Title: Evaluating the efficacy of a consumer-centric method for ecological sampling: Using bonobo (*Pan paniscus*) feeding patterns as an instrument for tropical forest characterization

Running headline: Consumer-centric sampling of food availability

Authors: Erin G. Wessling^{1*}, Liran Samuni^{1,2}, Roger Mundry^{3,4,5,6}, Miguel Adan Pascual⁷, Stefano Lucchesi^{1,2}, Bienfait Kambale⁸, Martin Surbeck^{1,7}

Author details:

¹ Harvard University, Cambridge, MA, USA

² Max Planck Institute for Evolutionary Anthropology, Leipzig, Germany

³ Platform Bioinformatics and Biostatistics, VetMedUni, Vienna, Austria

⁴ Cognitive Ethology Laboratory, German Primate Center, Leibniz Institute for Primate Research, Göttingen, Germany

⁵ Department for Primate Cognition, Georg-August-University Göttingen, Germany

⁶ Leibniz Science Campus Primate Cognition, Göttingen, Germany

⁷ Kokolopori Bonobo Research Project, Tshuapa, Democratic Republic of the Congo

⁸ Centre de Surveillance de la Biodiversité de l'Université de Kisangani, Kisangani, Democratic Republic of the Congo

* Corresponding author: Dr. Erin G. Wessling, Department of Human Evolutionary Biology, Harvard University, 11 Divinity Avenue, Cambridge, MA, USA, email: ewessling@fas.harvard.edu

Title: Evaluating the efficacy of a consumer-centric method for ecological sampling: Using bonobo (*Pan paniscus*) feeding patterns as an instrument for tropical forest characterization

Abstract

- 1 1. Characteristics of food availability and distribution are a key component of a species ecology. Objective
2 measurement of food resources, such as vegetation plot sampling, do not consider aspects of selection
3 by the consumer and therefore may produce imprecise measures of availability. Further, in most animal
4 ecology research, traditional ecological surveying often is time-intensive and supplementary to ongoing
5 behavioral observation. We propose a method to integrate ecological sampling of an animal's
6 environment into existing behavioral data collection systems by using the consumer as the surveyor.
7 Here, we introduce the consumer-centric method (CCM) of assessing resource availability for its ability
8 to measure food resource abundance, distribution, and dispersion. This method catalogues feeding
9 locations observed during behavioral observation and uses aggregated data to characterize these
10 ecological metrics.
- 11 2. We evaluated the CCM relative to traditional vegetation plot surveying using accumulated feeding
12 locations across three years visited by a tropical frugivore, the bonobo (*Pan paniscus*), and compared
13 it with data derived from over 200 vegetation plots across their 50km²+ home range.
- 14 3. We demonstrate that food species abundance estimates derived from the CCM are comparable to
15 those derived from traditional vegetation plot sampling after approximately 600 observation days or 60
16 spatially explicit feeding locations. The agreement between the methods further improved when
17 accounting for aspects of consumer selectivity in objective vegetation plot sampling (e.g., size minima).
18 Estimates of density from CCM correlated with plot-derived estimates and were relatively insensitive to
19 home range inclusion and other species characteristics, but were sensitive to sampling frequency (e.g.,
20 consumption frequency). Agreement between the methods in relative distribution of resources
21 performed better across species than expected by chance, although measures of dispersion correlated
22 poorly.
- 23 4. We demonstrate that while providing a robust measure to quantify local food availability, the CCM has
24 an advantage over traditional sampling methods as it incorporates sampling biases relevant to the
25 consumer. Therefore, as this method can be incorporated into existing observational data collection

and does not require additional ecological surveying, it serves as a promising method for behavioral ecological data collection for animal species who re-use space and consume immobile food items.

Keywords: dispersion | distribution | food availability | resource selection | species abundance | vegetation plot

1. Introduction

The abundance, dispersion, and distribution of food resources not only determines species distribution but also has a strong impact on many aspects of an animal's life-history, physiology, and sociality (e.g., Anholt and Werner 1995; Chapman et al. 2015; Davies and Deviche 2014; Hutto 1990; Lambert and Rothman 2005; Rogers 1987; van Schaik et al. 1993; Vogel and Janson 2007). Due to the core importance of food to an organism, the quantification of food availability and distribution are key considerations across studies and disciplines. Methods used to estimate food resource abundance, distribution, and dispersion are just as varied as the questions which necessitate these quantifications (Szigeti et al. 2016).

Measurement of food resource **abundance or density** (i.e., estimation of the amount of a resource available in a landscape) depends heavily on the type of resource and scale of interest (Bowering et al. 2018; Morrison 2016). Large scale analyses of abundance typically rely on remotely derived proxies via satellite imagery, but for questions related more immediately to the individual or social group scale, direct measurement of exploitable resources offer more direct insights into the resources available to a consumer (e.g., Foerster et al. 2016; Wessling et al. 2020). While mobile resources may be measured via consumer behavior (e.g., attack rates: Hutto 1990), for static food resources like plants, **abundance** is commonly estimated by sampling subsets of the area of interest. Example methods include transects or vegetation plots/quadrats (Baraloto et al. 2013; Ståhl et al. 2017; Vogel and Janson 2007), with the latter being the most common sampling method in studies of frugivorous or folivorous animals. To then estimate **distribution**, that is, a calculation of relative resource abundance or density across space within a landscape, sampling may be further stratified across a given area relevant from individuals to populations (e.g., home range, landscape, or region).

Measures of **dispersion** (i.e., patterns of clustering or patchiness), such as Morisita's index (Morisita 1962), are used to quantify the clustering of resources over space within a landscape (Krebs 1999; Stephens and Krebs 1986), often used in the contexts of understanding resource competition and socio-ecological behavior (e.g., Vogel and Janson 2011). Quantifications of food species dispersion are perhaps even more varied in practice and sensitive to the scale relevant to the consumer (e.g., Myers 1978; Vogel and Janson 2011). Dispersion metrics may also require distinct sampling methods tailored to specific questions (e.g., Vogel and Janson 2007), thus potentially requiring supplementary surveying effort to food abundance surveying.

