



Simple methods for improving the communication of uncertainty in species' temporal trends

O.L. Pescott^{a,*}, P.A. Stroh^b, T.A. Humphrey^c, K.J. Walker^d

^a UKCEH Wallingford, Benson Lane, Crowmarsh Gifford, OX10 8BB, UK

^b Botanical Society of Britain and Ireland, c/o Cambridge University Botanic Garden, CB2 1JE, UK

^c Botanical Society of Britain and Ireland, c/o UKCEH Wallingford, OX10 8BB, UK

^d Botanical Society of Britain and Ireland, Harrogate, UK

ARTICLE INFO

Keywords:

Information visualisation
Environmental monitoring
Risk-of-bias
Line ensembles
Model-based inference
Stakeholder communication
Bootstrap
Bayesian model

ABSTRACT

Temporal trends in species occupancy or abundance are a fundamental source of information for ecology and conservation. Model-based uncertainty in these trends is often communicated as frequentist confidence or Bayesian credible intervals, however, these are often misinterpreted in various ways, even by scientists. Research from the science of information visualisation indicates that line ensemble approaches that depict multiple outcomes compatible with a fitted model or data may be superior for the clear communication of model-based uncertainty. The discretisation of continuous probability information into frequency bins has also been shown to be useful for communicating with non-specialists. We present a simple and widely applicable approach that combines these two ideas, and which can be used to clearly communicate model-based uncertainty in species trends (or composite indicators) to stakeholders. We also show how broader ontological uncertainty can be communicated via trend plots using risk-of-bias visualisation approaches developed in other disciplines. The techniques are demonstrated using the example of long-term plant distributional change in Britain, but are applicable to any temporal data consisting of averages and associated uncertainty measures. Our approach supports calls for full transparency in the scientific process by clearly displaying the multiple sources of uncertainty that can be estimated by researchers.

1. Introduction

The monitoring of trends in species' distributions or populations is a fundamental activity within ecology and conservation (Lindenmayer and Likens, 2010). The resulting trends may have different uses depending on the rationale and design of the underlying monitoring program, but much "surveillance"-style monitoring is driven by both policy requirements and the curiosity of invested naturalists (Pescott et al., 2015; Schmeller et al., 2009). This means that feedback on trends to non-scientist stakeholders of various types is often a key program output. Species-level trends also form the basis of various multi-species composite indicators (e.g. van Strien et al., 2016). The literature on these has emphasised the importance of mathematical aspects of their construction (e.g. Lamb et al., 2009), including the development of methods for the propagation of model-based uncertainty from the species level to the multi-species trend line (Soldaat et al., 2017). Indeed, the accurate and full communication of uncertainty is now widely

considered to be fundamental for the development and maintenance of trust between scientists and the wider public (Fischhoff and Davis, 2014; Spiegelhalter, 2017), and considerable effort has been invested by information visualisation scientists in how best to achieve widespread understanding of technical scientific results (e.g. see the review of Padilla et al., 2022).

A standard approach to the visualisation of uncertainty in temporal trends is the use of frequentist confidence or Bayesian credible intervals to produce error ribbons or bands. Arguably, however, these are merely defaults (Gelman, 2014), and these types of presentations have not, to our knowledge, been critically examined within ecology in terms of whether they can be improved for the clear communication of uncertainty to stakeholders. Reviewing similar types of statistical visualisation based on conventional error bar types, Padilla et al. (2022) point to evidence that these can lead to misinterpretations of uncertainty, such as viewers assuming that points outside of error bars are impossible. Continuous probability information is mis-construed as categorical and

* Corresponding author.

E-mail address: olipes@ceh.ac.uk (O.L. Pescott).

<https://doi.org/10.1016/j.ecolind.2022.109117>

Received 3 May 2022; Received in revised form 22 June 2022; Accepted 26 June 2022

Available online 30 June 2022

1470-160X/© 2022 The Author(s). Published by Elsevier Ltd. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

deterministic. This is perhaps not surprising given that even researchers have trouble interpreting the information content of these conventions (Belia et al., 2005; Greenland et al., 2016; Hoekstra et al., 2014), and that the statistical meaning of similar graphics may vary between presentations (e.g. whether standard errors, confidence intervals, bootstrapped intervals etc.) When these practices are extended, as for a regression line presented with a continuous error ribbon, then additional interpretational issues, such as the potential for trends that may be in directional conflict with the average trend, may also present themselves (for examples, see Kay, 2021). Researchers have also found that the use of different graphical “marks” (e.g. types of line) to distinguish between average expectations and uncertainty in these, such as is common in the presentation of species’ trends and indicators (e.g. van Strien et al., 2016), can result in a bias of attention towards the expected value and away from its associated uncertainty (Hullman et al., 2015).

In the search for better visualisations, many different types of statistical and graphical strategies have been investigated (Padilla et al., 2022). These include ways of illustrating the variety of outcomes that are compatible with a fitted model or data, rather than just easily misinterpreted summary statistics (Greenland et al., 2016; Kale et al., 2019). Different graphical marks and “encodings” (e.g. colour and transparency) have also been widely explored. Whilst it is generally appreciated that it is unlikely that there is any one single, universal best practice for communicating uncertainty to viewers (Padilla et al., 2022), arguably enough experimental evidence has accumulated to indicate opportunities for improving practice in ecology. For example, the use of line ensembles, e.g. from multiple model fits derived from bootstrapping or Bayesian posteriors, that visualise the actual distribution of compatible outcomes, may offer “a more interpretable rendering of uncertainty [...], especially when viewers are unlikely to have statistical training” (Kale et al., 2019).

