

A Review of "Integrated Community Occupancy Models: A Framework to Assess Occurrence and Biodiversity Dynamics Using Multiple Data Sources"

Zachary Higa

University of Idaho

1 Introduction

In studies of wildlife and ecology, sometimes the topic of interest involves wanting to understand the factors that determine the occurrence of a species at a certain location. Knowing the presence or absence of a species at a location over a period time can provide insight into the behavioral patterns of a species, the conditions a species needs for survival, and many other pieces of information about a species and the habitat. As a result, occupancy models were developed to provide information on whether a species is ascertained to be present or absent from a particular site at a given time.

1.1 The Simplest Occupancy Model

To provide a basic idea of how occupancy models work, I will first state the simplest occupancy model below. Suppose we are looking at the presence or absence of a certain species at a given time t throughout n different sites. Let ψ represent the probability that a species is present at a site, and let p represent the probability of detecting the species. Define X_i as an indicator of whether the species is present or absent at the i th (where $i = 1, \dots, n$) site such that

$$X_i = \begin{cases} 1, & \text{if species is seen at site } i \\ 0, & \text{if species is not seen at site } i \end{cases}. \quad (1)$$

Hence, the probability that the species is seen at site i is

$$P(X_i = 1) = \psi p, \quad (2)$$

and conversely, the probability that the species is not seen at site i is

$$P(X_i = 0) = (1 - \psi) + \psi(1 - p) = 1 - \psi p. \quad (3)$$

Note that the probability that the species is not seen at site i includes the case where the species is not present at the site (probability $1 - \psi$), and the case where the species is present, but we were unable to detect it (probability $\psi(1 - p)$). Since we see that X_i acts as indicator for the presence of a species with success probability ψp , we have that $X_i \sim \text{Bernoulli}(\psi p)$.

Now, for X_1, X_2, \dots, X_n , where each X_i is the indicator of the presence or absence of a species at the i th site, define Y to be the number of sites where the species was present at time t such that

$$Y = X_1 + X_2 + \dots + X_n. \quad (4)$$

As we are taking the sum of n Bernoulli random variables, we have that Y is a binomial random variable, $Y \sim \text{Binomial}(n, \psi p)$. Therefore, the probability mass function of Y is

$$P(Y = y) = \binom{n}{y} (\psi p)^y (1 - \psi p)^{n-y}, \quad y = 0, 1, 2, \dots, n. \quad (5)$$

However, note that the above is only detecting the presence and absence of a species for each of the n sites at a single time t . If we wanted to assess the presence of a species at n sites for a given time period, then we would need to adjust the model above. Now let us redefine some of the random variables that we have previously defined above; define Y to represent the number of sites at which a species was seen one or more times, and define X_i to represent the number of times that a species was seen at site i , where $i \in \text{set of sites at which species was seen (denote as A)}$. Thus, we find that the likelihood function of the presence probability ψ and detectability probability p for a time interval from 1 to t is

$$L(\psi, p) = [\psi^y \prod_{i \in A} p^{x_i} (1 - p)^{t-x_i}] \times [(1 - \psi) + \psi(1 - p)^t]^{n-y}. \quad (6)$$

From the given likelihood function presented in (6) above, we can employ maximum likelihood methods to obtain estimates for ψ and p . Therefore, we have covered a basic case of estimating the presence probability and detection probability from a simple occupancy model.

1.2 Paper to be Reviewed

Now that we covered a basic introduction to the simplest occupancy model, we can proceed to the review of the paper of interest. "Integrated Community Occupancy Models: A Framework to Assess Occurrence and Biodiversity Using Multiple Data Sources" was written by Jeffrey Doser, Wendy Leuenberger, and three other collaborators (names in references below). The authors of this paper all come from fields such as Ecology, Wildlife Biology, Forestry, and Environmental Science.

I was interested in reviewing this particular paper because I wanted to further explore occupancy models that expand on the simplest model that we have

reviewed above. In this paper, the integrated community occupancy models is developed to provide inferences on both species-specific and community occurrence dynamics by combining multiple data sources. Hence, in this review, I will evaluate the integrated community occupancy model, and I will determine whether the use of this model is beneficial in studies that wish to make inferences on occurrence dynamics.