Despite its centrality to animal ecological research, ecological sampling design frequently does not conform to recommended standards nor is adequately validated by animal ecologists (Mortelliti et al., 2010; Szigeti et al. 2016). For example, sampling effort can substantially impact measures of resource abundance, but it is rarely validated whether efforts are sufficient to adequately measure the intended metrics. Further, ecological data collection often requires research effort additional to ongoing behavioral observations and is time intensive and thus infrequently conducted. Snapshots of abundance derived from these efforts in a landscape may therefore be used even many years after they have been collected or may fail to account for temporal variation.

The problem of insufficient quantifications of resource availability may also extend to sampling design. While traditional sampling methods in animal ecology may offer an objective measure of the resources potentially accessible to a consumer, these methods are by design blind to aspects of resource selection by the consumer. Given these disadvantages, the question arises whether there is a way to conduct ecological sampling that is time-efficient within existing behavioral data collection systems and also integrates resource selection criteria of the consumer? Behavioral observation has been used previously as a measure of food availability (Lovette and Holmes 1995; Hutto 1990), dispersion (Vogel and Janson 2011), and preference (Forester et al. 2009), however these methods are either limited in application or still necessitate ecological data collection.

We therefore introduce a consumer-centric method (CCM) for animal behavioral ecology studies which uses the consumer as the survey vehicle to potentially quantify food resources in a landscape. With this method, researchers catalogue food resource locations as they are consumed during the process of

behavioral observation. Here, we evaluate the CCM relative to traditional habitat plot data collection using accumulated feeding locations from two social groups of a tropical frugivore, the bonobo (*Pan paniscus*), as a case study. Specifically, we investigated whether behavioral data on feeding locations (trees and lianas) provide a reliable dataset allowing inference about food species' (1) densities, (2) distribution and (3) dispersion. We additionally assess (4) the minimum sampling effort required and (5) for what characteristics of a food species this method can be considered most reliable.

2. Methods

2.1 Study Species and Behavioral Observation

Data were collected at the Kokolopori Bonobo Reserve (Fig. 1) on two social groups of bonobos (Ekalakala: EKK, Kokoalongo: KKL) between May 2016 and December 2019. Groups were followed daily for behavioral data collection, during which we collected group feeding locations using a GPS (Garmin GPSMAP 62), and circumference at breast height (synonymous with and hereafter referred to as DBH) of feeding trees ≥ 20 cm diameter and lianas ≥ 5 cm DBH (SI 2.1, 2.2). Due to GPS measurement error and consequently an inability to distinguish individual trees on a small scale, we summarized feeding tree locations of each group into presence or absence of each species in 50 x 50 m cells. We used location data collected with the GPS tracklog function to calculate the home range of both bonobo groups using kernel density estimates (see SI 2.1). These groups share overlapping areas of their home ranges, including 64% and 66% of the home ranges of EKK and KKL, respectively (Samuni et al. 2020). We evaluated whether feeding location datasets were sufficiently sampled and stable by considering accumulation patterns of data per species over time (SI 2.3).

Fig. 1. (Left) Location of the study site relative to global bonobo distribution. (Right) 50 x 50m habitat plots (black dots; not to scale) within 1 km² grid cells (black square) overlaid upon all visited 50x50 cells within the 95% home range kernels for Ekalakala (red) and Kokoalongo (blue) bonobo groups.

2.2 Vegetation plots

We conducted vegetation plot sampling by overlaying 1x1 km grid cells over the whole ranging area and aimed to conduct plot sampling in every grid cell utilized by at least one of the groups (Fig. 1; SI 1.2). Like the observational cells, all habitat plots were 50 x 50m in size, within which data were collected on all trees meeting the minima defined for observational cells. In total, we sampled 236 plots within these grid cells, of which 214 plots fell within the 95% home range of either group, with 162 and 170 within the 95% range of EKK and KKL, respectively (Fig. 1). Plot sampling averaged 4.1 ± 1.6 (SD) plots per km² (range: 1 to 7) and was determined to be of sufficient sampling depth (SI 1.3).

2.3 Comparison of datasets

2.3.1 Density

To compare estimated species abundances derived from each dataset (CCM or vegetation plots), we derived three different indices. 1) We used the bonobo observational data to create a ‘presence index’ based on bonobo feeding locations for each food species, estimated as the number of 50x50m cells in which each species was present divided by the total number of cells within the 95% kernel home range of each group (see Fig. S5 for an example). 2) We calculated species density estimations using the vegetation plot data as the total number of individuals observed per area surveyed (num. individuals / km², hereafter “Plot Density”). 3) We calculated the number of 50x50m plots in which each species was present per total number of plots sampled for more direct comparison with the CCM (hereafter “Plot Presence”).

To evaluate method agreement, we created pair-wise sets of comparisons of the three density indices by means of Pearson’s correlation tests and used the correlation coefficient (r) as a measure of strength of agreement between methods. We conducted the pair-wise comparisons while assessing the influence of sampling effort on method agreement by varying levels of home range usage (kernel % range from 20 until 95 % in increments of 1%) and dietary inclusion (top 10 most consumed species until full diet) for each group. We only considered comparisons with at least 10 species in at least 10 vegetation plots. We additionally created a moving window over the kernel home range from 20% to 95% for which to compare methods more directly according to home range location. This window accounts for variation in area coverage by adjusting window radius to impose similarly sized datasets for comparison over the range of % kernel inclusion (i.e., for agreement from home range core to periphery; SI 3.1).

Finally, to identify potential dataset minima required for reliable and stable density indices derived from the CCM, we evaluated the pattern of correlation strength between indices from each method as the dataset grew over time (i.e., day of data collection), and set the minimum as the point from which the correlation coefficient remains relatively stable. We describe p -values for these correlations in our summaries below, however as these correlations require independent data and because we evaluated thousands of correlation coefficients per group ($n_{\text{EKK}}=15075$ and $n_{\text{KKL}}=12834$), we do not draw inference based on p -values but instead focus only upon correlation coefficients.

