'true' rate of change and associated uncertainty, the simple simulation above was extended to consider a broader range of 'true' rates of change on a continuous scale from a decline of -3% a year to an increase of +3% a year, presented on a log scale. For each measure of 'true' change and for each of the three levels of uncertainty in that change, we estimate the probability that the test will correctly conclude whether or not the target has been met (Figure 6).

When uncertainty around the 'true' change is low, we can reliably differentiate between 1% decline and zero change (the pale blue, indicating the likelihood of a true negative, extends all the way along the y-axis at -1% 'true' change (Figure 6a)), but this becomes challenging as trend uncertainty increases. When uncertainty is high, the incorrect test outcome may still be observed (6% likelihood) even when the rate of 'true' change is substantial, at +/-2% per annum (Figure 6c).

Figure 6: Visualisation of the balance of likelihood between drawing the conclusion that reflects the 'true' trend (correct outcome) or otherwise for a range of values of 'true' trend and three levels of trend uncertainty. Pale colours illustrate the likelihood that the test will correctly conclude whether the target has been met and bold colours show the likelihood the test will draw the incorrect conclusion (the error rate). The test in each case is that set out in the secondary target legislation, assessing whether the 'observed' indicator value in the final year is at or above the value in the penultimate year.



4. Recommendations

This report has focussed on an initial exploration of assessment options that could be used to test the 2030 species abundance target and set out certain attributes of each option that make them more or less appropriate for use. Given these attributes vary within each option depending on precise analytical decisions, this report is not able to recommend one particular option over another. Here, we outline a range of potential extensions from the current project that would build more of the supporting evidence needed to inform the developments required to effectively test the 2030 species abundance target. This evidence could advise a discussion as to whether the current secondary target legislation is sufficient, whether a supplementary document setting out how the secondary target legislation should be interpreted is required - for example, setting out a statistical test to compare the indicator values in 2029 and 2030 - or whether changes are recommended to the secondary target legislation and alternative or additional tests developed.

4.1 Analytical suggestions

4.1.1. Understanding the implications of the secondary target legislation

Repeat simulations similar to those presented in Section 3.3 aligned to the most up-to-date ESAI to further explore the implications of secondary target legislation. This could set out the likely level of false positives/negatives for a range of different levels of true change using realistic levels of trend uncertainty derived from the ESAI dataset. It would be important to include in the simulation levels of uncertainty that mimic those observed around the penultimate trend estimate and those observed around the final estimate. There will always be greater uncertainty in the final trend estimate and this simulation would give a measure of the extent to which this greater uncertainty influences how likely it is that we will correctly conclude whether or not the target has been met. This would allow a more informed choice of whether it would be appropriate to delay testing the species abundance target until data are available subsequent to the assessment period (e.g. for 2031).

4.1.2. Explore implications of alternative thresholds and test statistics

Assessing whether the point estimate of the rate of change is at or above zero can also be framed as asking whether at least 50% of the posterior distribution is at or above zero. A useful extension to Simulation 4.1.1 would be to consider a range of alternative statistical thresholds and estimate the level of false positives and negatives under different levels of true trend and trend uncertainty.

As well as capturing the proportion of the posterior distribution at or above zero, this simulation could also consider the range of alternative metrics set out in Figure 4. For example, the proportion of the posterior distribution at or above a pre-set ‘declining’ threshold, or a Bayes Factor indicating the level of support for either the null hypothesis of no change or the alternative hypothesis of a movement away from zero change. This would offer an assessment of the complementarity of the different metrics and whether more than one could be used in concert to allow greater power to differentiate between different levels of true changes.

4.1.3. Application to the ESAI

Building upon the previous simple simulations, explore how a set of candidate tests behave using the observed ESAI dataset or using species time-series data simulated with levels of interannual and intraspecific variation similar to the ESAI dataset. This could build on the work of Atwood et al. [13], who investigated the impact of different levels of smoothing, number of species and underlying rate of change on the proportion of the posterior distribution of the final rate of change that was at or above zero. A new simulation could repeat this work using the ESAI dataset and extend it to consider some of the alternative test metrics (Figures 3, 4), and further explore the implications of the smoothing level and time-period assessed. This work could also start to quantify which analytical choices or data biases are the greatest sources of uncertainty and which tests are most sensitive to particular variables.

