

When does modelling dependence change the interpretation of biodiversity indicators?

O.L. Pescott¹

1. UK Centre for Ecology & Hydrology Wallingford, OX10 8BB, olipes@ceh.ac.uk

Abstract

Recent biodiversity trend analyses have modelled uncertainty arising from temporal, spatial and phylogenetic dependence. For primarily descriptive indicators such as the Living Planet Indices (LPI) however, a prior question is whether dependence modelling improves estimation of a fixed quantity or actually changes the quantity being reported. Using the high-profile case of Johnson et al. (2024), the current paper argues that covariance-rich hierarchical models can make the relationship between a reported model coefficient and an intended estimand ambiguous. When aggregate trends are estimated from common coefficients in latent models with structured covariance, observations contribute unequal independent information to the estimator through assumed redundancy among related units. The reported collective trend is then a model-based coefficient (e.g. a mean over a latent-effect distribution) shaped by partial pooling and dependence; this is not automatically an average under declared estimand or “aggregation” weights that indicator trend users may assume. A further ambiguity is whether the reported coefficient summarises the analysed sample or supports claims about a broader spatial, taxonomic or other policy-relevant domain. One defensible reply, that the model is merely a more efficient estimator of a sample average, requires explicit statement of the estimand, weighting scheme, relationship between the model output and that estimand, and evidence of performance for that target; predictive success alone is not sufficient for this descriptive claim. Claims concerning broader generalisability require accompanying risk-of-bias assessments where evaluation against the true target (e.g. a time trend in some facet of global biodiversity) is unobtainable. The implication is that covariance-rich hierarchical indicator analyses should state the estimand, the role of dependence modelling, missing data assumptions and domain of inference clearly before interpreting the resulting trend for scientific or policy applications.

Introduction

In recent work on biodiversity indicators increasing emphasis has been placed on uncertainty, filtering decisions and model specification rather than on headline trajectories alone. Within that broader movement, Johnson et al. (2024) showed that biodiversity trend analyses can substantially understate modelled uncertainty when temporal, spatial and phylogenetic dependence are ignored. A related point was made by Wilkes et al. (2025), where biodiversity trend assessments were shown to contain substantial “hidden” uncertainty arising from collection biases, taxonomic incompleteness, data preparation and model specification. A broader methodological seriousness is therefore arguably emerging around how model-based outputs are supported by data and their underpinning, often silent, assumptions (Boyd et al., 2022, 2023; Pescott et al., 2026).

This problem sits within a wider issue in ecology in which the final inferential goal¹, e.g. descriptive, predictive or causal, is not always sharply defined or distinguished by practitioners (Carlin & Moreno-Betancur, 2025; Hodges, 1996). For example, in observational studies of nitrogen deposition effects, predictive modelling strategies poorly suited to coefficient estimation have frequently been used in settings where causal claims were being made (Pescott & Jitlal, 2020). In occupancy-detection modelling, Stewart et al. (2023) showed that information criteria can favour models that predict site-level occupancy more accurately while yielding markedly worse causal inference about a focal covariate because of collider bias. In broad-scale biodiversity monitoring, Boyd et al. (2023) reviewed how unrepresentative samples are often presented as though they supported representative descriptive claims about wider unsampled populations. The present issue is analogous to these: some elements of uncertainty may be modelled elaborately while the actual inferential target remains poorly specified.

Uncertainty, unanchored?