2.3.2 Dispersion

To evaluate agreement between methods in characterizing food species dispersion, we used Morisita's index (Morisita 1962). To allow for standardized and directly comparable sample units from which to calculate this index for both methods, we aggregated number of individuals per species visited by the bonobos across three different grid cell sizes (500x500m cells, 1000x1000m cells, and 1500x1500m cells), and calculated the average number of individuals for each species in each of these grid cells using the vegetation plot dataset. For both datasets, we then calculated the Morisita's index using the *dispindmorisita* function of the package 'vegan' (Oksanen et al. 2019) for each species. We further accounted for an unusual distribution of Morisita's indices deriving from the vegetation plot dataset by transforming the data to allow for a more normal distribution (SI 3.2).

2.3.3 Distribution

To evaluate the efficacy of the CCM to reliably quantify the distribution of food species in a landscape, we aggregated data by grid cell as in our dispersion comparison. We compiled the abundance data for both bonobo and vegetation plot datasets in two ways: by either i) aggregating (CCM) or averaging (plot dataset) the number of individuals per species per grid cell or ii) by marking the presence/absence of a given species per grid cell size. We chose to average rather than aggregate plot data because greater plot sampling in a grid cell will inherently increase species abundances, whereas sampling biases in CCM could be accounted for by controlling for location within each group's home range (i.e., % kernel home range). We then fitted model sets separately for each cell size and group (six sets of up to 70 species each), using each food species as a dataset and each cell as a datapoint. We used the estimated bonobo feeding data abundance per cell (a measure of distribution) as the response and the plot abundance as the test predictor using zero inflated Poisson models (500x500m grid size) or simple linear models for (1000x1000 and 1500x1500m grid sizes). Within these models, to account for variation in home range utilization by the bonobos we controlled for the % kernel home range of each cell by averaging the % kernel value assigned to each of the vegetation plots used to estimate the species abundance within that cell. We then calculated average Nagelkerke's R^2 (500x500m) or r^2 (1000x1000m and 1500x1500m) for each model set across levels of dietary inclusion (see SI 3.3 for detailed descriptions of the fitted models and model checks).

To also evaluate agreement between methods on simple presence of a species in a cell, we fitted a generalized linear mixed model with binomial error structure (Baayen 2008) for each grid cell size and each social group. The response in this model was the presence or absence of a species in a given cell as predicted by the bonobo observational data (with one datapoint per species per cell), and presence as measured by vegetation plot and % kernel as test predictors. In these (six total) binomial models we included cell ID and species as random effects and included random slopes for presence/absence in the plots and their correlation within the random effect of species (SI 3.3 for details and model checks). As a last validation of distribution agreement, we identified when bonobos missed the presence of a species in a cell that had been identified in habitat plots and calculated a proportion of missed species occurrences out of all cells per species, as well as evaluated potential sources of biases in likelihood to miss a species in a cell (see 2.4).

2.4 Identifying sources of bias

If a consumer is selective in which resources it uses within a landscape, then measurements from vegetation plots may not accurately measure the relevant resources to that consumer. To evaluate these potential discrepancies, we compared food tree and liana sizes (strongly tied to variability in food crop production: Chapman et al. 1992; SI 4) between CCM and vegetation plot data as an example of a potential selective characteristic. We then quantified seven characteristics of each species to evaluate how they contribute to rates of data accumulation and agreement between our sampling methods. Specifically, we considered the lifeform (tree or liana), patterns of dispersion, consumed food item (fruit or non-fruit), seasonality of consumption, density in the landscape, DBH variability, and frequency of consumption (SI 4.1) as test predictors in models with the following responses (SI 4.2): (1) the speed at which data accumulate in the CCM dataset (2) a measure of difference between estimates of density between the methods, and (3) likelihood for bonobos to miss the presence of a species in a cell (SI 4.2).

2.4 General Analyses

All data analyses were conducted in R (version 4.0.2; R Core Team 2020), and models were fitted using functions of the 'lme4' package (1.1.23; Bates et al. 2015). We report *p*-values between 0.05 and 0.1

as a 'trend' for all models to ease issues of dichotomization of significance (Stoehr 1999). To avoid issues of multiple testing when identical models were run across responses which varied only in their summary method (e.g., grid cell size) or dataset (e.g., social group), we describe only patterns which are stable and significant or trending across at least half of each model set; full results for all models as well as further description of all methods and model checks can be found in the SI. We used log transformation to help return predictor (e.g., species density, consumption frequency) and response (all density indices) variables to a roughly normal or symmetric distribution when they were right-skewed.

3. Results

3.1 Consumer Centric Dataset

The bonobo groups visited (i.e., fed in) a total of 12430 (EKK) and 13827 (KKL) 50x50m cells, amounting to an area 'surveyed' of 31.1 (EKK) and 34.6km² (KKL). This amounts to 58.6km² total area surveyed, as 46.7% of this area occurred within the home range overlap of both communities. Bonobos from EKK and KKL fed on a total of 78 tree and liana species (88.6% occurring in the diets of both groups) from trees and lianas, of which 96% of feeding occasions could be identified to a local name. These observations amounted to 8818 (EKK) and 9140 (KKL) unique feeding tree/liana locations (50x50m) consisting of 76 (EKK) and 72 (KKL) species, of which 58 (EKK) and 55 (KKL) species were consumed in at least 10 locations. The diets of both groups were strongly skewed towards a few frequently consumed species (SI 2.4). The groups visited a similar number of locations each day, with a mean of 10.0 ± 5.5 (KKL) and 8.9 ± 5.0 (EKK) locations visited. On average, 4.5 ± 2.0 (KKL) and 4.3 ± 1.8 (EKK) species were consumed per day by the bonobos.

Bonobos visited 60% (EKK) and 56% (KKL) of all visited cells within the first year of data collection, with gradual declines in the accumulation of newly visited cells over the 3+ year study period in both groups and a clear approach towards an asymptote for most of the top 30 species (Fig. S2). We found that the speed at which new feeding locations were added to the dataset also decreased across species (i.e., longer accumulation times) with each passing year for both groups, and that much of the observed decrease in new locations visited over time was likely driven by significant gains early within the dataset (Figs. S2, S3; SI 2.3). Data on species more variable in size (DBH) accumulated slower in EKK than species more uniform

in size (but no such relationship was found in KKL), and accumulation was also slower in species consumed for their fruits and in more abundant species in the landscape in both groups (SI 4.3, Tables S2 and S3).

