Improved Estimation of Bird Abundance and Energy Consumption by Combining Point and Transect Observations

Maria Jose Ayala Bolagay[1], Aaron Brothers[2], Marta Quitián[3], Eike Lena Neuschulz[4], Matthias Schleuning[4], Katrin Böhning-Gaese[4], Orlando Alvarez[1], Vinicio Santillán[1], and Christopher Thron[*2]

[1]Universidad Católica de Cuenca, Ecuador
[2]Texas A&M University-Central Texas, Killeen TX USA
[3]Instituto Mediterráneo de Estudios Avanzados (CSIC-UIB)
[4]Senckenberg Biodiversity and Climate Research Centre, Frankfurt Germany

April 16, 2023

Abstract

Point counts and transect surveys are both common methods for estimating bird abundance (by species, genus, or family) in a given environment. Sometimes both methods are used in the same study. Point counts reflect abundance directly, while transect observations of bird feeding instances reflect both the bird abundance and the mean energy consumption rate per bird.

In this research we develop an improved estimator of bird abundance based on both sets of measurements, and derive error bars for these abundance estimates. The improved estimator is a linear combination of the two individual abundance estimates obtained from the two data types. The combination weights are obtained through a comparison of two separate energy estimates. The mathematical methods developed are applied to a dataset consisting of point counts and transect surveys of Ecuadorian frugivore bird species in six different types of environment (three elevations and two degrees of forestation). The improvement in accuracy achieved by this combined estimator is verified by simulations applied to bird family data. We also obtain 90% confidence intervals for the bird family relative abundance estimates, and verify the confidence intervals using simulations.

Keywords: Energy consumption, Abundance, Estimation, Survey, Point counts, Transect sampling, Bird, Ecosystem, Frugivore, Species, Family, Simulation

[*]Corresponding author: thron@tamuct.edu

1 Introduction

Two common methods for estimating abundance of bird types (whether species, genera, or families) in a given area are point counts and transect surveys. Point counts are obtained when the observer chooses a set of fixed observation points within the area of interest, and counts bird detections (by sight or by sound) during multiple sessions over an extended period of several days or months. Transect surveys are conducted by the observer specifying a transect (i.e. narrow section) within the area and repeatedly walking through and recording bird observations at the same time of day over an extended period. Transect measurements can target bird feedings, so that the frequency of feeding on each different type of plant is recorded.

Each of the two survey types has distinct advantages and limitations. Point counts are straightforward to conduct and are effective for detecting a broad cross-section of species at fixed locations, but they do not directly record feeding behavior. Transect surveys, by contrast, capture behavioral data including feeding events and plant-bird interactions, but coverage of the survey area depends on the observer's path and pace, which can introduce systematic biases. Both methods have been widely adopted because neither alone provides a complete picture of bird community structure and energy dynamics; combining them allows investigators to cross-validate abundance estimates and extract behavioral inferences that neither method supports independently.

The energy consumption of different bird types can be expressed in two different ways. The total energy consumption of a bird type can be expressed as the number of individuals of that type times the mean consumption per individual. On the other hand, the same quantity may be expressed as the total number of feedings by birds of that type times the amount consumed per feeding. Point counts can be used to estimate bird abundance required for the first expression, while transect surveys can provide feeding data for the second expression.

We may make use of these two alternative estimates of energy consumption in several different ways. First, by comparing the two estimates, we may gain information about bird consumption rates. Second, we may combine the two estimates to obtain a more accurate estimate of energy consumption per bird type, and hence of bird type relative abundance. Finally, given that count measurements are Poisson distributed (which is commonly recognized as valid for count measurements), we can use the results obtained and simulations to obtain confidence intervals for the relative abundance of different bird types.

The methodologies that we have developed may be applied to other surveys where multiple forms of data related to species abundance are present. Our results can be used both to obtain more accurate abundance estimates, and to rigorously specify accuracies of the estimates thus obtained. We have also demonstrated a technique for estimating an important biological parameter, namely the exponent in the functional dependence of bird energy consumption on bird weight. Using this technique, other datasets may provide similar estimates.

The remainder of this paper is organized as follows. Section 2 surveys previous similar studies. Section 3 consists of two parts. In the first part, we present the theoretical derivations for the statistical methods used for the data analysis. These methods include the construction of point count-based and transect observation-based estimators for energy consumption and for bird abundance; and the combination of bird abundance estimators and log abundance estimators into more accurate estimators. In the second part, the methods used to evaluate the theoretical results are described. These methods include the use of an empirical dataset obtained from several regions in southern Ecuador; this dataset is used both to estimate model parameters from bird species data, and as a basis for simulations to test the accuracy of the different theoretical estimators using bird family data. In Section 4 the results of the evaluation are presented; all figures displaying these results are placed at the end of the paper, after the bibliography. Section 5 discusses the implications of our results in relation to prior work. Finally, in Section 6 our conclusions from this study are summarized.

2 Previous studies

Frugivore bird species play a critical role in tropical and subtropical ecosystems, functioning as primary agents of seed dispersal and thereby enabling plant population regeneration, maintaining plant diversity, and sustaining ecological balance [1]. The ecological importance of these interactions has motivated a substantial body of empirical research across diverse geographic regions. Studies in the Neotropics, Africa, and Asia have documented that frugivore bird assemblages differ markedly in composition and foraging behavior across habitats and elevational gradients, and that the intensity of fruit consumption and seed dispersal is closely tied to local bird abundance [2, 3, 4, 5, 6, 7, 8]. Several of these studies have specifically examined the relationship between bird body size and dietary preferences, noting that larger frugivores tend to consume larger fruit and disperse seeds over greater distances, while smaller species are restricted to smaller fruit and contribute to fine-scale seed rain patterns [7, 8]. Work in Andean systems in particular has highlighted how frugivory and seed dispersal effectiveness vary with elevation and habitat fragmentation, with fragmented landscapes typically supporting depauperate communities of large-bodied frugivores and altered plant-bird interaction networks [6, 5]. Understanding how bird abundance responds to these environmental gradients is therefore essential for predicting the cascading effects of habitat change on ecosystem function.

Accurately quantifying bird abundance in natural environments requires careful choice of census methodology, and the merits of different field approaches have been extensively debated in the literature [9, 10, 11]. Point count methods, in which observers record all detections within a fixed radius or time window from a stationary position, are widely used because they are standardizable, repeatable, and efficient over large spatial extents. However, point counts depend heavily on observer skill and may miss secretive or cryptic species. Transect

surveys, in which observers walk a predetermined route and record detections along the way, are better suited for capturing behavioral information such as feeding events and plant-bird interactions, but require careful control of walking pace and can introduce distance-related detection biases. Because the two methods sample different aspects of bird community activity, combining them in a single study can provide richer information than either method alone. This complementarity motivates the analytical framework developed in the present paper, which explicitly links the two data types through a shared model of energy consumption, allowing each to inform and constrain the other.

Statistical methods play an increasingly important role in extracting reliable estimates from field survey data, particularly when sample sizes are modest or data are missing. Bayesian approaches offer a principled framework for incorporating prior biological knowledge—such as the expected dependence of metabolic rate on body mass—into parameter estimation, and for propagating uncertainty through hierarchical models of abundance and detection [12, 13]. The methods proposed in the present paper operate within this general philosophy: by exploiting the theoretical relationship between point counts and transect feeding observations, we derive estimators that borrow strength from both data types, and we use simulation-based bootstrapping to characterize estimation uncertainty without requiring distributional assumptions that may not hold in practice. This approach yields abundance estimates with formally justified confidence intervals, contributing to a more rigorous quantification of bird community structure than is possible from either data source alone.

3 Methodology

This section consists of two subsections dealing with theoretical and empirical methods, respectively. The first section derives statistical estimators that make use of point counts and transect observations to estimate model parameters and demographics. The second section describes the application of these theoretical results to a specific ecological survey.

3.1 Theoretical equations

3.1.1 Transect observation-based expression for energy consumption per bird type

The energy consumption per bird type over a given period of time at a given location may be expressed as the total number of feedings times the mean energy transfer per feeding. We may rewrite this statement in the form of an equation:

$$\mathcal{E}_n = \sum_{m=1}^{N} \mathcal{F}_{nm}\mathcal{T}_{nm} \qquad (n = 1 \dots N, m = 1 \dots M), \tag{1}$$

where

- M, N: Number of fruit and bird types, respectively.
- $\mathcal{E}_n$: Total energy consumption of all birds of type n at the specified location during the specified period.
- $\mathcal{F}_{nm}$: Total number of feedings of bird type n on fruit type m; $\mathcal{T}_{nm}$: Mean energy transfer per feeding for bird type n on fruit type m.

Under the assumption that the transect observations are representative and capture the same proportion of all bird-fruit interactions, we have:

$$X_{nm} = K_{\mathcal{F}}\mathcal{F}_{nm} - u_{nm} \qquad (n = 1 \dots N, m = 1 \dots M), \tag{2}$$

where

- $K_{\mathcal{F}}$: Constant of proportionality, assumed to be independent of bird and fruit types;
- X_{nm}: Number of transect observations of bird type n consuming fruit type m;
- u_{nm}: Observational uncertainty for bird type n consuming fruit type m.

We furthermore assume initially that energy transfer per feeding depends on both the mean fruit weight and the mean bird weight. This assumption presumes that larger fruit may require less foraging, and larger birds may be able to consume more per feeding. This relationship is summarized in the equation:

$$\mathcal{T}_{mn} = K_{\mathcal{T}} f_m^{p_f} w_n^{p_{b1}} \qquad (n = 1 \dots N), \tag{3}$$

where

- $K_{\mathcal{T}}$: Constant of proportionality, assumed to be independent of bird type.
- f_m: mean fruit weight for type m.
- w_n: Mean weight of bird type n;
- p_f, p_{b1}: Exponents of fruit weight and bird weight respectively, in power model.

Substituting $\mathcal{F}_{nm}$ and $\mathcal{T}_{nm}$ in (1) with their expressions from (2) and (3) respectively, we have finally:

$$\mathcal{E}_n = \kappa_1 \sum_{m=1}^{M} (X_{nm} + u_{nm}) f_m^{p_f} w_n^{p_{b1}} \qquad (n = 1 \dots N), \tag{4}$$

where $\kappa_1 \equiv (K_{\mathcal{F}} K_{\mathcal{T}})^{-1}$ is a constant of proportionality that is independent of bird and fruit types. In order to simplify the model and make it amenable to analysis, we replace f_m in (4) with $\bar{f}_n$, which is defined as the mean fruit weight consumed by bird species n. This simplification will be accurate when

each bird species mostly consumes fruit species with similar average weights. An additional simplification concerns the magnitude of the measurement errors u_{nm}. We assume that these errors are nearly constant in percentage across species: that is to say, the error for species n will be roughly proportional to the number of observations X_n.

With the two simplifications described above, the final simplified model is given by the equation:

$$\mathcal{E}_n = \kappa_1 X_n (1 + u_n)(\bar{f}_n)^{p_f} (w_n)^{p_{b1}}, \tag{5}$$

where

- X_n: Sum of transect observations for bird type n. The formula for X_n is $X_n = \sum_{m=1}^{M} X_{nm}, n = 1 \ldots N$.
- u_n: Combined proportionate observational uncertainty for bird type n. The variables $\{u_n\}, n = 1, \ldots N$ are modeled as independent, identically distributed random variables with mean 0 and variance $\sigma^2_{X,n}$;
- $\bar{f}_n$: Average weight of fruit consumed by bird type n, given by the formula:

$$\bar{f}_n \equiv \frac{\sum_{m=1}^{M} X_{nm} f_m}{\sum_{m=1}^{M} X_{nm}}. \tag{6}$$

3.1.2 Point count-based estimate of energy consumption per bird type

A second expression for a bird type's energy consumption at a given location during a prescribed period is the product of the number of birds times the mean energy intake per bird, as expressed by the following formula:

$$\mathcal{E}_n = \mathcal{N}_n \mathcal{M}_n \qquad n = 1 \ldots N, \tag{7}$$

where $\mathcal{N}_n$ and $\mathcal{M}_n$ represent respectively the total number and mean energy consumption of birds of type n.

As with transect observations, we assume that point counts are representative samples across all bird types, so we have:

$$Z_n = K_{\mathcal{N}} \mathcal{N}_n - \tilde{v}_n \qquad n = 1 \ldots N, \tag{8}$$

where

- $K_{\mathcal{N}}$: Constant of proportionality, assumed to be independent of bird types;
- Z_n: Number of point count observations of bird type n, $n = 1 \ldots N$;
- $\tilde{v}_n$: Uncertainty in point counts for bird type n.

As with transect observations, we assume that the uncertainty is proportional to the number of observations:

$$\tilde{v}_n = Z_n v_n \qquad n = 1 \dots N, \tag{9}$$

where $\{v_n, n = 1 \dots N\}$ are independent, identically distributed random variables with mean 0 and variance $\sigma^2_{Z,n}$.

In the literature, bird type mean energy consumption is modeled as proportional to mean bird weight raised to a power as follows [14]:

$$\mathcal{M}_n \equiv K_{\mathcal{M}} w_n^{p_{b2}}, \tag{10}$$

where

- $K_{\mathcal{M}}$: Constant of proportionality, independent of bird type n;
- w_n: Mean weight of bird type n, $n = 1, \dots N$;
- p_{b2}: Weight exponent in energy model.

By substituting (8)-(10) into (7) we obtain finally:

$$\mathcal{E}_n = \kappa_2 Z_n (1 + v_n)(w_n)^{p_{b2}}, \tag{11}$$

where $\kappa_2 \equiv K_{\mathcal{M}}/K_{\mathcal{N}}$ is independent of n.

3.1.3 Parameter estimation based on energy consumption equations

In practice, equations (5) and (11) developed in the previous two sections cannot be used to estimate total energy consumption of bird types, because the constants κ_1 and κ_2 are impossible to obtain. However, by setting the two expressions equal to each other we obtain a relationship that enables us to estimate the exponents p_{b1}, p_{b2} and p_f based on observed point count and transect data. Equating $\mathcal{E}_n$ in (5) and (11) gives:

$$\kappa_1 X_n (1 + u_n)(\bar{f}_n)^{p_f} (w_n)^{p_{b1}} = \kappa_2 Z_n (1 + v_n)(w_n)^{p_{b2}}. \tag{12}$$

Taking logarithms and rearranging, we find:

$$\log(X_n) = \log(Z_n) + C_0 + p_b \log(w_n) - p_f \log(\bar{f}_n) + e_n, \tag{13}$$

where:

$$\begin{aligned} C_0 &\equiv \log\left(\frac{\kappa_2}{\kappa_1}\right); \\ p_b &\equiv p_{b2} - p_{b1} \\ e_n &\equiv \log\left(\frac{1 + u_n}{1 + v_n}\right) \text{(combined error term)} \end{aligned} \tag{14}$$

Note that if $u_n, v_n \ll 1$ then $e_n \approx u_n - v_n$. We will assume that $\{e_n, n = 1, ...N\}$ are independent, identically distributed random errors. Then if we have transect data $\{X_1, \dots, X_N\}$ and point count data $\{Z_1, \dots, Z_N\}$, we may use (13) to perform ordinary unweighted least squares regression and find values for C_0, p_b, and p_f that give the best fit to the data.

3.1.4 Estimates of relative abundance

In this section we make use of the results from previous sections to develop a pair of estimates for relative abundances of bird types, where the relative abundance a_n of species n is defined as type abundance divided by total bird abundance:

$$a_n = \frac{\mathcal{N}_n}{\sum_{k=1}^{N} \mathcal{N}_k}. \tag{15}$$

The two estimates are based on point counts and transect observations, respectively. We also show that in the case where the estimators are unbiased and uncorrelated, a linear combination of the two estimators can be formed that is unbiased and has smaller variance than either of the two individual estimators.

First we develop an estimator for relative abundance based on point counts. From (8) and (9) and the definition (15) we have:

$$\begin{aligned} a_n &= \frac{Z_n(1+v_n)}{\sum_{k=1}^{N} Z_k(1+v_k)} \\ &= \frac{Z_n}{\sum_{k=1}^{N} Z_k}(1+v_n) \cdot \left(1 + \frac{\sum_{k=1}^{N} v_k Z_k}{\sum_{k=1}^{N} Z_k}\right)^{-1}. \end{aligned} \tag{16}$$

The final multiplicative term in (16) is independent of n, and involves the sum of independent random errors, which will tend to cancel each other. We may therefore approximate this term as:

$$\left(1 + \frac{\sum_{k=1}^{N} v_k Z_k}{\sum_{k=1}^{N} Z_k}\right)^{-1} \approx (1-\nu)^{-1} \approx 1+\nu, \tag{17}$$

where $\nu \ll 1$ is a mean-zero random variable. This gives finally

$$\begin{aligned} a_n &\approx \frac{Z_n}{\sum_{k=1}^{N} Z_k}(1+v_n)(1+\nu) \\ &\approx \frac{Z_n}{\sum_{k=1}^{N} Z_k}(1+\nu+v'_n), \end{aligned} \tag{18}$$

where $\{v'_n = v_n(1+\nu), n = 1, \ldots N\}$ are statistically independent errors. It follows that an estimator $\widehat{a}_n$ for a_n that is based on point counts $\{Z_n\}$ can be obtained from (18) with the random error terms removed:

$$(\widehat{a}_n)_Z \equiv \frac{Z_n}{\sum_{k=1}^{N} Z_k}. \tag{19}$$

Next, an alternative expression for a_n based on transect observations can be obtained. Combining (8), (9) and rearranging, we have:

$$Z_n(1+v_n) \equiv K_{\mathcal{N}} \mathcal{N}_n \qquad n = 1 \ldots N, \tag{20}$$

and substituting into (12) and rearranging gives:

$$\mathcal{N}_n = \frac{\kappa_1}{\kappa_2 K_{\mathcal{N}}} X_n (1+u_n)(\bar{f}_n)^{p_{f,n}} (w_n)^{-p_{b,n}}. \tag{21}$$

(Note that we are generalizing (12) to the case where the fruit weight and bird weight exponents $p_{f,n}$ and $p_{b,n}$ depend on bird type.) This enables us to give an expression for relative abundance in terms of X_n instead of Z_n:

$$a_n = \frac{\mathcal{N}_n}{\sum_{k=1}^{N} \mathcal{N}_k} = \frac{X_n (1+u_n)(\bar{f}_n)^{p_{f,n}} (w_n)^{-p_{b,n}}}{\sum_{k=1}^{N} X_k (1+u_k)(\bar{f}_k)^{p_{f,n}} (w_k)^{-p_{b,n}}}. \tag{22}$$

By a similar argument as before, we obtain the following X_n-based estimator for relative abundance a_n:

$$(\widehat{a}_n)_X \equiv \frac{(\bar{f}_n)^{\widehat{p}_{f,n}} (w_n)^{-\widehat{p}_{b,n}} X_n}{\sum_k (\bar{f}_k)^{\widehat{p}_{f,n}} (b_k)^{-\widehat{p}_{b,n}} X_k} \tag{23}$$

where $\widehat{p}_{f,n}, \widehat{p}_{b,n}$ are estimates for the fruit and bird weight coefficients $p_{f,n}$ and $p_{b,n}$, respectively.