Against this background, Johnson et al. (2024) provide a prominent biodiversity trend application of covariance-rich hierarchical modelling developed using a number of datasets, including the Living Planet Database underlying the well-known Living Planet Index (LPI), a “leading global biodiversity indicator based on vertebrate population time series” (McRae et al., 2025). The Johnson et al. (2024) approach to modelling biodiversity trend data is especially relevant to the question of inferential goals because its framing straddles descriptive and predictive aims: the paper offers “revised average abundance trends that appropriately recognize uncertainty” alongside “improved local-scale prediction accuracy” (Johnson et al., 2024). These aims are achieved via a Bayesian correlated-effects hierarchical model in which temporal, spatial and phylogenetic covariances are incorporated simultaneously. Across ten datasets both the estimated average abundance trend and its uncertainty were altered relative to simpler models (Johnson et al., 2024). Improved prediction of withheld values and time trends was also reported. On those terms, the paper is a useful demonstration that familiar dependence structures can materially affect biodiversity trend summaries and their associated uncertainty. However, for descriptive trend indicators such as the policy-relevant LPI (a two-stage geometric mean of relative abundance in its most basic form; Deinet et al., 2024; McRae et al., 2025), a prior question remains: has uncertainty been better quantified around a specified descriptive target (e.g. averaging annual rates of change across populations within species and then across species; Deinet et al., 2024), or is the new reported coefficient now a model-based quantity whose inferential relationship to previously published methodologies remains unstated? This ambiguity is reinforced by the language of Johnson et al. (2024): its title frames the contribution as revealing uncertainty in the “status” of biodiversity change, and elsewhere in the manuscript the focus is stated to be on “deriving the collective abundance trend” or “trend inference”. These all sound descriptive, even though part of the model’s appeal is argued in terms of improved local-scale prediction (Johnson et al., 2024). This ambiguity is, however, not merely semantic. Johnson et al. (2024) define their collective trend as an average rate of change across species and locations, but also present their correlated-effects model as improving prediction beyond the fitted observations. These two claims can be made to be compatible, but they potentially correspond to different inferential tasks unless the domain (and grain) to which inference is intended to generalise (i.e. the target population *sensu* Lohr, 2022) and aggregation rule (i.e. the approach to averaging sampled units implied by an explicit choice of estimand) have been clearly laid out. Note that none of this

¹ Here we refer to all three types of goal as “inferences”, given that they all involve inferring things about larger populations or relationships using some combination of statistical theory and sampled data. We note that some schools of academic thought reserve the word “inference” for a narrower set of scientific conclusions more closely tied to hypothesis testing (e.g. Tredennick et al., 2021), but, following a number of other practitioners (e.g. Carlin & Moreno-Betancur, 2025; Hernán et al., 2019; Hodges, 1996; Shmueli, 2010, 2025), we find the triumvirate of inferential tasks used here clearer, whilst also recognising that tasks may not always sit neatly within one of the three (Greenland, 2025; Pescott et al., 2026).

challenges the validity of prediction as a scientific target in its own right: the estimand issue arises when the outputs of predictive assessments are being used to make broader descriptive claims.

To take the example of the LPI, this measure has normally been framed descriptively; that is “the objective is simply to obtain certain information about large groups” (Cochran, 1963); closer to the language we use here, Greenland (2025) defines this type of goal as an “inferred description of a target population”. In technical materials the LPI has been described as the average change in population size since 1970 (Deinet et al., 2024), while McRae et al. (2025) described the index as measuring the average rate of change in monitored vertebrate populations from the Living Planet Database, explicitly stating that it “describ[es] a state or recent trends”. These descriptions suggest some summary of change in a realised panel (i.e. a set of repeatedly measured sampling units), as is the case for the simplest “unweighted” version of the LPI (the LPI-U; Deinet et al., 2024)². A geo-taxonomically weighted version of the LPI also exists (LPI-D), but this does not appear to have involved predictive evaluative criteria to date, and can still be described according to a defined hierarchical aggregation pathway using explicit weights and transparently-stated target populations (McRae et al., 2017). These indicators do not obviously refer to latent parameters selected for their predictive performance. If any descriptive reading of the Johnson et al. (2024) abundance trends is intended, then the central issue is not just whether various forms of nonindependence are detectable, but whether modelling these changes the inferential target rather than simply improving estimation. In other words, are Johnson et al. (2024) offering an improved estimator of a panel-average quantity, a different model-based summary of the Living Planet Database panel, or a modelled basis for inference beyond the data-providing panel (cf. McRae et al., 2017)?

Please come bearing estimands

This general point, that different inferential aims carry different burdens of proof, has been made explicitly in the statistical literature: for descriptive inference, one proof has been suggested as “showing that the inferential statement has the properties claimed for it” (Hodges, 1996; see also Pescott et al., 2026). On a descriptive reading, the unweighted LPI (LPI-U) target is a fixed summary of the realised panel involving aggregation within and then across species; the diversity-weighted version (LPI-D) used for global indicators also uses geographic and taxonomic species richness weights at higher levels (Deinet et al., 2024). These weights, whether equal or not (McRae et al., 2017), are a part of the estimand (i.e. the precise real-world quantity being estimated; Lundberg et al., 2021), not an inevitable by-product of a model. It is useful to distinguish these aggregation or estimand weights from the effective information weightings induced by the use of any particular statistical estimation algorithm (i.e. the estimator). Information or observation weights can depend on partial pooling or covariance structure without necessarily changing the declared population target of substantive interest (Gelman, 2007).