2 The Review and Summary of the Paper

2.1 Basis for ICOM

The integrated community occupancy model (denoted as ICOM) combines hierarchical community modelling with data integration to find inferences across multiple species through multiple data sources. Data integration is used to combine replicated and non-replicated detection-nondetection data since the replicated data can account for the imperfections of detection methods, and the non-replicated data can provide information from varying data collection methods (for example, different detection rates). Then, the hierarchical community occupancy model allows us to make inferences on the occurrence dynamics at the community-level. In addition, with this hierarchical structure, the species-specific parameters act as random effects based on the community-level distribution that is normal with mean and variance parameters; the mean parameter represents the average effect across species in the community of interest, and the variance parameter represents the variation of species-specific effects within the community (Doser et al., 2022).

With the power to find inferences about occurrence dynamics across multiple species using multiple sources of data, the authors aim to use this model to determine whether global and environmental shifts are causing a negative impact on biodiversity processes within communities of species. For example, from prior research, the authors cited that there have been declines in the abundance and distribution of birds, bats, and insects due to climate change and other external factors. Therefore, the authors emphasize that they want to be able to make inferences on a broader level than assessing a single species, and they assert that ICOM will be effective in achieving this (Doser et al., 2022).

2.2 Building the ICOM

In the following sections, I will introduce the models that make up ICOM that will aid in the discussion of data that is collected from the simulated studies and the case study.

2.2.1 Ecological Process Model

Expanding upon the simplest occupancy model that we described in the introduction above, let $x_{i,j,t}$ represent the presence or absence of the i th species (where $i = 1, \dots, I$) at the j th site (where $j = 1, \dots, J$) during year t (where

$t = 1, \dots, T$). We know that $X_{i,j,t} \sim \text{Bernoulli}(\psi_{i,j,t})$, where ψ is the presence probability for the i th species at the j th site during year t . For the $t = 1$ year, the authors model $\psi_{i,j,t}$ such that

$$\text{logit}(\psi_{i,j,1}) = \beta_{0i,1} + \beta_i \cdot \mathbf{z}_{j,1}, \quad (7)$$

where $\beta_{0i,1}$ represents the species-specific occurrence probability (on the logit scale) in the first year and β_i represents a vector of species-specific regression coefficients that describe the effect of standardized covariates $\mathbf{z}_{j,1}$ on the occurrence probability of species i .

For the following years, the occurrence probability for the i th species at the j th site during year t will depend on if the i th species was present or absent from the j th site during year $t - 1$ (the prior year). Hence, they add species-specific autologistic parameter ϕ_i into the model such that

$$\text{logit}(\psi_{i,j,t}) = \beta_{0i,t} + \beta_i \cdot \mathbf{z}_{j,t} + \phi_i \cdot x_{i,j,t-1}, \quad (8)$$

where $\beta_{0i,t} + \phi_i$ represents the species-specific intercept in year t when species i occurred at site j during year $t - 1$, while $\beta_{0i,t}$ is the intercept in year t when the species i does not occur at site j during year $t - 1$ (as $x_{i,j,t-1} = 0$ due to the absence of species i) (Doser et al., 2022).

2.2.2 Observation Model: Replicated Detection-Nondetection data

As discussed in the basis of why ICOM was developed, the importance of using replicated detection-nondetection data was to account for the imperfection that exists within detection methods. This replicated data accounts for this because if we fail to detect a species the first time, conducting the data collection in a short period from the first time may detect the species.

For this replicated data, let $y_{i,r,k,t}$ denote the presence or absence of species i for the k th sampling replicate at site r (where $r = 1, \dots, R$ and R is a subset of the total J sites) during the year t . Hence, we have that $y_{i,r,k,t} \sim \text{Bernoulli}(p_{i,r,k,t} \cdot x_{i,j|r|,t})$, where $p_{i,r,k,t}$ represents the detection probability of species i at the r th site during the k th visit in year t , and $x_{i,j|r|,t}$ represents the true presence or absence of the i th species during year t at the j th site (as we know $x_{i,j,t}$ from the Ecological Process section above) that corresponds to the r th replicated data site.