3.2 Vegetation Plot Dataset

In total, 14855 trees and lianas were measured across 214 habitat plots (SI 1.1), thus exceeding plot surveying minima (124 plots, SI 1.3). Plot surveying required a cumulative total of 146 team days, averaging to 1.7 ± 0.6 (SD) plots completed per team day (range: 1 – 4). Trees comprise the majority (66.9%) of the individuals measured. This dataset averages to 277.7 individual tree and lianas /ha in across the habitat of these two groups, with 196.1 indiv./ha for food species, and 168.2 indiv./ha for potential food trees that meet bonobo size minima (see below) for the EKK and KKL home ranges collectively.

Seventy-five of the 200 taxa identified in the plots were consumed by at least one of the two groups, with 67 of 72 (EKK) and 70 of 75 species (KKL) in the diet occurring in the plots. Like the bonobo diet, the forest is heavily biased towards a few species, with one species accounting for over 10% of the dataset ('Bofili', local name for *Scorodophloeus zenkeri*), and the top 10 most common tree species accounting for almost 40% of all trees and lianas (n=6375, 39.2%). Correspondingly, only 16 species account for over 50% of the individuals in the plots, of which 11 occur in the diet of both groups. Species in the bonobo diet accounted for 67% of the total number of trees or lianas observed in the Kokolopori landscape.

3.3 Dataset comparison

3.3.1 Consumer selectivity of tree sizes

Trees visited by bonobos were significantly larger on average than trees measured in the plots (EKK: $t=-17.71$, $p<0.001$; KKL: $t=-20.38$, $p<0.001$), but by only an average of less than 1cm in both groups (Table S1). For 23.1% of consumed species, we found more individuals in the plots that did not reach the minimum size consumed than those who did exceed this minimum threshold. We subsequently restricted all analyses to trees/lianas that met this threshold, consequently reducing the number of individuals included in plot dataset by approximately 18% in both groups (8891 individuals in EKK and 8685 in KKL; SI 4). Reducing the dataset had a measurable effect on the correlation strengths between estimates of density (see below), with an average improvement of 0.04 for comparison (r) of the CCM estimate with the

Plot Presence estimates, and 0.07 improvement in correlation coefficient in the comparison of the CCM estimate with Plot Density.

3.3.2 Density

We found that the density estimates from the CCM and vegetation plots were comparable in both groups (Table 1 and Fig. 2). Patterns of correlational strength between the methods stabilized and smoothed from approximately 50% kernel home range inclusion and above, and when approximately a minimum of 15 species was included in the dataset of both groups. Statistical significance of the correlation was reached in both groups when including ca. 20 of the top species or more. Inclusion of less frequently used areas of the home range to the comparison did not appear to considerably affect the strength of agreement between methods but correlation strength decreased with greater number of species included in the comparison (Fig. 2; Table 1). While we did observe that peripheral areas of the home range generally resulted in lower methodological agreement (Fig. S6), bonobo data appeared largely insensitive to inclusion of the outer reaches of the home range in both groups when included alongside more intensively surveyed areas (i.e., the core range).

Broadly, the CCM more closely matched estimates of Plot Density relative to Plot Presence. However, for both comparisons we observed a decrease in the correlation coefficient the greater the number of species included in the EKK dataset (Fig. 2, blue lines in bottom left panel). For both groups, we found highest agreement between methods when restricting the comparison to the top 36-40 species (i.e., approximately half of the species in the diet), with one exception that only slightly outcompeted the r of the same range (KKL CCM vs. Plot Density). As expected, comparison between Plot Density and Plot Presence remained consistently high regardless of location within the home range of the bonobo groups, although correlations were lower when fewer species were included.

Table 1. Summary of correlation coefficients (*r*) between density estimates derived from the CCM and vegetation plot sampling for all significant comparisons above 50% kernel home range and of at least 10 species.

(a) Kokoalongo						
	$r_{\text{mean}} \pm \text{SD}$ (all combinations)	r_{range} (all combinations)	Number of species with highest r_{mean} (r_{mean})*	Number of species at r_{max}	% kernel with highest r_{mean} **	% kernel at r_{max}
Plot Presence and Plot Density	0.96 ± 0.02	0.82 - 0.99	-	-	-	-
Plot Presence and CCM	0.48 ± 0.04	0.31 - 0.55	36 species (0.53)	10	94 % (0.50)	58
Plot Density and CCM	0.58 ± 0.04	0.46 - 0.69	36 species (0.65)	71	69 % (0.60)	51
(b) Ekalakala						
	$r_{\text{mean}} \pm \text{SD}$ (all combinations)	r_{range} (all combinations)	Number of species with highest r_{mean} (r_{mean})*	Number of species at r_{max}	% kernel with highest r_{mean} **	% kernel at r_{max}
Plot Presence and Plot Density	0.97 ± 0.06	0.88 - 0.99	-	-	-	-
Plot Presence and CCM	0.52 ± 0.02	0.31 - 0.63	40 species (0.58)	48	51 % (0.56)	95
Plot Density and CCM	0.54 ± 0.03	0.42 - 0.70	19 species (0.66)	26	50 % (0.58)	95

*averaged across all combinations of % home range inclusion per number of species included

**averaged across all combinations of number of species included per % home range inclusion

Fig. 2. Correlation coefficients of density estimates between sampling methods (i.e., CCM and vegetation plots) for EKK (left) and KKL (right), according to home range percentage (top) and dietary inclusion (bottom). Color groups depict the three comparisons in this study (see legend), with numbers in brackets indicating number of species included (top legend) or percent home range included (bottom legend).

Once our moving window reached the dataset minimum of 20 plots at ca. 30% kernel, the correlation coefficient of the CCM with plot estimates increased until they reached a maximum around 60% kernel home range in both groups (Fig. S6). Peripheral areas of the home range were generally lower in agreement than more central areas but did not show persistent decreases with increasing peripheralization in a manner that would suggest consistently poorer sampling in peripheral areas. Sampling agreement was

strongest within our moving windows for the most frequently consumed species (e.g., 15 or 30 species) relative to more comprehensive subsets of the two groups' diets (e.g., 55 and 70 species).