Algebraically, the general point can be sketched schematically as follows: define the descriptive quantity of interest as some set of biological population trends θ_i , then a panel average can be defined as $\tau_a = \sum_i a_i \theta_i$, where the weights a_i sum to one. The information contributed by sampled units towards estimation using any given model could jointly depend on their precision (via hierarchical partial pooling) and dependence (via covariance structure), as in Johnson et al. (2024). The roles of these types of weightings do not have to coincide: estimates of individual θ_i parameters could involve

² Note that these descriptions gloss over the role of imputation in filling in missing time points within species population time series in the Living Planet Database (Deinet et al., 2024); however, given that Johnson et al. (2024) used extensive missing data filtering in their case studies, our argument throughout this paper conditions on the assumption that there are no such internal gaps. In reality, the imputation of within-panel missing data is another analytical task that also affects estimand interpretation; see the section “What remains unobserved?” below.

different amounts of shrinkage for example, even while the final aggregation of these to a population target average remains an equal-weighted mean. A simplified outline of the Johnson et al. (2024) approach can then be laid out as $\theta_i = \beta + d_i$, where θ_i is the latent trend of population i , with common coefficient β and population-specific deviations d_i (here these incorporate the hierarchical and correlated random-slope deviations used by Johnson et al., 2024 with $E(d_i = 0)$). For declared estimand or aggregation weights a_i (which could be equal), the panel-average quantity is $\tau_a = \sum_i a_i \theta_i = \beta + \sum_i a_i d_i$. Under the assumed zero-mean latent distributions, the expectation of τ_a given the common coefficient β (over hypothetical realisations of the latent random effects) is β , i.e. $E(\tau_a | \beta) = \beta$. However, as can be seen from the preceding formula, τ_a can only equal β exactly when the summation $\sum_i a_i d_i = 0$ (i.e. the weighted population-specific deviations cancel out). Figure 1a illustrates the basic distinction that a model common coefficient and a particular average of population trends need not coincide; Figure 1b formalises this and shows that zero-mean latent-effect distributions imply equality in expectation, not necessarily equality for some particular sampled panel. Alternatively, if the weights a_i are specified in order to represent the panel or some broader target (see *From coefficient to estimand* below), then the panel quantity can be estimated directly by aggregating the fitted θ_i trends, without requiring it to equal β explicitly (Fig. 1c).

Covariance-rich hierarchical models may therefore return a model-based latent collective trend whose interpretation depends on the assumed dependence structure; this may or may not link to an intended estimand. This distinction is also at the core of recent work on descriptive inference from nonprobability samples: here, recovering a population quantity requires assumptions linking the observed sample to the target population, such as independence of the response from sample inclusion after conditioning on measured auxiliary variables (Boyd et al., 2024; Dumelle et al., 2025). The fundamental question is not whether a complex model can be fitted, but what quantity is being targeted and under what assumptions an estimator is expected to recover it well.

Seen in these terms, the model of Johnson et al. (2024) could do more than simply widen uncertainty intervals around a clearly specified descriptive average. Their collective trend is estimated as the fixed abundance-year coefficient within a model containing hierarchical random effects and fitted temporal, spatial and phylogenetic dependence (Johnson, 2024), and so estimates are determined by the model's latent structure and covariance assumptions, rather than by nominal sample size alone. In the generalised least-squares approach related to such Bayesian covariance-based estimation, information is aggregated through the inverse covariance matrix rather than by treating all observations as independent (Aitken, 1936). When substantial positive dependence is estimated, observations in correlated parts of the panel contribute less new independent information about a common coefficient than observations from less correlated parts. This does not necessarily show that estimand weights are not as intended or implied, but it does mean that a reported coefficient cannot be interpreted as a descriptive panel average merely because a description of the process uses the language of averaging. The issue is whether the fixed abundance-year coefficient of Johnson et al. should be understood as a model-defined mean of a latent random-effects distribution, or as an estimator of a finite-panel average under explicit weights. The model-based interpretation does imply a target parameter: the principal model of Johnson et al. (2024) is additive on the log-abundance scale and their latent slope deviations are assigned zero-mean distributions, meaning that the common abundance-year coefficient can be seen as the expected trend under the modelled latent-effect distributions as noted in the schematic outline above (Fig. 1). This quantity (β) can be given a superpopulation-like interpretation if the latent distributions are treated as the data-generating process for potentially observable units, but the ecological unit and composition of that implied superpopulation are not defined. A model can offer a quantification of latent variation based on its assumptions, but this does not by itself establish the real-world referent of that model. Perhaps more importantly, the modelled latent quantity is not necessarily identical to a weighted panel average (e.g. McRae et al., 2017), and neither does it necessarily

correspond to an average calculated over a broader target population, even though Johnson et al. (2024) seem to invite both readings at different points.