4.1.4 Sensitivity analysis

A final aspect of the potential analytical work would be a sensitivity analysis. Using a small set of assessment methods identified in the previous simulations, investigate how sensitive the test outcomes are to the species or taxa groups included and if the test is repeated in subsequent years.

4.1.5 Compatibility with the 2042 target

The secondary target legislation does not mention what has to happen between the 2030 target of halting the decline in species abundance and the 2042 target, which states that the indicator value in 2042 must be 10% greater than that in 2030 and higher than the 2022 value. It would be useful to set the discussion in this report and the results of any future simulations in context, by considering the potential envelope of trends that would allow the 2042 target to be met. For example, would meeting the 2042 target become unlikely unless the 2030 target was met for every intervening year between the two targets, meaning that the decline was halted throughout the 2030s.

4.2 Broader suggestions

Conducting the recommended simulations would support informed decisions of how to effectively test the 2030 species abundance target, however, as mentioned earlier in the report, some decisions, such as the acceptable balance between false negatives and false positives, are a societal and value-based decision. Therefore, the analytical pieces of work would be valuably complemented by a broader workshop bringing together researchers, policy and decision makers.

The workshop could consider aspects such as the acceptable level of risk in any test and the related acceptable levels of statistical confidence. It could also consider the potential benefits of requiring that the test is met in multiple years or across multiple subsets of the data, or that it is delayed until at least one year of data subsequent to the assessment period is available. Finally, it could consider the benefits of looking at the long-term as well as recent change and asking first whether the rate of recent change differs from a previous period and then whether the recent rate of change is zero or positive.

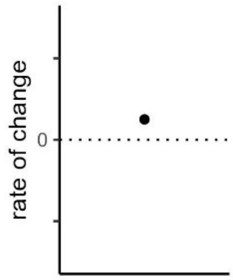
Acknowledgements

We thank the nine ecological analysts who reviewed our work and took part in discussion sessions representing the RSPB Centre for Conservation Science, University of Cambridge, Institute of Zoology and UK Centre for Ecology and Hydrology. We also thank Andy Gill and Robbie McDonald of the Office from Environmental Protection (OEP), for insightful discussion and the OEP for commissioning and funding this work.

Appendix 1: Detailed description of each assessment option

Option 1. Final rate of change (Δ year = 1) without uncertainty [7]

2	Simplifying step	ii) – GAM
	Test	A. – estimated final rate of change



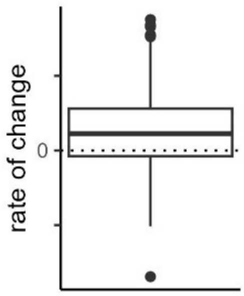
Description: Bayesian indicator created in a GAM framework[9]. If the absolute change between the penultimate and ultimate years of the indicator is zero or positive then the decline is considered to have halted. This is intended assessment method for testing to the Environment Act target using the ESAI, set out the secondary target legislation.

Years testing over	2 years
Years contributing to the assessment	Variable depending on smoothing
Likelihood of false positives	Low where change is zero
Likelihood of false negatives	Medium

Comments: Increased smoothing and testing to the penultimate, rather than the ultimate year of data would reduce incidence of false positives and negatives.

Option 2. Final rate of change (Δ year = 1) Bayesian test using UI [13]

3	Simplifying step	ii) – GAM
	Test	A – Credible interval of final rate of change



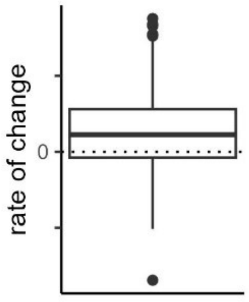
Description: Bayesian indicator creation in GAM framework[24], including a similar dataset to that underpinning the ESAI, tested with variable levels of smoothing. Takes the proportion of the credible interval of the final rate of change (between the final and penultimate year) at or above zero as the likelihood that decline has halted.

Years testing over	2 years
Years contributing to the assessment	Variable depending on smoothing
Likelihood of false positives	Low when change is close to zero
Likelihood of false negatives	NA – No threshold set

Comments: One or more additional years of data subsequent to the assessment period would increase the precision of the final change estimate and mean that it was subject to less change as additional years of data were added.