The density of the species in the landscape and the variability in size significantly impacted agreement between the methods (Table S4); specifically, lower species density in the plots (estimate average: 0.57 ± 0.11 [SE]) and lower size variability (-1.29 ± 0.62 [SE]) improved method agreement. Further, in KKL only, greater seasonality, non-fruit item consumption, and greater consumption frequency decreased agreement between methods.

Correlation strength between the two methods reached significance and stabilized across methods and groups once exceeding 600 days (i.e., ca. 5300 [KKL] to 6000 [EKK] total visited locations) and continued to improve as data was collected until the end of our data period (Fig. 3; $EKK_{\max}$: 1222 days, $KKL_{\max}$: 1151 days).

Fig. 3. Pearson's r (left) and p -value (right; dashed line indicates 0.05 alpha level) of all three methods comparisons (see legend) for Ekalakala (full line) and Kokoalongo (dashed lines) over the duration of the dataset.

3.3.3 Dispersion

Overall, Morisita's indices from the CCM correlated weakly and non-significantly to vegetation plot indices, regardless of grid cell size used or bonobo group (Table 2a).

Table 2. Average (a) correlation coefficients (*r*) and (b) proportion of variance explained (*r*²; 500x500m) or Nagelkerke's *R* (1000x1000m and 1500x1500m) between the CCM and plot datasets across three different grid cell sizes for (a) dispersion and (b) distribution estimates.

(a) Dispersion					
	EKALAKALA		KOKOALONGO		
Cell size	Mean +SD (range)		Mean +SD (range)		
500	0.08 + 0.17 (-0.54, 0.55)		0.00 + 0.16 (-0.35, 0.61)		
1000	0.00 + 0.19 (-0.8, 0.25)		-0.03 + 0.14 (-0.65, 0.13)		
1500	-0.17 + 0.14 (-0.86, 0.07)		-0.20 + 0.14 (-0.83, -0.01)		
(b) Distribution					
	EKALAKALA		KOKOALONGO		
Cell size	Mean ± SD (range)	Num species p<0.05 (% of total species)	Mean ± SD (range)	Num species p<0.05 (% of total species)	Significant species in both groups
500	0.25 ± 0.05 (0.21, 0.38)	15 (29%)	0.23 ± 0.04 (0.20, 0.36)	13 (28%)	11
1000	0.23 ± 0.02 (0.20, 0.30)	13 (19%)	0.24 ± 0.02 (0.20, 0.31)	8 (11%)	7
1500	0.23 ± 0.03 (0.14, 0.27)	8 (12%)	0.24 ± 0.02 (0.21, 0.27)	6 (9%)	3

3.3.4 Distribution

Across both bonobo groups and all three grid cell sizes, we found that more species significantly correlated between the two methods for individual abundances across cells than would be expected by chance, with an average of 18% of species significantly correlated between methods across the three cell sizes (Table 2b). Percentage of species with significant correlations across methods declined as grid cell sizes increased, as did the number of significant species which remained consistent across both groups. Generally, proportion variance explained (*r* or Nagelkerke's *R*) by abundance per cell based on plots

averaged 0.25 ± 0.32 [SD] across species in all grid cell sizes and groups for predicting abundance per cell based on CCM. Average r did not vary substantially with cell size or between groups (Fig. 5).

The presence of a species in a cell as measured by plots significantly predicted the presence of that species in the cell as identified with CCM (estimate: 0.60 ± 0.20 (SD), range: 0.32 – 0.81; Table S5). The location of a cell within the home range appeared to play a consistent role, with food species less likely to be identified by CCM in more peripheral cells (average estimate: -0.05 ± 0.01 (SD), range: -0.05 – -0.04; Table S5). Bonobos missed presence of a species on average in $17.5\% \pm 16.3\%$ (SD; range: 0 – 68.4%) of the 500x500m cells and in $18.4\% \pm 16.5\%$ (SD; range: 0 – 61.2%) of the 1000x1000m cells. Increases in species abundance correlated with an increase in the likelihood for bonobos to miss the presence of species in a cell irrespective of cell size or group but species were less likely to be missed in a cell if they were more frequently consumed. We additionally found some support for species consumed for their fruits to be more likely to be missed in smaller cell sizes (Table S6).

Fig. 5. Averaged proportion variance explained (r) or Nagelkerke's R (top) and p -values for the estimate (bottom; dashed line indicates 0.05 alpha level) for correlations between estimated abundances per cell of

species (i.e., distribution agreement) as derived from CCM and vegetation plots for EKK (left) and KKL (right) and for three grid cell sizes (red: 500x500m, green: 1000x1000m, blue: 1500x1500m).

4. Discussion

Here we demonstrate the applicability of the consumer-centric method (CCM) for measuring resource density and distribution in an animal's landscape. We demonstrate that food species estimates derived from the CCM method are comparable to estimates derived from traditional vegetation plot sampling following a relatively short data collection timeframe, including before data have reached saturation. The method also seems promising for characterizing distribution of food patches within a landscape. Furthermore, we demonstrated that the CCM has an advantage over traditional sampling methods as it incorporates sampling bias important to the consumer into quantification of the ecological landscape.

4.1 Robustness of the CCM

The CCM estimates of abundance showed strong similarity to estimates from traditional ecological sampling. Behavioral ecologists have previously used consumption rates to infer about the abundance of food resources (Hutto 1990; Lovette & Holmes 1995; Watts & Mitani 2015). These methods are particularly susceptible to handling time, consumer motivation, and/or dependence of preference from resource availability and are subsequently difficult to validate (Lovette & Holmes 1995). The key advantage of the CCM is that rather than quantifying availability from occurrences of consumption (frequency dependent) it depends on independent locations (spatially dependent), thereby allowing validation with traditional vegetation plot sampling.

Although we found a significant but minor periphery effect on agreement between methods in the presence/absence of species, the correlation of abundance estimates between methods were unaltered by % of home range inclusion. The lack of a spatial effect on agreement between the methods is in some part likely to be a result of home range selection on the part of the consumer (e.g., second-order selection *sensu* Johnson 1980), i.e., bonobos may have already selected their home range based upon resource availability hence no sampling biases therewithin. In absence of home range use biases, the CCM therefore reliably

estimates resource availability across the entirety of a group's space use, although future studies should verify absence of sampling biases on agreement between the CCM and traditional methods in their own study species.

