Unifying occupancy-detection and local frequency scaling (Frescalo) models

Oliver L. Pescott*

[a] *UK Centre for Ecology and Hydrology, Benson Lane, Crowmarsh Gifford, OX10 8BB, Oxfordshire, UK*

Abstract

Frescalo's "local frequency scaling" and classical occupancy-detection models both seek to recover true species-occurrence signals from imperfect data. In this paper, we show that the two approaches rest on the same underlying detection mathematics. Occupancy models treat each site's repeat visits as independent detection trials and separately estimate occupancy probability and per-visit detectability. Frescalo, by contrast, pools data across ecologically defined neighbourhoods, standardises for uneven effort, and infers a single discovery rate per species plus a species-specific "time-factor" to capture trends. We show that these two approaches rest on the same detection mathematics: the occupancy–detection formulation can be linked directly to Frescalo's discovery framework, where occupancy and detectability combine into one rate parameter (which, when sampling is light, closely matches the product of occupancy and per-visit detectability). This connection clarifies how Frescalo's neighbourhood-scale and time corrections function as a coarser-scale analogue of repeat-visit models. By casting Frescalo in occupancy modelling terms, we hope to promote further investigation into the adoption of occupancy model diagnostics, extensions and other tests within Frescalo analyses, improving transparency and rigour when working with less-structured biodiversity data.

Keywords: occupancy models, sampling effort, effort correction, citizen science, unstructured data, Hill numbers

*Corresponding author
Email address: olipes@ceh.ac.uk (Oliver L. Pescott)

1. Introduction

Occupancy-detection models [17] and the Frescalo "local frequency scaling" method [14] both aim to correct raw biological records (i.e. species occurrence) data for imperfect sampling. Classical occupancy models do this at the scale of repeated visits to individual sites, explicitly estimating true presence probabilities (ψ) and detectability (p) via a hierarchical likelihood. Frescalo was designed to work at larger spatio-temporal scales, exploiting emergent patterns of relative frequency in "neighbourhoods" to derive Poisson-process-based scaling factors (α) and species' relative "time-factors" (x) indexing true fluctuations in site occupancy. Given that many datasets lack repeat-visit structure, and/or may exhibit variation in the detection process that is only poorly explained by available covariates [19], understanding how Frescalo recovers effort-adjusted trends from aggregated data can broaden the toolkit of ecologists.

Whilst the place of occupancy-detection models in the quantitative ecologist's armoury is well-established (e.g. MacKenzie et al. [17] has almost 6000 citations according to Google Scholar, May 2025, ~260 per year since 2002, a figure that is almost certainly a large underestimate of actual applications), Frescalo has only seen occasional use by comparison (143 citations, around 11 per year since 2012). This may be due partly to the broader application of occupancy models, covering both small-scale monitoring and applications to less structured data at coarser scales [e.g. 24], but, even so, the scope for the use of Frescalo to derive time trends and other metrics from unstructured data is likely to be larger than currently realised: within the outputs that have utilised the method feature a number of national species distribution Atlases [3, 22, 1], Red Lists [21, 7] and national biodiversity "status" reports [6]. Arguably then, an increase in the familiarity of ecologists with the approach would lead to even more such successful applications.

Although the two model types can appear quite different, Pescott et al. [19] informally suggested that Frescalo could be seen as a type of occupancy-detection model "where an adjustment for overlooked species is made in relation to spatial rather than temporal replication, whilst simultaneously adjusting for variable regional effort". We here show that this suggestion can be formalised due to the two model types' reliance on the same core mathematics of Bernoulli versus Poisson detections [cf. 20]. Below we (1) recall each framework, (2) write down their key equations, and (3) algebraically map one onto the other, demonstrating that Frescalo time trends are based on an implicit occupancy-detection model whose "occupancy" and "visits" are folded into a single site/species discoverability rate parameter λ and standardised neighbourhood effort index $s_{i(N)}$.

