Accounting for detection bias increases trait-matching in a tropical plant-hummingbird network

Ben Weinstein - Stony Brook University

Monday, March 16, 2015

# Introduction

Network ecology is a rapidly developing field that has reinvorgorated research on assemblage stability and complexity. By harnessing graph theory to represent ecological interactions, networks have informed the evolution of specialization (), the importance of complex interactions () and predicting population decline due to shared partners(). While considerable work has focused on sampling and measuring network properties (Bluthgen, Bascompte, Jordano), there has been much less attention to a central problem of network ecology: our ability to detect interactions is imperfect in time and space. This limitation can greatly bias results by assuming that non-detection is synomonous with non-interaction.

The importance of detection bias is widely appreciated in wildlife sciences(Mackenzie), but only rarely incoporated into evolutionary and theoretical ecology (xy, but see). However, the same issues that limit inference in determining population dyanimcs (), site selection(), and breeding ecology (), also undermine our ability to make inference about the evolution of ecological systems. We can harness the tools developed in widlife sciences to strengthen the realism of network observations. The chief improvement is to incorporate detection bias directly into our estimates of species interactions using occupancy models. In occupancy modeling, repeated surveys are used to estimate the probability of detection given a species probability of occurrence. The probability of occurrence is a latent variable which cannot be directly infered from the data. If we did not see a species at a site, was it absent or undetected? By analogy, if we failed to find a pollinator interacting with a plant, is it due to detection or is it a 'forbidden link' due mismatch in species timing or functional morphology? The goal of this paper is to use a large dataset on plant-hummingbird interactions from a tropical montane forest to evaluate the effect of imperfect detection on estimating species interactions.

In the tropical Andes, up to thirty hummingbirds can co-occur in local assemblages. Hummingbirds are largley nectarivorous and have a diversity of bill morphologies. In general, species with longer and more curved bills tend to be more specialized (Temeles,Magliansi). This trait relationship may help promote community stability through niche partitioning (Weinstein). To test the strength of trait-matching between bird bill and flower corolla lengths, studies use the number of visits to a flower as interaction strength among birds and plants. These data are used to represent links between the nodes in the network graph. While randomization approaches have been purposed to test for the strength of specialization (Bluthgen, Schueling), few studies have evaluated the detection rates of different interactions, and included these coefficients into estimates of trait-matching (Bartolomeau).

To estimate detection, observers return multiple times to a site within a sampling window. Human observation is expensive, and it is often prohibitively expensive to design studies to accomodate occupancy modeling using human observations. However, the hummingbird dataset in this paper was collected by using time-lapse cameras which turn on at dawn and dusk automatically. In addition, more traditional hummingbird transects were used to survey flower visitation. Combining these sampling types, we have a very large network with over 3,000 interactions (Weinstein XXXX). By embracing automated monitering of plant-animal interactions, we can begin to determine and account for detection bias.

To illustrate the utility of this approach, I begin with simulations to demonstrate the importance of accounting for detection. In addition, occupancy models are complex and require fitting a number of latent variables. Before I fit models to an observed dataset, it is crucial to simulate data with known results to show that the models perform as expected. In this analysis, I will use a Bayesian hierarcichal approach. We are ultimately interested in estimating three quantities, the number of interactions between a hummingbird species and plant species, the probability of detection, and the frequency (or 'intensity') of the interaction. To do this, i will divide our model into three components: an observation model, a process model, and a hierarcichal set of priors.

### Observation Model

The observation model describes the sampling process that yields our data (Elison). These are the actual number of observations observed by reviewing the cameras and transects. We know there is some true, but unknown intensity of interactions among hummingbird i and plant j, and that we are likely to see some, but not all, interactions. We assume that we are equally likely to see any one interaction, so the number of detections of hummingbird i on plant j in month k is modeled as as a binomial trail where the probability of detecting a species is a species-specific parameter *detect* with *N* trials (Royle).

### Process Model

To estimate the parameter *N*, which is the unknown true rate of interactions, we need a process model. The process model is the mechansism that generates interactions among plants and pollinators in a given time period. In our analysis, our sampling period is one day. We can model the daily intensity as a poisson distribution with . The hypothesis is that should increase with greater matching between plant corolla and hummingbird bill lengths. Since lambda is a intensity of interactions, it must be a positive number. We therefore model as a function of linear predictors with a log link function. In our simulation, we give equal penalty for having a bill too long, or too short for a given corolla. The process model can be written as:

### Hierarcichal structure

Each interaction is not independent. Due to foraging behavior (Sandlin), physiological requirements (Powers), or difference in available resources (Weinstein), species may respond individualistically to differences in bill or corolla lengths. However, our goal is not to study any one species, but to capture the overall effect of trait-matching on hummingbirds and plants. We therefore use a hierarcichal model to account both for within species variability, but to estimate a group-level mean (Bolker). Similiarly, due to differences in abundance, behavior, and foraging strategy (Colwell), some hummingbird species will be more detectable than others. We allow a fixed species level effect of hummingbird species on detection probability. Each of these levels could be fit to covariates as well depending on vegetation height, observer distance and quality (Royle). By combining an observation model, process model, and hiericichal structure, I am accounting for known obstacles in intepretting my data. To illustrate the added information based on this setup, I compare my results to a direct poisson generalized linear model that does not account for detection or non-independence of observations.