Further, we found consumption frequency to impact likelihood to miss species presences. Consequently, restricting estimation to only the top half of the consumed species (by frequency) appears to offer a compromise between maintenance of dietary relevance while maximizing fidelity with density estimates as assessed by objective plot measurements. This minimum translated to species consumed in approximately at least 60 locations over our three-year dataset. A general consequence of sampling frequency by a consumer is that estimates improve in precision as data accumulate over time. While species in our dataset were variable in "saturation level", rates of new locations sampled by the bonobos slowed over the course of data collection and inter-method correlation of species abundances stabilized after fewer than two years of data collection (approx. 600 days). As our results indicate that sampling rate affects stability of estimates (e.g., frequency of consumption), we anticipate that this general minimum will be higher for species with slower sampling frequency, i.e., for less frequently consumed species, as well as masting species or species which are consumed aseasonally.

Generally, species distribution (i.e., relative abundance) correlated weakly between the methods across species regardless of scale of comparison (i.e., cell size). A greater proportion of species reached significant agreement between methods in smaller rather than larger cell sizes, potentially as a function of proximity, i.e., the larger the cell size used the greater the potential distance between bonobo feeding locations and comparatively small plot areas. Nevertheless, our finding that correlations of distribution within species was significant across a greater proportion of food species than expected by chance (i.e., 5%) and that the rates at which bonobos missed the presence of a species in a cell is likewise better than common rates of species misses between multiple observers in single plot (Millberg et al. 2008) provides hope that reliable estimates of sub-landscape abundances may improve with greater sampling depth.

While detectability is rarely 100% in either method (Morrison 2016), the miss rates by a consumer in the CCM may rather carry additional information about the nature of resource selection (and the individuals which are subsequently ignored). This is especially likely to be the case in consumers who have the capacity to keep track of spatiotemporal patterns of resource availability. Bonobos likely have a concept

of where and when resources become available, and therefore are also capable of targeting resources that are rare (Janmaat et al. 2013; Normand et al. 2013). Consequently, the CCM mimics *ad hoc* sampling (Foster et al. 1998, Gordon & Newton 2006, Hopkins 2007), and our results indicate that the CCM more closely matches plot density estimates at capturing rare species relative to more abundant species.

Nonetheless, in absence of full censusing, we cannot differentiate between which sampling method produced a more precise representation of food species availability, dispersion, and distribution patterns. Ideally, methodological sampling biases could be identified by simulating both sampling schemes from a simulated ‘forest’. Unfortunately, we rarely understand the complexity of consumer movement and resource selection patterns (Buskirk & Millspaugh 2006). Therefore, subsequent conclusions drawn from simulated sampling behavior would be just as arbitrary as the decisions made to simulate them (Johnson 1980).

4.2 Measuring different phenomena

We argue that the CCM, with adequate evaluation, may be a more appropriate tool for most applications in behavioral ecology than traditional inventory methods. Traditional plot sampling quantifies the total amount of potential resources which also include inaccessible, unattractive, or otherwise unpalatable resources to a consumer. Only a subset of these resources comprises true resource availability, i.e., resources with potential to be selected (Aldredge et al. 1998; Buskirk & Millspaugh 2006; Johnson 1980), and although correlated, each represents inherently two separate phenomena (Hutto 1990). Because we rarely understand the processes of food selection by which consumers filters objective resource abundance into availability, the CCM provides the advantage of using the consumer as a means to avoid arbitrary decisions as to how to best sample the landscape (Johnson 1980). We detail examples of this selectivity and the resulting advantages of the CCM below.

First, we observed significant differences between average sizes of trees/lianas visited by bonobos relative to what was available in the landscape of consumed species (as measured in vegetation plots). Reducing our plot dataset to sizes selected by the consumer increased the correlations between CCM and vegetation plot measures and demonstrates the inadequacies of consumer-objective plots in mirroring consumer behavior. Second, that bonobos missed or ignored certain food resources in cells identified to contain them underlines further how researchers are likely unaware of relevant selection criteria that impact

measurement of true resource availability. Because apes possess mental maps of their environments and are known to adjust travel to target preferred food sources (Janmaat et al. 2013; Lucchesi et al. 2021), they are unlikely to consistently miss potentially important food resources within their home range at the scale observed in our study. Third, we found that CMM estimates of density and distribution differed between bonobo social groups, even with largely overlapping home ranges. This conforms to previous findings of group-specific feeding selection criteria in bonobos (Samuni et al. 2020), independent of local abundance. If resource availability for a consumer in a given landscape is dependent on group identity, then only methods like the CCM incorporating these criteria allow comparable estimates for comparative studies across social groups.

Altogether, by accounting for consumer selection, the accumulation of data on food patch location are inherently less subjective than datasets dependent upon arbitrary decisions by the investigator (Johnson 1980). Biases in resource measurement occur via multiple sources including selection of sampling method, metric, effort, as well as through unavoidable systematic or random measurement errors (Baraloto et al. 2013; Milberg et al. 2008; Morrison 2016; Ståhl et al. 2017; Wessling et al. 2020). The CCM, however, accounts for several of these issues because consumers are knowledgeable and motivated surveyors who actively target resources, with apparently negligible impact of scale variation (e.g., cell size) or abundance on fidelity of CCM estimates to plot-derived estimates.

Therefore, estimates derived from the CCM provide accurate measures of availability once data have reached a sufficient depth. Our spatially explicit CCM further allows for data accumulation and consequential improvement of the accuracy of estimates over time until otherwise removed due to irrelevance (e.g., patch loss). Nevertheless, if rapid abundance assessment is preferable for a project, traditional ecological sampling may remain a preferable method due to a 600 person-day burn-in time required (this study) by the CCM before estimates become reliably stable per social group relative to 150 person days of plots for both groups. However, these 150 person days are supplementary to observational data, inasmuch as person-days necessary to collect both sets of data must be considered additive to observational data collection. Yet, if databases of feeding locations are already available, adapting these data to CCM estimation of resource density or distribution save researchers from needing to collect additional data to quantify resource abundance.