2. Occupancy-detection models

2.1. Basic single-species, single-season model

Following MacKenzie et al. [17], at site i for species j let $z_{ij} \sim \text{Bernoulli}(\psi_{ij})$. Conditional on presence ($z_{ij} = 1$) and visits $v = 1, \ldots, V$ with visit-specific detectabilities p_{ijv},

$$y_{ijv} \mid z_{ij} = 1 \sim \text{Bernoulli}(p_{ijv}), \qquad y_{ijv} = 0 \text{ if } z_{ij} = 0 \text{ (for all } v). \tag{1}$$

Assuming conditional independence across visits, the probability of at least one detection across V visits is

$$\Pr(\max_v y_{ijv} = 1) = \psi_{ij}\Big[1 - \prod_{v=1}^{V}(1 - p_{ijv})\Big]. \tag{2}$$

If detectability is homogeneous across visits (i.e. $p_{ijv} \equiv p_{ij}$), this reduces to

$$\Pr(\max_v y_{ijv} = 1) = \psi_{ij}\big[1 - (1 - p_{ij})^V\big]. \tag{3}$$

(Where we below write p without indexing, we mean either the homogeneous case or a model in which p_{ijv} is governed by shared parameters.) The model therefore estimates $\psi_{ij} = \Pr(\text{occupied})$ and the visit-level detection probability $\Pr(\text{detect on a visit} \mid \text{occupied})$ (either as p_{ij} under homogeneity, or via parameters that generate the p_{ijv}). Inference proceeds via the full likelihood over all sites and detection histories.

We also note that it is often convenient to parameterise occupancy on the complementary log-log (cloglog) scale [20],

$$\psi_{ij} = 1 - \exp\{-\exp(\eta_{ij})\}, \tag{4}$$

whose inverse link is $\eta_{ij} = \log[-\log(1 - \psi_{ij})]$. Equivalently, letting $\mu_{ij} = \exp(\eta_{ij}) > 0$ denote a latent Poisson "use" (presence) rate over the closure window, we can write

$$\psi_{ij} = 1 - \exp(-\mu_{ij}). \tag{5}$$

This places occupancy on the same probability-rate mapping used in Frescalo's species "discovery" model (see equation 6 below).

3. Frequency scaling using local occupancy (Frescalo)

3.1. Neighbourhood frequencies

Frescalo [14] pools presence-only data across an areal neighbourhood N around target site i. We denote the observed proportion of neighbourhood sites in which species j was recorded by f_{ij} (in practice this frequency may relate to a weighted neighbourhood as per Hill [14], but this detail is not crucial for what follows). Under a Poisson-process model of species discovery (conditional on presence) with rate λ_{ij} and unknown total neighbourhood-level sampling effort $s_{i(N)}$, one has

$$f_{ij} = 1 - \exp(-\lambda_{ij} s_{i(N)}). \tag{6}$$

Thus λ_{ij} is a combined availability-detectability rate at the neighbourhood scale (cf. the occupancy-detectability collapse in the "Bridging" section below) and

f_{ij} is the marginal probability of ≥ 1 record over effort $s_{i(N)}$. Subsequently, a frequency-weighted neighbourhood index (a single number that rises with sampling "depth" in a neighbourhood)

$$\phi_i = \frac{\sum_j f_{ij}^2}{\sum_j f_{ij}} \tag{7}$$

is standardised to a target value Φ by solving for a site-specific effort multiplier α_i such that

$$\phi_i(\alpha_i) = \frac{\sum_j [1 - (1 - f_{ij})^{\alpha_i}]^2}{\sum_j [1 - (1 - f_{ij})^{\alpha_i}]} = \Phi. \tag{8}$$

Mathematically, Φ is chosen so that every neighbourhood's weighted-mean frequency $\phi_i = \sum_j f_{ij}^2 / \sum_j f_{ij}$ equals Φ. Hill [14] showed that ϕ_i is equivalent to the ratio of the neighbourhood's mean species richness to the 'effective number of common species' (often labelled N_2, the reciprocal of Simpson's index; Hill [13]), which means that ϕ_i isolates neighbourhood sampling intensity from true differences in richness and evenness. By fixing $\phi_i = \Phi$, we align all neighbourhoods to the same effort scale without erasing real ecological differences (Chao and Ricotta [8] note some relevant qualifications concerning this metric type for very species-poor assemblages, but these are unlikely to be important at the scales at which Frescalo is envisaged to be useful).