Shrinkage to what?

Possible defences should be acknowledged: it could be argued that the estimand remains the average change across the monitored panel, with the covariance-enhanced model of Johnson et al. (2024) serving as a more efficient estimator of that same quantity, rather than as a broader population-level estimand or a different latent-process summary; perhaps chosen because expected error is reduced by “borrowing strength” (i.e. partial pooling) across space, time and phylogeny. Under exchangeability and correctly specified variance/covariance structure, partial pooling can reduce expected error for unit-level trends, especially when observations are noisy or sparse relative to the variation among the true trends (Efron & Morris, 1973; Morris, 1983). The same logic can also support the estimation of aggregated estimands, as has been shown for small-area estimation and multilevel regression and poststratification (Kennedy & Gelman, 2021; Rao & Molina, 2015). For a fixed panel average however, this is not automatic: shrinkage improves estimation only if a model’s output is explicitly linked to an estimand aggregation rule and shown to reduce bias, variance or mean squared error (MSE) for that declared target. The modelling approach of Johnson et al. (2024) does leave room for such an interpretation, because their correlated-effects model is parameterised around a common abundance-year coefficient with structured deviations; however, that interpretation is not established. Predictive improvement is demonstrated, but lower MSE, bias or interval calibration for a clearly defined descriptive target are not. Nor does predictive accuracy for withheld data define the domain or weights over which a broader descriptive average should be taken (i.e. it says nothing quantitative about generalisability). Moreover, even if the point estimate could be defended on these grounds, the corresponding uncertainty interval would remain model-based, since it depends on the fitted covariance structure and associated redundancy. The issue of inferential clarity would therefore not disappear, it would need to be made explicit as a statement that an appropriate model-based estimator is being used for a fixed descriptive quantity.

From coefficient to estimand

In support of the “efficient estimator of the panel average” defence, or the stronger claim of a broader descriptive estimand, further steps would be required. A panel-average estimand would have to be defined explicitly first, rather than only implied in the phrase “averaged across all species and locations” (Johnson et al., 2024). If the intended target were broader than the analysed panel, the target population would also have to be clearly defined: for example, all monitored populations in the source Living Planet Database, all locations in some geographic area, all species in a taxonomic group, or possibly a policy-relevant combination of these. Second, the relationship between the estimand and the reported collective-trend parameter would have to be made explicit (ideally algebraically). If the intended estimand is the panel average, this could be done via a target-preserving step such as posterior aggregation of estimated unit-level trends under declared estimand weights, in the spirit of poststratification (Gelman & Little, 1997; Kennedy & Gelman, 2021; Fig. 1). A different possibility would be to parameterise the model so that the reported fixed effect is the desired marginal mean, by imposing relevant centring constraints for example. A centring constraint can ensure that the (weighted) average of the unit-level deviations is zero, forcing an overall fixed effect to equal the panel average (e.g. Kaufman & Sain, 2010).

The former posterior aggregation approach treats the aggregate as a derived quantity (Fig. 1c); the latter builds the aggregate into the model’s parameterisation. Both can be principled modelling strategies, but neither can be identified automatically with a common abundance-year coefficient unless the

equivalence to a transparent target can be shown. The precise weighting scheme used or implied would also need to be stated: equal weight across populations, equal weight across species, or some other construction. The discussion of alternative weightings for biomass decline or individuals lost in Johnson et al. (2024) underlines this point. Complex spatial, temporal, and phylogenetic covariance structure then raises the further question of how information is pooled across units in the panel. For broader-than-sample inference it also raises additional questions: what distribution of unsampled species, sites or populations is being averaged over, and with what weights? Most importantly, the estimator would need to be evaluated against the stated target itself, through assessment of bias, variance, MSE and/or interval calibration, rather than through predictive accuracy for withheld information.

Boyd et al. (2025) here provide a useful exemplar, because in that paper the declared target is not the sample itself but a wider landscape mean under biased data collection. The adjustment away from the “naive” sample summary is coherent: the estimand, inferential goal and biasing assumptions have been stated *a priori*. The message here is not that latent or superpopulation models are always appropriate, but that a shift from sample summary to model-based target requires an explicit estimand and clear argument for the required assumptions.

What remains unobserved?