3.1.5 Minimum variance combination of relative abundance estimators

Now consider the linear combination of estimators:

$$(\widehat{a}_n)_{\beta_n,\gamma_n} \equiv \beta_n (\widehat{a}_n)_Z + \gamma_n (\widehat{a}_n)_X, \tag{24}$$

where $\{\beta_k, \gamma_k, k = 1 \ldots N\}$ are positive coefficients. It follows by the linearity of distribution mean that:

$$\text{mean}\left[(\widehat{a}_n)_{\beta_n,\gamma_n}\right] = \beta_n \text{mean}\left[(\widehat{a}_n)_Z\right] + \gamma_n \text{mean}\left[(\widehat{a}_n)_X\right]; \tag{25}$$

In the case where the estimators $(\widehat{a}_n)_Z$ and $(\widehat{a}_n)_X$ are unbiased, we have:

$$\text{mean}\left[(\widehat{a}_n)_Z\right] = \text{mean}\left[(\widehat{a}_n)_X\right] = a_n; \tag{26}$$

The requirement that $(\widehat{a}_n)_{\beta_n,\gamma_n}$ be an unbiased estimator of a_n is expressed mathematically as:

$$\text{mean}\left[(\widehat{a}_n)_{\beta_n,\gamma_n}\right] = a_n. \tag{27}$$

Combining (25), (26), and (27) gives the requirement:

$$\gamma_n = 1 - \beta_n. \tag{28}$$

Given that $(\widehat{a}_n)_Z$ and $(\widehat{a}_n)_X$ are based on independent observations and are thus uncorrelated, it follows:

$$\text{Var}\left[(\widehat{a}_n)_{\beta_n,\gamma_n}\right] = \beta_n^2 \text{Var}\left[(\widehat{a}_n)_Z\right] + \gamma_n^2 \text{Var}\left[(\widehat{a}_n)_X\right], \tag{29}$$

and a standard calculation shows that the unbiased minimum-variance estimator for a_n is obtained by choosing:

$$\gamma_n \equiv (1-\beta_n); \quad \beta_n \equiv \frac{\mathrm{Var}[(\widehat{a}_n)_X]}{\mathrm{Var}[(\widehat{a}_n)_Z] + \mathrm{Var}[(\widehat{a}_n)_X]}, \tag{30}$$

which has variance

$$\min_{\beta_n} \mathrm{Var}\left[(\widehat{a}_n)_{\beta_n, 1-\beta_n}\right] = \frac{\mathrm{Var}[(\widehat{a}_n)_X]\mathrm{Var}[(\widehat{a}_n)_Z]}{\mathrm{Var}[(\widehat{a}_n)_Z] + \mathrm{Var}[(\widehat{a}_n)_X]}. \tag{31}$$

3.1.6 Combined estimates of log relative abundance

Sometimes it is more advantageous to consider the logarithms of relative abundances of bird species, genera, or families, instead of relative abundances themselves. Typically the logarithms have a more symmetrical distribution which is closer to normal. By taking natural logarithms of $(\widehat{a}_n)_Z$ and $(\widehat{a}_n)_X$, we obtain estimators for the logarithm of relative abundance, which we denote as $(\widehat{\ell}_n)_Z$ and $(\widehat{\ell}_n)_X$ respectively:

$$\begin{aligned}
(\widehat{\ell}_n)_Z &\equiv \log\left((\widehat{a}_n)_Z\right) = \log Z_n - \log\left(\sum_k Z_k\right); \\
(\widehat{\ell}_n)_X &\equiv \log\left((\widehat{a}_n)_X\right) \\
&= \log X_n - \widehat{p}_{b,n}\log w_n - \widehat{p}_{f,n}\log f_n - \log\left(\sum_k (f_k)^{-\widehat{p}_{f,n}}(b_k)^{-\widehat{p}_{b,n}} X_k\right).
\end{aligned} \tag{32}$$

As with the relative abundance estimators, we may form linear combinations of the log relative abundance estimators:

$$(\widehat{\ell}_n)_{\widetilde{\beta}_n, \widetilde{\gamma}_n} \equiv \widetilde{\beta}_n(\widehat{\ell}_n)_Z + \widetilde{\gamma}_n(\widehat{\ell}_n)_X, \tag{33}$$

where $\{\widetilde{\beta}_k, \widetilde{\gamma}_k, k = 1 \ldots N\}$ are positive coefficients. In the case where $(\widehat{\ell}_n)_Z$ and $(\widehat{\ell}_n)_X$ are unbiased and uncorrelated, it follows as before that:

$$\mathrm{Var}\left[(\widehat{\ell}_n)_{\widetilde{\beta}_n, \widetilde{\gamma}_n}\right] = (\widetilde{\beta}_n)^2 \mathrm{Var}\left[(\widehat{\ell}_n)_Z\right] + (\widetilde{\gamma}_n)^2 \mathrm{Var}\left[(\widehat{\ell}_n)_X\right], \tag{34}$$

and an unbiased minimum-variance estimator for ℓ_n is obtained by choosing:

$$\widetilde{\gamma}_n \equiv (1-\widetilde{\beta}_n); \quad \widetilde{\beta}_n \equiv \frac{\mathrm{Var}[(\widehat{\ell}_n)_X]}{\mathrm{Var}[(\widehat{\ell}_n)_Z] + \mathrm{Var}[(\widehat{\ell}_n)_X]}. \tag{35}$$

3.1.7 Calculation of coefficients for combined estimators

The combined estimators of relative abundance and log relative abundance defined in Sections 3.1.5 and 3.1.6 respectively require the specification of coefficients $\{\beta_n, \widetilde{\beta}_n, n = 1 \ldots N\}$. These coefficients are not determined a priori, but

rather must be supplied by the user. In this section, we describe a simulation-based procedure for determining coefficients which will give estimators that have reduced variance.

The proposed estimation procedure for β_n is based on a bootstrapping approach, and runs as follows. Suppose the user has recorded point counts $z'_1, \dots z'_N$ and transect observations $x'_1, \dots x'_N$ for N different species.

1. Compute abundance estimates based on $\{z'_1 \dots z'_N\}$, according to (19):

$$(\widehat{a}_n)_{z'} \equiv \frac{z'_n}{\sum_{k=1}^N z'_k}. \tag{36}$$

2. Generate M simulated sets of point counts and transect observations $(z''_{1m} \dots z''_{Nm}, x''_{1m} \dots x''_{Nm}), m = 1 \dots M$ based on the user's observed data. Simulation values are independently generated according to the following distributions:

$$z''_{nm} \sim Pois(z'_n); \qquad x''_{nm} \sim Pois\left(\frac{z'_n(\bar{f}_n)^{p_{f,n,m}} w_n^{p_{b,n,m}}}{\sum_{k=1}^N z'_k(\bar{f}_k)^{p_{f,n,m}} w_j^{p_{b,n,m}}} \sum_{k=1}^N x'_k\right), \tag{37}$$

   where $Pois(\lambda)$ denotes a Poisson random variable with mean event rate λ; and ${p'}_{f,n,m}$ and ${p'}_{b,n,m}$ are randomly generated according to probability distributions reflecting the user's knowledge of fruit and weight exponents in the model (3). (Note that Poisson distributions are commonly used to model survey counts [15].) This expression generates simulated point counts with expected values equal to the observed point counts $z'_1 \dots z'_N$, and simulated transect observations whose expected values yield energy consumption rates that agree with those calculated from the point counts, and whose sum agrees with the sum of the user's observed transect observations.

3. For each of the M simulated sets of observations, the user calculates abundance estimators according to (19):

$$(\widehat{a}_{nm})_{z''} \equiv \frac{z''_{nm}}{\sum_k z''_{km}}; \qquad (\widehat{a}_{nm})_{x''} \equiv \frac{(\bar{f}_n)^{\bar{p}_{f,n}}(w_n)^{-\bar{p}_{b,n}} x''_{nm}}{\sum_k (\bar{f}_k)^{\bar{p}_{f,n}}(b_k)^{-\bar{p}_{b,n}} x''_{km}}. \tag{38}$$

4. The user calculates values of $\beta_n, n = 1 \dots N$ that minimize root mean squared error (RMSE) of the combined estimator $(\widehat{a}_n)_\beta \equiv \beta_n(\widehat{a}_n)_Z + (1 - \beta_n)(\widehat{a}_n)_X$:

$$\beta_n = \text{argmin}_\beta \sqrt{\frac{1}{N}\sum_{m=1}^M \left(\beta(\widehat{a}_{nm})_{z''} + (1-\beta)(\widehat{a}_{nm})_{x''} - (\widehat{a}_n)_{z'}\right)^2}. \tag{39}$$

The minimized RMSE values in (39) can also be used to estimate confidence intervals for $(\widehat{a}_n)_\beta$, since the RMSE value should approximate one standard deviation. So if the RMSE value obtained is denoted as $\Delta_{\beta,n}$, then $(\widehat{a}_n)_\beta \pm 1.65\Delta_{\beta,n}$ should give a 90% confidence interval for the true value of a_n.

A similar procedure can be used to determine values of $\widetilde{\beta}_n, n = 1 \ldots N$ and confidence intervals for an optimal linear combination of log abundance estimators, defined as $(\widehat{\ell}_n)_{\widetilde{\beta}} \equiv \widetilde{\beta}_n(\widehat{\ell}_n)_Z + (1 - \widetilde{\beta}_n)(\widehat{\ell}_n)_X$.

3.2 Empirical evaluation and verification of theoretical equations

In this section, we describe how the equations derived in Section 3.1 are evaluated and verified using data from an actual survey of frugivore birds conducted in several locations in Ecuador. Both the data collected and the verification methods applied are described.

3.2.1 Dataset used in methodology evaluation

Data was collected from 18 one-hectare plots spread across three elevations: approximately 1000 m above sea level (asl) (970–1092 ± 47.2 m, 4°6′S, 78°58′W); 2000 m asl (2025–2119 ± 32.4 m, 3°58′S, 79°4′W); and 3000 m asl (2874–2966 ± 32.9 m, 4°6′S, 79°10′W). These plots represented two types of habitat—natural forest and fragmented forest—within and around Podocarpus National Park (PNP) and the San Francisco Biological Reserve (BRSF) on the southeastern slope of the Andes in southern Ecuador. Podocarpus National Park is recognized as one of Ecuador's most biodiverse protected areas and encompasses a steep elevational gradient supporting distinct vegetation zones, from evergreen premontane forest at low elevations through evergreen lower montane forest at mid elevations to upper montane forest dominated by *Weinmannia* shrubs, vascular epiphytes, and epiphytic mosses at high elevations [16, 17]. The fragmented sites were located in working landscapes surrounding the park, where the primary land use separating forest patches consists of cattle pasture and small-scale agriculture, creating a mosaic of remnant forest fragments embedded in open matrix habitat. The six sites are characterized in Table 1. Study plots were selected jointly by the German Platform for Biodiversity and Ecosystem Monitoring and Research in South Ecuador (RESPECT) to ensure representative coverage of each forest type across the elevational gradient. The area exhibits a humid tropical montane climate [18] with a bimodal rainfall pattern, with peak precipitation from May through June and minimum precipitation from October through November [19].

Location	Approx. altitude (m)	Environment type
Bombuscaro	1000	Natural
Copalinga	1000	Fragmented
ECSF	2000	Natural
Finca	2000	Fragmented
Cajanuma	3000	Natural
Bellavista	3000	Fragmented

Table 1: Summary of characteristics for the six study locations. Natural sites are located within protected forest; fragmented sites are located in working landscapes where remnant forest patches are separated by cattle pasture and small-scale agriculture.

Point count surveys. To assess bird diversity and abundance, nine observation points were established within each one-hectare plot, spaced at least 50 m apart to minimize double-counting. At each point, a trained observer recorded all bird species detected by sight or sound within a 20-m radius during a fixed 10-minute observation window [20]. Surveys were conducted during the first three hours after sunrise, when bird activity is highest, and were suspended during rain or strong wind. Each plot was surveyed eight times over a two-year period ($\frac{1}{6}$ h $\times$ 9 points $\times$ 8 replicates $\times$ 18 plots = 216 total sampling hours). To minimize observer bias, all surveys at a given plot were conducted by the same trained observer. The total number of bird species (species richness) and overall number of individual birds (abundance) at each plot were quantified by aggregating records across all observation points and replicates. The point count dataset comprises a total of [N_{pc}] individual detection records across all plots and replicates. Bird species were classified as frugivores if their diet consisted of 40% or more fruit [21].

Transect surveys. Transect surveys were conducted along fixed linear transects of 200 m established within each plot. Observers walked each transect at a slow, steady pace (approximately 1 km/h) during the same morning activity window used for point counts. All frugivore feeding events observed within 10 m of the transect centerline were recorded, including the identity of the bird (to species level where possible, otherwise to genus or family), the plant species being consumed, and the number of fruit items ingested per feeding bout. Each transect was surveyed on the same eight occasions as the point count plots. The transect dataset comprises a total of [N_{tr}] feeding event records. Mean body masses for each bird species or family were obtained from published sources [21], and mean fruit masses for each plant species consumed were obtained from field measurements or the literature.

3.2.2 Parameter estimation using energy models

Best fits for the parameters p_f and p_b were found by applying linear regression to (13), for several different datasets. Fits were found for species data from each of the six regions separately; for species data grouped by altitude; for species data grouped by environment type (natural versus fragmented); and for species data for all six regions combined. For each fit, scatter plots were made for actual versus predicted transect observations, based on (13). R^2 and Akaike information criterion (AIC) values and 95 percent confidence intervals for p_f and p_b were also computed. We use AIC rather than likelihood-ratio tests or p-values for model comparison because AIC penalizes model complexity in a way that is directly interpretable in terms of information loss, and because the one- and two-parameter models compared here are not strictly nested in a manner that satisfies the assumptions of a formal likelihood-ratio test; a smaller AIC value indicates that the added parameter does not justify the increased model complexity [22].

Scatter plots, R^2 and AIC values and 95 percent confidence intervals for p_b were also generated for the same datasets, but with a simplified single-parameter model that assumes $p_f = 0$ in (13). The 2-parameter and 1-parameter models are compared for goodness of fit. The purpose is to determine whether p_f has a significant effect on mean energy consumption per feeding. All computations were performed in Python (version 3.9) using the `NumPy`, `SciPy`, and `statsmodels` libraries. Simulation code implementing the bootstrapping procedure described in Section 3.1.7 is available at [repository URL].

3.2.3 Performance evaluation of abundance estimation using empirically-based simulations

Statistical characteristics of the abundance and log abundance estimators defined in Sections 3.1.4 and 3.1.6 were evaluated using simulations as follows. In this evaluation, the purpose was not to estimate true relative abundances of the regions from which the dataset was taken—since we don't know these true relative abundances, the accuracy of such estimates cannot be ascertained. Rather, the regional data is used to supply realistic scenarios; the regional data from the six regions is used as "ground truth" for six hypothetical regions, and simulations are randomly generated based on the data, and used to predict this "ground truth". In this way we can evaluate the accuracy of estimators using simulations with realistic characteristics. We base simulations on per-family data instead of per-species data, because per-species point counts are too small to establish differences in accuracies between estimators.

Testing for estimator accuracy and precision

Simulations were conducted to determine the accuracy and precision of point count-based, transect observation-based, and combined estimators as follows.

10000 simulated scenarios were generated based on the bird family data from each region. The data from the six actual regions was used to generate "ground

truth" for six hypothetical regions as follows. If $PC_1, PC_2, \ldots PC_N$ are the observed point counts in a given actual region, then the true relative abundance of family n in the corresponding hypothetical region is given by:

$$a_n = \frac{PC_n}{\sum_{j=1}^{N} PC_j}. \tag{40}$$

Next, we simulated 10,000 surveys in each hypothetical region, where each survey includes point counts and transect observations. For a given simulation, the point count for each bird family n is generated by a Poisson random variable whose mean is the point count from the observed data. We denote these Poisson random variables as $z_n, n = 1, \ldots N$, where

$$z_n \sim Pois(PC_n). \tag{41}$$

Transect observations for each simulation are randomly generated as follows. First, for each bird family n we generate a weight exponent $p_{b,n}$ as a Gaussian random variable with 95% confidence interval $[-0.6, -0.1]$:

$$p_{b,n} \sim Gauss(\mu = -0.35, \sigma^2 = 0.125^2). \tag{42}$$

The choice of distribution (42) will be motivated in Section 4.1 (see the final line of Table 3): this represents the uncertainty in the parameter p_b which is estimated as described in Section 3.2.2. The transect observation for birds in family n (denoted as x_n) is then simulated as follows:

$$x_n \sim Pois\left(PC_n w_n^{p_{b,n}} \frac{\sum_{j=1}^{N} TO_j}{\sum_{j=1}^{N} PC_j w_j^{p_{b,j}}}\right), \tag{43}$$

where TO_j denotes the number of transect observations of bird family j in the original data ($j = 1 \ldots N$). The choice of Poisson parameter in (43) was made so that the expected sum of simulated observations over bird families at a given region is equal to the sum of transect observations in the original data. The Poisson parameter also reflects a 1-parameter model of energy consumption that presumes $p_f = 0$ in (4). This choice will be justified by the results in Section 4.1, which show there is no statistical evidence of energy consumption's dependence on mean fruit weight.

Given these simulated point counts $\{z_n\}$ and transect observations $\{x_n\}$, we then compute estimates $(\widehat{a}_n)_z$ and $(\widehat{a}_n)_x$ for a_n based on $\{z_n\}$ and $\{x_n\}$ respectively according to (19) and (23):

$$(\widehat{a}_n)_z \equiv \frac{z_n}{\sum_k z_k}; \quad (\widehat{a}_n)_x \equiv \frac{(w_n)^{-\widehat{p}_{b,n}} x_n}{\sum_k (b_k)^{-\widehat{p}_{b,n}} x_k}, \tag{44}$$

where $\widehat{p}_{b,n}$ is the investigator's best determination of the bird weight exponent $p_{b,n}$. Two cases were considered: (a) the investigator does not have detailed knowledge of the bird weight exponent, and sets $\widehat{p}_{b,n}$ equal to the mean value

$\bar{p}_b$ of the distribution in (42); and (b) the investigator knows exactly the values of $p_{b,n}$ for each n, and uses $\widehat{p}_{b,n} = p_{b,n}$.