This process yields the standardised neighbourhood frequencies

$$\tilde{f}_{ij} = 1 - (1 - f_{ij})^{\alpha_i} \tag{9}$$

which are independent of time (i.e. they are calculated with respect to the entire time period under consideration, rather than any subdivisions of this used for trend calculations), and serve as a proxy for the "true" discoverability- or effort-standardised neighbourhood species rank-frequency curve.

3.2. Temporal correction

Within each time period t, one chooses a set of local "benchmark" species [16] and computes the proportion recorded per site and time period (Hill's s_{it}) as an index of site-level recording effort. (Note that there are potentially many ways to choose one's site benchmarks, but Hill [14] proposed a fixed proportion R^* of the standardised neighbourhood species rank-frequency curve after an additional normalisation step involving the division of species' ranks by the expected species count $\sum_j \tilde{f}_{ij}$; however, the precise method of choosing site benchmarks does not affect what follows.) For each species j in period t, Hill defines a baseline Poisson mean (rate) as

$$Q_{ijt} = -\ln[1 - s_{it}\tilde{f}_{ij}], \tag{10}$$

i.e. the cloglog transform of the baseline discovery probability. The modelled discovery probability is then

$$P_{ijt}(x_{jt}) = 1 - \exp(-Q_{ijt}x_{jt}). \tag{11}$$

Hill [14] estimates the time-factor x_{jt} by matching the total modelled to total observed presences y_{ijt}:

$$\sum_i y_{ijt} = \sum_i P_{ijt}(x_{jt}). \tag{12}$$

In practice one can iterate x_{jt} in the exact Poisson form above until those sums coincide (e.g. see the *R* code of Pescott [18]), although analytical solutions are also possible (J.M. Yearsley, pers. comm.). The difference between the (summed) observed presences y_{ijt} and the model's baseline expectation after standardising time-independent neighbourhood effort α_i and adjusting for site/time specific effort s_{it} is therefore captured by the time factor x_{jt}, with $x_{jt} = 1$ corresponding to no temporal deviation from the baseline ($s_{it}f_{ij}$) expectation. Frescalo can thus deliver detection-corrected trends from unstructured data when its core assumptions are met.

4. Bridging the gap

4.1. Static occupancy and detection

We can compare the static (i.e. single-season) single-species occupancy-detection model probability of at least one detection in V visits

$$\psi[1 - (1-p)^V] \tag{13}$$

with the Poisson-process discovery probability (conditional on a species' presence in the all-time frequency curve) used in Frescalo

$$1 - e^{-\lambda s_{i(N)}}. \tag{14}$$

For small p and moderate V (so that pV is small and Vp^2 remains negligible), we may use the standard Taylor series approximation:

$$\ln(1-p) = -p - \frac{p^2}{2} - \frac{p^3}{3} \cdots \approx -p \quad \text{when } p \ll 1. \tag{15}$$

Using the facts that $(1-p)^V = \exp\{V\ln(1-p)\}$, and $\ln(1-p) \approx -p$ (for $p \ll 1$), consequently $(1-p)^V \approx e^{-pV}$. Therefore $\psi[1-(1-p)^V] \approx \psi[1-e^{-pV}]$. On the other hand, setting $\lambda = \psi p$ and $V = s_{i(N)}$, Frescalo's Poisson form $1 - e^{-\psi pV}$ expands in the same way. Recalling the first two terms of the Taylor series for the exponential function, $1 - e^{-x} \approx x - (x^2/2)$, we have

$$1 - e^{-\psi pV} \approx \psi pV - \frac{(\psi pV)^2}{2}, \tag{16}$$

and,

$$\psi(1 - e^{-pV}) \approx \psi\left[pV - \frac{(pV)^2}{2}\right] = \psi pV - \frac{\psi(pV)^2}{2}. \tag{17}$$

Therefore to first order both are $\psi(pV)$ with only $O((pV)^2)$ differences: only quadratic, and higher, terms differ. We recover Frescalo's $1 - e^{-\lambda s_{i(N)}}$ approximately whenever pV is small. For larger pV, the neglected higher-order terms no longer agree so the approximation is lost. However, one can always recover the exact Poisson rate by solving:

$$1 - e^{-\lambda V} = \psi[1 - (1-p)^V] \implies \lambda = -\frac{1}{V}\ln[1 - \psi(1 - (1-p)^V)], \qquad (18)$$

but that formula reduces to $\lambda = \psi p$ in the limit $pV \to 0$.