A second issue, separate from the estimand question but relevant to covariance-rich models, concerns internal (i.e. within-panel) missing data (this could be equated with “nonresponse” in the broader survey sampling literature) and broader target population coverage issues (Groves, 2004). Rich covariance structure does not remove these problems, it relocates them. A model still makes a claim about what renders unobserved values ignorable or learnable conditional on the included temporal, spatial, phylogenetic or other auxiliary structure (Dumelle et al., 2025). Prediction of hold-out observations or trends may support interpolation within the observed panel and provide within-sampled domain evidence of improved predictive performance, but it does not by itself justify inferences for unmonitored species, sites or time periods (i.e. out-of-Data(set)-Generating-Process inference). This is important because Johnson et al. (2024) address prediction within an already filtered dataset: Living Planet Database time series with missing annual abundance values were removed, consecutive abundance estimates were required, and in most datasets only the longest half of time series was retained. The paper is therefore strongest as a contribution to dependence modelling and prediction within filtered (or otherwise “complete-for-inference”) datasets and where missing value prediction is justified based on stated ignorability assumptions together with an adequate model for the observed dependence structure (Dumelle et al., 2025). It cannot be said, however, to offer a general solution to missing data and representativeness in biodiversity monitoring panels without strong assumptions (Dumelle et al., 2025; Pescott, 2023).

From within the LPI literature McRae et al. (2025) provide a useful contrast. As already noted, these authors generally treat the LPI as a descriptive measure of the average change in monitored vertebrate populations, whilst also presenting temporal and spatial extrapolation (i.e. prediction) of LPI-like quantities as a possible future extension of the index and its underlying database. That separation is pertinent, because it suggests that predictive use is best viewed as an additional task rather than as the already settled core meaning of LPI indicators. This distinction is also useful for reading Johnson et al. (2024): improved prediction over spatial or phylogenetic structure may be valuable in its own right, but it should not be allowed to retrospectively determine what a collective trend was supposed to mean or to define its “true” uncertainty.

Anchoring the target

Taken together, these considerations point to a simple reporting consequence: in papers using partial pooling and/or covariance-rich hierarchical models for biodiversity indicators, at least four things should be stated explicitly: (i) the estimand; (ii) the role played by dependence modelling; (iii) the missing data assumptions required; and, (iv), the domain to which inference is intended to generalise. The logic of risk-of-bias assessment tools for description is obviously relevant here, since there the inferential goal and statistical target population need to be specified before bias can be assessed (Boyd et al., 2022; Pescott et al., 2026). More generally, similar ordering has been formalised in other disciplines through estimand-oriented frameworks, in which the target quantity is defined prior to the choice of missing data strategy and sensitivity analysis (e.g. Kahan et al., 2024). A similar ordering would improve biodiversity indicator practice more generally.

It need not be denied that Johnson et al. (2024) have shown something important about model-based uncertainty under correlated structure; their paper also supports their stated narrow methodological aim of demonstrating trend sensitivity to model choice (cf. Wilkes et al., 2025). An essential corollary, however, is that for descriptive indicators such as the LPI, the question is not only whether dependence has been modelled well (or whether uncertainty statements are now considered “appropriate”; Johnson et al., 2024), but what inferential target the resulting coefficient represents: a summary of the realised panel, a model-implied latent mean, or a defined basis for inference to some wider domain. Read alongside the other ecological papers reviewed here, and the broader statistical literature on estimands, model-assisted estimation and generalisation, the main caution is that model sophistication should not be allowed to silently redefine the target of inference. For descriptive environmental monitoring, this is part of the substantive claim being made rather than a mere technical footnote.

Funding

This work was partially supported by NERC, through the UKCEH National Capability for UK Challenges Programme NE/Y006208/1, and by the Joint Nature Conservation Committee funding to the UK National Plant Monitoring Scheme.

Acknowledgements

Sincere thanks to Dr K.M. Irvine and an anonymous reviewer whose comments significantly improved the paper.