Thus, we find that the species detection probability of $p_{i,r,k,t}$ varies by site and sampling covariates such that

$$\text{logit}(p_{i,r,k,t}) = \alpha_{0i,t} + \boldsymbol{\alpha}_i \cdot \mathbf{w}_{r,k,t}, \quad (9)$$

where $\alpha_{0i,t}$ is the species-specific detection probability (on the logit scale) in year t with average covariate values and $\boldsymbol{\alpha}_i$ represents a vector of parameters that describe the effect of standardized covariates $\mathbf{w}_{r,k,t}$ on the detection probability of species i (Doser et al., 2022).

2.2.3 Observation Model: Nonreplicated detection-nondetection data

For the nonreplicated detection-nondetection data, let $v_{i,m,t}$ denote the presence or absence of species i at site m (where $m = 1, \dots, M$ and M is a subset of the total J sites) during year t . Hence, we have that $v_{i,m,t} \sim \text{Bernoulli}(\pi_{i,m,t} \cdot x_{i,j|m|,t})$, where $\pi_{i,m,t}$ represents the probability of detecting species i at site m during year t , and $x_{i,j|m|,t}$ represents the true occurrence status of species i at the j th site corresponding to m th nonreplicated data site during the year of t .

Thus, we find that the species detection probability of $\pi_{i,m,t}$ varies by species, site, and time such that

$$\text{logit}(\pi_{i,m,t}) = \gamma_{0,i,t} + \boldsymbol{\gamma}_i \cdot \mathbf{s}_{m,t}, \quad (10)$$

where $\gamma_{0,i,t}$ is a species-year-specific intercept and $\boldsymbol{\gamma}_i$ represents a vector of parameters that describe the effect of standardized covariates $\mathbf{s}_{m,t}$ on the detection probability of species i (Doser et al., 2022).

2.2.4 Linking Species Models Across the Community

As explained in the basis of ICOM section, through the hierarchical community occupancy model, the species-specific parameters in the ecological process model and the observation models are treated as random effects from a community-level normal distribution with mean and variance parameters. As a result, we find that $\beta_{0,i,t} \sim \text{Normal}(\mu_{\beta_{0,t}}, \sigma_{\beta_{0,t}}^2)$, where $\mu_{\beta_{0,t}}$ is the mean of the occurrence probability across all species in the community during year t , and $\sigma_{\beta_{0,t}}^2$ is the variance of the occurrence probability across all species in the community during year t .

For the ICOM model, the authors wanted to make inferences on occupancy dynamics on a community level. Consequently, to make these inferences, they want to be able to be able to easily calculate some biodiversity metrics from the latent occurrence state ($x_{i,j,t}$) while accounting for the imperfection in the detection of species. Thus, the authors use a Bayesian Monte Carlo Markov Chain (MCMC) to obtain a posterior distribution to obtain estimates for the parameters; the posterior distribution is acquired through the calculation of the biodiversity metric for each iteration of a MCMC (Doser et al., 2022).

2.2.5 Data Integration using Joint Likelihood

From the replicated and nonreplicated detection-nondetection data that we discussed in the sections above, we combine these two data sources using the joint likelihood method. Hence, under the assumption that the likelihoods for each dataset are independent and conditioned by the ecological process model, we have our joint likelihood function to be:

$$\text{L}_{\text{ICOM}}(\boldsymbol{\alpha}_0, \boldsymbol{\alpha}, \boldsymbol{\gamma}_0, \boldsymbol{\gamma} | \mathbf{x}, \boldsymbol{\beta}_0, \boldsymbol{\beta}, \mathbf{y}, \mathbf{v}) = \text{L}_{\text{REP}}(\boldsymbol{\alpha}_0, \boldsymbol{\alpha} | \mathbf{x}, \boldsymbol{\beta}_0, \boldsymbol{\beta}, \mathbf{y}) \cdot \text{L}_{\text{NREP}}(\boldsymbol{\gamma}_0, \boldsymbol{\gamma} | \mathbf{x}, \boldsymbol{\beta}_0, \boldsymbol{\beta}, \mathbf{v}), \quad (11)$$

where $L_{\text{REP}}(\alpha_0, \alpha | x, \beta_0, \beta, y) \cdot L_{\text{NREP}}(\gamma_0, \gamma | x, \beta_0, \beta, v)$ is the product of the individual conditional likelihoods for the replicated and nonreplicated data sources (Doser et al., 2022).