We introduce a simple method for communicating uncertainty in regression fits for species’ temporal trends. The approach presented here is based on bootstrapped linear regression line ensemble plots, combined with a frequency-based discretised summary of the ensemble slopes. It could also easily be applied to the posterior distribution of the slope parameter from a single Bayesian linear model. Whilst we use the example of ordinary linear regression here for simplicity, the line ensemble plot idea can be applied to many other types of linear model, such as those using link functions and/or random effects (Kay, 2021). In some of these cases, however, more thought would be required for the discretised visualisation: for example, for generalised linear models, discretisations of parameter distributions would likely be more interpretable on the original scale, rather than on that of the link function (Gelman and Hill, 2007).

We argue that the visualisation of multiple outcomes compatible with our model/data combination, combined with a discretised summary of these, clearly demonstrates model-based uncertainty in complementary ways, with the discretisation providing a frequency-based presentation that is likely to be more easily understood by non-specialist viewers (Hullman et al., 2018). We also demonstrate how broader ontological uncertainty (Spiegelhalter, 2017)—i.e. non-model based uncertainty—can be included in such plots, acknowledging that model-based uncertainty alone can be very misleading for model/data combinations with a high risk-of-bias (Boyd et al., 2022; Greenland, 2017; van der Bles et al., 2019).

2. Methods and results

2.1. Case study

Here we use plant distribution data collected by the Botanical Society of Britain and Ireland (BSBI) to demonstrate our approach. The frequency scaling using local occupancy method (“Frescalo”; Hill, 2012; Pescott et al., 2019) is used to produce temporal relative occupancy estimates for each species (see Supplementary data 1). The uncertainty

visualisation method developed here, however, is sufficiently general to be applied to any dataset or model that can be made to yield averages and associated measures of uncertainty per time period (cf. Soldaat et al., 2017). The four example species used here are *Allium vineale* L., *Hornungia petraea* (L.) Rchb., *Hypochaeris maculata* L., and *Parnassia palustris* L. (names follow Stace, 2019), and were chosen to provide different temporal trends and levels of uncertainty.

2.2. Monte Carlo simulation bootstrapping and trend classification

For a given species, 100 simulated relative occupancy estimates were drawn for each of the four time periods based on their Frescalo-estimated means and standard deviations. For each set of four estimates, a linear regression fit was calculated. Line ensemble plots providing the 100 simulated linear regression fits for each species are given in Fig. 1, along with the original means and standard deviations from Frescalo. Density plots showing the distribution of the 100 linear regression slope estimates for each species are given in Fig. 2, along with the cut-points for our discretisation scheme. For this example, the cut-points shown were developed by the authors based on temporal trends estimated for around 1,700 taxa modelled. The result was a five-class scheme, with category labels: strong decline (–), moderate decline (–), stable (0), moderate increase (+), and strong increase (++)). The 100 simulated slope estimates for each species were classified based on these cut-points, and are displayed as frequency bar charts in Fig. 3. A link to the R code and data is in Supplementary data 2. Ultimately any scheme of cuts could be used, and these could be specified and labelled however is thought best for the data, model, and communication aims. Research in this area suggests that discretisations based on fewer categories can lead to more consistent viewer estimates of the underlying probability of events, as compared to having more categories that begin to visually approach continuous displays such as density plots (Kay et al., 2016).

2.3. Broader ontological uncertainty

An additional species, *Potamogeton polygonifolius* Pourr., was chosen to demonstrate the fact that model-based uncertainty alone can often be highly misleading, particularly where observational data with potentially serious biases are being used (Boyd et al., 2022; Greenland, 2017). For this species, the Frescalo estimates have low uncertainty, and suggest that an increase in the species’ 10 km distribution over the last one-hundred years is strongly supported. However, the authors of the current paper assessed this conclusion to have a high risk-of-bias (Boyd et al., 2022), due to external knowledge of how this species was treated by plant recorders in the first time period (1930–69; Braithwaite et al., 2006). Risk-of-bias tools typically consist of a set of “domains” against which expert judgement is used to come to some evidence-supported conclusion on the potential for bias within a study. For example, within medical research, risk-of-bias tools exist for assessing randomised controlled trials against domains which are known to have the potential to cause bias in the causal estimand of interest (McGuinness & Higgins, 2021). An example is the strength of the randomisation mechanism used to assign patients to treatment or control groups (Higgins and Altman, 2008). Until very recently such tools were unknown within ecology and evolution; within the past year, however, tools for assessing the risk-of-bias within causal inference focused experiments (i.e. “internal validity”; Konno et al., 2021), and for assessing the risk-of-bias of studies focused on broader descriptive inference (i.e. “external validity”; Boyd et al., 2022) have been published. General guidelines for producing such tools within the area of environmental management have also been produced (Frampton et al., 2022). Here we used the “Risk-of-Bias in Temporal Trends in ecology” (ROBITT) tool of Boyd et al. (2022) to assess the exemplar temporal trend for *P. polygonifolius* against risk-of-bias domains relevant to the task of biodiversity-focused descriptive inference. Briefly, these are geographic, environmental, taxonomic, and