Besides the estimators $(\widehat{a}_n)_z$ and $(\widehat{a}_n)_x$, we also computed the combined estimate $(\widehat{a}_n)_\beta = \beta(\widehat{a}_n)_z + (1-\beta)(\widehat{a}_n)_x$ as a function of β for $0 \leq \beta \leq 1$ for each simulation. Taking these results over all simulations, we were able to compute the means and variances of $(\widehat{a}_n)_z$, $(\widehat{a}_n)_x$, and $(\widehat{a}_n)_\beta$. This enables us to compute the accuracy and precision of all estimators for each family in each of the six hypothetical regions, for the two cases (a) and (b) defined in the previous paragraph.

A similar procedure was used to gauge the accuracy and precision of log abundance estimators $(\widehat{\ell}_n)_z, (\widehat{\ell}_n)_x$, and $(\widehat{\ell}_n)_\beta$ defined as follows:

$$(\widehat{\ell}_n)_z \equiv \log(\widehat{a}_n)_z; \qquad (\widehat{\ell}_n)_x \equiv \log(\widehat{a}_n)_x; \qquad (\widehat{\ell}_n)_\beta \equiv \beta(\widehat{\ell}_n)_z + (1-\beta)(\widehat{\ell}_n)_x. \quad (45)$$

Combined estimator parameter determination and confidence interval estimation

The second set of simulations is intended to test the procedure for determining variance-reducing linear combinations of estimators described in Section 3.1.7.

- First, for each of the six different regions observed in the study, 2500 simulated scenarios for point count and transect observation values $\{z_1 \dots z_n, x_1, \dots x_n\}_m$, $m = 1 \dots 2500$ were generated according to (41). Each observation vector $z_1 \dots z_N$ was used to create relative abundance vectors $a_n = z_n / \sum_{k=1}^n z_k$, $n = 1 \dots N$. These relative abundances are taken as "ground truth" for the scenario (these are generated so as to increase the variety in the "ground truth" scenarios, making the test more robust.)

- Next, for each scenario, point counts and transect observations (simulating the dataset collected by a single hypothetical "ecologist" studying the scenario) were randomly generated as z'_n and x'_n respectively according to (compare (41)-(43))

$$\begin{aligned} z'_n &\sim Pois(z_n), n = 1 \dots N; \\ p_{b,n} &\sim Gauss(\mu = -0.35, \sigma^2 = 0.125^2), n = 1 \dots N; \\ x'_n &\sim Pois\left(z_n w_n^{p_{b,n}} \frac{\sum_{k=1}^N x_k}{\sum_{k=1}^N z_k w_k^{p_{b,k}}} \right), n = 1 \dots N. \end{aligned} \quad (46)$$

- Each simulated ecologist uses her own point counts and transect observations to compute estimators $(\widehat{a}_n)_{z'}, (\widehat{a}_n)_{x'}, (\widehat{\ell}_n)_{z'}$ and $(\widehat{\ell}_n)_{x'}$ according to (38). Two alternative situations are modeled: (a) the ecologist does not have detailed knowledge of the coefficients $p_{b,n}$, and uses the mean value $\bar{p}_b = 0.35$ to estimate $(\widehat{a}_n)_{x'}$ and $(\widehat{\ell}_n)_{x'}$; (b) the ecologist has precise knowledge of $p_{b,n}$, and uses the same values that are used in (46) for these estimates.

- Each simulated ecologist then uses the prescription in Section 3.1.7 with $M = 2500$ to estimate the linear combination parameters β and $\widetilde{\beta}$ for the estimators $(\widehat{a}_n)_\beta$ and $(\widehat{\ell}_n)_\beta$, as well as 90% confidence intervals for these estimates. As in the previous step, two scenarios apply: (a) the ecologist uses the mean value $\bar{p}_b = 0.35$ to estimate $(\widehat{a}_n)_\beta$ and $(\widehat{\ell}_n)_\beta$; (b) the ecologist uses the same values of $p_{b,n}$ that are used in (46) for these estimates.

To compare the accuracies of the point count-based, transect-based, and combined estimators, 95% confidence intervals for the relative error in relative abundance (estimate − ground truth relative abundance)/(relative abundance) are plotted. In order to verify the confidence intervals computed in the last bullet above, we also recorded the proportion of time that the ground truth relative abundances were within the 90% confidence intervals.

4 Results

4.1 Application of energy consumption estimation methodology to dataset

The six panels in Figure 1 show 95% confidence intervals (±2 standard deviations) for the model coefficients p_b (*left*) and p_f (*right*) as estimated from different data subsets using the two-coefficient model (13). Confidence intervals are shown for the coefficient estimates using per-species data from six separate locations (yellow bars); using data grouped by altitude (purple bars); using data grouped by fragmentation (blue bars); and using all data combined (green bar). In 11 of the 12 cases (including the overall data model), the confidence intervals for p_f contain 0, indicating a lack of statistical evidence for a nonzero value of p_f. Equivalently, this result is consistent with a one-sided test of $H_0 : p_f = 0$ at the 5% significance level in 11 of 12 cases. On the other hand, in 6 of the 12 cases (including the model for overall data), the confidence interval for p_b lies entirely below 0, indicating that p_b is negative at 5% significance.

Scatter plots of actual versus predicted transect observations for the two-coefficient model, by location, elevation, habitat type, and combined, are provided in Supplement A (Figures A1–A4). For most regions there is a generally positive trend, albeit with numerous errant values particularly on the low side, and there is no discernible influence of bird weight on the accuracy of the predictions. A summary of the best-fit parameters, R^2 values, and AIC values for the two-coefficient model across all data subsets is given in Table 2. R^2 values range from 0.03 to 0.40, indicating relatively poor fits.

The fact that p_f is not significantly different from 0 in virtually all cases motivates replacing $p_f = 0$ in (13). The model then has only the coefficient p_b

Location	C_0 est.	C_0 CI	p_b est.	p_b CI	p_f est.	p_f CI	R^2	AIC
By region								
Bombuscaro	3.6	[2.4,4.9]	-0.6	[-0.9,-0.3]	0.2	[-0.2,0.5]	0.30	92
ECSF	3.7	[0.4,7.1]	-0.6	[-1.4,0.2]	0.2	[-0.8,1.0]	0.19	64
Cajanuma	3.6	[-7.2,14.5]	-0.7	[-4.0,2.1]	0.3	[-1.0,1.6]	0.12	24
Copalinga	2.0	[0.3,3.7]	-0.2	[-0.5,0.2]	0.2	[-0.2,0.5]	0.03	141
Finca	3.0	[0.6,5.3]	-0.7	[-1.3,-0.04]	0.02	[-0.4,0.5]	0.25	60
Bellavista	3.7	[-4.0,11.4]	-0.7	[-2.4,0.9]	0.7	[-0.3,1.8]	0.40	23
By altitude								
Altitude 1000	3.7	[-4.0,11.4]	-0.7	[-2.4,0.9]	0.7	[-0.3,1.8]	0.40	23
Altitude 2000	3.2	[1.4,5.1]	-0.7	[-1.2,0.9]	-0.2	[-0.6,0.2]	0.3	86
Altitude 3000	4.2	[-0.8,9.1]	-0.8	[-2.0,0.3]	-0.7	[-1.4,-0.0]	0.40	33
By environment								
Natural	3.4	[2.2,4.5]	-0.6	[-0.8,-0.3]	-0.1	[-0.4,0.1]	0.40	166
Fragmented	1.9	[0.6,3.1]	-0.3	[-0.6,-0.0]	0.0	[-0.2,0.2]	0.08	208
All birds	2.2	[1.2,3.1]	-0.4	[-0.6,-0.1]	-0.1	[-0.3,0.1]	0.11	245

Table 2: Summary of parameter estimations from regression for the two-coefficient model (13), which includes both bird weight and fruit weight. CI = 95% confidence interval; AIC = Akaike information criterion.

plus the constant C_0 that are fitted to the transect and point count data:

$$\log(X_n) = \log(Z_n) + C_0 + p_b \log(w_n) + e_n. \tag{47}$$

Figure 2 summarizes confidence intervals for the bird weight coefficient in the single-coefficient model (47). Confidence intervals are consistent across all regions, and overlap in the range of -0.3 to -0.5. As with the two-coefficient model, the bird weight coefficients tend to be negative, and some are negative with 95% significance (including the overall estimate). Confidence intervals for high altitudes are smaller, which may reflect the fact that bird species are more variable at low altitudes. Taken together, the data gives a 95% confidence interval for p_b as $[-0.6, -0.1]$. Scatter plots for the one-coefficient model, by location, elevation, habitat type, and combined, are provided in Supplement A (Figures A5–A8). Visually, the one-coefficient model appears to have accuracies comparable to the two-coefficient model.

Table 3 summarizes the fit parameters for the single-coefficient model. The AIC values for all fits are reduced relative to the two-coefficient model (Table 2), thus confirming that the exclusion of fruit weight as a significant parameter is justified: a smaller AIC indicates that the loss of information from dropping p_f is more than offset by the reduction in model complexity.

Recall that p_b is defined in (14) as the difference of the two exponents p_{b1} and p_{b2}, where p_{b1} is related to bird weight's influence on energy consumption per bird, while p_{b2} is related to bird weight's influence on energy consumption per observed feeding. Estimates of p_{b1} in [14] for most families are close to 0.7 (hummingbirds are an exception, with $p_{b1} > 1$). Our finding of $p_b \approx -0.4$ is consistent with a value of $p_{b1} \approx 1$. This would imply that bird energy consumption per feeding is roughly proportional to bird weight. This is not unreasonable, because heavier birds proportionately have larger anatomical features, including larger stomachs.

Figure 3 shows a scatter plot of the ratio of feeding observations to point count observations versus point count frequency. The figure indicates a negative correlation between the feeding/observation ratio and the number of observations. This is an indication of regression to the mean: overpredictions (or underpredictions) based on transect observations will tend to be revised downwards (or upwards) based on feeding observations—and vice versa. The presence of regression to the mean indicates that point counts and transect observations tend to counterbalance each other, and estimators based on both factors may achieve greater accuracy. (This indication is confirmed in the following sections.)

4.2 Performance evaluation of estimators

The results of the performance evaluation tests described in Section 3.2.3 are summarized in Figures 4 and 5–6 in the main body, with full supporting detail

Location	C_0 est.	C_0 CI	p_b est.	p_b CI	R^2	AIC
By region						
Bombuscaro	3.4	[2.2,4.5]	-0.6	[-0.9,-0.2]	0.27	91
ECSF	3.4	[0.4,7.1]	-0.6	[-1.3,0.2]	0.18	63
Cajanuma	1.7	[-3.9,7.3]	-0.3	[-1.9,1.4]	0.03	22
Copalinga	1.5	[0.28,2.7]	-0.14	[-0.5,0.2]	0.01	140
Finca	2.9	[0.8,5.0]	-0.7	[-1.3,-0.0]	0.25	58
Bellavista	-0.9	[-5.4,3.6]	0.09	[-1.15,1.33]	0.00	25
By altitude						
Altitude 1000	1.9	[0.9,2.9]	-0.3	[-0.5,0.0]	0.06	156
Altitude 2000	2.9	[1.2,4.6]	-0.6	[-1.1,-0.2]	0.27	85
Altitude 3000	0.2	[-3.7,4.1]	-0.1	[-1.1,1.0]	0.00	37
By environment type						
Natural	3.1	[2.1,4.0]	-0.5	[-0.8,-0.25]	0.22	166
Fragmented	1.9	[0.9,2.9]	-0.3	[-0.6,0.0]	0.08	206
All	2.0	[1.2,2.8]	-0.3	[-0.6,-0.1]	0.1	244

Table 3: Summary of parameter estimations from regression for the single-coefficient model with $p_f = 0$ in (13). CI = 95% confidence interval; AIC = Akaike information criterion. Reduced AIC values relative to Table 2 confirm that dropping the fruit weight parameter improves model parsimony.

provided in Supplements B and C. Supplement B contains results for estimators computed using the mean value $\bar{p}_b$ of the bird weight exponent (case a), while Supplement C contains corresponding results using the exact values of $p_{b,n}$ (case b), as defined in Section 3.2.3.

Case (a): estimators computed using mean value of bird weight exponent $\bar{p}_b$

The mean values of the estimators $(\widehat{a}_n)_z$ and $(\widehat{a}_n)_x$ over 10,000 simulations, compared to the true relative abundance, are shown in Figure B1 of Supplement B. The figure shows that there is very little bias in most estimators, with estimators for birds with lower relative abundance tending to have the most bias. Figure B2 of Supplement B shows similar results for the log abundance estimators $(\widehat{\ell}_n)_z$ and $(\widehat{\ell}_n)_x$.

Figure 4 shows the relative standard deviations for $(\widehat{a}_n)_\beta$ for values of β between 0 and 1, for all different bird families in the six different regions. The standard deviations are based on estimates of 10,000 simulated ecologists. The triangles on the horizontal axis mark the optimal values of β as calculated by (30). The figure shows that (30) gives a very accurate estimate of the β that minimizes standard deviation. All of the optimal β values are between 0.5 and 1, reflecting the fact that $(\widehat{a}_n)_z$ (corresponding to $\beta = 1$) is consistently more accurate than $(\widehat{a}_n)_x$ (corresponding to $\beta = 0$). Figure B3 of Supplement B shows similar results for the standard deviations for $(\widehat{\ell}_n)_\beta$.

Figures B4 and B5 of Supplement B show the relative RMSE of $(\widehat{a}_n)_\beta$ and $(\widehat{\ell}_n)_\beta$ respectively as functions of β. These graphs are nearly identical to the corresponding standard deviation figures, confirming that bias error is negligible compared to standard deviation for both estimators.

Figures B6 and B7 of Supplement B show RMSE-minimizing β values for $(\widehat{a}_n)_\beta$ and $(\widehat{\ell}_n)_\beta$ respectively, versus the true relative abundance of each bird family. It is clear that lower β values (between 0.4 and 0.9) only occur for birds with relative abundance less than about 0.1, with only two exceptions from the Cajanuma dataset (which had only 2 families).

Case (b): Estimators computed using exact values of bird weight exponent $p_{b,n}$

Supplement C contains figures corresponding to those in Supplement B, except they show results obtained when the investigator has accurate information about $p_{b,n}$. It is evident from Figures C3–C6 of Supplement C that the estimates $(\widehat{a}_n)_X$ and $(\widehat{\ell}_n)_X$ (corresponding to $\beta = 0$) are more accurate than the corresponding estimates when $p_{b,n}$ is not known exactly. As a result the optimal values of β tend to be lower when $p_{b,n}$ is known exactly, as shown in Figures C7 and C8 of Supplement C (compare Figures B6 and B7 of Supplement B for the inexact case).

4.3 Verification of accuracy of combined estimators

Figures 5 and 6 show 95% confidence intervals for the relative errors for point count-based, transect observation-based, and combined estimators of relative

abundance obtained from the simulations described in Section 3.1.7; results are also summarized numerically in Tables 4 and 5. Both figures show that relative abundance estimates using the combined estimator $(\widehat{a}_n)_\beta$ are always at least as accurate as the best of the two estimators $(\widehat{a}_n)_Z$ and $(\widehat{a}_n)_X$. Comparing between the two figures shows that combined estimates using the exact values for $p_{b,n}$ are much more accurate than those using $\overline{p_b}$.

Figures B8 and C9 of Supplement B and Supplement C respectively show similar results for relative errors of different estimators of log relative abundance. Taken together, the results indicate that more accurate knowledge about individual bird species' energy consumption habits and requirements can enable improved estimates of relative abundance and log relative abundance. The degree of improvement is displayed quantitatively in Table 4 (for relative abundance estimators) and in Table 5 (for log relative abundance estimators).

	$\widehat{a}_Z$ rel. err.		$\widehat{a}_X$ rel. err.		$\widehat{a}_\beta$ rel. err.	
	(mean ± std)		mean $\overline{p}_b$	exact p_b	mean $\overline{p}_b$	exact p_b
Bombuscaro	0.28 ± 0.21		0.55 ± 0.18	0.16 ± 0.11	0.22 ± 0.12	0.14± 0.10
Copalinga	0.30± 0.19		0.64 ± 0.27	0.23 ± 0.17	0.25 ± 0.15	0.18 ± 0.13
ECSF	0.23 ± 0.10		0.66 ± 0.29	0.21 ± 0.12	0.21 ± 0.15	0.15 ± 0.08
Finca	0.16 ± 0.10		0.54 ± 0.28	0.16 ± 0.14	0.15 ± 0.09	0.11 ± 0.08
Cajanuma	0.18 ± 0.11		0.33 ±0.21	0.14 ± 0.09	0.16 ± 0.10	0.10 ± 0.07
Bellavista	0.20 ± 0.16		0.60 ± 0.34	0.49 ± 0.50	0.20 ± 0.16	0.18 ± 0.15
Altitude 1	0.29 ± 0.20		0.61 ± 0.24	0.20 ± 0.15	0.24 ± 0.14	0.16 ± 0.12
Altitude 2	0.20 ± 0.10		0.61± 0.29	0.19 ± 0.13	0.18 ± 0.09	0.13 ± 0.08
Altitude 3	0.19 ± 0.15		0.51 ± 0.33	0.37 ± 0.44	0.18 ± 0.14	0.16 ± 0.13
Natural	0.25 ± 0.18		0.57 ± 0.26	0.26 ± 0.27	0.21 ± 0.11	0.16 ± 0.13
Fragmented	0.25 ± 0.18		0.61 ± 0.29	0.26 ± 0.27	0.22 ± 0.15	0.16 ± 0.13
TOTAL	0.25 ± 0.18		0.59 ± 0.27	0.22 ± 0.22	0.21 ± 0.13	0.15 ± 0.11

Table 4: Means and standard deviations of RMSE for relative error of point count-based, transect-based, and combined relative abundance estimators, for individual regions and grouped by altitude, environment type, and total.