References

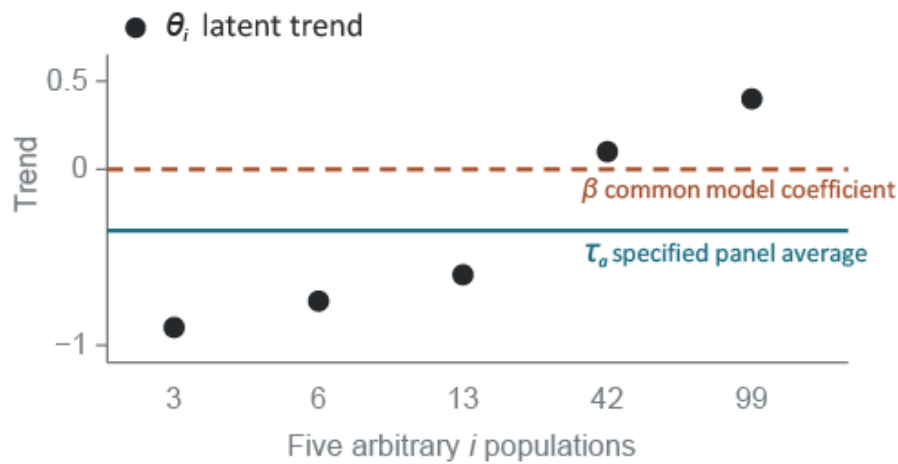
- Aitken, A. C. (1936). IV.—On Least Squares and Linear Combination of Observations. *Proceedings of the Royal Society of Edinburgh*, 55, 42–48. <https://doi.org/10.1017/S0370164600014346>
- Boyd, R. J., Botham, M., Dennis, E., Fox, R., Harrower, C., Middlebrook, I., Roy, D. B., & Pescott, O. L. (2025). Using causal diagrams and superpopulation models to correct geographic biases in biodiversity monitoring data. *Methods in Ecology and Evolution*, 16(2), 332–344. <https://doi.org/10.1111/2041-210X.14492>
- 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 in Ecology and Evolution*, 13(7), 1497–1507. <https://doi.org/10.1111/2041-210X.13857>
- Boyd, R. J., Powney, G. D., & Pescott, O. L. (2023). We need to talk about nonprobability samples. *Trends in Ecology & Evolution*. <https://doi.org/10.1016/j.tree.2023.01.001>