Option 3. Final rate of change (Δ year = 1) frequentist test using UI [14]

4	Simplifying step	ii) – GAM
	Test	A – Confidence interval of final rate of change



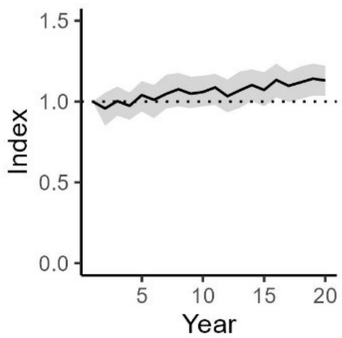
Description: There was a Defra Public Service Agreement target to halt the decline in the farmland bird indicator for England by 2020. The target was considered to be met if the 95% confidence interval of the final change in the indicator (between the final and penultimate year) was at or above zero.

Years testing over	2 years
Years contributing to the assessment	Variable depending on smoothing
Likelihood of false positives	Low when change is zero
Likelihood of false negatives	High

Comments: There were significant subsidiary tests associated with this target. As well as the test being met across the Farmland Bird Indicator, the test needed to be met for all 19 constituent species and be met for several consecutive years.

Option 4. Final to baseline comparison (Δ year > 1) [15]

5	Simplifying step	ii) – GAM
	Test	A – Final to baseline comparison



Description: Many well established species abundance indicators assess change over different time-periods by comparing the UI of the final index value to the baseline. Only where the final year 95% uncertainty intervals are entirely above or below the baseline is the indicator considered to be increasing or decreasing respectively.

Years testing over	Variable, often 5, 10, 25 years
Years contributing to the assessment	Variable depending on smoothing
Likelihood of false positives	Low when change is zero
Likelihood of false negatives	High

Comments: One recent paper [24] used uncertainty in the first data point to project a comparator, rather than assuming no change.

Option 5. Distribution of species level growth rates cf. to one centred on zero [17]

8	Simplifying step	iii) – Linear model
	Test	A

Description: Uses a Bayesian framework to assess the likelihood that the distribution of population average annual growth rates, estimated from the slope of linear regressions over the full span of each population time-series, differs to a distribution centred on zero but with the same level of variability. This mimicking a scenario where conservation had been successful in halting declines in populations on average. Uses the Living Planet dataset at the scale of taxa groups within countries ('systems'), sample size >10.

Years testing over	Variable, but mean time-series length is ~5 years and comparator period assumed to be of similar duration
Years contributing to the assessment	Variable, on average 10-20
Likelihood of false positives	Low when change is zero
Likelihood of false negatives	High

Comments: The authors suggest lowering the accepted level of confidence that a target has been met (i.e. lower than 95% likelihood) and setting a 'reference' threshold for the rate of change to be above that is lower than zero (the example given is -1.5% per year). Related work: Overlapping index - Pastore & Calcagni (2019) [25].

Option 6. Linear model cf. to intercept only plus density dependence test [18]

14	Simplifying step	iii) – Linear model over a single time period
	Test	A - Is rate of change at or above zero

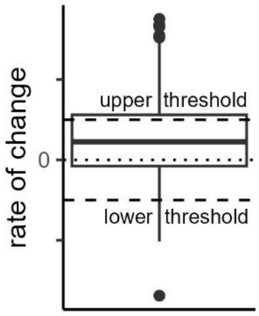
Description: Compare relative support (AIC weights) for a linear model with explanatory variable of time (Index ~ time), to one without (Index ~ 1). Then use a secondary 'density dependence' test, which asks if the size of a population affects its subsequent growth rate ($\ln(Index_{t+1}/Index_t) \sim Index_t$). If the relationship is negative, the population fluctuates around a return point, a type of long-term mean population density. The AIC weights of the two tests are combined to give a single probability for 'no trend'.

Years testing over	10 – 20 years
Years contributing to the assessment	Density dependence test works best with ≥16 years of data
Likelihood of false positives	Low when change is zero
Likelihood of false negatives	Low

Comments: This is an interesting approach and explicitly tackles the issue of differentiating between low power and negligible trend.