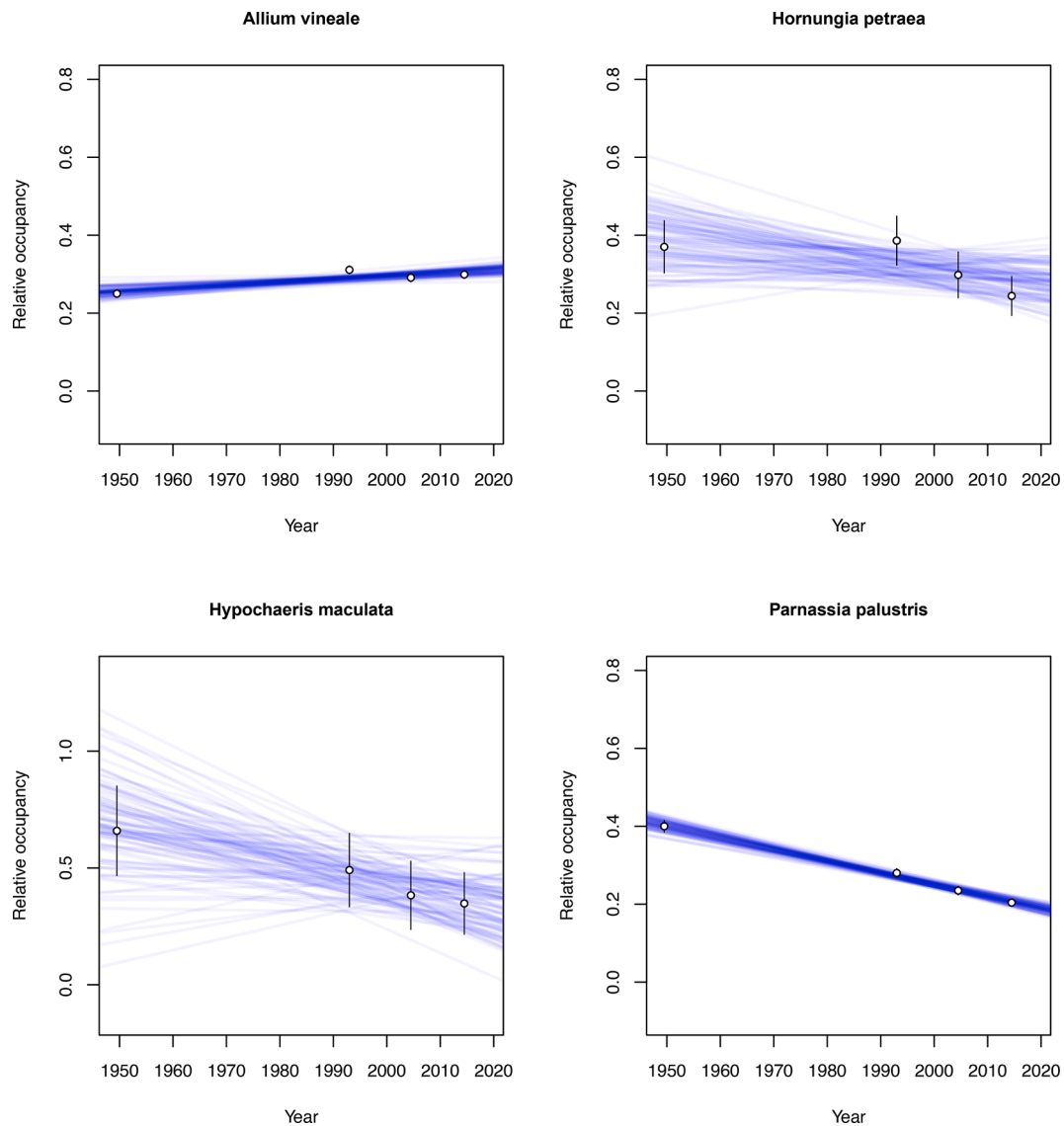


Fig. 1. Temporal trend line ensemble plots for four plant species. In each case 100 linear regression fits to Monte Carlo-simulated data are given; transparent lines are used in order to further communicate model-based certainty. The filled white points and black bars are the Frescalo means and standard deviations for each time period, plotted at the median of each date-class. Note the different y-axis scale for the species (*Hypochaeris maculata*) with the less certain relative occupancy estimates.

“other” biases; more detail on these, including a guidance document, can be found in [Boyd et al. \(2022\)](#). Within the ROBITT structure, the bias identified in the current example fits best into the “other” category, as we identified a systematic temporal bias in an aspect of the observation process giving rise to our data. We consider this to result in a high potential risk-of-bias given previous commentaries on this case ([Braithwaite et al., 2006](#)). We have therefore added this information as a risk-of-bias bar ([McGuinness & Higgins, 2021](#)) to the plot to alert the viewer ([Fischhoff and Davis, 2014](#); [van der Bles et al., 2019](#)). Note that in existing applications of this method within the medical sciences, visualisations summarise assessments of the risk-of-bias in domains across the studies included in a systematic review. Here, we apply the approach to a single study. Our overall summary assessment is based on a “weakest link” approach across ROBITT domains—that is to say, the highest risk-of-bias assessed is the summary conclusion displayed ([Fig. 4](#)). Weighted options are also possible ([McGuinness & Higgins, 2021](#)), and existing advice states that users should make it clear how results are summarised into an overall risk-of-bias assessment, irrespective of the approach ([Frampton et al., 2022](#)).

3. Discussion

Understanding uncertainty is a fundamental part of science, but uncertainty itself is often poorly communicated by scientists ([Greenland, 2017](#); [Hullman, 2020](#)). The subject is complicated by the many types of uncertainty that researchers encounter ([Regan et al., 2002](#)), and by the fact that subtle statistical and philosophical concepts overlay scientists’ attempts to characterise reality from samples ([Rafi and Greenland, 2020](#); [Spiegelhalter, 2017](#)). Whilst here we mainly deal with the communication of uncertainties that are conditional on the chosen model, as opposed to those that relate to the internal or external validities of chosen models ([Boyd et al., 2022](#)), research suggests that even this aspect of scientific communication can be improved ([Hullman et al., 2015](#)), particularly where non-scientist stakeholders are the target audience ([van der Bles et al., 2019](#)). Techniques have been developed for propagating error from species-level models to composite indicators (e. g. [Soldaat et al., 2017](#)), but within ecology there has been little consideration of alternative techniques for the visual communication of trend uncertainty, outside of simply presenting error ribbons around an