While this method is best applied to estimate the availability of discrete, immobile, and spatially explicit resources, these advantages transcend application beyond bonobos and allow researchers to evaluate the strengths of the method for their investigations across all potential consumers who meet these criteria (further discussed in Table S7). Functionally, assumptions of the CCM are similar to studies investigating resource preference, a method which also combines objective habitat measures with subjective animal-centric data (Manly et al. 2002). For example, this method can only be applied to consumers which re-use space over time, like a consistent home range, and assumes that consumers have equal access to all the areas of this space (Alldredge et al. 1998). Nonetheless, researchers must verify CCM sampling is of sufficient sampling depth and absent of biases (e.g., sampling biases or characteristics of food items) for their consumer before the CCM can be applied as a means of resource availability. When applied correctly, the CCM will enable many behavioral ecologists to quantify aspects of food availability by using existing data, in a manner that is more suitable to its application as well as allows for more precise comparison ways that make this data comparable across social groups, subsequently promising new insights in the interplay between an animal and its environment.

Author contributions

EW, LS conceptualized the manuscript, EW, LS, MS designed the analyses, EW, LS, RM performed the analyses; EW, MAP, SL, BK, MS provided data; EW wrote the manuscript with input from all co-authors.

Statement of where you intend to archive your data

All necessary data to reproduce this study will be submitted to the Dryad Digital Repository.

Acknowledgements

This work is dedicated to the late Dr. Deborah Moore, who is deeply missed at Kokolopori. We are grateful to the Ministry of Scientific Research and Technology (MSRT) and the ICCN of the Democratic Republic of the Congo, as well as to residents of the surrounding communities for facilitating our research in the Kokolopori Bonobo Reserve. We thank Leonard Nkanga, Paulin Lokutshu Lokole, Serge Betamba Bokelakela, Valerie Beraud, Vincent Rison, and the team of local assistants for their invaluable contribution

to data collection and bonobo tracking. This work was supported by Harvard University, and meets ethical standards for animal research of the Max Planck Society and the MSRT. The authors have no conflict of interest to declare.

References

- Aldredge, J. R., Thomas, D. L., & McDonald, L. L. (1998). Survey and comparison of methods for study of resource selection. *Journal of Agricultural, Biological, and Environmental Statistics*, 237-253. <https://doi.org/10.2307/1400580>
- Anholt, B. R., & Werner, E. E. (1995). Interaction between food availability and predation mortality mediated by adaptive behavior. *Ecology*, 76(7), 2230-2234. <https://doi.org/10.2307/1941696>
- Baayen, R. H. (2008). Analyzing Linguistic data: A Practical Introduction to Statistics Using R. Cambridge: Cambridge University Press. <https://doi.org/10.1017/CBO9780511801686>
- Baraloto, C., Molto, Q., Rabaud, S., Hérault, B., Valencia, R., Blanc, L., Fine, P.V. & Thompson, J. (2013). Rapid simultaneous estimation of aboveground biomass and tree diversity across Neotropical forests: a comparison of field inventory methods. *Biotropica*, 45(3), 288-298. <https://doi.org/10.1111/btp.12006>
- Bates, D. Maechler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, 67(1), 1-48. <https://doi.org/10.18637/jss.v067.i01>
- Bowering, R., Wigle, R., Padgett, T., Adams, B., Cote, D., & Wiersma, Y. F. (2018). Searching for rare species: A comparison of Floristic Habitat Sampling and Adaptive Cluster Sampling for detecting and estimating abundance. *Forest Ecology and Management*, 407, 1-8. <https://doi.org/10.1016/j.foreco.2017.10.016>
- Buskirk, S. W., & Millsaugh, J. J. (2006). Metrics for studies of resource selection. *The Journal of Wildlife Management*, 70(2), 358-366.
- Chapman, C. A., Chapman, L. J., Wingham, R., Hunt, K., Gebo, D., & Gardner, L. (1992). Estimators of fruit abundance of tropical trees. *Biotropica*, 527-531. <https://doi.org/10.2307/2389015>
- Chapman, C. A., Schoof, V. A., Bonnell, T. R., Gogarten, J. F., & Calmé, S. (2015). Competing pressures on populations: long-term dynamics of food availability, food quality, disease, stress and animal

abundance. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 370(1669), 20140112. <https://doi.org/10.1098/rstb.2014.0112>

Davies, S., & Deviche, P. (2014). At the crossroads of physiology and ecology: food supply and the timing of avian reproduction. *Hormones and Behavior*, 66(1), 41-55. <https://doi.org/10.1016/j.yhbeh.2014.04.003>

Dobson, A. J., and Barnett, A. (2008). An Introduction to Generalized Linear Models, 3rd Ed. Boca Raton: CRC Press.

Foerster, S., Zhong, Y., Pintea, L., Murray, C. M., Wilson, M. L., Mjungu, D. C., & Pusey, A. E. (2016). Feeding habitat quality and behavioral trade-offs in chimpanzees: A case for species distribution models. *Behavioral Ecology*, 27(4), 1004–1016. <https://doi.org/10.1093/beheco/arw004>

Forester, J.D., Im, H. K., & Rathouz, P. J. (2009). Accounting for animal movement in estimation of resource selection functions: sampling and data analysis. *Ecology*, 90(12), 3554-3565. <https://doi.org/10.1890/08-0874.1>

Foster, R.B., Hernandez, N.C., Kakudidi, E.K., & Burnham, R.J. (1998). Rapid assessment of tropical plant communities using variable transects: an informal and practical guide. Field Museum of Chicago.

Gordon, J.E., & Newton, A.C. (2006). Efficient floristic inventory for the assessment of tropical tree diversity: A comparative test of four alternative approaches. *Forest Ecology and Management*, 237(1-3), 564-573. <https://doi.org/10.1016/j.foreco.2006.10.002>

Hopkins, M.J.G. (2007). Modelling the known and unknown plant biodiversity of the Amazon Basin. *J. Biogeogr.* 34: 1400–1411. <https://doi.org/10.1111/j.1365-2699.2007.01737.x>

Hutto, R. L. (1990). Measuring the availability of food resources. *Studies in Avian Biology*, 13, 20-28.