Option 7. Categorise final to baseline comparison (Δyear > 1) [15]

6	Simplifying step	ii) – GAM
	Test	B



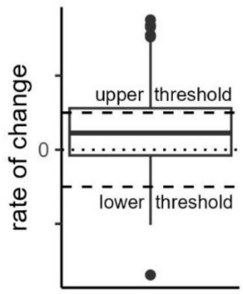
Description: Many species indicators use pre-defined thresholds to place species level change into categories based on the annual average growth across a particular time-period, estimated using a comparison of the final to baseline value. The categories used by the Wild Bird Indicators have been adopted by several other indicators, such as the UK[26] and England priority species indicators[8]. In each case the Stable, or Little Change category lies between -1.14 % and +1.16% per annum.

Years testing over	Variable, often 5, 10 or 25 years
Years contributing to the assessment	Variable depending on smoothing
Likelihood of false positives	Low when change is zero
Likelihood of false negatives	Low

Comments: These categories are most often used as contextual information around an indicator, to illustrate species level variation around the average trajectory shown in the indicator, rather than a formal assessment method.

Option 8. Categorise linear model trend: Equivalence test, two one-sided t-tests [19]

15	Simplifying step	iii) – Linear model over a single time period
	Test	B - Trend within stable threshold



Description: The null hypothesis here is that the trend is not zero, but outside an a priori specified equivalence region defining trends that are considered to be negligible. This null hypothesis can be tested with two one-sided tests or by testing whether an uncertainty interval lies entirely within the equivalence region. The equivalence region for trends in population size tested was a log-linear regression slope of (−0.0346, 0.0346). This corresponds to a half-life or doubling time of 20 years for population size.

Years testing over	~20 years
Years contributing to the assessment	~20 years
Likelihood of false positives	Medium when change is zero
Likelihood of false negatives	Medium

Comments: This feels a sensible approach, barring discussion on how to set the thresholds of the equivalence region, which are likely to be case specific.

Options 9. Categorise linear model trend using its magnitude and significance [6]

16	Simplifying step	iii) – Linear model over one time period
	Test	B – Stable threshold

Description: The MSI tool runs a linear model (Index ~ time) across each bootstrap of the indicator. The linear trend and uncertainty interval are categorised by magnitude and precision. The stable category is defined at: CI includes 1.00 AND 0.95 ≤ lower CL AND upper CL ≤ 1.05 (no significant increase or decline, likely that changes are smaller than 5% per year). There is also an Uncertain category, where magnitude of change is similar but CI are larger.

Years testing over	10+
Years contributing to the assessment	10+
Likelihood of false positives	Medium when change is zero
Likelihood of false negatives	Medium

Comments: These categories are largely used for interpretation rather than formal testing to targets.

Option 10. Change point analysis using 2nd derivatives [16]

7	Simplifying step	ii) – GAM
	Test	C – 2nd derivative

Description: The 2nd derivative (or change in the rate of change) is estimated between adjacent trend estimates. This is repeated across all bootstraps of the index and change-points are considered to occur where the 95% uncertainty interval of the 2nd derivative omits zero. Change points above zero indicate change to a more positive slope (open circles in figure).

Years testing over	three years (change between two first derivatives)
Years contributing to the assessment	Variable depending on smoothing
Likelihood of false positives	Low when change is zero
Likelihood of false negatives	High

Comments: Change points vary considerably with the level of smoothing and only abrupt changes are detected.

Option 11. Break-point regression: Assessing whether simulated future trends can be detected in relation to historic change [20]

9	Simplifying step	iii) – Linear model over >1 time period
	Test	C - Break-point model separate intercepts

Description: Data were simulated illustrating different rates of historical decline and levels of recovery based on the Living Planet Index database and linear model (Index ~ time * period) with a range of putative change points used to test whether different types of future trend could be reliably identified at a range of future time points.

Years testing over	Minimum of five years historic data compare to different future five-year periods
Years contributing to the assessment	10+
Likelihood of false positives	Low when change is zero
Likelihood of false negatives	High

Comments: This method was able to detect future trajectories reliably after ten years if there was a substantial change between the historic trend and projected future rate of change.