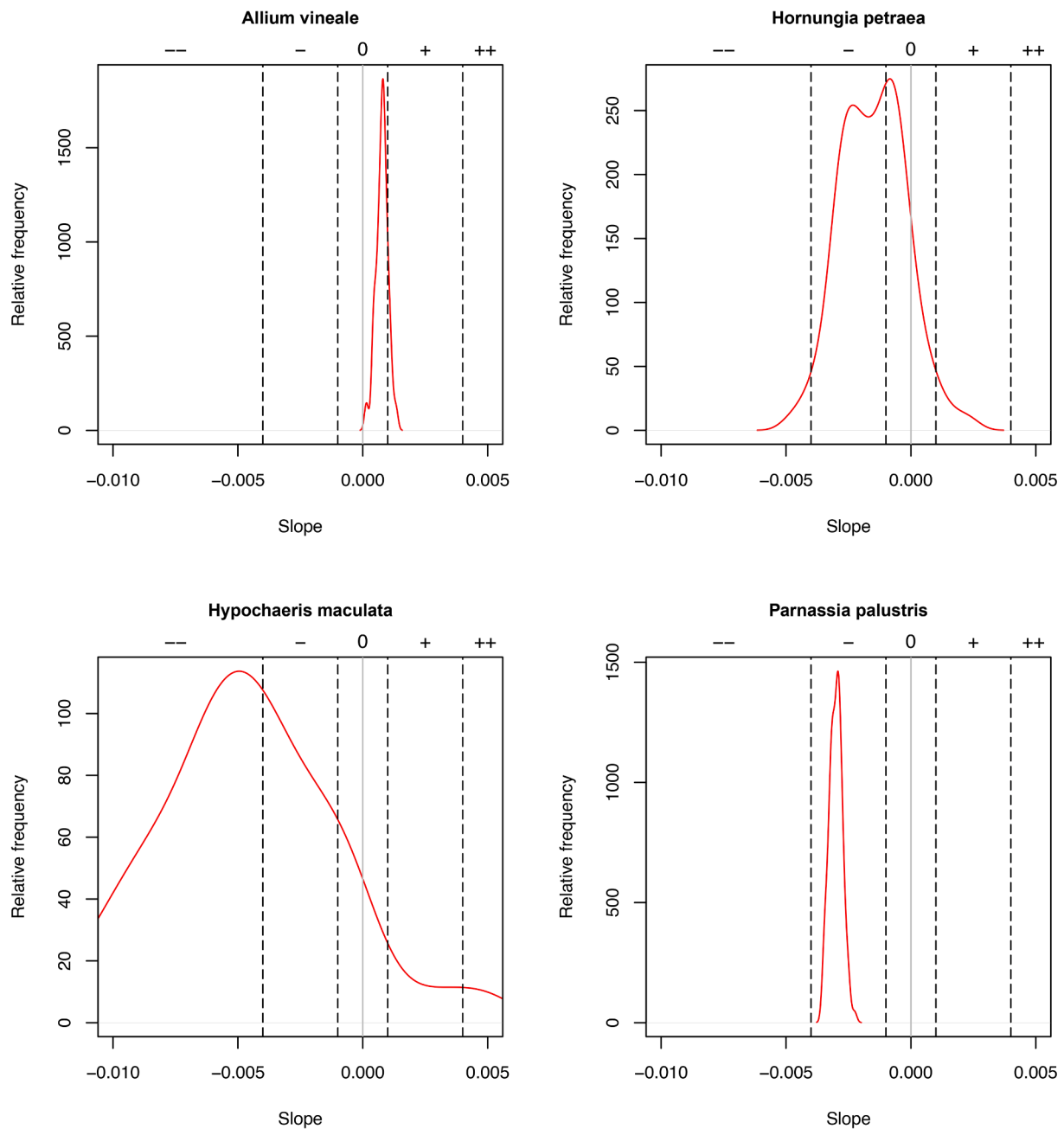


Fig. 2. Density plots for the 100 simulated linear regression slope estimates for each species. The black vertical broken lines indicate the cut-points used; a grey vertical solid line is plotted at zero. The trend categories used in this case are given along the top of the plots as: – (strong decline); - (moderate decline); 0 (stable); + (moderate increase); and, ++ (strong increase).

average trend.

Research within information visualisation science suggests that the use of “visual boundaries” (e.g. error ribbons) can be a useful technique (Padilla et al., 2022); however, ribbons can also serve to emphasise the slope of the average trend, rather than indicating all the possible trajectories that are compatible with a fitted model (cf. Fig. 1). The development of static line ensembles and dynamic hypothetical outcome plots (i.e. animations of outcomes compatible with a model; Hullman et al., 2015) has sought to overcome this limitation. For example, the psychologist John Kruschke presented a technique for visualising ensembles of linear regression posterior fits within the first edition of his book on Bayesian methods (Kruschke, 2011). More recently, Kay (2021) released an R package, “tidybayes”, that includes functions for the creation of both ensemble and hypothetical outcome plots from parameter

posterior distributions estimated using the Hamiltonian Monte Carlo-based Bayesian modelling framework Stan. Such technical developments, coupled with empirical explorations of the experienced information content of such displays by user groups (Kay et al., 2016; Kale et al., 2018), suggests that their use is likely to increase in the coming years.