Frescalo's Poisson rate λ is therefore exactly the function of occupancy, detectability and visit count that makes the first part of equation 18 true [cf. 20]. Whilst in Frescalo we do not use information on V directly, we infer it via the continuous neighbourhood effort index $s_{i(N)}$, standardised across all neighbourhoods by the spatial scaler α_i. Frescalo can therefore be interpreted as an occupancy-detection analogue at the neighbourhood scale: it replaces the two parameters (ψ, p) and known V with a Poisson rate λ and a continuous effort-multiplier (α) equalising variable survey effort (inferred by the neighbourhood-level $s_{i(N)}$) across sites.

Underpinning all of this is the assumption that, within any neighbourhood and time period, species discovery behaves like a Poisson discovery model. It is the assumption which justifies the complementary log-log link in equation 10 above. The frequency-weighted mean ϕ_i of the neighbourhood frequency curve increases monotonically with effort under this discovery model, and, via its link to N_2, allows effort standardisation across neighbourhoods. When sampling intensifies (so that pV is no longer small), the simple $\lambda \approx \psi p$ approximation breaks down, and the nonlinear cloglog mapping then implies the exact relation given in equation 18 above.

A key step in recognising the equivalent elements of these models is to appreciate that Frescalo applies its discoverability standardisation at a large scale: not only is the adjustment done with respect to the multi-site neighbourhood and across all species, but it is also calculated across all time periods in the analysis. The standardised neighbourhood frequencies $\tilde{f}_{ij}$ and the species rank-frequency curve they form is estimated once, independently of time, before temporal change is examined. In contrast to multi-species occupancy models, where cross-species dependence is modelled explicitly via hierarchical community structure—shared site-level random effects and hyperparameters (and sometimes residual covariance/latent factors) that induce correlation among species [9]—cross-species dependence is handled implicitly through shared effort and the benchmark/standardisation steps; Frescalo does not fit any explicit multivariate covariance structure, but relies on the Poisson discovery model per species and the monotone response of ϕ_i to effort.

4.2. Time trend interpretation

A time trend in occupancy derived from a classical occupancy-detection model is modelled simply by letting ψ_{ij} vary linearly or non-linearly over time,

conditional on both model-specific [25, 26] and other standard survey sampling assumptions [5] being reasonable relative to the inferential target. Frescalo, by contrast, posits a single time-independent set of discoverability-adjusted baseline frequencies $\tilde{f}_{ij}$, and then uses benchmarks and the site/period effort index s_{it} to compute standardised frequencies under an assumption of stasis, subsequently letting the time-factors x_{jt} absorb any residual differences as true ecological change.

This underscores a key difference in how effort-adjustment processes function in each model type. Occupancy-detection models assume that true site occupancies, and so trends in these, are directly recoverable from visit-level information; Frescalo assumes that fine-scale visit data is generally unavailable and/or uninformative for all or part of the time series of interest, and so models species' discoverability at a much larger scale. The main aim of this adjustment is to ensure a common scale across which neighbourhoods, and therefore sites, can be compared: without the harmonisation of effort across neighbourhoods, the time-factors estimated for each site for a species would not be comparable, making average time-factors and trends in these meaningless.

Another fundamental difference is the meaning of the site occupancy values produced. As noted, ψ_{ij} has the simple meaning of predicted species' site occupancy under the classical model (notwithstanding debates around usage versus occupancy when these types of models are applied at different scales; [23]). The Frescalo time-factor x_{jt} is, however, defined relative to the benchmark average, and values > 1 or < 1 indicate that a species is at a higher or lower average frequency relative to the common species where it occurs, rather than in absolute occupancy probability. This may be an important limitation to inferring effort via observable recording outcomes, as opposed to having knowledge of those factors that directly map onto effort, such as the actual number of visits and covariates that are known to explain an important portion of observed variance in species' visit-level detectability [14, 24, 15].