Option 12. Break-point regression: Does the slope differ before and after change point a decade prior to final year? [21]

10	Simplifying step	iii) – Linear model over >1 time period
	Test	C – Break-point model with common intercept

Description: A break-point regression was run on each bootstrap on the Index (Index ~ time + time:period). A change point was considered present when the 95% uncertainty interval of the interaction estimated omitted zero. The model incorporated an autocorrelation term.

Years testing over	50 years
Years contributing to the assessment	50 years
Likelihood of false positives	Low when change is zero
Likelihood of false negatives	High

Comments: Since the breakpoint regression is conducted on an unsmoothed indicator the test is sensitive to whether the breakpoint year happens to be a particularly low or high point within the interannual variation.

Option 13. Break-point regression: Difference in slope of linear models for two time periods [6]

11	Simplifying step	iii) – Linear model over >1 time period
	Test	C – Difference of two linear models

Description: The MSI tool runs a linear model (Index ~ time) across each bootstrap of the index separately before and after a specified year and estimates the difference in slope per bootstrap. A change point is considered when the 95% uncertainty interval of the difference in slope omits zero.

Years testing over	Variable, but often at least ten before and after a change point
Years contributing to the assessment	All within the time-series, usually several decades
Likelihood of false positives	Low when change is zero
Likelihood of false negatives	High

Comments: Incorporates uncertainty in index values via bootstrapping procedure.

Option 14 Progressive partial regression [22]

12	Simplifying step	iii) – Linear model over >1 time period
	Test	C

Description: Attempts to ascertain values of the dependent variable over which the slope of a regression does not differ from zero.

Years testing over	Variable
Years contributing to the assessment	Variable
Likelihood of false positives	Low when change is zero
Likelihood of false negatives	Low

Comments: The mechanics of this technique are unclear in the documentation available. It seems there would be a high chance of concluding 'stability' where there was ongoing decline. It is designed for use where, for example, a nutrient limits crop yield up until a threshold value of the nutrient above which no further increase in yield occurs.

Option 15. Anomalies versus change points [12]

1	Simplifying step	i) – None
	Test	D

Description: Uses an unsupervised machine learning model to try to differentiate between 'anomalies' (an unusual data point(s), but where the underlying trend or data distribution doesn't shift) and 'change-points' (where changes to data distribution persist downstream). Since an anomaly does not change the data distribution, if an anomaly occurs in the time series, the time series data must continue to follow the same data distribution as before the anomaly occurred for a certain period of time. Otherwise, the data distribution is considered to have changed and this is a change point.

Years testing over	Assessing each year
Years contributing to the assessment	Requires a long time-series to train the model and datapoints surrounding the assessments period to advise categorisation.
Likelihood of false positives	Medium when change is zero
Likelihood of false negatives	Medium

Comments: This paper is focussed on anomaly detection, rather than change-point. It also relies on a lot of data to build the classifiers and is likely more suited to time-series where the frequency of data points is higher than annual. There are likely similar approaches that would be worth investigation.

Option 16. Piecewise regression of smoothed indicator [1]

13	Simplifying step	iii) – Linear model over >1 time period (following GAM)
	Test	D

Description: A piecewise linear model was run on a smoothed indicator (Index ~ time * period), with each period representing a decade. The model was run across each bootstrap of the indicator and the estimate and associated CI for each decadal slope were extracted and compared to the average slope estimate across decades to test whether each decadal trend was outside of the full population of changes. Used in the State of Nature report 2023.

Years testing over	50 years
Years contributing to the assessment	50
Likelihood of false positives	Low when change is zero
Likelihood of false negatives	High