Whilst much of the work on ensemble plots has been within a Bayesian framework, the principle can be applied to any model parameter for which probabilistic outcomes can be generated, either via parametric or non-parametric methods (Padilla et al., 2022). Here we used a Monte Carlo simulation-based approach to produce bootstrapped linear models to propagate uncertainty from an earlier analysis yielding time period-specific relative occupancy mean and standard deviation estimates (Hill, 2012). Such ensembles contain more information than a

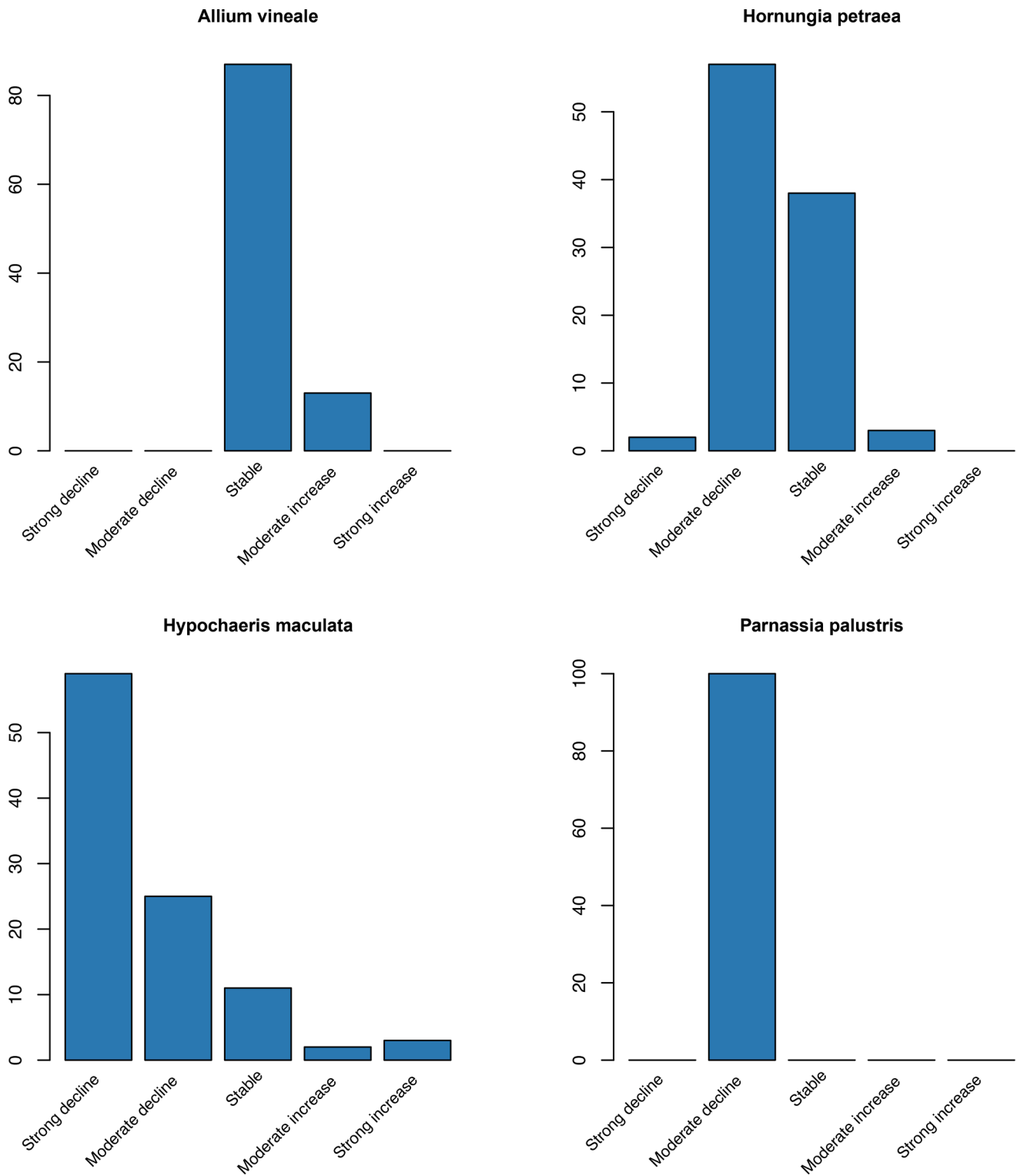


Fig. 3. Discretised slope magnitude frequency plots based on the distribution of the 100 simulated linear regression slope estimates shown in Fig. 2.

frequentist confidence interval (better termed a “compatibility” interval; Amrhein and Greenland, 2022; Rafi and Greenland, 2020) or a Bayesian credible interval (even if displayed with multiple percentile bands), as they clearly visualise the range of possible outcomes that are compatible with a fitted model. However, ensembles still communicate information in the visual and numerical terms of the statistical model used, and this places a burden on the viewer to translate model-based expectations into verbal understanding. In some cases, but particularly for those where

non-scientist stakeholders are an important target audience, we suggest that a simple classification of this uncertainty will make the information transmitted by ensembles easier to understand (cf. Hullman et al., 2018). Indeed, whilst writing this paper, we discovered that educators in psychology have demonstrated benefits of discretising continuous probability information into frequency formats when teaching Bayesian reasoning (Gigerenzer and Hoffrage, 1995; Sedlmeier and Gigerenzer, 2001).

