When does modelling dependence change the target of biodiversity indicators?

O.L. Pescott[1]

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 descriptive indicators such as the Living Planet Index (LPI) however, a prior question is whether dependence modelling improves estimation of a fixed quantity or risks changing the quantity being reported. Using the high-profile case of Johnson et al. (2024), this paper argues that covariance-rich hierarchical models can make this relationship ambiguous. When aggregate trends are estimated as common coefficients in latent models with structured covariance, observations contribute unequal independent information through assumed redundancy among related units. The reported collective trend may therefore be a model-based coefficient shaped by partial pooling and dependence structure, rather than the sampled panel average under declared weights that indicator users may assume. A further ambiguity is whether the reported coefficient summarizes the analysed panel or supports claims about a broader spatial, taxonomic or policy-relevant domain. The defensible reply, that the model is merely a more efficient estimator of a panel 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 insufficient for this descriptive claim. The implication is that covariance-rich indicator analyses should state the estimand, the role of dependence modelling, missing data assumptions and domain of inference before interpreting the resulting trend for scientific 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 analysis 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).

This problem sits within a wider issue in ecology in which inferential goals, i.e. descriptive, predictive and causal, are not always sharply defined or distinguished by practitioners (Carlin & Moreno-Betancur, 2025). For example, in observational studies of nitrogen deposition effects, modelling strategies poorly suited to coefficient estimation have 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 biodiversity monitoring, Boyd et al. (2023) reviewed how unrepresentative samples are often presented as though they supported representative descriptive claims about wider populations. The present issue is analogous: some elements of uncertainty may be modelled elaborately while the actual inferential target remains under-specified.

Against this background, Johnson et al. (2024) provide a prominent biodiversity indicator application of covariance-rich hierarchical modelling developed using a number of datasets, including the well-known Living Planet Index (LPI) indicator (McRae et al., 2025). The Johnson et al. paper is especially relevant here because its framing straddles descriptive and predictive aims: a correlated-effects model was fitted in which temporal, spatial and phylogenetic covariances were incorporated simultaneously, and across ten datasets both the estimated collective trend and its uncertainty were altered relative to simpler models. Improved prediction of withheld values and trends was also reported. On those terms, the paper is a useful and high-profile demonstration that familiar dependence structures can materially affect biodiversity trend summaries and their associated uncertainty. However, for descriptive indicators, a prior question remains: has uncertainty been better quantified around a fixed descriptive target, or is the reported coefficient instead a model-based quantity whose relationship to that target has not been demonstrated? This ambiguity is reinforced by the language of the paper: its title frames the contribution as revealing uncertainty in the “status” of biodiversity change and “trend inference”, which 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 not merely semantic. Johnson et al. (2024) define their collective trend as an average rate of change across species and locations, but also present the correlated-effects model as improving prediction beyond the fitted observations. These two claims can be compatible, but they correspond to different inferential tasks unless the target population and aggregation rule have been cleared laid out.

To take the example of the LPI, this is normally framed descriptively. In technical materials the LPI is 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 point to a summary of change in a realised panel (i.e. a set of repeatedly measured sampling units). They do not obviously refer to a parameter of a latent process selected on its predictive performance. If a descriptive reading is intended, then the central issue is not just whether various forms of nonindependence exist, 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 the same panel-average quantity, a different model-based summary of the LPI panel, or a modelled basis for inference beyond the panel?

This general point, that different inferential aims carry different burdens of proof, has been made explicitly in the statistical literature: for descriptive inference, the proof is “showing that the inferential statement has the properties claimed for it” (Hodges, 1996; Pescott et al., 2026). On a descriptive reading, the LPI target is a fixed summary of the realised panel, an equal-weight mean across monitored population time-series for example. These weights are a part of the estimand, not a natural by-product of a covariance model. Covariance-rich hierarchical models may instead 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). The fundamental question is not whether a complex model can be fitted, but what quantity is being targeted and under what assumptions an estimator (i.e. algorithm applied to a sample) is expected to recover it well.

Seen in these terms, the model of Johnson et al. (2024) could do more than widen uncertainty intervals around a 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), 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 have changed, but it does mean that a reported coefficient cannot be interpreted as a descriptive panel average merely because a verbal description of the process uses the language of averaging.