2.2.6 A Critique on the Notation Used for the Models

Though it is not a major critique on the ideas or the mathematics behind the models that are provided above, I had a slight issue with the notation that was used for the species-specific intercepts. For example, the species-specific intercept for the i th species during year t is denoted by $\beta_{0_{i,t}}$. It may be more of a personal preference, but it is confusing having the 0 immediately next to the β as it would look as if we are multiplying the two were it not for the subscripts provided. Hence, I would correct the intercepts to instead use $\beta_{0_{i,t}}$ to denote the species-specific intercept for species i during year t . As a result, I will instead use this notation for the remainder of this review.

In addition to the notation issue that I addressed above, the equation for (11) was originally,

$$L_{\text{ICOM}}(\alpha_0, \alpha, \gamma_0, \gamma | x, \beta_0, \beta, y, v) = L_{\text{REP}}(\alpha_0 \alpha | x \beta_0 \beta y) \cdot L_{\text{NREP}}(\gamma_0 \gamma | x \beta_0 \beta v),$$

with no commas separating the parameters on the right-hand side of the equation. When viewing the paper in pdf form, it appears that the commas were emitted from the right-hand side of the equation to fit the equation in one line. However, without the parameters being separated with commas, it makes it seem as if $\alpha_0 \alpha$ is one parameter and $x \beta_0 \beta y$ is another one. Therefore, I went ahead and put my corrected version as (11) instead of what was originally in the paper.

2.3 Studies Conducted

2.3.1 Simulated Study One: Integration

In the first simulation study, the authors aimed to demonstrate how the data integration of replicated and nonreplicated detection-nondetection data sources performs better than a single data set. To do this, the authors simulated a replicated data source with $K = 3$ replicates with a detection probability of 0.5 and two nonreplicated data sources; one nonreplicated data source with high detection probability (0.78), and one data source with low detection probability (0.22). The reasoning for using varying these detection probabilities was to be able to see the benefits of integration when using different nonreplicated data sources.

Then, the authors generated 100 replicates of each of these data sources with $I = 25$ species, $T = 6$ years, and $J = 150$ (50 for each data source) total sites under ICOM with varying community-level parameters. Also, the species' occurrence probability was generated using equations (7) and (8), and the detection processes were generated as functions of the species- and year-specific intercepts along with a species-specific effect from a covariate as seen in the observation model sections above (Doser et al., 2022).

2.3.1.1 Results of First Simulation Study

Now examining the results of the first simulation study through Figure 1 seen below, we see that the bias estimates for both the species-level occurrence intercepts and covariates are the smallest when the replicated and the two non-replicated data sources are combined together. In addition, we see that the bias estimates also decrease in the case where two data sources are integrated together (with the exception being the bias in the intercept between the two nonreplicated data sources).