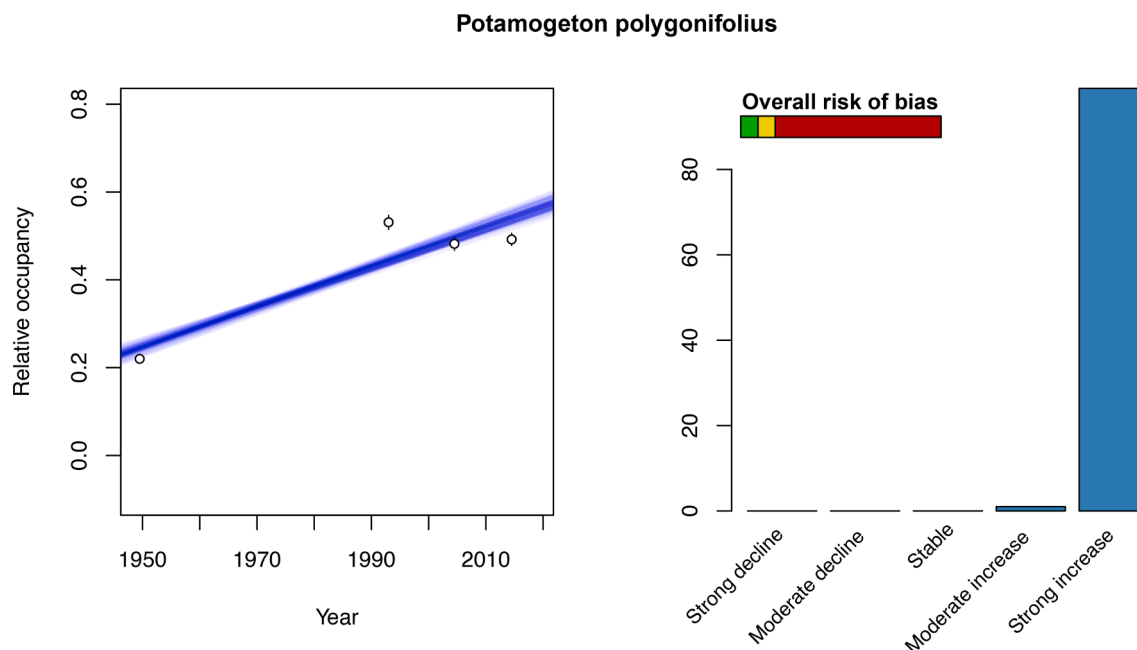


Fig. 4. Line ensemble and discretised slope magnitude frequency plots for *Potamogeton polygonifolius*. Here, a “risk-of-bias” visualisation bar has been added to the discretised frequency plot to emphasise the presence of high non-model-based uncertainty (McGuinness & Higgins, 2021), where green = “Low risk”, yellow = “Some concerns”, and red = “High risk”. Risk levels were assessed using a version of the ROBITT scheme of Boyd et al. (2022), and the overall high risk evaluation relates to a strong expert belief in important variation in how the species was identified by recorders over the time period considered (Braithwaite et al., 2006).

We recognise, and indeed emphasise, that model-based uncertainty is only one aspect of the overall uncertainty associated with statistical inference (Rafi and Greenland, 2020; Regan et al., 2002; Spiegelhalter, 2017). Multiple models of reality may fit data equally well by some metric, but provide different conclusions (Copas and Eguchi, 2020; Steegen et al., 2016); samples may also lack external validity (i.e. be unrepresentative of the statistical target population; Boyd et al., 2022). Model-based uncertainty is uncertainty conditional on a chosen model (or multiple models, for model-based averaging approaches) combined with a dataset, and may actually still miss the true parameter at which science aims. This is a wider issue, and, at least for the description of species’ trends or composite indicators based on these, relates to the numerous steps between the observation of a species in the field and the creation of some statistical model to estimate a temporal trend (Boyd et al., 2021). Fully accounting for, and clearly communicating, this broader uncertainty is a much larger project, and research in this area continues to develop. Current areas that are developing rapidly include techniques designed to visualise the effects of “forking paths” (Gelman and Loken, 2014) in research (Liu et al., 2021), frameworks for visually communicating risk-of-bias effectively (McGuinness and Higgins, 2021), and the body of work on the visualisation of multi-model ensemble outcomes. The latter has hitherto largely been the preserve of those working with complex, process-based, numerical simulations, e.g. climate, weather, and fisheries stock modellers (Potter et al., 2009).

For the broader trend creation exercise used here as a case study, we have found species where the model-based uncertainty is low, but for which the estimated trend is considered unlikely by taxon group experts. For example, the temporal trend for Bog Pondweed (*P. polygonifolius*; Fig. 4) suggests an increase in relative occupancy over the period modelled. However, expert opinion has previously considered that this is likely to be an artifact of changes in recorders’ approaches to the identification of this species in Britain over the twentieth century, and we agree with this assessment. This is a case of low model-based uncertainty coupled with an expert-assessed high risk-of-bias. The current distribution atlas project of the BSBI (Walker et al., 2010) is therefore also considering the use of an expert-assessed risk-of-bias classification (McGuinness and Higgins, 2021) to present alongside a discretised line

ensemble approach (Fig. 4).

Accurately communicating the full uncertainty in species’ temporal trends, or indicators based on these, is a complex matter that has arguably not been well addressed by the ecological literature to date. There is, however, much to learn from other disciplines, both in terms of visualisation technique (Padilla et al., 2022), and in terms of careful thought about the assumptions underlying typical statistical practice in our field (Boyd et al., 2022; Greenland, 2021, 2017; Rafi and Greenland, 2020). Despite the challenges, we believe that the clear communication of as much of the estimable uncertainty as possible is the most ethical and honest way forward for science in terms of how it relays its findings to the rest of society (Fischhoff, 2012; Spiegelhalter, 2017; van der Bles et al., 2019).

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Link to data/code in [Supplementary data 2](#)

Acknowledgements

Our sincere thanks to all the recorders who have contributed to the BSBI plant distribution dataset underlying the case study used in this paper. Thanks also to Rob Boyd, Rob Cooke, and David Roy for comments on the manuscript. The constructive comments of two anonymous reviewers also improved the paper. The contribution of OLP to this work was supported by the Natural Environment Research Council award number NE/R016429/1 as part of the UKCEH UK-SCAPE programme delivering National Capability.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.ecolind.2022.109117>.