One way around this issue is the observation of Bijlsma [2] that site occupancy probabilities can be back-calculated from Frescalo via the combination of the standardised species' frequencies $\tilde{f}_{ij}$, the species' time-factors x_{jt}, and by setting $s_{it} = 1$ across all sites and time periods (i.e. constant effort), and this has been exploited in at least one published analysis [see 10]. However, this requires a note of caution: whilst sensitivity analyses published in Hill [14] suggest that trends in time-factors estimated by Frescalo can be relatively insensitive to the choice of R^*, the benchmark threshold (variation in this parameter changing the intercept of estimated trends but not their slope), the same is not true of back-calculated site occupancy probabilities (P_{ijt} in Frescalo terms). Because the relationship between time-factors and species' frequencies is non-linear, the shifts in time trend intercept seen using different values of R^* will not translate into the same proportional changes in predicted site occupancies over time. This may be particularly important when these trends are used to classify species' into risk categories, as for example happens in Red Listing exercises [e.g. 21, 7].

5. Conclusions

Unstructured species occurrence data are too valuable to ignore, especially for historical periods where little or no information about the visit-level data collection process survives [19, 11]. Hill's "frequency scaling using local occupancy" or Frescalo method allows the careful analyst to infer a large-scale discoverability/effort index that can be used to place neighbourhoods on a common footing for the estimation of distribution trends. The large-scale formulation of this approach not only allows for the potential inclusion of more data sources (e.g. records extracted from Atlases or museums), but may also act to reduce the actual error in species' trends intrinsically [4, 22]. By demonstrating how Frescalo represents the classical occupancy-detection model's ψ and p with λ, and how it infers visit-related effort via an emergent community-level mean rate ϕ, the approach performs an occupancy-detection-type correction even when explicit or informative temporal repeat-visit data are lacking. Beyond this fundamental use-case, Goury et al. [12] have recently highlighted a number of promising uses and extensions of the Frescalo method that could emerge from a greater appreciation of the approach. We hope that the parallels and clarifications described here will support this expansion.

6. Acknowledgements

OLP was supported by the UKCEH National Capability for UK Challenges programme NE/Y006208/1. Thank you also to the two anonymous reviewers whose comments improved the manuscript.

7. Conflict of Interest Statement

None.

References

[1] Balmer, D.E., Gillings, S., Caffrey, B., 2013. Bird Atlas 2007-11: The Breeding and Wintering Birds of Britain and Ireland. BTO Books, Thetford.

[2] Bijlsma, R.J., 2013. The Estimation of Species Richness of Dutch Bryophytes between 1900 and 2011. Technical Report 15. Dutch Bryological and Lichenological Society. Gouda, the Netherlands.

[3] Blockeel, TL., Bosanquet, SDS., Hill, MO., Preston, CD. (Eds.), 2014. Atlas of British and Irish Bryophytes. Pisces Publications, Newbury.

[4] Boyd, R.J., Bowler, D.E., Isaac, N.J.B., Pescott, O.L., 2024. On the trade-off between accuracy and spatial resolution when estimating species occupancy from geographically biased samples. Ecological Modelling 493, 110739. doi:10.1016/j.ecolmodel.2024.110739.

[5] Boyd, R.J., Powney, G.D., Pescott, O.L., 2023. We need to talk about nonprobability samples. Trends in Ecology & Evolution doi:10.1016/j.tree.2023.01.001.

[6] Burns, F., Mordue, S., al Fulaij, N., Boersch-Supan, PH., Boswell, J., Boyd, RJ., Bradfer-Lawrence, T., de Ornellas, P., de Palma, A., de Zylva, P., et al., 2023. State of nature 2023 .

[7] Callaghan, D.A., 2022. A new IUCN Red List of the bryophytes of Britain, 2023. Journal of Bryology 44, 271–389. doi:10.1080/03736687.2023.2185393.

[8] Chao, A., Ricotta, C., 2019. Quantifying evenness and linking it to diversity, beta diversity, and similarity. Ecology 100, e02852. doi:10.1002/ecy.2852.

[9] Devarajan, K., Morelli, T.L., Tenan, S., 2020. Multi-species occupancy models: Review, roadmap, and recommendations. Ecography 43, 1612–1624. doi:10.1111/ecog.04957.

[10] Eichenberg, D., Bowler, D.E., Bonn, A., Bruelheide, H., Grescho, V., Harter, D., Jandt, U., May, R., Winter, M., Jansen, F., 2021. Widespread decline in Central European plant diversity across six decades. Global Change Biology 27, 1097–1110. doi:10.1111/gcb.15447.