5 Discussion

5.1 Fruit weight and bird weight as predictors of energy consumption

A central finding of this study is that mean fruit weight consumed does not significantly influence the number of feeding observations, once bird weight is accounted for. This result held consistently across 11 of the 12 data subsets examined (individual regions, elevational groups, habitat types, and the combined dataset), and was corroborated by the systematic reduction in AIC values when the fruit weight parameter p_f was dropped from the model. This finding

	$\widehat{\ell}_Z$ rel. err.		$\widehat{\ell}_X$ rel. err.		$\widehat{\ell}_\beta$ rel. err.	
	(mean ± std)		mean $\overline{p}_b$	exact p_b	mean $\overline{p}_b$	exact p_b
Bombuscaro	0.26 ± 0.15		0.53± 0.15	0.18 ± 0.13	0.22 ± 0.11	0.14 ± 0.09
Copalinga	0.29 ± 0.17		0.58 ± 0.18	0.23 ± 0.16	0.25 ± 0.18	0.13 ± 0.11
ECSF	0.24 ± 0.12		0.63 ± 0.20	0.22 ± 0.13	0.22 ± 0.10	0.16 ± 0.09
Finca	0.17 ± 0.11		0.52 ± 0.18	0.16 ± 0.14	0.16± 0.10	0.12 ± 0.09
Cajanuma	0.20 ± 0.14		0.33 ± 0.20	0.15 ± 0.10	0.17 ±0.11	0.12 ± 0.08
Bellavista	0.21± 0.18		0.53± 0.11	0.34 ± 0.20	0.18 ± 0.12	0.17 ± 0.12
Altitude 1	0.28± 0.16		0.56 ± 0.17	0.21 ± 0.15	0.24 ± 0.12	0.16 ± 0.11
Altitude 2	0.21 ± 0.12		0.58 ± 0.20	0.19 ± 0.13	0.20 ± 0.10	0.14 ± 0.09
Altitude 3	0.21 ± 0.16		0.46± 0.17	0.17 ± 0.19	0.18 ± 0.12	0.15 ± 0.11
Natural	0.25 ± 0.14		0.55 ± 0.20	0.19 ± 0.13	0.22 ± 0.11	0.15 ±0.09
Fragmented	0.25± 0.17		0.56± 0.17	0.23 ± 0.17	0.22 ± 0.13	0.16 ± 0.11
Total	0.25± 0.15		0.57± 0.22	0.23 ± 0.22	0.22 ± 0.12	0.15 ± 0.10

Table 5: Means and standard deviations of RMSE for error of point count-based, transect-based, and combined log relative abundance estimators, for individual regions and grouped by altitude, environment type, and total.

	$\widehat{a}_\beta$ 90% C.I. proportions		$\widehat{\ell}_\beta$ 90% C.I. proportions	
	using $\overline{p}_b$	using exact p_b	using $\overline{p}_b$	using exact p_b
Bombuscaro	0.906 ± 0.007	0.900 ± 0.004	0.907 ± 0.007	0.901 ± 0.004
Copalinga	0.905 ± 0.008	0.901 ± 0.004	0.906 ± 0.001	0.901 ± 0.004
ECSF	0.905 ± 0.007	0.903 ± 0.002	0.905 ± 0.007	0.904 ± 0.002
Finca	0.903 ± 0.005	0.905 ± 0.008	0.903 ± 0.005	0.906 ± 0.008
Cajanuma	0.902 ± 0	0.907 ± 0	0.902 ± 0	0.908 ± 0
Bellavista	0.904 ± 0.005	0.901 ± 0.009	0.904 ± 0.005	0.902 ± 0.009
Altitude 1	0.906 ± 0.008	0.901 ± 0.004	0.906 ± 0.008	0.901 ± 0.004
Altitude 2	0.904 ± 0.006	0.904 ± 0.005	0.904 ± 0.006	0.904 ± 0.005
Altitude 3	0.903 ± 0.004	0.903 ± 0.008	0.9034 ± 0.004	0.904 ± 0.008
Natural	0.905 ± 0.007	0.902 ± 0.004	0.905 ± 0.007	0.9028 ± 0.004
Fragmented	0.904 ± 0.007	0.902 ± 0.006	0.905 ± 0.007	0.902 ± 0.006
TOTAL	0.905 ± 0.007	0.902 ± 0.005	0.905 ± 0.007	0.903 ± 0.005

Table 6: Proportion of instances where the true relative abundance (respectively log relative abundance) lay within the 90% confidence intervals computed by the bootstrapping technique described in Section 3.1.7. Estimates using the mean value of the bird weight exponent $\bar{p}_b$ and the exact value of p_b are shown. The error bars are for one standard deviation calculated from the proportions obtained for different families in the region. Proportions consistently near 0.90 across all regions confirm that the bootstrapping procedure produces well-calibrated confidence intervals.

contrasts with the intuition that larger fruit should require fewer foraging bouts due to higher energy yield per item, but is consistent with field observations

suggesting that frugivores often adjust bite size rather than foraging frequency when consuming fruit of varying sizes [7]. It also aligns with the well-established view that frugivore foraging rates are constrained primarily by digestive processing time, which scales with gut volume and hence with bird body size rather than with fruit size per se [8].

The bird weight exponent p_b, by contrast, was estimated to be consistently negative (overall 95% confidence interval $[-0.6, -0.1]$), indicating that heavier bird families have fewer transect feeding observations per individual relative to what their point count abundance would predict. As discussed in Section 4, this is consistent with $p_{b1} \approx 1$, implying that energy consumption per feeding event scales approximately linearly with bird body mass. This interpretation aligns with prior allometric studies reporting that basal metabolic rate scales with body mass raised to a power near 0.7 for most avian families [14], while absolute food intake per bout scales more steeply. The larger confidence intervals observed at low elevations, relative to higher elevations, are consistent with the greater phylogenetic and body-size diversity of frugivore assemblages in warmer lowland forests, which would introduce more variability in the relationship between body size and foraging rate. Future work using species-level body mass data, rather than family-level averages, could reduce this variability and yield tighter parameter estimates.

5.2 Combining point counts and transect observations for improved abundance estimation

The simulation results demonstrate that point count-based and transect observation-based estimators of relative abundance are complementary in their error structure, with the point count estimator consistently showing lower variance under the conditions simulated. The negative correlation between the feeding-to-observation ratio and the number of transect observations (Figure 3) provides empirical confirmation of regression to the mean in the transect-based estimates, which is the mechanism that makes the two estimator types counterbalancing. This counterbalancing effect is precisely what the linear combination exploits: by assigning weight β to the point count estimator and $1-\beta$ to the transect estimator, the combined estimator achieves a variance lower than either component alone, as guaranteed by the minimum-variance formula (30).

The practical implication is that field studies collecting both point count and transect data need not choose between them at the analysis stage. Instead, both sources can be integrated using the framework developed here to obtain abundance estimates with formally smaller uncertainty. The improvement is most pronounced when the investigator has accurate prior knowledge of the bird weight exponent $p_{b,n}$: in this case, the transect-based estimator becomes substantially more informative, and the combined estimator achieves relative RMSE values roughly 40% lower than the point count estimator alone (Table 4). When only the mean value $\bar{p}_b$ is available, the improvement is more modest but still consistent across all regions and groupings. This suggests that investing effort in characterizing the allometric energy scaling of the focal bird families—

whether from prior literature or from empirical calibration within the study system—pays dividends in the accuracy of final abundance estimates.

These findings have direct relevance for ongoing monitoring programs in Andean forests, where both survey methods are commonly deployed but rarely integrated analytically. The Podocarpus region in particular supports well-documented gradients in frugivore community composition across elevation and habitat fragmentation [6, 16], and the present framework could be applied to quantify how relative abundance of key seed-dispersing families changes across these gradients with greater precision than point counts alone would allow. Similar applications are conceivable in any system where ecologists collect concurrent point count and behavioral observation data, including insectivore foraging studies and waterbird monitoring programs.

5.3 Confidence intervals and their calibration

A notable practical contribution of this paper is the bootstrapping procedure for computing confidence intervals for the combined abundance estimator. Table 6 shows that the 90% confidence intervals produced by this procedure contain the true relative abundance in approximately 90% of simulation instances across all regions, groupings, and parameterizations tested. This calibration is remarkably consistent, with proportions ranging from 0.900 to 0.908 regardless of whether the mean or exact bird weight exponent was used. The robustness of the calibration across these two scenarios suggests that the bootstrapping procedure is not highly sensitive to uncertainty in the weight exponent, which is an important practical property since this parameter must typically be estimated from data.

Rigorous confidence intervals for abundance estimates are rarely reported in avian community monitoring studies, which tend to rely on informal descriptions of survey effort as a proxy for precision. The present framework offers a path toward more formal uncertainty quantification that is grounded in the statistical properties of the underlying observation process. This is particularly valuable in the context of long-term monitoring, where changes in abundance over time must be distinguished from sampling variability—a distinction that requires well-calibrated confidence intervals.

5.4 Limitations and future directions

Several limitations of the present study should be acknowledged. First, the theoretical framework assumes that point count and transect observations are made independently and that the detection probability within the specified radius or distance band is constant across species and habitats. Violations of the constant detection probability assumption—which are common in heterogeneous landscapes—could bias both the point count and transect estimators in ways not captured by the current model. Distance sampling methods [10] or N-mixture models could in principle be integrated into the framework to relax this assumption.

Second, the empirical application combines observations across bird families, which span a wide range of body sizes and behavioral repertoires. Family-level aggregation, while necessary given the relatively small per-species sample sizes in the dataset, smooths over ecologically meaningful interspecific variation. Future applications with larger datasets could operate at the species or genus level, potentially yielding more informative estimates of $p_{b,n}$ and hence more accurate transect-based abundance estimators.

Third, the simulation-based evaluation uses the observed data as "ground truth," which means the simulations inherit whatever systematic biases may be present in the original field data. Independent validation using replicated field datasets from different sites or time periods would provide a stronger test of the framework's generalizability.

Despite these limitations, the methods developed here represent a step toward more rigorous integration of complementary survey data types in avian ecology, and the general approach of linking energy consumption models to multiple observation processes could be extended to other taxa and monitoring contexts.

6 Conclusions

The principal conclusions of this study are as follows:

- For the dataset drawn from six habitats in Ecuador described in Section 3.2.1, there is no statistical evidence that bird species' energy consumption depends on the weight of fruit consumed.
- On the other hand, there is statistical evidence from this dataset supporting existing literature that energy consumption of bird species depends on bird weight.
- There is also statistical evidence from this same dataset that energy consumption per transect observation is larger for heavier birds than for lighter birds.
- Simulations based on the Ecuador data demonstrate that point counts and transect observations can be used to obtain independent estimates of relative abundance and log of relative abundance. Typically, the point count-based estimates are more accurate.
- The two estimates mentioned in the previous item may be linearly combined to form another estimate which has lower RMSE. The user may run simulations to estimate the parameter β needed to form a linear combination of point count based and transect observation based abundance estimators with smaller root mean squared error.
- Error bars may be computed for the linearly combined relative abundance estimates that accurately reflect the uncertainty in these estimates.

- With more accurate information concerning the dependence of bird type energy consumption as a function of weight, a better estimate of β can be made that leads to combined estimators of abundance with lower root mean squared error.

CRediT statement

Conceptualization: V.S., M.J.A.B., A.B., C.T.; **Data curation**: V.S., M.Q., E.L.N., M.S.; **Formal analysis**: M.J.A.B., A.B., C.T.; **Funding acquisition**: K.B.S., O.A.; **Investigation**: V.S., M.Q., E.L.N., M.S.; **Methodology**: M.J.A.B., A.B., C.T.; **Project administration**: V.S., K.B.S., O.A.; **Resources**: K.B.S., O.A.; **Software**: M.J.A.B., A.B., C.T.; **Supervision**: K.B.S., V.S., C.T.; **Validation**: M.J.A.B., C.T.; **Visualization**: M.J.A.B., A.B., C.T.; **Writing – original draft**: M.J.A.B., A.B., C.T.; **Writing – review & editing**: V.S., E.L.N., C.T.

Acknowledgements

We are grateful to the Fulbright Program administered by the US Department of State, the Catholic University of Cuenca and to Texas A&M University-Central Texas for providing travel funding and logistical support during the conduct of this research.

References

[1] Cagan H Sekercioglu. Increasing awareness of avian ecological function. *Trends in Ecology & Evolution*, 21(8):464–471, 2006.

[2] Rudi Partasasmita. Role of frugivorous birds in the dispersal of shrub species in a submontane forest of west java, indonesia. *Biodiversitas Journal of Biological Diversity*, 16(1):35–41, 2015.

[3] Pedro A Pereira, Dione M Faria, and Marco AR Mello. Bird-plant interactions in a forest fragment in southwestern brazilian amazonia. *Acta Amazonica*, 38(2):307–314, 2008.

[4] Laury Cullen Gomes, Luiz Octavio Marcondes-Machado, and Jacques ME Vielliard. Potential role of frugivorous birds in seed dispersal of plants in a restinga habitat (southeastern brazil). *Revista de Biología Tropical*, 56(2):845–856, 2008.

[5] Gilbert Barrantes and Gustavo A Pereira. Seed dispersal of a dry-fruited tree, casearia nitida, by frugivorous birds in a tropical forest of costa rica. *Revista de Biología Tropical*, 50(3-4):937–946, 2002.

[6] John G Blake and Bette A Loiselle. Bird assemblages in second-growth and old-growth forests, costa rica: Perspectives from mist nets and point counts. *The Auk*, 118(2):304–326, 2001.

[7] Timothy C Moermond and Julie S Denslow. The influence of fruit morphology on fruit choice of tropical birds. *Biotropica*, 17(4):319–330, 1985.

[8] Carlos M Herrera. A study of avian frugivores, bird-dispersed plants, and their interaction in mediterranean shrublands. *Ecological Monographs*, 54(1):1–23, 1984.

[9] William J Sutherland. *Ecological census techniques: a handbook*. Cambridge university press, 2006.

[10] Stephen T Buckland, David R Anderson, Kenneth P Burnham, and Jeffrey L Laake. *Introduction to Distance Sampling: Estimating Abundance of Biological Populations*. Oxford University Press, 2001.

[11] C. John Ralph, John R. Sauer, and Sam Droege. Monitoring bird populations by point counts. Technical report, 1995.

[12] Andrew Gelman, John B Carlin, Hal S Stern, David B Dunson, Aki Vehtari, and Donald B Rubin. *Bayesian Data Analysis*. CRC Press, 3 edition, 2013.

[13] Roderick JA Little and Donald B Rubin. *Statistical Analysis with Missing Data*. Wiley, 2002.

[14] J. Gurney D. Crocker, A. Hart and C. McCoy. Project pn0908: Methods for estimating daily food intake of wild birds and mammals. Technical report, Central Science Laboratory, 2002.

[15] Stefany Coxe, Stephen G West, and Leona S Aiken. The analysis of count data: A gentle introduction to poisson regression and its alternatives. *Journal of personality assessment*, 91(2):121–136, 2009.

[16] Carsten Rahbek, Hanne Bloch, Michael Koie Poulsen, and Jan Fischer Rasmussen. The avifauna of the podocarpus national park-the" andean jewel in the crown" of ecuador's protected areas. *Ornitologia Neotropical*, 6(2):7, 1995.

[17] Leonardo Ordóñez-Delgado, Cristian Ramón-Vivanco, and Verónica Ortiz-Chalán. Systematic review of the state about the knowledge of the vertebrates of the Podocarpus National Park. *La Granja*, 30(2):7–19, 2019.

[18] Markus Kottek, Jürgen Grieser, Christoph Beck, Bruno Rudolf, and Franz Rubel. World map of the köppen-geiger climate classification updated. *Meteorologische Zeitschrift*, 15(3):259–263, 07 2006.

[19] Paul Emck. *A climatology of south Ecuador-with special focus on the major Andean ridge as Atlantic-Pacific climate divide*. PhD thesis, Friedrich-Alexander-Universität Erlangen-Nürnberg (FAU), 2007.

[20] Colin J Bibby, Neil D Burgess, David M Hillis, David A Hill, and Simon Mustoe. *Bird census techniques*. Elsevier, 2000.

[21] Hamish Wilman, Jonathan Belmaker, Jennifer Simpson, Carolina de la Rosa, Marcelo M Rivadeneira, and Walter Jetz. Eltontraits 1.0: Species-level foraging attributes of the world's birds and mammals: Ecological archives e095-178. *Ecology*, 95(7):2027–2027, 2014.

[22] Kenneth P Burnham and David R Anderson. *Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach*. Springer, 2002.

Figures

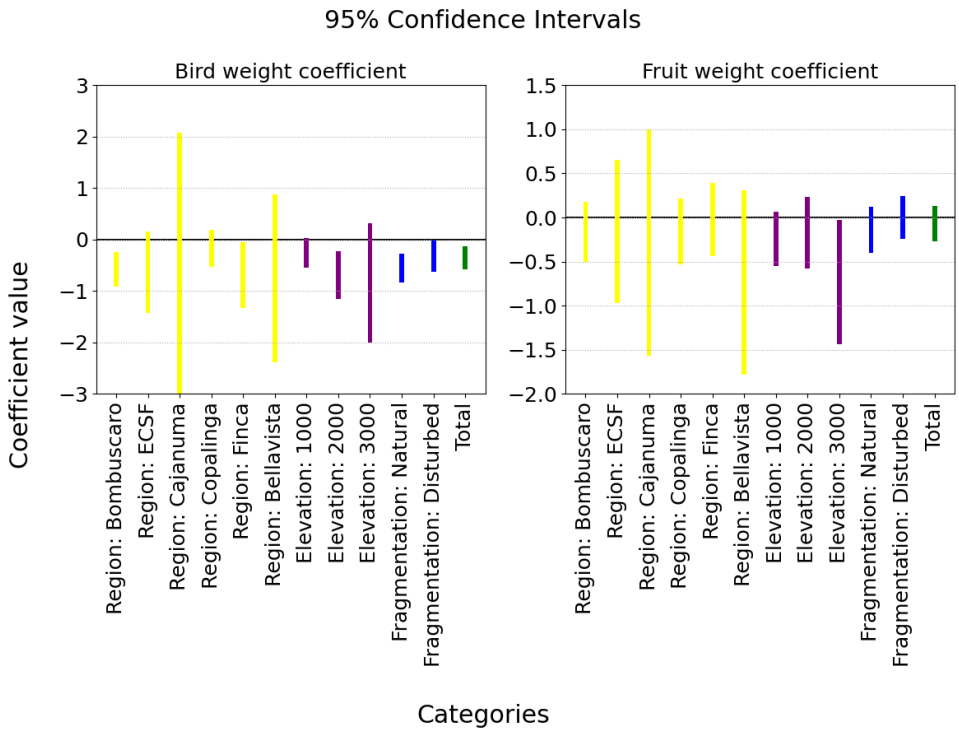

Figure 1: 95% confidence intervals for the bird weight coefficient p_b (*left panel*) and the fruit weight coefficient p_f (*right panel*) from the two-coefficient model (13), estimated from different data subsets: by site (yellow bars, $n = 6$), by elevation (purple bars, $n = 3$), by habitat type (blue bars, $n = 2$), and for all data combined (green bar, $n = 1$). Confidence intervals for p_f that contain zero indicate no statistically significant effect of fruit weight on feeding frequency at $\alpha = 0.05$; this was the case in 11 of 12 subsets. Confidence intervals for p_b lying entirely below zero indicate a significant negative effect of bird weight, consistent with larger birds consuming more energy per feeding event.