A similar warning comes from the marginal versus conditional effects literature on multilevel and other correlated data models. The relevance of that literature here is not that Johnson et al. (2024) must be reporting a conditional quantity when it should be marginal, nor that this is a non-collapsibility problem of the kind problematic to nonlinear generalized mixed models (Muff et al., 2016). The main models of Johnson et al. are fitted to log-abundance with additive hierarchical and covariance structure where this distinction is less important. The issue is rather whether their fixed abundance-year coefficient can be understood as a model-implied mean of a latent random-effects distribution, or as an estimator of a realised finite-panel average under explicit weights. Work on marginalised multilevel models and related ecological discussions of subject-specific versus population-averaged inference has demonstrated that latent dependence modelling and preservation of a marginal descriptive target are compatible, but not necessarily automatic (Fieberg et al., 2009; Heagerty & Zeger, 2000; Muff et al., 2016; Ver Hoef et al., 2024).

A defence should however 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 time, space and phylogeny. Under exchangeability and correctly specified variance or 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 some aggregate estimands, as has been highlighted 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 of that target only if a model's output is explicitly linked to an aggregation rule and shown to reduce bias, variance or mean squared error (MSE). The modelling approach of Johnson et al. (2024) leaves 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 observations define the domain or weights over which a broader descriptive average should be taken. Even if the point estimate could be defended on these grounds, the corresponding uncertainty interval would remain model-based, since its 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 claim that a model-assisted estimator is being used for a fixed descriptive quantity.

In support of the "efficient estimator of the panel average" defence, or the stronger claim of a broader (i.e. statistical population-level) 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 collective-trend parameter would have to be made explicit (ideally algebraically). If the intended estimand is the panel average, this would require a target-preserving step such as posterior aggregation of estimated unit-level trends under declared weights, in the spirit of poststratification (Gelman & Little, 1997; Kennedy & Gelman, 2021). A different possibility would be to parameterise the model so that the reported fixed effect is the desired marginal mean, by imposing relevant centering or marginalisation constraints for example (Heagerty & Zeger, 2000).

The former approach treats the aggregate as a derived quantity; the latter builds the aggregate into the model's parameterisation. Both can be principled modelling strategies, but neither way is equivalent to simply reporting a common abundance-year coefficient unless the equivalence can be shown. The implied weighting scheme 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 for example, 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 or superior, but that a shift from sample summary to model-based target requires an explicit estimand and clear argument for the required assumptions.

A second issue, separate from the estimand question but relevant to covariance-rich models, concerns missing data and noncoverage. Rich covariance structure does not remove this problem, it relocates it. The model still makes a claim about what renders unobserved values ignorable or learnable conditional on the included temporal, spatial, phylogenetic and auxiliary structure (Dumelle et al., 2025). Prediction of withheld observations may support interpolation within the observed modelling frame, but it does not by itself justify descriptive inference to unmonitored species, sites or time periods. This is important because Johnson et al. (2024) address prediction within an already filtered modelling frame: 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 datasets, not as a general solution to missing data, noncoverage and representativeness in biodiversity monitoring panels.

From within the LPI literature McRae et al. (2025) provide a useful internal contrast. As noted above, these authors 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 the LPI. 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.

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 clearly relevant here, since there the inferential goal and statistical target population need to be specified before bias is 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). The same ordering would improve biodiversity indicator practice more generally.

It need not be denied that Johnson et al. (2024) have shown something important about uncertainty under correlated structure. An essential corollary, however, is that for descriptive indicators such as the LPI, the question is not only whether dependence has been modelled well, but what inferential target the resulting coefficient represents: a summary of the filtered realised panel, a model-implied latent mean, or a basis for inference to some wider domain. Read alongside the other ecological papers reviewed here, and the broader statistical literature on marginal versus conditional targets, 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.

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

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

Fieberg, J., Rieger, R. H., Zicus, M. C., & Schildcrout, J. S. (2009). Regression modelling of correlated data in ecology: Subject-specific and population averaged response patterns. *Journal of Applied Ecology*, *46*(5), 1018–1025. https://doi.org/10.1111/j.1365-2664.2009.01692.x

Gelman, A., & Little, T. C. (1997). Poststratification into many categories using hierarchical logistic regression. *Survey Methodology*, *23*(2), 127–135.

Heagerty, P. J., & Zeger, S. L. (2000). Marginalized Multilevel Models and Likelihood Inference. *Statistical Science*, *15*(1), 1–19.

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

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

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

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

Muff, S., Held, L., & Keller, L. F. (2016). Marginal or conditional regression models for correlated non-normal data? *Methods in Ecology and Evolution*, *7*(12), 1514–1524. https://doi.org/10.1111/2041-210X.12623

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.

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

Ver Hoef, J. M., Blagg, E., Dumelle, M., Dixon, P. M., Zimmerman, D. L., & Conn, P. B. (2024). Marginal inference for hierarchical generalized linear mixed models with patterned covariance matrices using the Laplace approximation. *Environmetrics*, *35*(7), e2872. https://doi.org/10.1002/env.2872

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