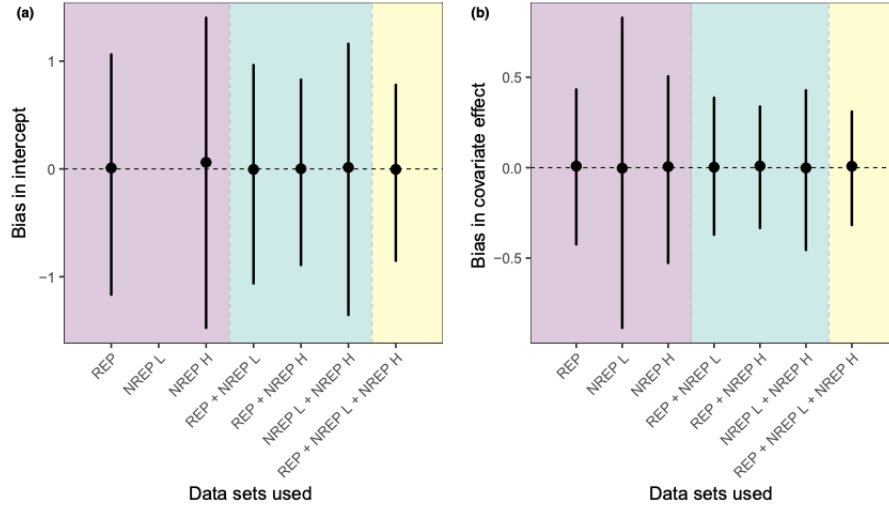


Figure 1: The Sampling Distribution of Estimated Biases in (a) the Species-Level Occurrence intercepts and (b) Covariates Effects (Doser et al., 2022)

2.3.2 Simulated Study Two: Community Modelling

In the second simulation study, the paper aimed to demonstrate that assessing the species-level parameter estimates from ICOM is better than assessing the parameter estimates from a single-species integrated distribution model (denoted as IDM). In IDM, the species-specific parameters were estimated individually for each species instead of obtaining as random effects of the community-level distribution. Similar to the simulation study one, the authors simulated 100 datasets containing one replicated data source with $K = 3$ replicates and two unreplicated data sources with the same detection probability (0.5). Also, they simulated a community of $I = 25$ species and $T = 6$ years, and the species-specific occurrence, detection intercepts, and the effect of covariates were generated from a uniform distribution to compare ICOM and IDM in cases where the species-level effects are not normal (Doser et al., 2022).

2.3.2.1 Results of Second Simulation Study

In the second simulation study, referencing Table 1 seen below, we see that the precision seems to show an improvement from IDM to ICOM by at least 9% for each parameter. However, we do see that the IDM bias is lower than the ICOM bias for all of the parameters. As a result, the authors note that the IDM is more accurate for estimating species-specific parameters, but they state that the loss in accuracy is negligible because the true species-specific parameter estimates were within the 95% Bayesian confidence interval through 94.9% of the simulations (Doser et al., 2022).

Parameter	Precision improvement (%)	ICOM bias	IDM bias	Parameter
$\gamma 0_i$	41.9	0.235	0.101	NREP detection intercept
$\gamma 1_i$	33.4	0.070	0.027	NREP detection covariate
ϕ_i	30.0	0.173	0.121	Auto-logistic
$\beta 0_i$	29.2	0.173	0.108	Occurrence intercept
$\beta 1_i$	22.5	0.042	0.017	Occurrence covariate
$\alpha 0_i$	18.8	0.104	0.044	REP detection intercept
$\alpha 1_i$	9.63	0.023	0.012	REP detection covariate

Table 1: Precision and Accuracy of Species-Specific Parameter Estimates from ICOM and IDM

2.3.3 Case Study: Foliage-Gleaning Birds

For the case study, the authors applied ICOM to analyze some biodiversity metrics of the community of 12 foliage-gleaning birds in the White Mountains from years 2010 to 2018. The datasets that were utilized in this case study were two replicated datasets from the Hubbard Brook Experimental Forest (HBEF) and the National Ecological Observatory Network (NEON) at Bartlett Experimental Forest (Barnett et al., 2019), and they used a unreplicated dataset from the North American Breeding Bird Study (BBS; Pardieck et al., 2019). A summary of these three datasets are provided in Table 2 below.

The model used in this case study follows the ecological process Equation (7) with occurrence probability in the first year is represented by $\psi_{i,j,1}$ such that

$$\text{logit}(\psi_{i,j,1}) = \beta_{0,i,1} + \beta_{1,j} \cdot \text{ELEV}_j + \beta_{2,j} \cdot \text{ELEV}_j^2 + \beta_{3,i} \cdot \text{FOR}_j, \quad (12)$$

where $\beta_{0,i,1}$ is the species-specific intercept at year 1, $\beta_{1,j}$ and $\beta_{2,j}$ are the species-specific effects of Elevation (linear and quadratic effects, respectively), and $\beta_{3,i}$ is the species-specific effect of forest cover within a 250 meter radius. Then, the occurrence in the following years are derived from Equation (8) (Doser et al., 2022).