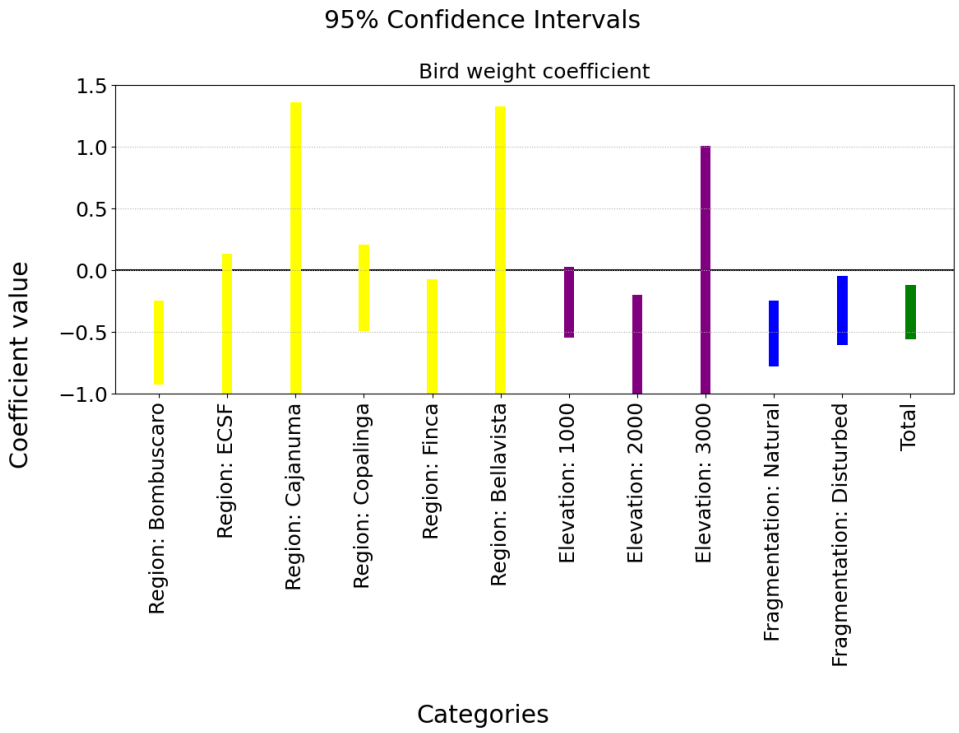

Figure 2: 95% confidence intervals for the bird weight coefficient p_b in the one-coefficient model (47) ($p_f = 0$), estimated from different data subsets: by site (yellow), by elevation (purple), by habitat type (blue), and for all data combined (green). Confidence intervals are broadly consistent across subsets, overlapping in the range $[-0.3, -0.5]$. The combined dataset yields a 95% confidence interval of $[-0.6, -0.1]$, used to parameterize the simulation distributions in Section 3.2.3.

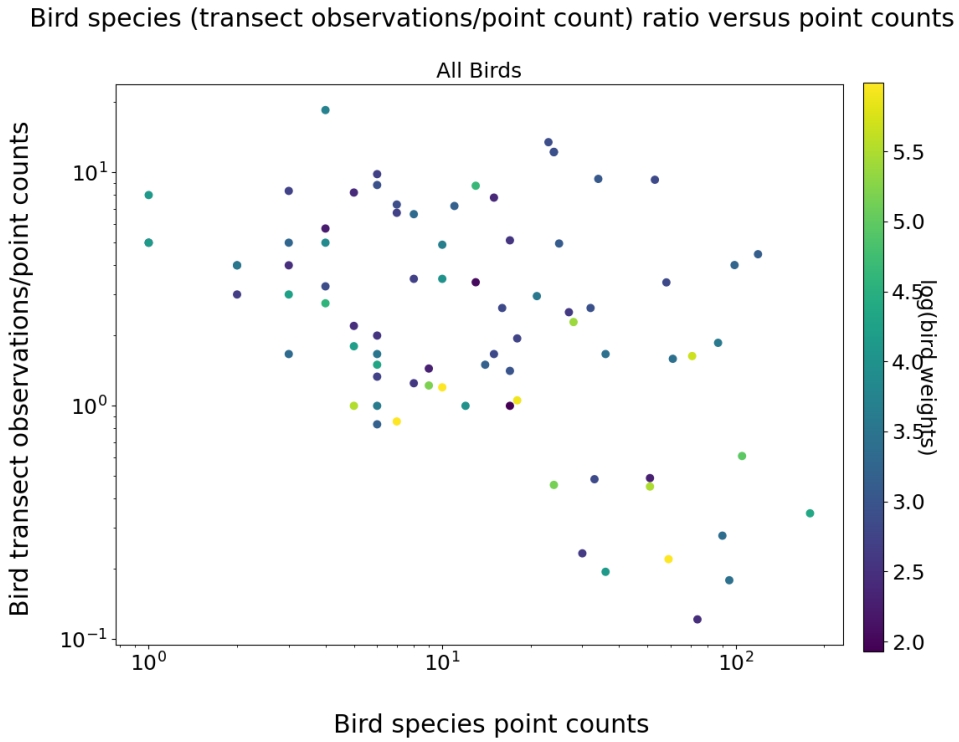

Figure 3: Scatter plot of the per-family ratio of feeding observations to point count observations versus total point count observations, using data pooled across all six sites. Each point represents one bird family at one site. The negative trend indicates regression to the mean: families with high point counts tend to have proportionally fewer feeding observations, and vice versa, suggesting the two data types provide complementary and partially counterbalancing information about abundance.

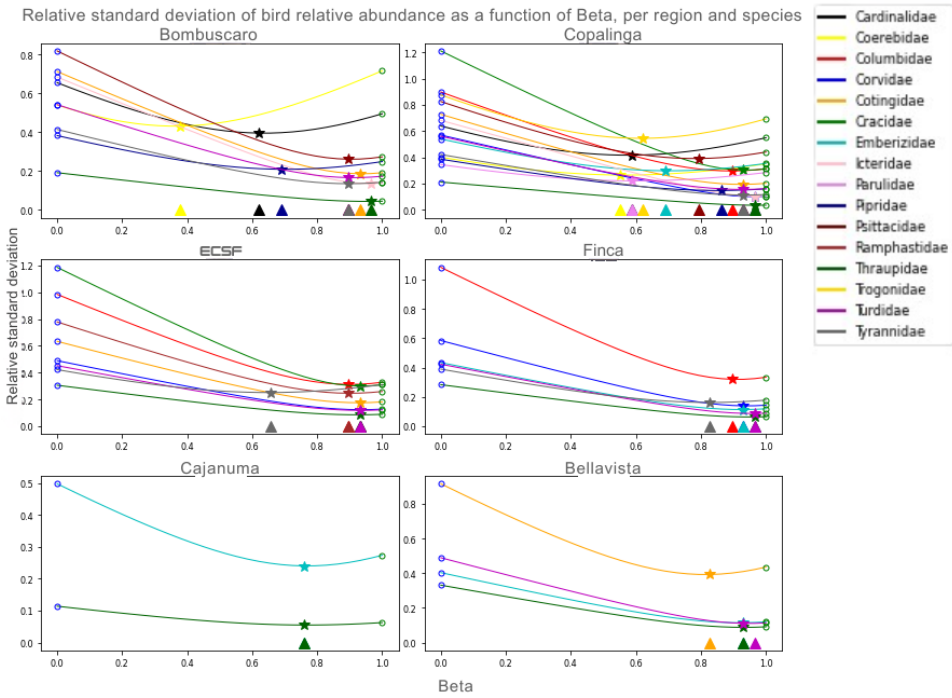

Figure 4: Relative standard deviation (standard deviation divided by true relative abundance) of the combined estimator $(\widehat{a}_n)_\beta$ as a function of the mixing parameter $\beta \in [0, 1]$, for each bird family in each of the six study sites (one curve per family). $\beta = 0$ corresponds to the pure transect-based estimator $(\widehat{a}_n)_X$; $\beta = 1$ corresponds to the pure point count-based estimator $(\widehat{a}_n)_Z$. Stars mark the minimum of each curve; triangles on the x-axis indicate the theoretically optimal β from (30). The transect estimator uses the mean bird weight exponent $\bar{p}_b$ (case a). The close agreement between stars and triangles confirms the accuracy of the analytical formula.

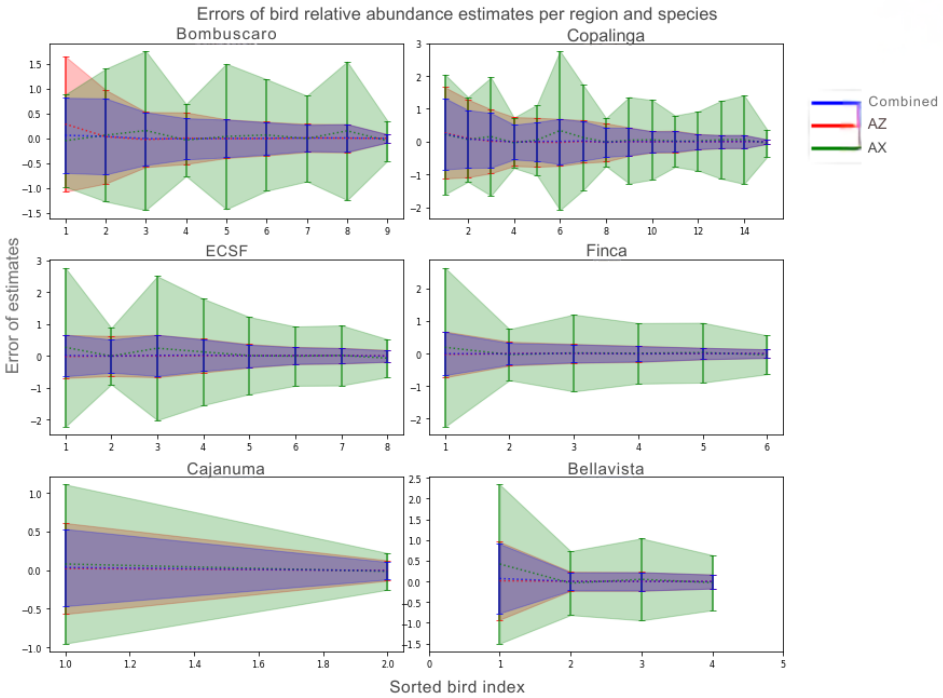

Figure 5: 95% confidence intervals for the relative error (estimated minus true relative abundance, divided by true relative abundance) of three estimators: the point count-based estimator $(\widehat{a}_n)_Z$ (labeled AZ), the transect-based estimator $(\widehat{a}_n)_X$ (labeled AX), and the combined estimator $(\widehat{a}_n)_\beta$. Results are shown for each bird family in each of the six study sites. The AX and combined estimators use the mean bird weight exponent $\bar{p}_b$ (case a). Families are ordered by increasing true relative abundance within each site panel. The combined estimator confidence interval is always at least as narrow as the better of the two individual estimators.

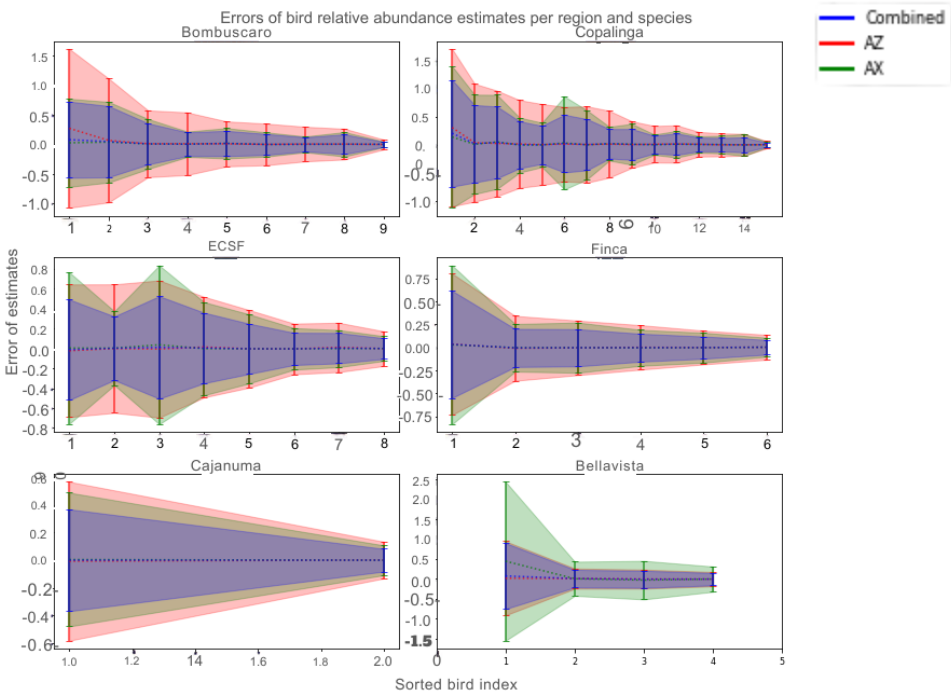

Figure 6: As Figure 5, but with the transect-based and combined estimators using the exact bird weight exponent $p_{b,n}$ (case b). Confidence intervals for AX and the combined estimator are substantially narrower than in Figure 5, particularly for the transect estimator, demonstrating the large benefit of accurate knowledge of the weight-energy relationship.

Supplement A: Model fit diagnostics

This supplement contains scatter plots of actual versus predicted transect observations for the two-coefficient model (Figures A1–A4) and the one-coefficient model (Figures A5–A8), organized by location, elevation, habitat type, and combined.

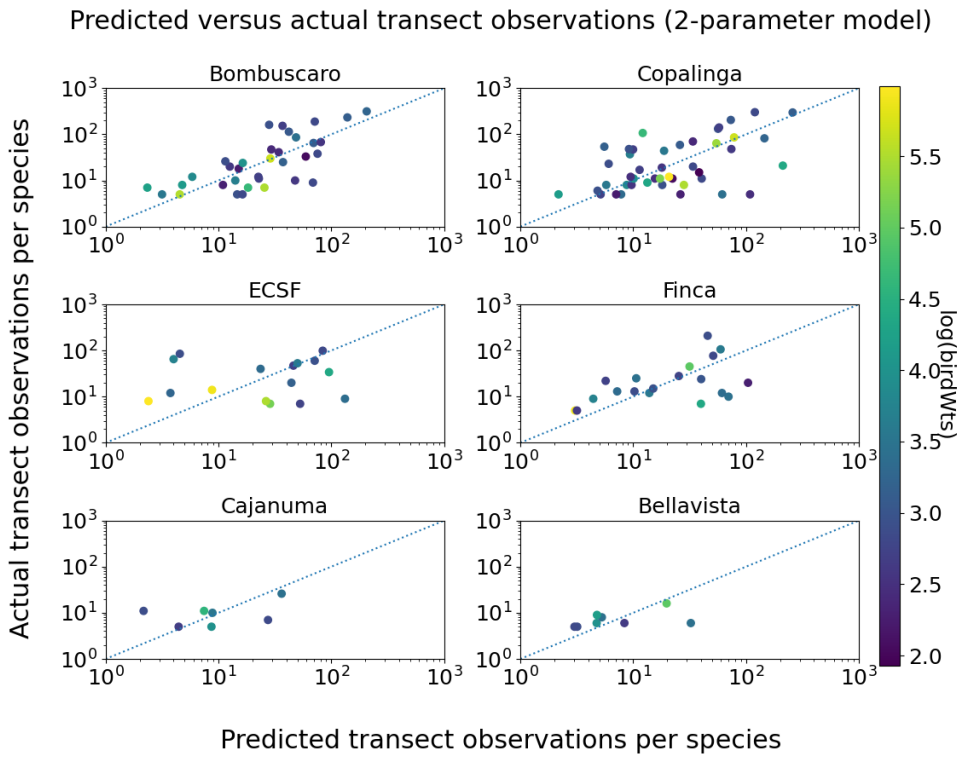

Figure A1: Transect observation prediction accuracy by location for the two-coefficient model (13) that includes both bird weight and fruit weight as predictors. Each panel shows log-scale actual versus log-scale predicted transect observations for one of the six study sites, with point color indicating mean bird species weight. The 45-degree line corresponds to zero prediction error. Scatter around the line reflects residual variability not explained by bird weight or fruit weight.

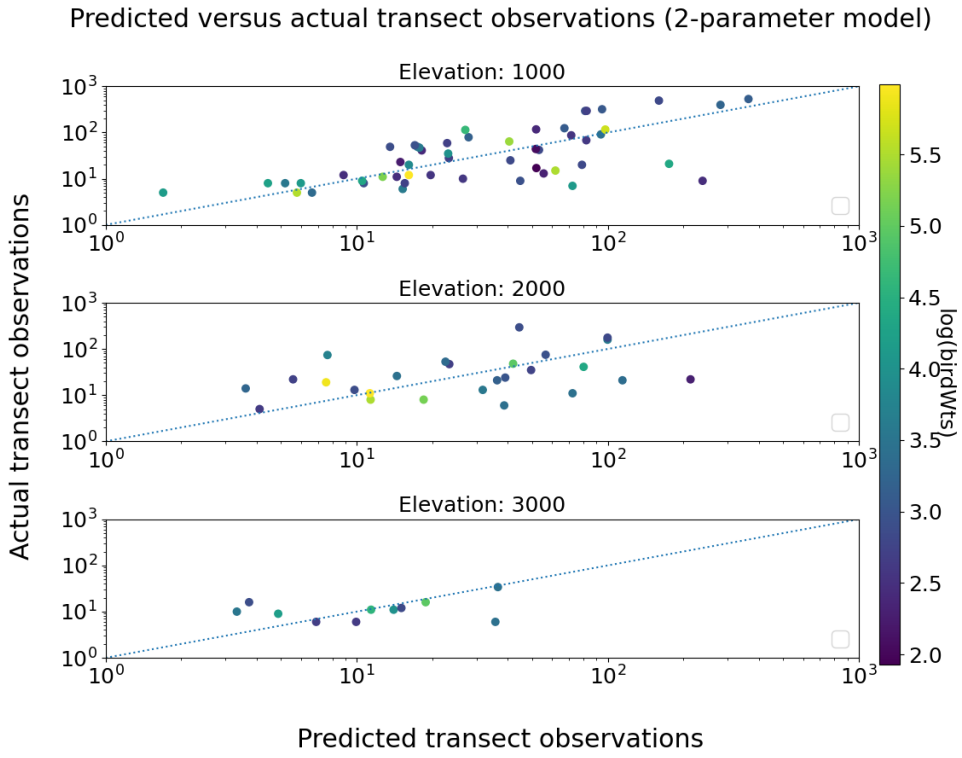

Figure A2: Transect observation prediction accuracy of the two-coefficient model (13) (bird weight + fruit weight), with species data pooled by elevation. Low (1000 m), mid (2000 m), and high (3000 m) elevation panels each combine data from the two sites at that elevation. Axes and line as in Figure A1.