# Methods

## Simulation

I first create data with constant detection for each hummingbird speices, and the interaction frequency is a function of a species hierarcichal model and corolla similarity. There is uneven sampling among species.

### Parameters

* 8 hummingbird species
* Range of hummingbird bill sizes (mm) ~ Pois(20)
* Twenty plants
* Range of corolla sizes (mm) ~ Pois(20)
* Mean frequeny () for each hummingbird is drawn from U(0,10)
* For each plant the occupancy is N(,0.2) (truncated 0-1)
* Trait matching (minimizing Bill-Corolla difference) is drawn from a hierarcichal distribution
* Imperfect detection
* 24 month replicates
* Phenology = .5 (plants are in flower/present .5 of surveys)

# Results

**View simulated strength and form of trait matching**

### True Interaction Matrix

The left hand panel is the true state for each hummingbird plant combination. The middle and right hand panel are two realizations that might result from a month of sampling given incomplete detection and phenology.

## Observed Data

### Data Collection

* Data were collected at the Maquipucuna Research Station and Santa Lucia Ecolodges, Ecuador (0.118 S,-78.612 W) between June 2013 and July 2015 along an elevation gradient from 1300 m to 2500 m. The area contains regenerating secondary and primary cloud-forest, and has a warmer dry season from June to September (Webster and Rhodes 2005).
* We used time-lapse cameras to monitor hummingbird visitation. Flowers in bloom were noted during each transect, and cameras were placed 1-3 meters from focal flowers. Cameras turned on automatically at dawn and recorded an image every second for at least two days.
* We used the computer vision program MotionMeerkat to identify images where hummingbirds were observed (Weinstein 2014).

### Model Fitting

## Hierarcichal Occupancy Model Formulation

### Observation

The number of observations observed between and in is a binomial trail where the probability of detecting a species is a species specific parameter detect with N trials.N is the underlying frequency of interactions among each pair of plants and hummingbirds drawn from a poisson with , which is modeled as a lognormal link. The detection probability is fixed for each species. This could account for differing abundances or foraging intensity. Species level priors for the strength of trait-matching is drawn from a normal distribution of all hummingirds.

**Priors**

**Hyperpriors**

**Derived quantities**

### Posterior Checks

To measure the relative fit of each model, we performed posterior goodness of fit checks. This approach allows a standardized way of comparing models of different types. We choose ^2 as a fit statistic calculated as Expected-Observed/Expected^2 + 0.5. We add 0.5 to add fits to avoid dividing by 0 where we expectThe goodness if of fit is a measured as chi-squared. For the occupancy model, the expected value for each month is the detection rate \* the estimate intensity of interactions. The expected value is compared to the observed value of the actual data. However, given that we are drawing from a stochastic processes, we draw a replicate dataset is generated from the posterior predicted intensity. Better fitting models are smaller values and closer to the 1:1 line. A perfect model would be 0 discrepancy. This is unrealsitic given the stochasticity in the sampling processes.

# Results

## Poisson Estimate of trait-matching without detection bias or non-independence.

Fit a model of observations without repect to their hierarcichal structure or with detection bias.

Observations between each bird and flower is modeled as a poisson with a lognormal link.

## Hierarcichal Bayesian Model

## [1]   
## [2] sink("Bayesian/NMixture.jags")   
## [3]   
## [4] cat("   
## [5] model {   
## [6] for (i in 1:Birds){   
## [7] for (j in 1:Plants){   
## [8]   
## [9] # True state model for the only partially observed true state   
## [10] log(lambda[i,j])<-alpha[i] + beta[i] \* traitmatch[i,j]   
## [11] N[i,j] ~ dpois(lambda[i,j])   
## [12]   
## [13] for (k in 1:Months) {   
## [14]   
## [15] # Observation model for the actual observations   
## [16] Y[i,j,k] ~ dbin(detect[i],N[i,j])   
## [17]   
## [18] #Fit discrepancy statistics   
## [19] eval[i,j,k]<-detect[i]\*N[i,j]   
## [20] E[i,j,k]<-pow((Y[i,j,k]-eval[i,j,k]),2)/(eval[i,j,k]+0.5)   
## [21]   
## [22] y.new[i,j,k]~dbin(detect[i],N[i,j])   
## [23] E.new[i,j,k]<-pow((y.new[i,j,k]-eval[i,j,k]),2)/(eval[i,j,k]+0.5)  
## [24] }   
## [25] }   
## [26] }   
## [27]   
## [28] for (i in 1:Birds){   
## [29] detect[i] ~ dunif(0,1) # Detection for each bird species   
## [30] alpha[i] ~ dnorm(intercept,tau\_alpha)   
## [31] beta[i] ~ dnorm(gamma,tau\_beta)   
## [32] }   
## [33]   
## [34] #Hyperpriors   
## [35] gamma~dnorm(0.001,0.001)   
## [36] intercept~dnorm(0.001,0.001)   
## [37]   
## [38] tau\_alpha ~ dgamma(0.001,0.001)   
## [39] sigma\_int<-pow(1/tau\_alpha,0.5) #Derived Quantity   
## [40] tau\_beta ~ dgamma(0.001,0.001)   
## [41] sigma\_slope<-pow(1/tau\_beta,0.5)   
## [42]   
## [43]   
## [44] #derived posterior check   
## [45] fit<-sum(E[,,]) #Discrepancy for the observed data   
## [46] fitnew<-sum(E.new[,,])   
## [47]   
## [48] }   
## [49] ",fill=TRUE)   
## [50]   
## [51] sink()