	HBEF	NEON	BBS
Data type	Replicated	Replicated	Nonreplicated
Years	2010–2018	2015–2018	2010–2018
Number of sites	373	81	200
Elevation (m)	607 (240, 932)	432 (268, 766)	352 (134, 917)
Forest cover (%)	97.7 (71, 100)	94.2 (75, 100)	70.6 (0, 92)
Survey location	Experimental forest	Experimental forest	Roadside

Table 2: Summary of the Datasets used in Case Study (Doser et al., 2022)

2.3.3.1 Results of Case Study

In the case study, the authors evaluated the species richness and JACCARD index values (a measure of how similar a site is with another site in terms of the species occurrence) among the community of birds. As the occurrence probabilities peaked at around the mid-level elevation, they also found that both the species richness and JACCARD index values maximized at around mid-level elevations (around 600 – 800 meters) and higher forest cover (Doser et al., 2022).

In Figure 2 below, we see that the average occurrence probabilities throughout the 12 species of foliage-gleaning birds seemed to vary depending the data sources that were used as indicated by the different trends seen in the scatter-plots for the same species.

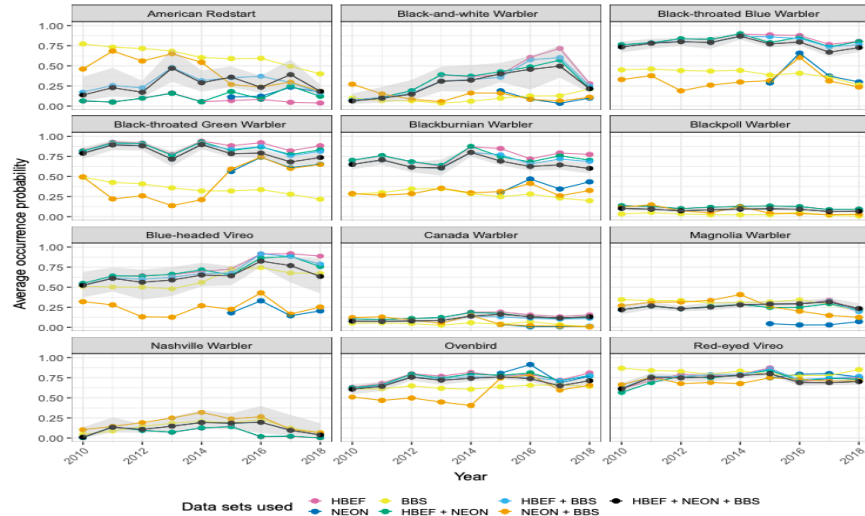


Figure 2: Average Occurrence Probabilities for the 12 species from 2010-2018 (Doser et al., 2022)

Most importantly (for the evaluation of ICOM), the authors found that the integration of all three data sources through ICOM yielded better predictive results for the community of birds. Looking at Table 3 below, a twofold cross-validation results of the predictive performances of combinations of the three data sources is provided. The values seen in the parentheses are indicative of the average rank of the model for an individual species across all the models; a value closer to 1 indicates that the model is the best for all species, and a value closer to 7 indicates that the model is the worst for all species. Thus, from Table 3, we see that the model that is the best for all species was the integrated data from the three data sources with the lowest value of 2.33.