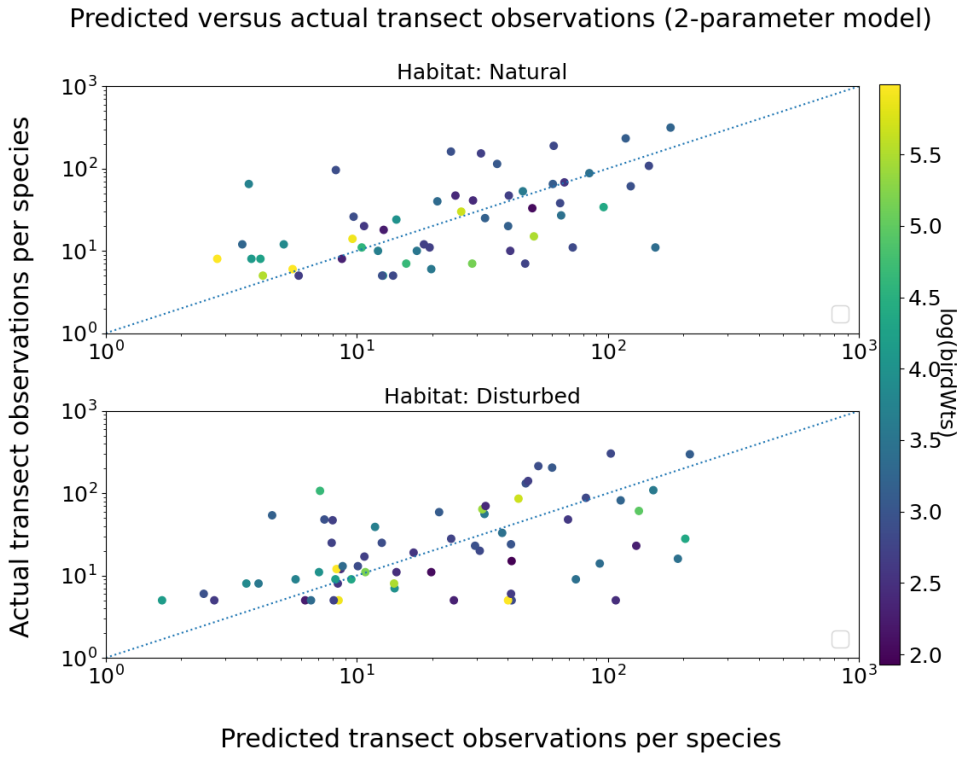

Figure A3: Transect observation prediction accuracy of the two-coefficient model (13) (bird weight + fruit weight), with species data pooled by habitat type. Natural forest and fragmented forest panels each combine data from the three sites of that habitat type. Axes and line as in Figure A1.

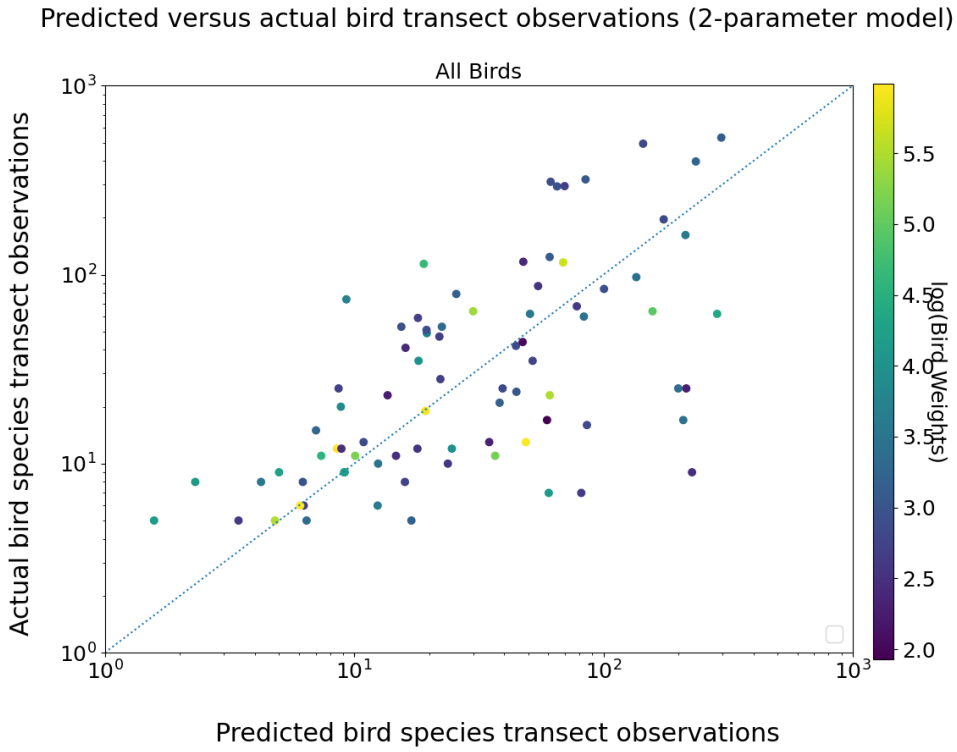

Figure A4: Transect observation prediction accuracy for the two-coefficient model (13) (bird weight + fruit weight) using all species observations from all six sites combined. Axes and line as in Figure A1.

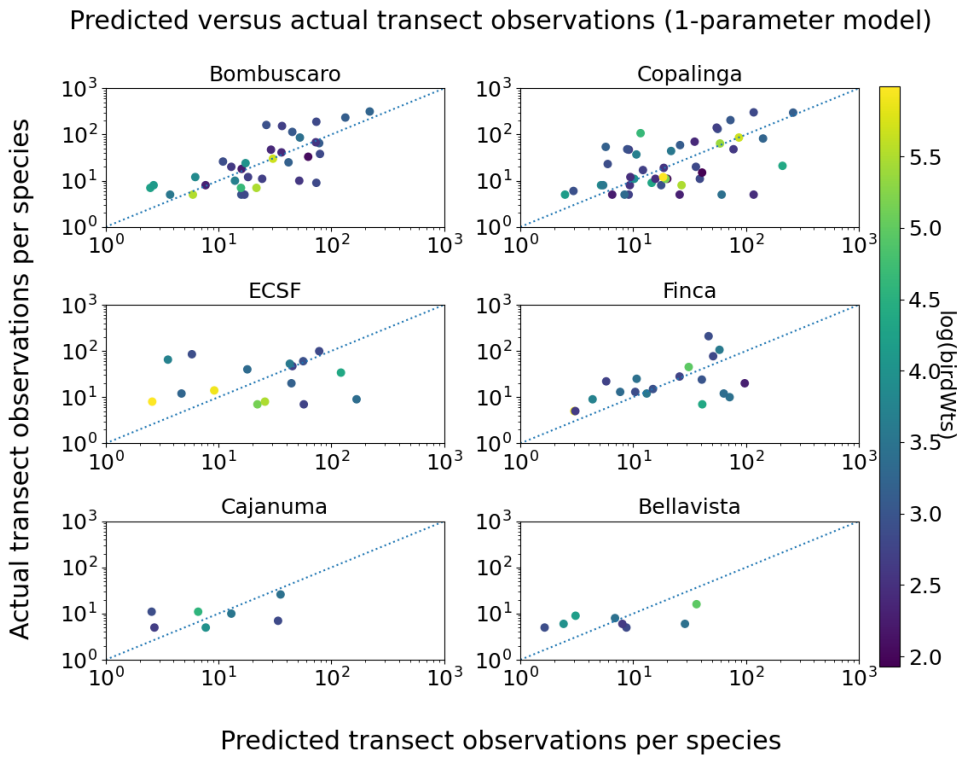

Figure A5: Transect observation prediction accuracy by location for the one-coefficient model (47), which includes bird weight only ($p_f = 0$). Compare with Figure A1 for the two-coefficient model. Axes and line as in Figure A1.

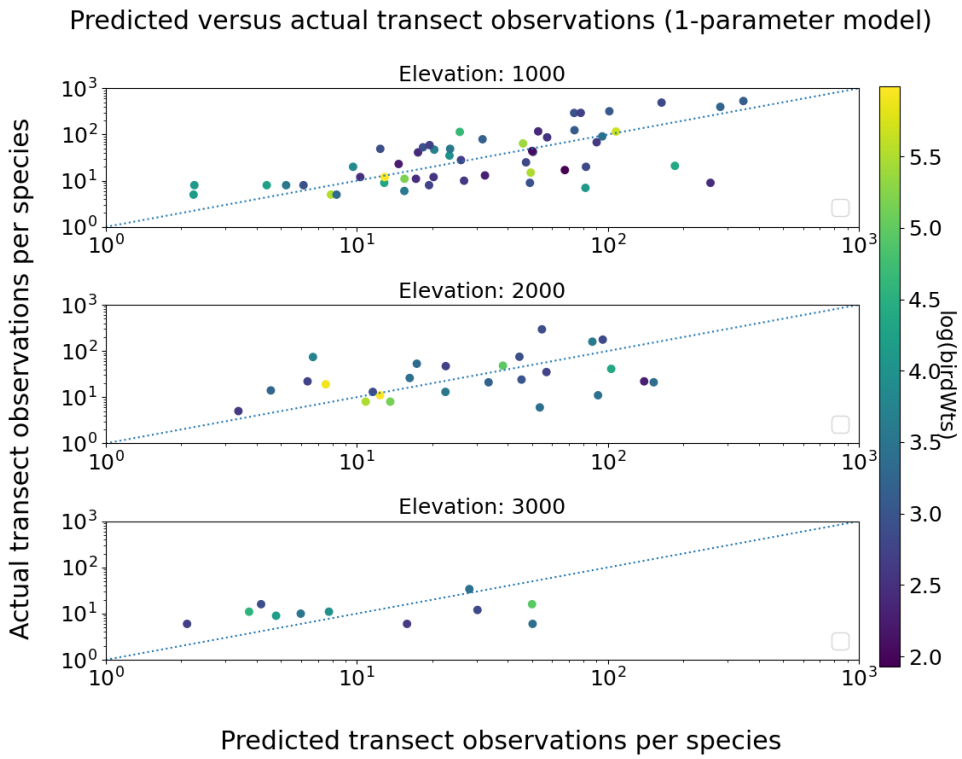

Figure A6: Transect observation prediction accuracy for the one-coefficient model (47) (bird weight only), with species data pooled by elevation. Compare with Figure A2. Axes and line as in Figure A1.

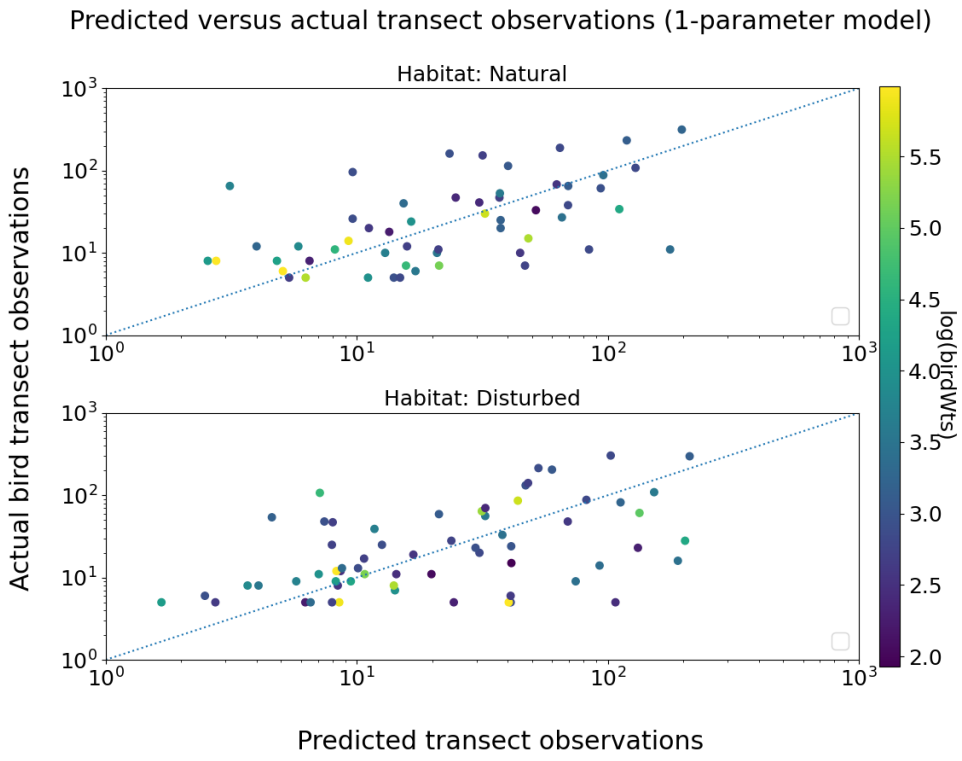

Figure A7: Transect observation prediction accuracy for the one-coefficient model (47) (bird weight only), with species data pooled by habitat type. Compare with Figure A3. Axes and line as in Figure A1.

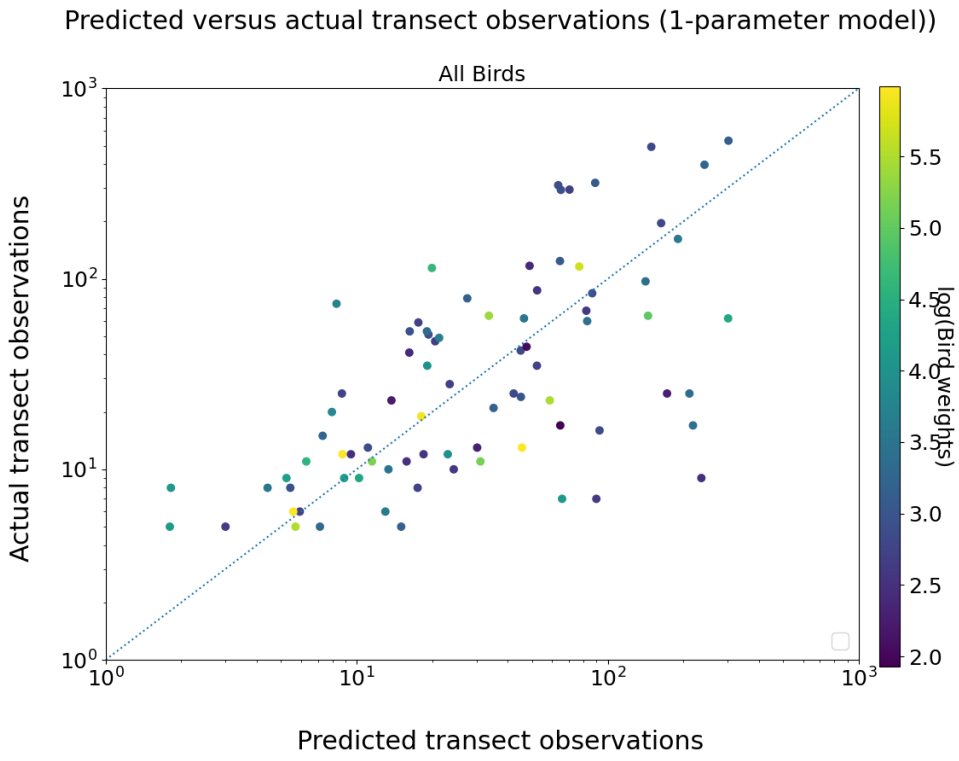

Figure A8: Transect observation prediction accuracy for the one-coefficient model (47) (bird weight only), using all species observations from all six sites combined. Compare with Figure A4. Axes and line as in Figure A1.

Supplement B: Estimator bias, variance, and RMSE using mean bird weight exponent $\bar{p}_b$ (case a)

This supplement contains detailed performance evaluation figures for the point count-based, transect-based, and combined estimators of relative abundance and log relative abundance, computed using the mean value $\bar{p}_b$ of the bird weight exponent (case a as defined in Section 3.2.3).

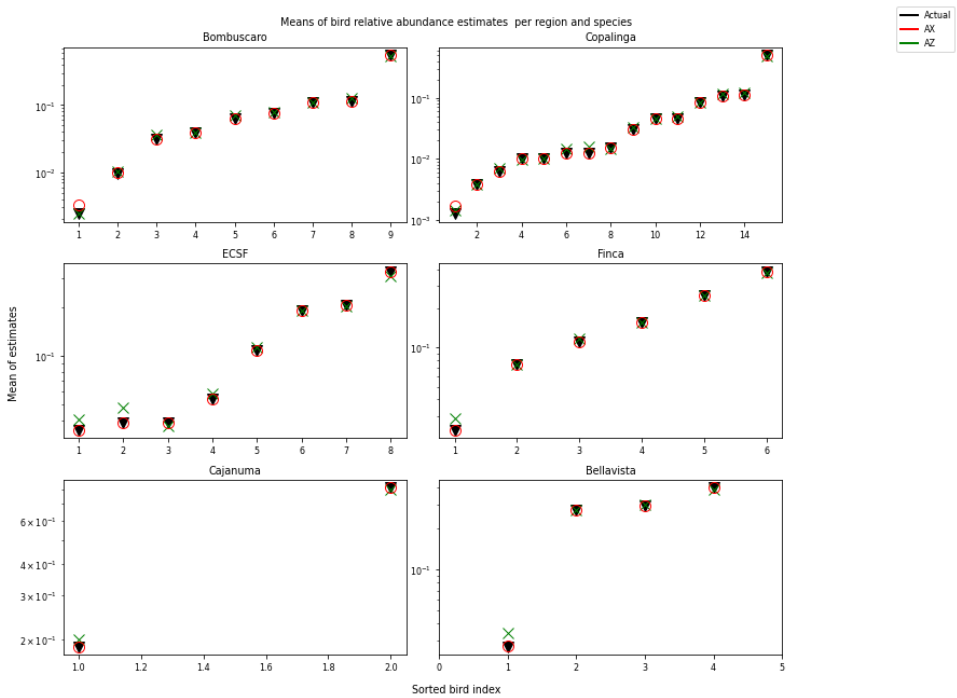

Figure B1: Mean values of the point count-based estimator $(\widehat{a}_n)_z$ (circles) and the transect-based estimator $(\widehat{a}_n)_x$ (triangles) over 10,000 simulations, compared to true relative abundance (dashed line), for bird families at each of the six study sites. The transect estimator is computed using the mean bird weight exponent $\bar{p}_b$ (case a). Within each site panel, families are sorted in order of increasing true relative abundance. Both estimators show little systematic bias, with slightly larger departures for rare families.