References

- Amrhein, V., Greenland, S., 2022. Rewriting results in the language of compatibility. *Trends Ecol. Evol.* 37 (7), 567–568.
- Belia, S., Fidler, F., Williams, J., Cumming, G., 2005. Researchers misunderstand confidence intervals and standard error bars. *Psychol. Methods* 10, 389–396. <https://doi.org/10.1037/1082-989X.10.4.389>.
- Boyd, R.J., Powney, G.D., Burns, F., Danet, A., Duchenne, F., Grainger, M.J., Jarvis, S.G., Martin, G., Nilsen, E.B., Porcher, E., Stewart, G.B., Wilson, O.J., Pescott, O.L., 2022. ROBITT: A tool for assessing the risk-of-bias in studies of temporal trends in ecology. *Methods Ecol. Evol.* <https://doi.org/10.1111/2041-210X.13857>.
- Boyd, R.J., Powney, G.D., Carvell, C., Pescott, O.L., 2021. occAssess: An R package for assessing potential biases in species occurrence data. *Ecol. Evol.* 11, 16177–16187. <https://doi.org/10.1002/ece3.8299>.
- Braithwaite, M.E., Ellis, R.W., Preston, C.D., 2006. Change in the British flora 1987–2004. Botanical Society of the British Isles, London.
- Copas, J., Eguchi, S., 2020. Strong model dependence in statistical analysis: goodness of fit is not enough for model choice. *Ann. Inst. Stat. Math.* 72, 329–352. <https://doi.org/10.1007/s10463-018-0691-8>.
- Fischhoff, B., 2012. Communicating Uncertainty: Fulfilling the Duty to Inform. *Issues in Science and Technology* 28, No. 4 (summer 2012). <https://issues.org/fischhoff/#.YrL2WRJQ-7s.link>.
- Fischhoff, B., Davis, A.L., 2014. Communicating scientific uncertainty. *Proc. Natl. Acad. Sci.* 111, 13664–13671. <https://doi.org/10.1073/pnas.1317504111>.
- Frampton, G., Whaley, P., Bennett, M., et al., 2022. Principles and framework for assessing the risk of bias for studies included in comparative quantitative environmental systematic reviews. *Environ Evid* 11, 12. <https://doi.org/10.1186/s13750-022-00264-0>.
- Gelman, A., 2014. How do we choose our default methods? In: Lin, X., Genest, C., Banks, D.L., Molenberghs, G., Scott, D.W., Wang, J.-L. (Eds.), *Past, Present, and Future of Statistical Science*. Chapman and Hall/CRC.
- Gelman, A., Hill, J. (Eds.), 2007. *Data Analysis Using Regression and Multilevel/Hierarchical Models*. Cambridge University Press.
- Gelman, A., Loken, E., 2014. The Statistical Crisis in Science. *Am. Sci.* 102 (6), 460.
- Gigerenzer, G., Hoffrage, U., 1995. How to improve Bayesian reasoning without instruction: Frequency formats. *Psychol. Rev.* 102, 684–704. <https://doi.org/10.1037/0033-295X.102.4.684>.
- Greenland, S., 2021. Dealing With the Inevitable Deficiencies of Bias Analysis—and All Analyses. *Am. J. Epidemiol.* 190, 1617–1621. <https://doi.org/10.1093/aje/kwab069>.
- Greenland, S., 2017. For and Against Methodologies: Some Perspectives on Recent Causal and Statistical Inference Debates. *Eur. J. Epidemiol.* 32, 3–20. <https://doi.org/10.1007/s10654-017-0230-6>.
- Greenland, S., Senn, S.J., Rothman, K.J., Carlin, J.B., Poole, C., Goodman, S.N., Altman, D.G., 2016. Statistical tests, P values, confidence intervals, and power: a guide to misinterpretations. *Eur. J. Epidemiol.* 31, 337–350. <https://doi.org/10.1007/s10654-016-0149-3>.
- Higgins, J.P.T., Altman, D.G., 2008. Assessing risk of bias in included studies. In: Higgins, J.P.T., Green, S. (Eds.), *Cochrane Handbook for Systematic Reviews of Interventions*. Wiley Online Library, New Jersey, United States, pp. 187–241. <https://doi.org/10.1002/9780470712184.ch8>.
- Hill, M.O., 2012. Local frequency as a key to interpreting species occurrence data when recording effort is not known. *Methods Ecol. Evol.* 3, 195–205.
- Hoekstra, R., Morey, R.D., Rouder, J.N., Wagenmakers, E.-J., 2014. Robust misinterpretation of confidence intervals. *Psychon. Bull. Rev.* 21, 1157–1164. <https://doi.org/10.3758/s13423-013-0572-3>.
- Hullman, J., 2020. Why Authors Don't Visualize Uncertainty. *IEEE Trans. Visual Comput. Graphics* 26, 130–139. <https://doi.org/10.1109/TVCG.2019.2934287>.
- Hullman, J., Kay, M., Kim, Y.-S., Shrestha, S., 2018. Imagining Replications: Graphical Prediction & Discrete Visualizations Improve Recall & Estimation of Effect Uncertainty. *IEEE Trans. Visual Comput. Graphics* 24, 446–456. <https://doi.org/10.1109/TVCG.2017.2743898>.
- Hullman, J., Resnick, P., Adar, E., Papaleo, E., 2015. Hypothetical Outcome Plots Outperform Error Bars and Violin Plots for Inferences about Reliability of Variable Ordering. *PLoS ONE* 10 (11), e0142444.
- Kale, A., Nguyen, F., Kay, M., Hullman, J., 2018. Hypothetical Outcome Plots Help Untrained Observers Judge Trends in Ambiguous Data. *IEEE Trans. Vis. Comput. Graph.* 25 (1), 892–902.
- Kay, M., 2021. *tidybayes: Tidy Data and Geoms for Bayesian Models*. R package version 3.0.2, <https://mjskay.github.io/tidybayes/>. DOI: 10.5281/zenodo.1308151.