Comments: Requires a long time-span of data

References

1. Burns, F., et al., *State of nature 2023*. 2023. Available at: www.stateofnature.org.uk
2. Brondizio, E.S., et al., *Global assessment report on biodiversity and ecosystem services of the Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services*. 2019. Available at: https://www.ipbes.net/
3. UK Government, *Environment Act 2021*. 2021. Available at: https://www.legislation.gov.uk/ukpga/2021/30/contents
4. UNEP, *DECISION ADOPTED BY THE CONFERENCE OF THE PARTIES TO THE CONVENTION ON BIOLOGICAL DIVERSITY. 15/4. Kunming-Montreal Global Biodiversity Framework*, UNEP, Editor. 2022. Available at: https://www.cbd.int/doc/decisions/cop-15/cop-15-dec-04-en.pdf
5. McRae, L., S. Deinet, and R. Freeman, *The diversity-weighted living planet index: controlling for taxonomic bias in a global biodiversity indicator.* PloS one, 2017. **12**(1): p. e0169156.
6. Soldaat, L.L., et al., *A Monte Carlo method to account for sampling error in multi-species indicators.* Ecological Indicators, 2017. **81**: p. 340-347.
7. UK Government, *The Environmental Targets (Biodiversity) (England) Regulations 2023, Statutory Instruments 2023 No. 91*. 2023. Available at: https://www.legislation.gov.uk/uksi/2023/91/made
8. Defra, *Indicators of species abundance in England*. 2025. Available at: https://www.gov.uk/government/statistics/indicators-of-species-abundance-in-england/indicators-of-species-abundance-in-england
9. Freeman, S.N., et al., *A generic method for estimating and smoothing multispecies biodiversity indicators using intermittent data.* Journal of Agricultural, Biological and Environmental Statistics, 2021. **26**: p. 71-89.
10. Johnson, T., et al., *Revealing uncertainty in the status of biodiversity change.* Nature, 2024. **628**(8009): p. 788-794.
11. Bane, M.S., et al., *An evidence-base for developing ambitious yet realistic national biodiversity targets.* Conservation Science and Practice, 2023. **5**(2): p. e12862.
12. Liu, J., et al., *Anomaly and change point detection for time series with concept drift.* World Wide Web, 2023. **26**(5): p. 3229-3252.
13. Atwood, M., et al., *Aiming higher to test a bend in the curve of biodiversity loss.* Ecology Letters, submitted.
14. Gregory, R.D., D. G. Noble, and J. Custance, *The state of play of farmland birds: population trends and conservation status of lowland farmland birds in the United Kingdom.* Ibis, 2004. **146**: p. 1-13.
15. Defra, *Wild bird populations in the UK and England, 1970 to 2023: Accredited official statistics*. 2025: https://www.gov.uk/government/statistics/wild-bird-populations-in-the-uk/wild-bird-populations-in-the-uk-and-england-1970-to-2023. Available at:
16. Fewster, R.M., et al., *Analysis of population trends for farmland birds using generalized additive models.* Ecology, 2000. **81**(7): p. 1970-1984.
17. Leung, B. and A. Gonzalez, *Global monitoring for biodiversity: uncertainty, risk, and power analyses to support trend change detection.* Science Advances, 2024. **10**(7): p. eadj1448.
18. Schmidt, B.R. and A.H. Meyer, *On the analysis of monitoring data: Testing for no trend in population size.* Journal for Nature Conservation, 2008. **16**(3): p. 157-163.
19. Dixon, P.M. and J.H. Pechmann, *A statistical test to show negligible trend.* Ecology, 2005. **86**(7): p. 1751-1756.
20. McRae, L., et al., *The utility of the Living Planet Index as a policy tool and for measuring nature recovery.* Philosophical Transactions B, 2025. **380**(1917): p. 20230207.
21. Burns, F., et al., *An assessment of the state of nature in the United Kingdom: a review of findings, methods and impact.* Ecological indicators, 2018. **94**: p. 226-236.

22. Oosterbaan, R., *Free software for the division of the relation between Y (response variable) and X (causal variable) into segments or ranges where X is influential and where X has no effect*. 2021. Available at: https://www.waterlog.info
23. Makowski, D., et al., *Indices of effect existence and significance in the Bayesian framework.* Frontiers in psychology, 2019. **10**: p. 2767.
24. Talis, E.J., et al., *Penguindex: a Living Planet Index for Pygoscelis species penguins identifies key eras of population change.* Polar Biology, 2023. **46**(8): p. 707-718.
25. Pastore, M. and A. Calcagnì, *Measuring distribution similarities between samples: A distribution-free overlapping index.* Frontiers in psychology, 2019. **10**: p. 1089.
26. Eaton, M.A., et al., *The priority species indicator: measuring the trends in threatened species in the UK.* Biodiversity, 2015. **16**(2-3): p. 108-119.