Janmaat, K. R., Ban, S. D., & Boesch, C. (2013). Chimpanzees use long-term spatial memory to monitor large fruit trees and remember feeding experiences across seasons. *Animal Behaviour*, 86(6), 1183-1205. <https://doi.org/10.1016/j.anbehav.2013.09.021>

Oksanen, J., Blanchet, F.G., Friendly, M., Kindt, R., Legendre, P., McGlinn, D., Minchin, P.R., O'Hara, R. B., Gavin, X., Simpson, L., Solymos, P., Henry, M., Stevens, H., Szoecs, E., & Wagner, H. (2019). *vegan: Community Ecology Package*. R package version 2.5-6.

- Johnson, D. H. (1980). The comparison of usage and availability measurements for evaluating resource preference. *Ecology*, 61(1), 65-71. <https://doi.org/10.2307/1937156>
- Krebs, C. (1999). *Ecological methodology* (2nd ed., pp. 114–115). Menlo Park: Addison-Wesley.
- Lambert, J. E., & Rothman, J. M. (2015). Fallback foods, optimal diets, and nutritional targets: primate responses to varying food availability and quality. *Annual Review of Anthropology*, 44, 493-512. <https://doi.org/10.1146/annurev-anthro-102313-025928>
- Lovette, I. J., & Holmes, R. T. (1995). Foraging behavior of American Redstarts in breeding and wintering habitats: implications for relative food availability. *The Condor*, 97(3), 782-791. <https://doi.org/10.2307/1369186>
- Lucchesi, S. Cheng, L., Wessling, E.G., Kambale, B., Lokasola, A.L., Ortmann, S., & Surbeck, M. 2021. Importance of subterranean fungi in the diet of bonobos in Kokolopori. *American Journal of Primatology*, e23308. <https://doi.org/10.1002/ajp.23308>
- Manly, B. F. L., McDonald, L., Thomas, D. L., McDonald, T. L., & Erickson, W. P. (2007). *Resource selection by animals: statistical design and analysis for field studies*. Springer Science & Business Media.
- Milberg, P., Bergstedt, J., Fridman, J., Odell, G., & Westerberg, L. (2008). Observer bias and random variation in vegetation monitoring data. *Journal of Vegetation Science*, 19(5), 633-644. <https://doi.org/10.3170/2008-8-18423>
- Morisita, M. (1962). I σ -Index, a measure of dispersion of individuals. *Researches on Population Ecology*, 4(1), 1-7.
- Morrison, L.W. (2016) Observer error in vegetation surveys: a review. *Journal of Plant Ecology*, 9, 367–379. <https://doi.org/10.1093/jpe/rtv077>
- Mortelliti, A., Amori, G. & Boitani, L. (2010) The role of habitat quality in fragmented landscapes: a conceptual overview and prospectus for future research. *Oecologia*, 163, 535–547
- Myers, J. H. (1978). Selecting a measure of dispersion. *Environmental Entomology*, 7(5), 619-621. <https://doi.org/10.1093/ee/7.5.619>
- Normand, E., Ban, S. D., & Boesch, C. (2009). Forest chimpanzees (*Pan troglodytes verus*) remember the location of numerous fruit trees. *Animal Cognition*, 12(6), 797-807.

- R Core Team (2020). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>.
- Rogers, L. L. (1987). Effects of food supply and kinship on social behavior, movements, and population growth of black bears in northeastern Minnesota. *Wildlife Monographs*, 3-72.
- Samuni, L., Wegdell, F., & Surbeck, M. (2020). Behavioural diversity of bonobo prey preference as a potential cultural trait. *eLife*, 9, e59191. <https://doi.org/10.7554/eLife.59191>
- Ståhl, G., Ekström, M., Dahlgren, J., Esseen, P. A., Grafström, A., & Jonsson, B. G. (2017). Informative plot sizes in presence-absence sampling of forest floor vegetation. *Methods in Ecology and Evolution*, 8(10), 1284-1291. <https://doi.org/10.1111/2041-210X.12749>
- Stephens, D., & Krebs, J. (1986). Foraging theory. Princeton University Press.
- Stoehr, A. M. (1999). Are significance thresholds appropriate for the study of animal behaviour? *Animal Behaviour*, 57(5), F22–F25. <https://doi.org/10.1006/anbe.1998.1016>
- Szigeti, V., Körösi, Á., Harnos, A., Nagy, J., & Kis, J. (2016). Measuring floral resource availability for insect pollinators in temperate grasslands—a review. *Ecological Entomology*, 41(3), 231-240. <https://doi.org/10.1111/een.12298>
- van Schaik, C. P., Terborgh, J. W., & Wright, S. J. (1993). The phenology of tropical forests: adaptive significance and consequences for primary consumers. *Annual Review of Ecology and Systematics*, 24(1), 353-377. <https://doi.org/10.1146/annurev.es.24.110193.002033>
- Vogel, E. R., & Janson, C. H. (2007). Predicting the frequency of food-related agonism in white-faced capuchin monkeys (*Cebus capucinus*), using a novel focal-tree method. *American Journal of Primatology*, 69(5), 533-550. <https://doi.org/10.1002/ajp.20368>
- Vogel, E. R., & Janson, C. H. (2011). Quantifying primate food distribution and abundance for socioecological studies: an objective consumer-centered method. *International Journal of Primatology*, 32(3), 737-754. <https://doi.org/10.1007/s10764-011-9498-7>
- Watts, D. P., & Mitani, J. C. (2015). Hunting and prey switching by chimpanzees (*Pan troglodytes schweinfurthii*) at Ngogo. *International Journal of Primatology*, 36(4), 728-748. <https://doi.org/10.1007/s10764-015-9851-3>

593 Wessling, E. G., Dieguez, P., Llana, M., Pacheco, L., Pruetz, J. D., & Kühl, H. S. (2020). Chimpanzee
594 (*Pan troglodytes verus*) density and environmental gradients at their biogeographical range edge.
595 *International Journal of Primatology*, 41(6), 822-848. [https://doi.org/10.1007%2Fs10764-020-](https://doi.org/10.1007%2Fs10764-020-00182-3)
596 [00182-3](https://doi.org/10.1007%2Fs10764-020-00182-3)