- Kay, M., Kola, T., Hullman, J.R., Munson, S.A., 2016. When (ish) is My Bus?: User-centered Visualizations of Uncertainty in Everyday, Mobile Predictive Systems. In: CHI '16: Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems, pp. 5092–5103. <https://doi.org/10.1145/2858036.2858558>.
- Konno, K., Livoreil, B., Pullin, A.S., 2021. *Collaboration for Environmental Evidence Critical Appraisal Tool*, Version 0.3 (Prototype). <https://environmentalevidence.org/cee-critical-appraisal-tool/>.
- Kruschke, J.K., 2011. *Doing Bayesian Data Analysis*, 1st ed. Academic Press, Burlington, MA.
- Lamb, E.G., Bayne, E., Holloway, G., Schieck, J., Boutin, S., Herbers, J., Haughland, D.L., 2009. Indices for monitoring biodiversity change: Are some more effective than others? *Ecol. Ind.* 9, 432–444. <https://doi.org/10.1016/j.ecolind.2008.06.001>.
- Lindenmayer, D.B., Likens, G.E., 2010. The science and application of ecological monitoring. *Biol. Conserv.* 143 (6), 1317–1328.
- Liu, Y., Kale, A., Althoff, T., Heer, J., 2021. Boba: Authoring and Visualizing Multiverse Analyses. *IEEE Trans. Visual Comput. Graphics* 27, 1753–1763. <https://doi.org/10.1109/TVCG.2020.3028985>.
- McGuinness, L.A., Higgins, J.P.T., 2021. Risk-of-bias VISualization (robvis): An R package and Shiny web app for visualizing risk-of-bias assessments. *Res. Synthet. Methods* 12, 55–61. <https://doi.org/10.1002/jrsm.1411>.
- Padilla, L., Kay, M., Hullman, J., 2022. Uncertainty visualization. In: Piegorisch, W.W., Levine, R.F., Zhang, H.H., Lee, T.C.M. (Eds.), *Handbook of Computational Statistics and Data Science*. Wiley.
- Pescott, O.L., Humphrey, T.A., Stroh, P.A., Walker, K.J., 2019. Temporal changes in distributions and the species atlas: How can British and Irish plant data shoulder the inferential burden? *British Irish Botany* 1, 250–282. <https://doi.org/10.33928/bi.2019.01.250>.
- Pescott, O.L., Walker, K.J., Pocock, M.J.O., Jital, M., Outhwaite, C.L., Cheffings, C.M., Harris, F., Roy, D.B., 2015. Ecological monitoring with citizen science: the design and implementation of schemes for recording plants in Britain and Ireland. *Biol. J. Linn. Soc.* 115, 505–521. <https://doi.org/10.1111/bij.12581>.
- Potter, K., Wilson, A., Bremer, P.-T., Williams, D., Doutriaux, C., Pascucci, V., Johnson, C.R., 2009. Ensemble-Vis: A Framework for the Statistical Visualization of Ensemble Data. In: in: *2009 IEEE International Conference on Data Mining Workshops*, pp. 233–240. <https://doi.org/10.1109/ICDMW.2009.55>.
- Rafı, Z., Greenland, S., 2020. Semantic and cognitive tools to aid statistical science: replace confidence and significance by compatibility and surprise. *BMC Med. Res. Method.* 20, 244. <https://doi.org/10.1186/s12874-020-01105-9>.
- Regan, H.M., Colyvan, M., Burgman, M.A., 2002. A Taxonomy and Treatment of Uncertainty for Ecology and Conservation Biology. *Ecol. Appl.* 12, 618–628. <https://doi.org/10.2307/3060967>.
- Schmeller, D.S., Henry, P.-Y., Julliard, R., Gruber, B., Clobert, J., Dziock, F., Lengyel, S., Nowicki, P., Déri, E., Budrys, E., 2009. Advantages of Volunteer-Based Biodiversity Monitoring in Europe. *Conserv. Biol.* 23 (2), 307–316.
- Sedlmeier, P., Gigerenzer, G., 2001. Teaching Bayesian reasoning in less than two hours. *J. Exp. Psychol.* 130, 380–400. <https://doi.org/10.1037/0096-3445.130.3.380>.
- Soldaat, L.L., Pannekoek, J., Verweij, R.J.T., van Turnhout, C.A.M., van Strien, A.J., 2017. A Monte Carlo method to account for sampling error in multi-species indicators. *Ecol. Ind.* 81, 340–347. <https://doi.org/10.1016/j.ecolind.2017.05.033>.
- Spiegelhalter, D., 2017. Risk and Uncertainty Communication. *Annu. Rev. Stat. Appl.* 4, 31–60. <https://doi.org/10.1146/annurev-statistics-010814-020148>.
- Stace, C., 2019. *New Flora of the British Isles*, 4th ed. C&M Floristics, Suffolk, UK.
- Steege, S., Tuerlinckx, F., Gelman, A., Vanpaemel, W., 2016. Increasing Transparency Through a Multiverse Analysis. *Perspect Psychol Sci* 11, 702–712. <https://doi.org/10.1177/1745691616658637>.
- van der Bles, A.M., van der Linden, S., Freeman, A.L.J., Mitchell, J., Galvao, A.B., Zaval, L., Spiegelhalter, D.J., 2019. Communicating uncertainty about facts, numbers and science. *R. Soc. Open Sci.* 6 (5), 181870.
- van Strien, A.J., Meyling, A.W.G., Herder, J.E., Hollander, H., Kalkman, V.J., Poot, M.J.M., Turnhout, S., van der Hoorn, B., van Strien-van Liempt, W.T.F.H., van Swaay, C.A.M., van Turnhout, C.A.M., Verweij, R.J.T., Oerlemans, N.J., 2016. Modest recovery of biodiversity in a western European country: The Living Planet Index for the Netherlands. *Biol. Conserv.* 200, 44–50. <https://doi.org/10.1016/j.biocon.2016.05.031>.
- Walker, K.J., Pearman, D., Ellis, B., McIntosh, J., Lockton, A., 2010. *Recording the British and Irish flora 2010–2020*. BSBI, London.