Confidence in trends and assessments based upon the UK and England Biodiversity Indicators

Peer-review by the ad-hoc Science Panel quality assuring the UK and England Species Trends

Professor Ken Norris, Institute of Zoology (Panel Chair)
Professor Stephen Buckland, University of St Andrews
Professor Rhys Green, University of Cambridge
Professor Helen Roy, Centre for Ecology & Hydrology
Dr Phil Stephens, University of Durham

January 2016

Background

- The Biodiversity Indicators Working Group devised a method for assigning a level of confidence in the trend and assessment of the trend for each individual biodiversity indicator. The method and confidence assessments were published in draft in 2014 (http://jncc.defra.gov.uk/pdf/biyp2014_ConfidenceStatements_final.pdf) subject to peer review by the ad hoc Science Panel peer reviewing the species trends.
- The method is based on a scoring system that assesses various aspects of the underlying data on which each indicator is based and the rigour with which trends in the data over time are assessed.
- Scores are then combined into a rating of each of these areas (high, medium or low) and an overall confidence rating for each indicator (high, medium or low).
- This short report provides the conclusions of the peer review by the ad hoc Science Panel.

Comments

- While we recognise the potential value in assigning confidence to indicators, we have significant concerns about the existing methodology.
- Our comments follow the structure of the confidence assessment.

Stage 1: Data Collection:

- Unless an indicator is based on complete, accurate counts of every species involved it will inevitably be unrepresentative and biased to some extent. This means that virtually all the indicators should receive a 'low' rating for 'Data Coverage', and hence a 'low' confidence rating overall according to the methodology. In reality, it is important to assess how unrepresentative and biased the data are likely to be. At the moment, the confidence assessment methodology makes no provision to do this. We regard this as a critically important issue.
- Representativeness and bias are strongly dependent on the sampling design (termed 'Method for Data Collection' within the methodology) and the consistency with which the sampling design has been applied across time and domains e.g. different species groups within an indicator (termed 'Comparability' within the methodology). We note that even a stratified, random sampling design does not completely preclude issues of representativeness and bias. For example, the Breeding Bird Survey (BBS) fails to adequately survey certain species groups (e.g. nocturnal species) and includes potentially significant biases (e.g. changes in detection over time). The important point is that there is a hierarchical relationship between components of the confidence assessment methodology –

a poor and inconsistent sampling design increases unrepresentativeness and bias; a rigorous sampling design reduces these issues. The methodology at present does not explicitly recognise these dependencies.

Stage 1: Rigour of assessment of change:

- Known biases in the data can sometimes be addressed in the data analysis thereby reducing the impact of the biases on the trend assessment. At the moment, no assessment is made within the methodology of the extent to which bias is being addressed by the data analysis. We see this as a key issue.
- We are concerned by the rather vague language used in parts of this section. For example, "Uncertainty is quantifiable and a statistical test is used to assess change" says nothing about how uncertainty is quantified and how a trend in the data is statistically modelled. Generating bootstrapped confidence intervals based on 'sites' or 'species' make rather different assumptions; different statistical modelling frameworks and model structures would be appropriate for different data types and biases. Furthermore, the methodology makes no mention of the degree of uncertainty involved. Clearly, we should be more confident in a trend the less uncertainty there is in the trend.

General:

- A number of the indicators are composite in that they combine very different data, collected using an array of different methods and hence contain a range of representativeness and bias issues. For example, the bats indicator (C8a) consists of multiple datasets, including surveys of known roost and hibernation sites that are unrepresentative and introduce significant observer bias into the trend analysis. It is unclear how such variation in indicator components are treated by the assessment, and worryingly in the case of bats the indicator received a 'high' confidence assessment even though components of the indicator should score 'low' on the 'Data Coverage' criterion. At the moment, it is unclear how composite indicators are being treated with respect to the methodology.

<u>Conclusions</u>

- We see little merit in continuing with the methodology in its present form – it fails to consider representativeness and bias properly in terms of data collection or analysis; the aggregate scoring system fails to recognise the dependencies between the different components of the assessment framework; and there is a lack of clarity in how parts of the framework are worded or applied particularly around trend estimation and associated uncertainty.
- Rather than work with confidence statements we suggest that the Biodiversity Indicators Working Group develops instead evidence standards that relate to:
    - Representativeness
        - How representative is the indicator of the wider community?
    - Bias
        - Is there significant observation, reporting, detection or geographical bias in the survey data on which the indicator is based?
    - Statistical modelling for trend analysis and bias reduction
        - Are trends estimated using an appropriate statistical modelling framework, and are significant biases addressed by the analysis so that the trends have low bias?

    - Uncertainty
        - Is uncertainty estimated using appropriate methods and are levels of uncertainty within acceptable limits?
- In practice, it is very difficult to weight the interpretation of an indicator by a confidence measure. Either the indicator is robust enough to interpret trends in it or it is not. A robust indicator should have high representativeness, low bias, trends estimated using an appropriate statistical modelling framework that reduces any bias, and low levels of uncertainty in the trends. Establishing evidence standards would enable indicators to be assessed, and steps taken to improve indicators that fall below acceptable standards.