[11] Gotelli, N.J., Booher, D.B., Urban, M.C., Ulrich, W., Suarez, A.V., Skelly, D.K., Russell, D.J., Rowe, R.J., Rothendler, M., Rios, N., Rehan, S.M., Ni, G., Moreau, C.S., Magurran, A.E., Jones, F.A.M., Graves, G.R., Fiera, C., Burkhardt, U., Primack, R.B., 2023. Estimating species relative abundances from museum records. Methods in Ecology and Evolution 14, 431–443. doi:10.1111/2041-210X.13705.

[12] Goury, R., Bowler, D.E., Harrower, C., Münkemüller, T., Vallet, J., Yearsley, J., Thuiller, W., Pescott, O.L., 2025. A practical guide to species trend detection with unstructured data using local frequency scaling (Frescalo). ecoEvoRxiv doi:10.32942/X2WS8N.

[13] Hill, M.O., 1973. Diversity and evenness: A unifying notation and its consequences. Ecology 54, 427–432.

[14] Hill, M.O., 2012. Local frequency as a key to interpreting species occurrence data when recording effort is not known. Methods in Ecology and Evolution 3, 195–205.

[15] Kéry, M., Royle, J.A., Schmid, H., Schaub, M., Volet, B., Häfliger, G., Zbinden, N., 2010. Site-Occupancy Distribution Modeling to Correct Population-Trend Estimates Derived from Opportunistic Observations. Conservation Biology 24, 1388–1397. doi:10.1111/j.1523-1739.2010.01479.x.

[16] Latour, J., van Swaay, C., 1992. Dagvlinders als indicatoren voor de regionale milieukwaliteit. De Levende Natuur 93, 19–22.

[17] MacKenzie, D.I., Nichols, J.D., Lachman, G.B., Droege, S., Royle, J.A., Langtimm, C.A., 2002. Estimating Site Occupancy Rates When Detection Probabilities Are Less Than One. Ecology 83, 2248–2255. doi:10.1890/0012-9658(2002)083[2248:ESORWD]2.0.CO;2.

[18] Pescott, O., 2025. An R translation of Hill's Fortran code for Frescalo. doi:10.5281/zenodo.15305437.

[19] 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. doi:10.33928/bib.2019.01.250.

[20] Royle, J.A., Nichols, J.D., 2003. Estimating Abundance from Repeated Presence–Absence Data or Point Counts. Ecology 84, 777–790. doi:10.1890/0012-9658(2003)084[0777:EAFRPA]2.0.CO;2.

[21] Stroh, P., Leach, S., August, T., Walker, K., Pearman, D., Rumsey, F., Harrower, C., Fay, M., Martin, J., Pankhurst, T., Preston, C., Taylor, I., 2014. A Vascular Plant Red List for England. Botanical Society of Britain and Ireland, Bristol.

[22] Stroh, P., Walker, K., Humphrey, T., Pescott, O., Burkmar, R. (Eds.), 2023. Plant Atlas 2020. Mapping Changes in the Distribution of the British and Irish Flora. Botanical Society of Britain and Ireland & Princeton University Press, Princeton.

[23] Valente, J.J., Jirinec, V., Leu, M., 2024. Thinking beyond the closure assumption: Designing surveys for estimating biological truth with occupancy models. Methods in Ecology and Evolution 15, 2289–2300. doi:10.1111/2041-210X.14439.

[24] van Strien, A.J., van Swaay, C.A., Termaat, T., 2013. Opportunistic citizen science data of animal species produce reliable estimates of distribution trends if analysed with occupancy models. Journal of Applied Ecology 50, 1450–1458. doi:10.1111/1365-2664.12158.

[25] Warton, D.I., Stoklosa, J., Guillera-Arroita, G., MacKenzie, D.I., Welsh, A.H., 2017. Graphical diagnostics for occupancy models with imperfect detection. Methods in Ecology and Evolution 8, 408–419. doi:10.1111/2041-210X.12761.

[26] Wright, W.J., Irvine, K.M., Rodhouse, T.J., 2016. A goodness-of-fit test for occupancy models with correlated within-season revisits. Ecology and Evolution 6, 5404–5415. doi:10.1002/ece3.2292.