- Boyd, R. J., Stewart, G. B., & Pescott, O. L. (2024). Descriptive inference using large, unrepresentative nonprobability samples: An introduction for ecologists. *Ecology*, 105(2), e4214. <https://doi.org/10.1002/ecy.4214>
- Carlin, J. B., & Moreno-Betancur, M. (2025). On the Uses and Abuses of Regression Models: A Call for Reform of Statistical Practice and Teaching. *Statistics in Medicine*, 44(13–14), e10244. <https://doi.org/10.1002/sim.10244>
- Cochran, W. G. (1963). *Sampling Techniques* (2nd ed.). John Wiley & Sons, Inc.
- Deinet, S., Marconi, V., Freeman, R., Puleston, H., & McRae, L. (2024). *Living Planet Report 2024 Technical Supplement: Living Planet Index*. ZSL. <https://doi.org/10.17605/OSF.IO/6TE9H>
- Dumelle, M., Trangucci, R., Nahlik, A. M., Olsen, A. R., Irvine, K. M., Blocksom, K., Ver Hoef, J. M., & Fuentes, C. (2025). Missing data in ecology: Syntheses, clarifications, and considerations. *Ecological Monographs*, 95(4), e70037. <https://doi.org/10.1002/ecm.70037>
- Efron, B., & Morris, C. (1973). Stein's Estimation Rule and its Competitors—An Empirical Bayes Approach. *Journal of the American Statistical Association*, 68(341), 117–130. <https://doi.org/10.1080/01621459.1973.10481350>
- Gelman, A. (2007). Struggles with Survey Weighting and Regression Modeling. *Statistical Science*, 22(2), 153–164.
- Gelman, A., & Little, T. C. (1997). Poststratification into many categories using hierarchical logistic regression. *Survey Methodology*, 23(2), 127–135.
- Greenland, S. (2025). Some Ways to Make Regression Modeling More Helpful Than Misleading. *Statistics in Medicine*, 44(13–14), e10313. <https://doi.org/10.1002/sim.10313>
- Groves, R. M. (2004). *Survey Errors and Survey Costs*. Wiley-Interscience.
- Hernán, M. A., Hsu, J., & Healy, B. (2019). A Second Chance to Get Causal Inference Right: A Classification of Data Science Tasks. *CHANCE*, 32(1), 42–49. <https://doi.org/10.1080/09332480.2019.1579578>
- Hodges, J. S. (1996). Statistical Practice as Argumentation: A Sketch of a Theory of Applied Statistics. In J. C. Lee, W. O. Johnson, & A. Zellner (Eds.), *Modelling and Prediction Honoring Seymour Geisser* (pp. 19–45). Springer. https://doi.org/10.1007/978-1-4612-2414-3_2
- Johnson, T. F. (2024). *GitTFJ/correlated_effect_model: V1.0.2* [Computer software]. Zenodo. <https://zenodo.org/records/10638241>
- Johnson, T. F., Beckerman, A. P., Childs, D. Z., Webb, T. J., Evans, K. L., Griffiths, C. A., Capdevila, P., Clements, C. F., Besson, M., Gregory, R. D., Thomas, G. H., Delmas, E., & Freckleton, R. P. (2024). Revealing uncertainty in the status of biodiversity change. *Nature*, 628(8009), 788–794. <https://doi.org/10.1038/s41586-024-07236-z>
- Kahan, B. C., Hindley, J., Edwards, M., Cro, S., & Morris, T. P. (2024). The estimands framework: A primer on the ICH E9(R1) addendum. *BMJ*, 384, e076316. <https://doi.org/10.1136/bmj-2023-076316>
- Kaufman, C. G., & Sain, S. R. (2010). Bayesian functional {ANOVA} modeling using Gaussian process prior distributions. *Bayesian Analysis*, 5(1), 123–149. <https://doi.org/10.1214/10-BA505>
- Kennedy, L., & Gelman, A. (2021). Know your population and know your model: Using model-based regression and poststratification to generalize findings beyond the observed sample. *Psychological Methods*, 26(5), 547–558. <https://doi.org/10.1037/met0000362>
- Lohr, S. L. (2022). *Sampling: Design and Analysis* (3rd ed.). CRC Press.
- Lundberg, I., Johnson, R., & Stewart, B. M. (2021). What Is Your Estimand? Defining the Target Quantity Connects Statistical Evidence to Theory. *American Sociological Review*, 86(3), 532–565. <https://doi.org/10.1177/00031224211004187>
- McRae, L., Cornford, R., Marconi, V., Puleston, H., Ledger, S. E. H., Deinet, S., Oppenheimer, P., Hoffmann, M., & Freeman, R. (2025). The utility of the Living Planet Index as a policy tool and for measuring nature recovery. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 380(1917), 20230207. <https://doi.org/10.1098/rstb.2023.0207>
- McRae, L., Deinet, S., & Freeman, R. (2017). The Diversity-Weighted Living Planet Index: Controlling for Taxonomic Bias in a Global Biodiversity Indicator. *PLOS ONE*, 12(1), e0169156. <https://doi.org/10.1371/journal.pone.0169156>
- Morris, C. N. (1983). Parametric Empirical Bayes Inference: Theory and Applications. *Journal of the American Statistical Association*, 78(381), 47–55. <https://doi.org/10.1080/01621459.1983.10477920>
- Pescott, O. L. (2023). Seek a Paradigm and Distrust It? Statistical Arguments and the Representation of Uncertainty. *Harvard Data Science Review*, 5(3). <https://doi.org/10.1162/99608f92.a02188d0>
- Pescott, O. L., Boyd, R. J., Powney, G. D., & Stewart, G. B. (2026). Towards a unified approach to formal “risk of bias” assessments for causal and descriptive inference. *Quality & Quantity*. <https://doi.org/10.1007/s11135-026-02687-0>

- Pescott, O. L., & Jitlal, M. (2020). Reassessing the observational evidence for nitrogen deposition impacts in acid grassland: Spatial Bayesian linear models indicate small and ambiguous effects on species richness. *PeerJ*, 8, e9070. <https://doi.org/10.7717/peerj.9070>
- Rao, J. N. K., & Molina, I. (2015). *Small Area Estimation*. John Wiley & Sons.
- Shmueli, G. (2010). To explain or to predict? *Statistical Science*, 289–310.
- Shmueli, G. (2025). To Explain, to Predict, or to Describe: Figuring out the Study Goal [Commentary on “On the Uses and Abuses of Regression Models” by Carlin and Moreno-Betancur]. *Statistics in Medicine*, 44(13–14), e10307. <https://doi.org/10.1002/sim.10307>
- Stewart, P. S., Stephens, P. A., Hill, R. A., Whittingham, M. J., & Dawson, W. (2023). Model selection in occupancy models: Inference versus prediction. *Ecology*, 104(3), e3942. <https://doi.org/10.1002/ecy.3942>
- Tredennick, A. T., Hooker, G., Ellner, S. P., & Adler, P. B. (2021). A practical guide to selecting models for exploration, inference, and prediction in ecology. *Ecology*, 102(6), e03336. <https://doi.org/10.1002/ecy.3336>
- Wilkes, M. A., McKenzie, M., Johnson, A., Hassall, C., Kelly, M., Willby, N., & Brown, L. E. (2025). Revealing hidden sources of uncertainty in biodiversity trend assessments. *Ecography*, 2025(5), e07441. <https://doi.org/10.1111/ecog.07441>

a. Population-specific latent trends

Illustrative latent population unit trends on a common scale



b. Common coefficient and specified panel target

$$\boldsymbol{\theta} = \beta \mathbf{1} + H\mathbf{b}, \quad \mathbf{b} \sim N(\mathbf{0}, G_\psi)$$