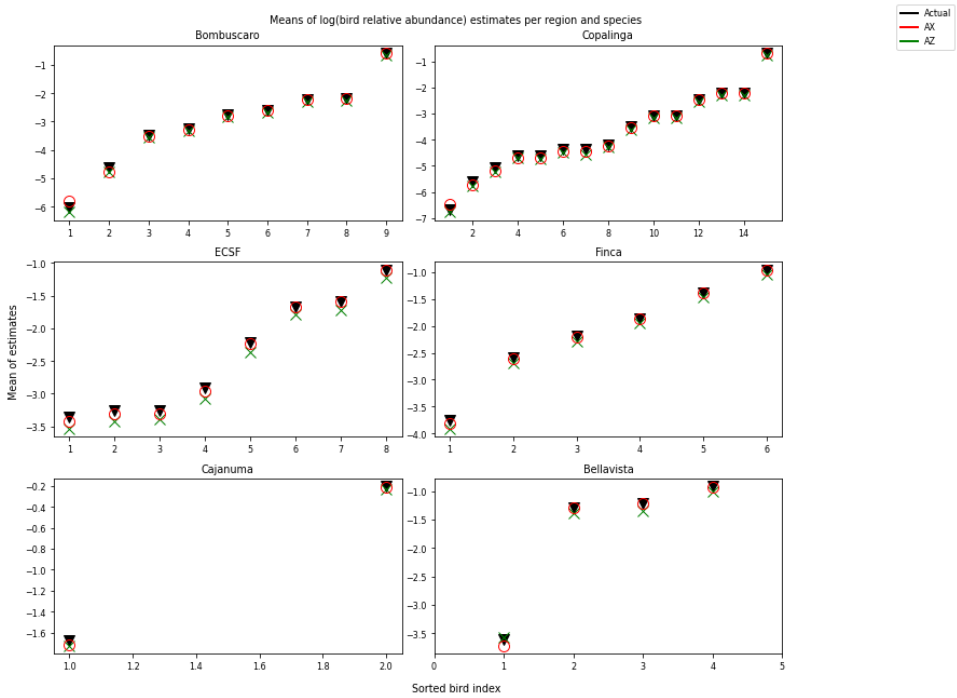

Figure B2: Mean values of the point count-based log abundance estimator $(\widehat{\ell}_n)_z$ (circles) and the transect-based log abundance estimator $(\widehat{\ell}_n)_x$ (triangles) over 10,000 simulations, compared to true log relative abundance (dashed line), for bird families at each of the six study sites. The transect estimator uses the mean bird weight exponent $\bar{p}_b$ (case a). Families are sorted as in Figure B1.

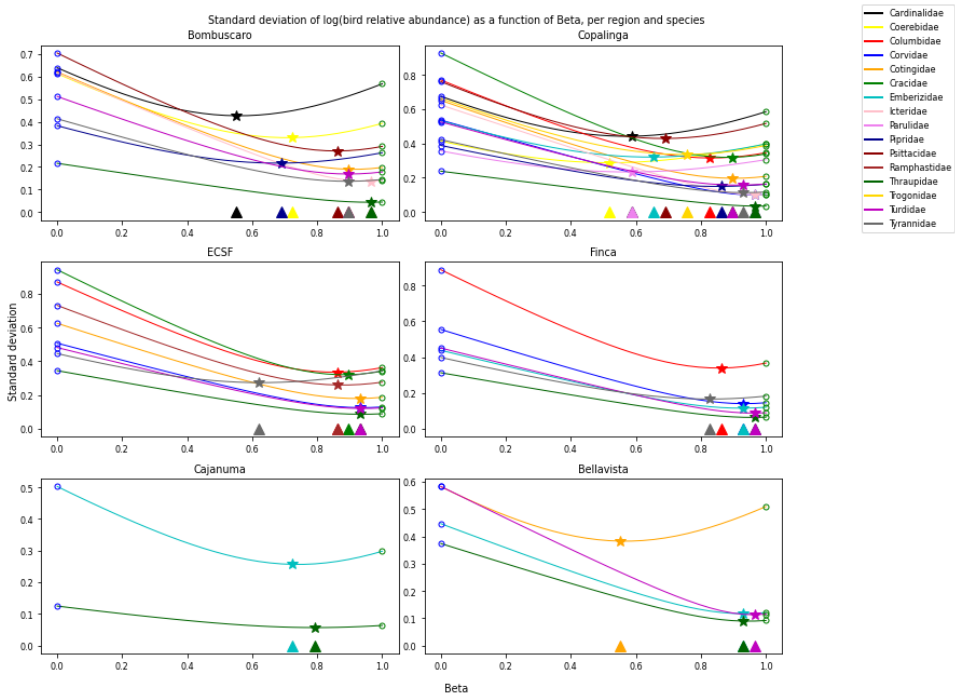

Figure B3: Standard deviation of the combined log abundance estimator $(\widehat{\ell}_n)_\beta$ as a function of $\beta \in [0, 1]$, for each bird family in each of the six study sites. $\beta = 0$ corresponds to $(\widehat{\ell}_n)_X$ (transect-based); $\beta = 1$ corresponds to $(\widehat{\ell}_n)_Z$ (point count-based). Stars mark the minimum of each curve; triangles on the x-axis indicate the theoretically optimal β from (35). Transect estimator uses $\bar{p}_b$ (case a).

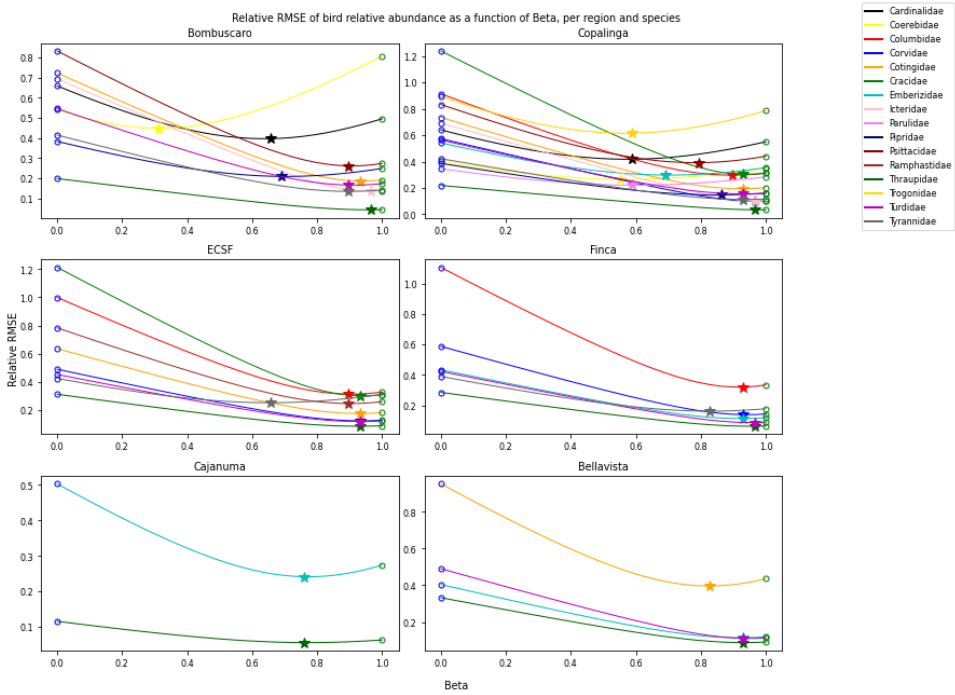

Figure B4: Relative root mean squared error (RMSE divided by true relative abundance) of the combined estimator $(\widehat{a}_n)_\beta$ as a function of $\beta \in [0, 1]$, for each bird family in each site. $\beta = 0$ corresponds to $(\widehat{a}_n)_X$; $\beta = 1$ to $(\widehat{a}_n)_Z$. Stars and triangles as in Figure 4 of the main body. The near-identical appearance of this figure and Figure 4 confirms that bias is negligible relative to variance. Transect estimator uses $\bar{p}_b$ (case a).

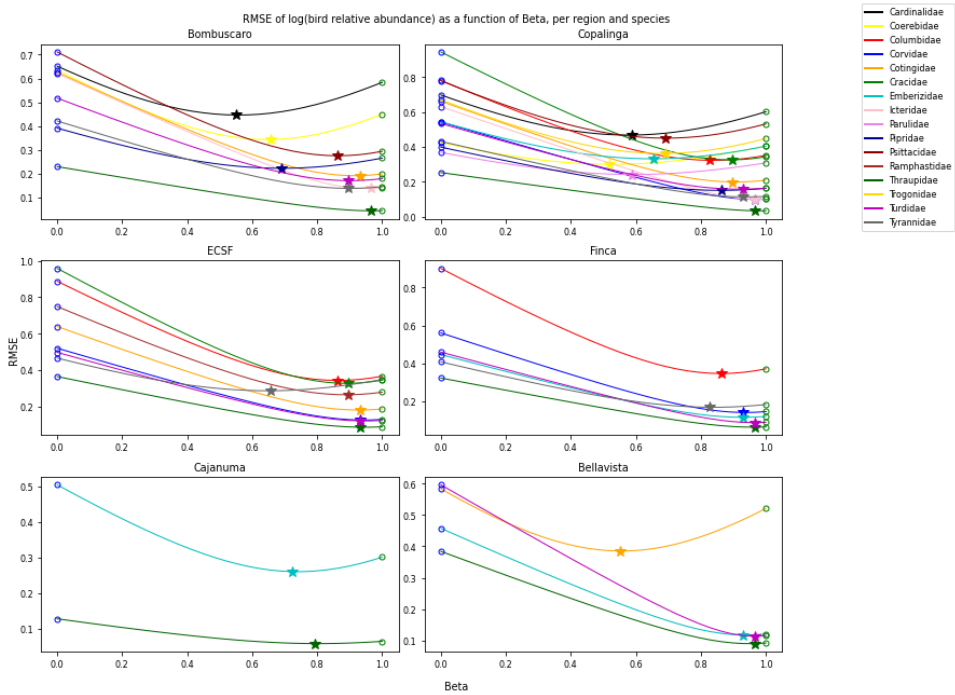

Figure B5: Root mean squared error (RMSE) of the combined log abundance estimator $(\widehat{\ell}_n)_\beta$ as a function of $\beta \in [0,1]$, for each bird family in each site. $\beta = 0$ corresponds to $(\widehat{\ell}_n)_X$; $\beta = 1$ to $(\widehat{\ell}_n)_Z$. Stars and triangles as in Figure B3. Transect estimator uses $\bar{p}_b$ (case a).

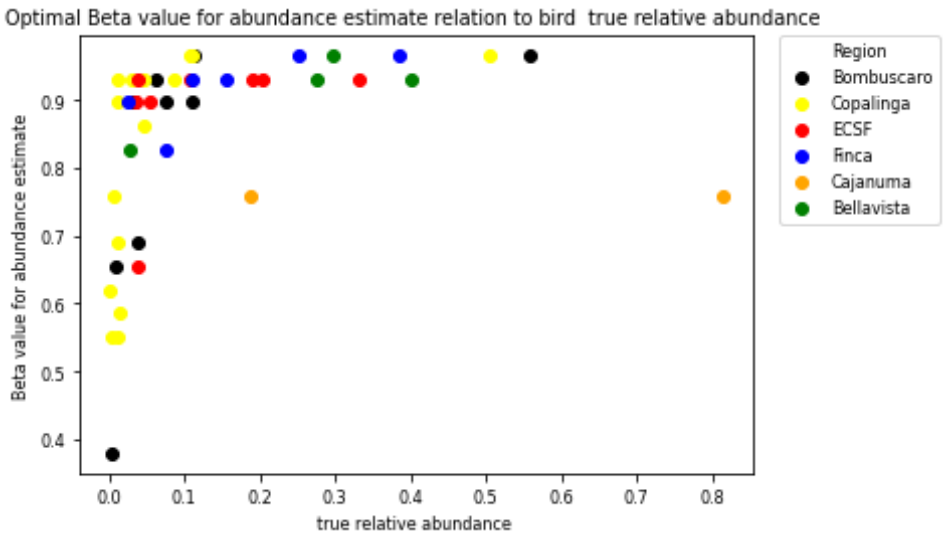

Figure B6: RMSE-minimizing value of β for the combined relative abundance estimator $(\widehat{a}_n)_\beta$, plotted against the true relative abundance of each bird family, for all families across all six sites. Each point represents one family-site combination. Lower β values (giving more weight to the transect estimator) occur predominantly for rare families (true relative abundance below 0.1), while common families consistently receive high β (closer to the point count estimator). Transect estimator uses $\bar{p}_b$ (case a).

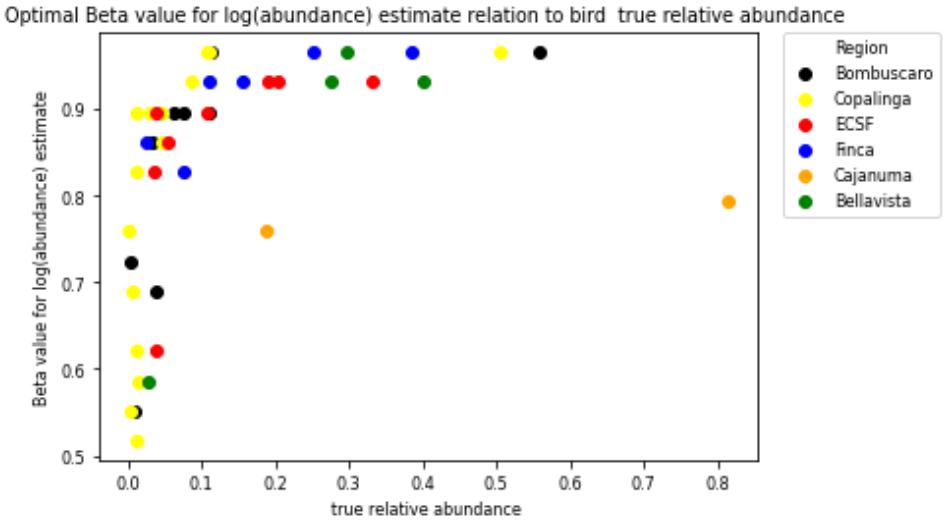

Figure B7: RMSE-minimizing value of β for the combined log abundance estimator $(\widehat{\ell}_n)_\beta$, plotted against true relative abundance. Layout and interpretation as in Figure B6. Transect estimator uses $\bar{p}_b$ (case a).

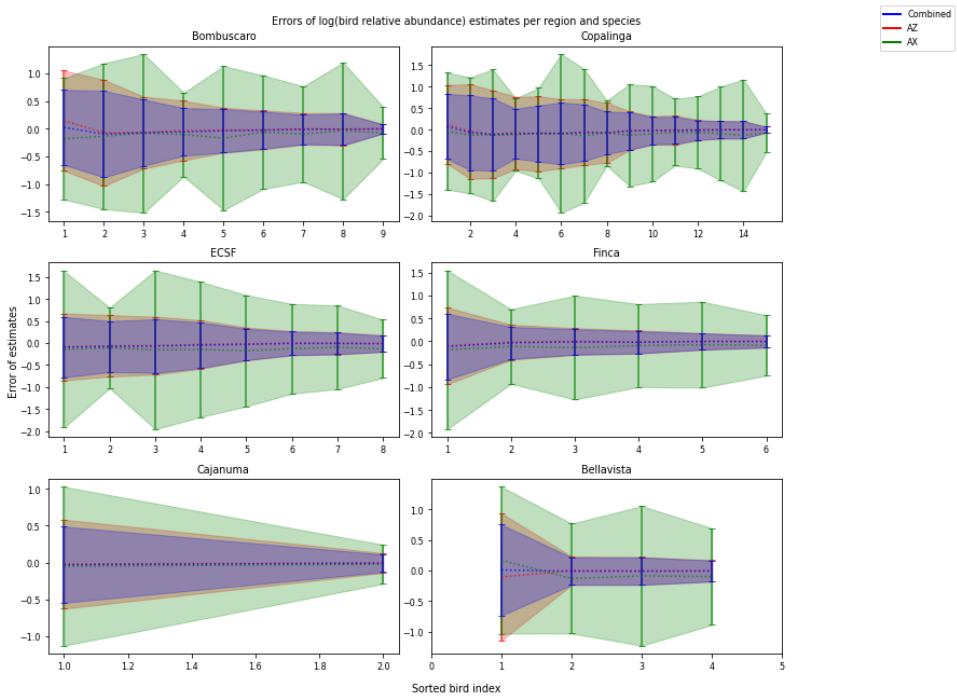

Figure B8: 95% confidence intervals for the error of three log abundance estimators: $(\widehat{\ell_n})_Z$ (labeled AZ), $(\widehat{\ell_n})_X$ (labeled AX), and the combined estimator $(\widehat{\ell_n})_\beta$. Results are shown for each bird family in each site. The AX and combined estimators use $\bar{p}_b$ (case a). Families are ordered by increasing true relative abundance within each site panel.

Supplement C: Estimator bias, variance, and RMSE using exact bird weight exponent $p_{b,n}$ (case b)

This supplement contains figures corresponding to those in Supplement B, but computed using the exact bird weight exponent $p_{b,n}$ for each family (case b as defined in Section 3.2.3). Comparison with Supplement B illustrates the improvement in estimator accuracy that results from precise knowledge of the weight-energy relationship.

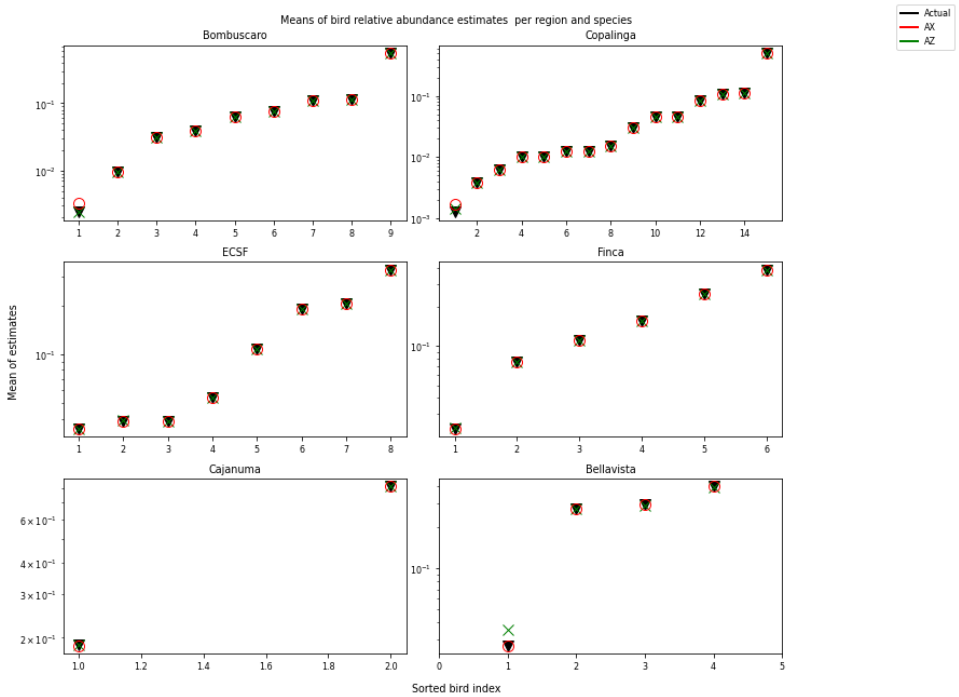

Figure C1: As Supplement B Figure B1, but with the transect-based estimator computed using the exact bird weight exponent $p_{b,n}$ for each family (case b). Compared to Figure B1, the transect estimator shows reduced bias and tighter agreement with the true values, demonstrating the benefit of accurate knowledge of the weight-energy relationship.