## Compiling model graph  
## Resolving undeclared variables  
## Allocating nodes  
## Graph Size: 27642  
##   
## Initializing model

### Assess Convergence

### Posteriors

**True values are given in the dashed lines.**

### Predicted Relationship

Does accounting for non-independence and detection change our estimate of trait matching?

Blue line is the true relationship. The red line is the posterior mean with confidible intervals in shaded grey for the proposed bayesian model. The poisson glm that does not account for detectability or non-independence is in green.

**Conclusion:** Accounting for detection and non-independence greatly increases the accuracy of the predicted state. The poisson glm underestimates the strength of trait matching among hummingbirds and their foodplants.

### Predicted Frequency

Validating the model

### Posterior Check

Since I have simualted the data, it should fit as well as any random dataset drawn from the estimated parameters. An ideal fit would be posterior values sitting along the 1:1 line.

# Observed dataset

## character(0)

### Elevation ranges

Create a binary variable whether each observation was in a low elevation or high elevation transect. We have some species that just occur at the top of the gradient, and are not present in the sampling window of flowers at the low elevation.

Start by a hand drawn table - then move to quantitative

### Assess Convergence

### Posteriors

### Species Predictions

### Overall predicted relationship

Does accounting for non-independence and detection change our estimate of trait matching?

The red line is the posterior median with confidible intervals in shaded grey for the proposed bayesian model. The poisson glm that does not account for detectability or non-independence is in green.

**Conclusion:** Accounting for detection and non-independence greatly increases the accuracy of the predicted state. The poisson glm underestimates the strength of trait matching among hummingbirds and their foodplants.

### Poisson regression without detection bias

##   
## Call:  
## glm(formula = Yobs ~ poly(x, 2, raw = T), family = poisson(),   
## data = niavedat)  
##   
## Deviance Residuals:   
## Min 1Q Median 3Q Max   
## -1.5650 -1.4603 -0.7953 0.3621 7.4013   
##   
## Coefficients:  
## Estimate Std. Error z value Pr(>|z|)   
## (Intercept) 1.243718 0.057838 21.504 <2e-16 \*\*\*  
## poly(x, 2, raw = T)1 -0.005239 0.102839 -0.051 0.959   
## poly(x, 2, raw = T)2 -0.056191 0.033760 -1.664 0.096 .   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for poisson family taken to be 1)  
##   
## Null deviance: 1055.8 on 406 degrees of freedom  
## Residual deviance: 1026.8 on 404 degrees of freedom  
## AIC: 2129.3  
##   
## Number of Fisher Scoring iterations: 5

### Predicted Frequency

### Discrepancy

The goodness if of fit is a measured as chi-squared. The expected value for each month is the detection rate \* the estimate intensity of interactions. The expected value is compared to the observed value of the actual data. In addition, a replicate dataset is generated from the posterior predicted intensity. Better fitting models will have lower discrepancy values and be Better fitting models are smaller values and closer to the 1:1 line. A perfect model would be 0 discrepancy. This is unrealsitic given the stochasticity in the sampling processes. Rather, its better to focus on relative discrepancy.

### Posterior p-value (worst term in the world.)

We are interested in the mean number of interactions per species.

# Discussion

Species are difficult to observe in the natural world. By allowing for observation error, we can better estimate the underlying mechanisms shaping ecological communities. In this paper, i compared poisson generalized linear models with a hierarcichal bayesian approach that accounted for detection bias and non-independence among species. Both the estimate of trait-matching, as well as the model fit increased using the hierarcichal approach. By using the raw number of interactions among plants and pollinators, we assume that each species sampled evenly, and that each interaction was equally detectable. The automated cameras used to moniter the majority of these interactions turn on and off at dawn and dusk and provide a much longer duration of observation than is possible with a human observer(). Nevertheless, hummingbird-plant interactions were detected less than 20% of the sampling periods. This suggest that the problem of detection bias would be even more severe when the sampling window is even shorter. The occupancy models presented here can be extended to include time-varying components (Royle), and covariates. For example, detection may decrease as plant height increases, since observations may be harder to confirm at greater distances. The underlyin #Similiar work

* <http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0069200>
* <http://onlinelibrary.wiley.com/doi/10.1111/j.2041-210x.2012.00249.x/abstract>