The vector of population latent trends $\boldsymbol{\theta}$ equals the combination of the common coefficient β applied to each population plus the vector of zero-centred random-effect (RE) components $\mathbf{b}$ with covariance matrix of latent REs G_ψ (conditional on covariance parameters ψ).

$$\tau_a = \mathbf{a}^T \boldsymbol{\theta} = \beta + \mathbf{a}^T H \mathbf{b}$$

With declared weights $a_i \geq 0$, $\sum_i a_i = 1$

A declared panel average τ_a is a weighted combination of the population latent trends (i.e. $\mathbf{a}^T \boldsymbol{\theta}$), which also equals the common coefficient β plus the combination of the weight vector $\mathbf{a}$, the RE deviations vector $\mathbf{b}$ and the matrix map between them H (i.e. $\mathbf{a}^T H \mathbf{b}$).

$$\begin{aligned} \vdots \quad E(\tau_a \mid \beta, \psi) &= \beta \\ \tau_a = \beta &\Leftrightarrow \mathbf{a}^T H \mathbf{b} = 0 \end{aligned}$$

The expectation of τ_a conditional on β and structure ψ is β , but the panel average will only equal β if, and only if (i.e. $\Leftrightarrow$), the weighted deviations equal zero.

c. Dependence affects estimation; the scientific target is declared separately

data $y \rightarrow$ hierarchical / covariance model $\rightarrow p(\beta, \boldsymbol{\theta} \mid y)$

Common coefficient summary: $p(\beta \mid y)$

Declared panel target: $p(\tau_a \mid y)$ via $\tau_a = \mathbf{a}^T \boldsymbol{\theta}$

G_ψ and observation covariance affect learning / pooling;
 $\mathbf{a}$ (with panel and scale fixed) specifies the aggregation weights.

Figure legend

Figure 1. A common trend coefficient need not equal a declared realised panel average. (a) Five illustrative latent population-specific trends θ_i , all defined on a common trend scale. The dashed line denotes the common model coefficient β , while the solid line denotes the realised panel average τ_a ; equal population weights are used here for illustration and are not intended to reproduce a particular Living Panel Index (LPI). The numerical separation between the lines is also illustrative and does not represent a simulation or modelled estimate. **(b)** A schematic representation writes the vector of latent population trends as $\boldsymbol{\theta} = \beta \mathbf{1} + H\mathbf{b}$, where $\mathbf{1}$ is a vector of ones, $\mathbf{b}$ contains zero-centred latent random-effect (RE) components, H maps those to the population-specific trends, and G_ψ is their covariance matrix under covariance parameters ψ . For declared aggregation weights $\mathbf{a}$, with $a_i \geq 0$ and $\sum_i a_i = 1$, the realised target is $\tau_a = \mathbf{a}^\top \boldsymbol{\theta} = \beta + \mathbf{a}^\top H\mathbf{b}$. Here $E(\tau_a | \beta, \psi) = \beta$ denotes an expectation over hypothetical realisations of the latent random effects, holding the common coefficient and model structure fixed; it is not a statement that the realised panel average necessarily equals β . Equality for the realised panel occurs if, and only if, $\mathbf{a}^\top H\mathbf{b} = 0$. **(c)** The fitted hierarchical/covariance model yields a joint posterior distribution $p(\beta, \boldsymbol{\theta} | y)$, where $p(\cdot | y)$ denotes the probability distribution of the unknown quantities after conditioning on the observed data y . From this joint posterior one may report the marginal posterior $p(\beta | y)$ for the common coefficient, or derive the posterior $p(\tau_a | y)$ for a declared panel target by applying $\tau_a = \mathbf{a}^\top \boldsymbol{\theta}$ to each posterior draw. Dependence and observation-level covariance therefore affect what is learned about β and $\boldsymbol{\theta}$, but they do not by themselves specify the scientific aggregation weights $\mathbf{a}$. The model shown is a schematic reduction of Johnson et al. (2024), not a reproduction of their complete workflow or any given LPI aggregation procedure.