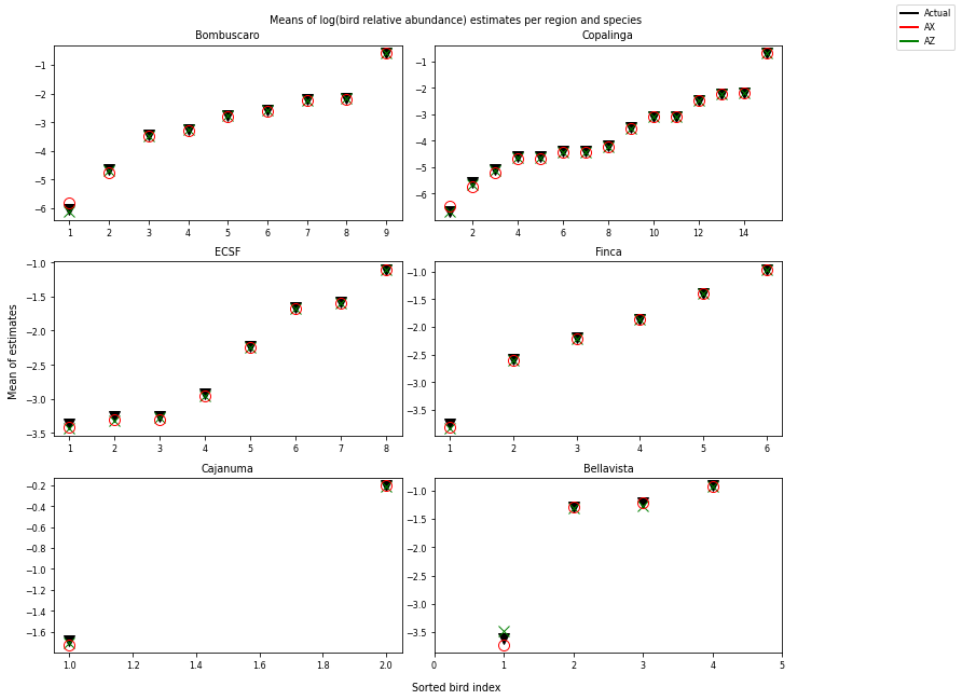

Figure C2: As Supplement B Figure B2, but with the transect-based log abundance estimator computed using the exact bird weight exponent $p_{b,n}$ for each family (case b). Families are sorted as in Figure B1.

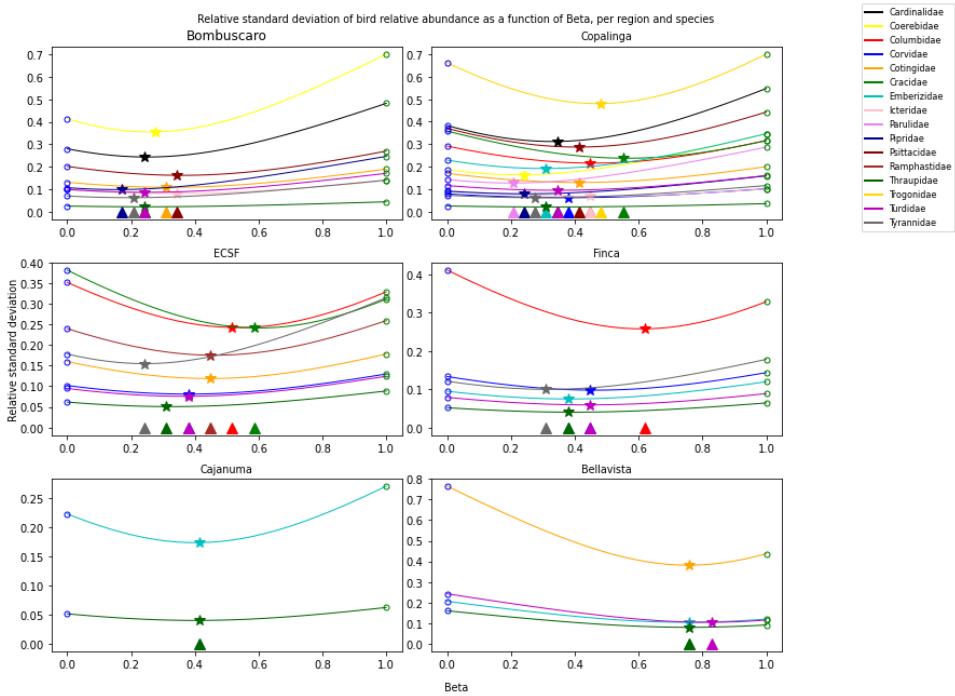

Figure C3: As main body Figure 4, but with the transect estimator using the exact bird weight exponent $p_{b,n}$ (case b). The minimum standard deviation at $\beta = 0$ (pure transect estimator) is markedly lower than in Figure 4, and optimal β values shift toward lower values, reflecting the improved accuracy of the transect estimator when $p_{b,n}$ is known precisely.

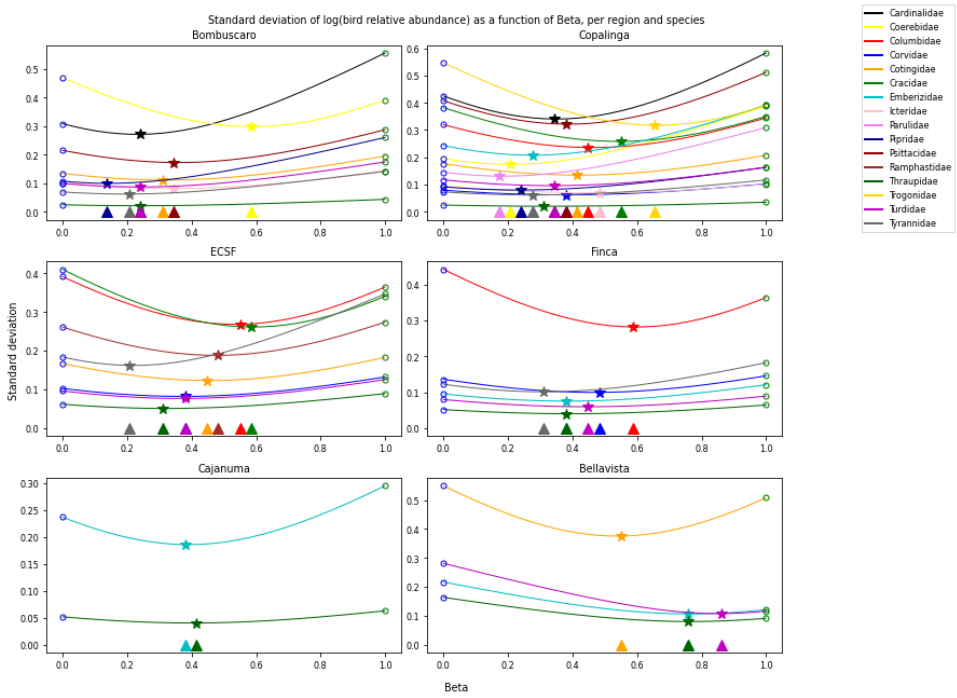

Figure C4: As Supplement B Figure B3, but with the transect log abundance estimator using the exact bird weight exponent $p_{b,n}$ (case b). Optimal β values shift toward lower values compared to Figure B3, reflecting the improved transect estimator accuracy.

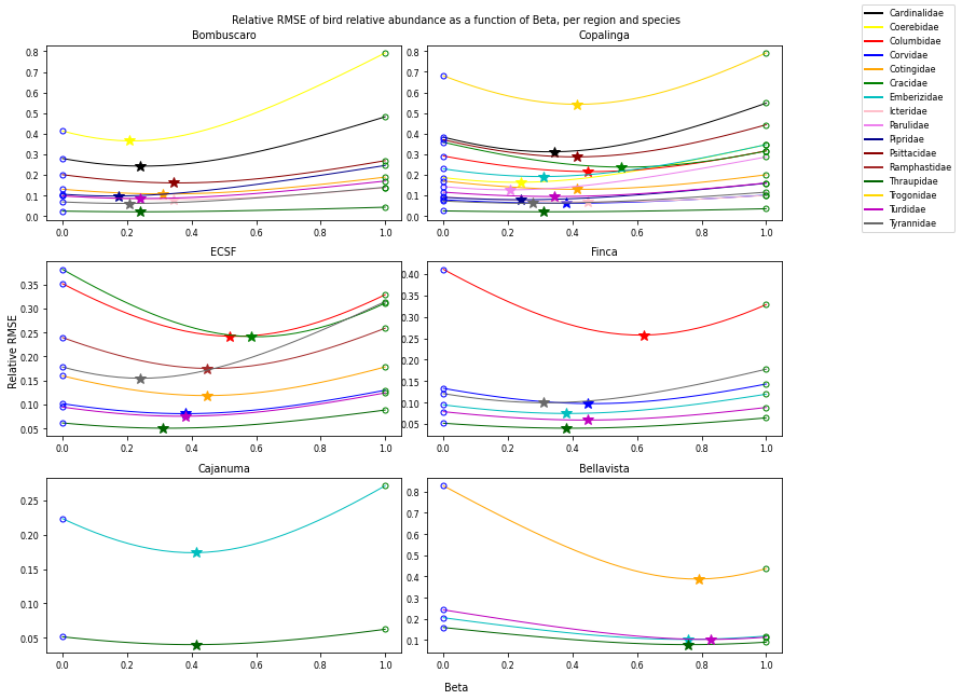

Figure C5: As Supplement B Figure B4, but with the transect estimator using the exact bird weight exponent $p_{b,n}$ (case b). Minimum relative RMSE values are substantially lower than in Figure B4, confirming the advantage of precise knowledge of the weight-energy exponent.

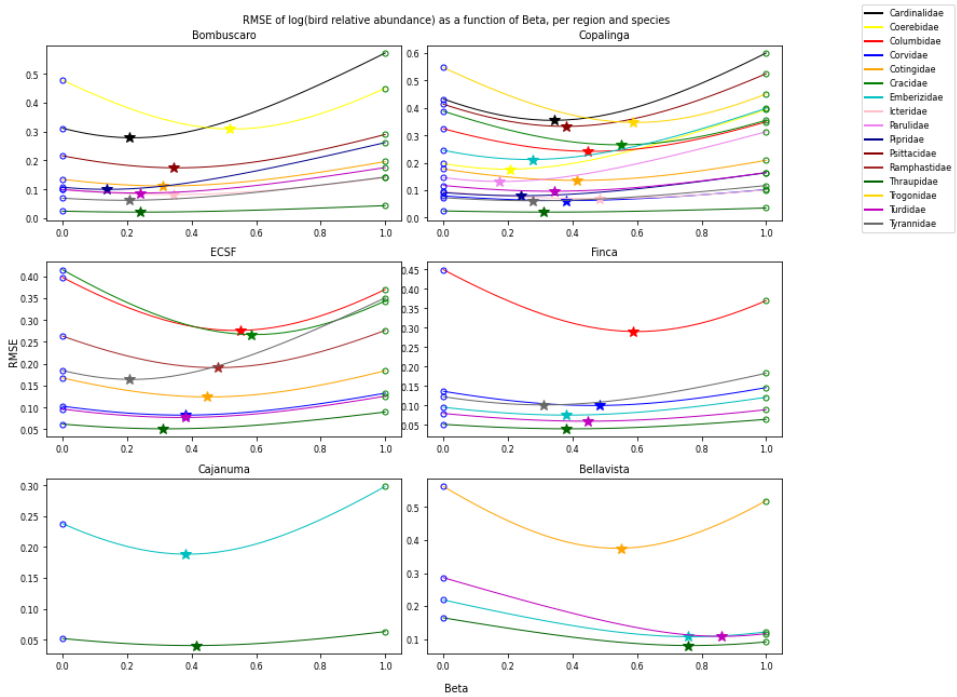

Figure C6: As Supplement B Figure B5, but with the transect log abundance estimator using the exact bird weight exponent $p_{b,n}$ (case b). Minimum RMSE values are reduced relative to Figure B5.

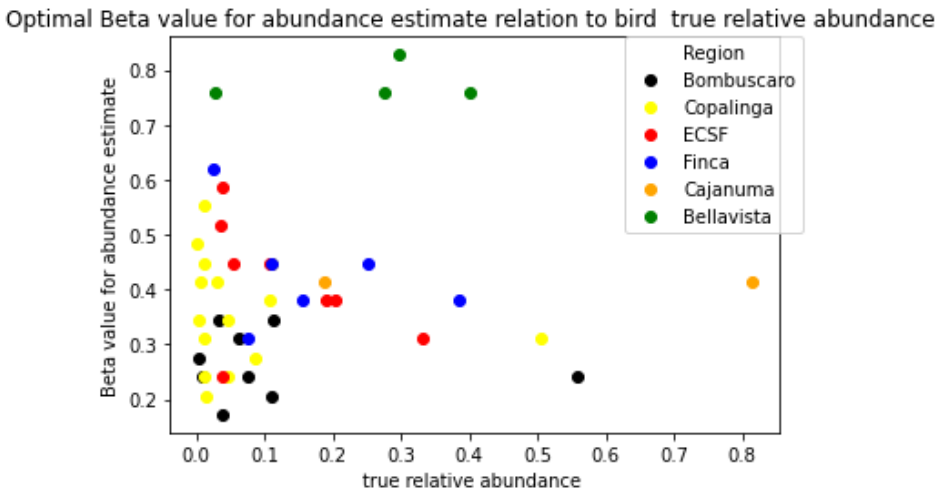

Figure C7: As Supplement B Figure B6, but with the transect estimator using the exact bird weight exponent $p_{b,n}$ (case b). Optimal β values are generally lower than in Figure B6, indicating that more weight is appropriately assigned to the transect estimator when $p_{b,n}$ is known precisely.

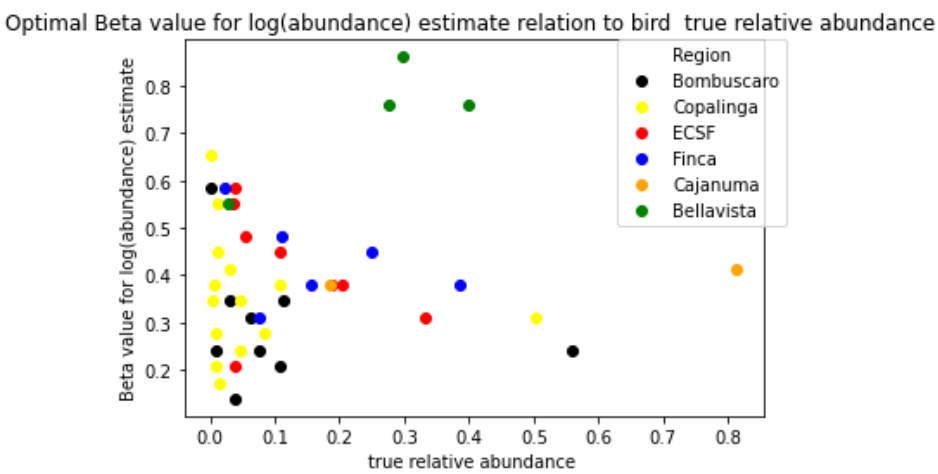

Figure C8: As Supplement B Figure B7, but with the transect log abundance estimator using the exact bird weight exponent $p_{b,n}$ (case b). Layout and interpretation as in Figure C7.

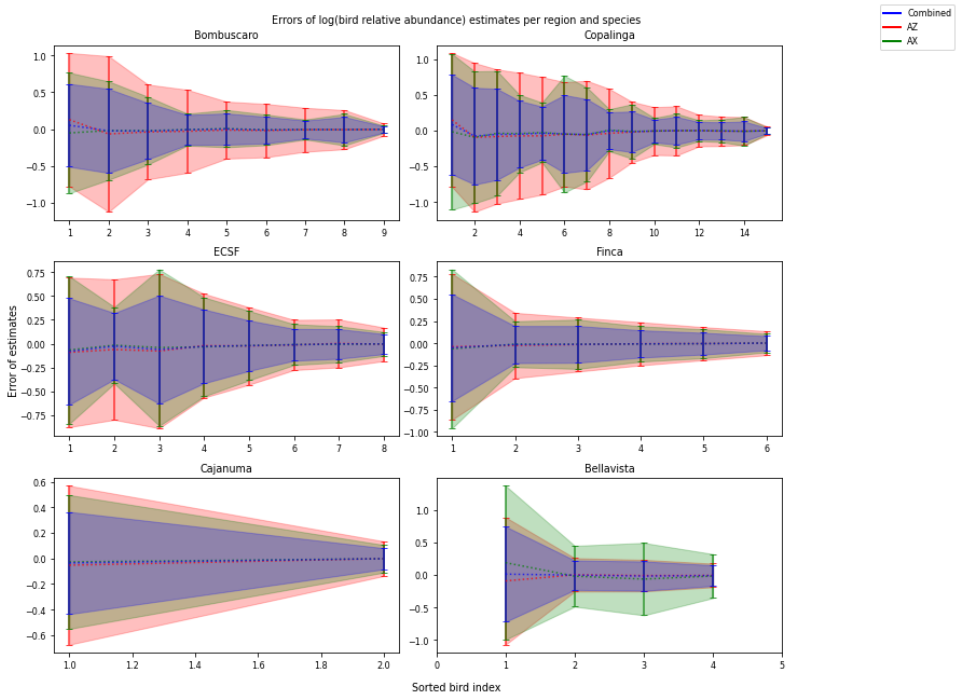

Figure C9: As Supplement B Figure B8, but with the transect-based and combined log abundance estimators using the exact bird weight exponent $p_{b,n}$ (case b). Confidence intervals are substantially narrower for the transect and combined estimators compared to Figure B8.