Model	HBEF	BBS	NEON	All
HBEF	-10,174 (3.75)	-6,200 (4.67)	-947 (4.75)	-17,321 (4.5)
NEON	—	—	-792 (2.83)	—
BBS	-13,876 (3.75)	-5,702 (3.5)	-1,148 (5.08)	-20,726 (3.83)
HBEF + NEON	-10,097 (3.58)	-5,878 (3.33)	-852 (3.67)	-16,829 (3.33)
HBEF + BBS	-9,732 (2.67)	-5,717 (3.67)	-934 (4.42)	-16,383 (2.67)
NEON + BBS	-12,560 (4.5)	-5,759 (3.25)	-801 (3.83)	-19,121 (4.33)
HBEF + NEON + BBS	-9,767 (2.75)	-5,691 (2.58)	-836 (3.42)	-16,294 (2.33)

Table 3: Twofold Cross-validation Results Comparing Predictive Performance Across Models using Different Combinations of the Three Datasets

3 My Evaluation of ICOM

3.1 A Critique of ICOM

The authors briefly touched upon this in the beginning of their paper, but the use of models to assess communities is scarce. In their perspective, the reasoning is due to the need of large quantities of observations to do this. Using data integration to combine data sources that are readily available does save time on the collection process of data; however, there will still be higher computational costs when combining data sources. For example, the authors stated that in their first simulation study, the ICOM model using the three data sources took twice as long to run compared to a community model that used one replicated data source (Doser et al., 2022).

As a result of this, choosing when to implement ICOM becomes complex, and a team of researchers would need to evaluate if their community model of interest would benefit from multiple data sources. In addition to this, though it may save time from collecting the data itself, selecting the data sources that will be integrated with one another is a meticulous task as well. Therefore, if you are a researcher who prioritizes reduced cost and time within their research, then the utilization of ICOM may not be for you.

3.2 The Positives of ICOM

Though the computational costs of ICOM may be higher, the ICOM did produce positive results in the studies that the authors conducted. The first simulation study showed that the integration of data increases the accuracy of the species-level parameter estimates, and the second study showed that using ICOM led to a substantial improvement in precision over IDM. When applying the data to the case study of the 12 species of foliage-gleaning birds, the predictive performance of ICOM of all the data sources was higher than any of the models that used a subset of all the data (Doser et al., 2022).

In addition to the performance of ICOM in the studies, being able to make inferences on both a community-level and on a species-level is quite powerful. As an example, suppose that we see that there is a lowering trend on the occurrence probabilities of a community of birds at a specific site with the exception of one species. From there, we can make an inference about the community of birds as a whole, and we could take a closer look at the one species to figure out why this particular bird in the community is not following the same trend.

3.3 The Verdict

Through the majority of this review, I was aiming to weigh the strengths and weaknesses of ICOM to determine whether ICOM was a beneficial model for the scientific community. In my opinion, I have come to the conclusion that the ICOM is a positive contribution to research in existing occupancy models. I feel that the ability to make inferences at both the community-level and species-level is valuable, and the improvements in performance from ICOM outweigh the computational costs. In addition, I had no criticisms about any of the mathematical models and ideas that were utilized in ICOM, though my limited knowledge in this field of study could potentially attribute to this.

However, even though I believe that ICOM is a beneficial model, I do think that additional research needs to be conducted with the application of ICOM. The paper was published recently in February of 2022, and I could not find additional sources that discussed or utilized ICOM. Thus, by gaining more usage of ICOM through other case studies, there may be other flaws to ICOM that was not present in this particular study, and it was the authors' intention to convey that ICOM was a breakthrough in this field of study.

Regardless, this paper was a challenging, yet insightful read on a topic that I was not familiar with. It is always exciting to learn about a model that is fairly new, and I am looking forward to reading more academic articles about ICOM being applied to studies in the future.

References

- Barnett, D. T., Duffy, P. A., Schimel, D. S., Krauss, R. E., Irvine, K. M., Davis, F. W., Gross, J. E., Azuaje, E. I., Thorpe, A. S., Gudex-Cross, D., Patterson, M., McKay, J. M., McCorkel, J. T., & Meier, C. L. (2019). The terrestrial organism and biogeochemistry spatial sampling design for the national ecological observatory network. *10*(2). <https://doi.org/https://doi.org/10.1002/ecs2.2540>
- Doser, J. W., Leuenberger, W., Sillett, T. S., Hallworth, M. T., & Zipkin, E. F. (2022). Integrated community occupancy models: A framework to assess occurrence and biodiversity dynamics using multiple data sources. *13*(4). <https://doi.org/https://doi.org/10.1111/2041-210X.13811>
- Pardieck, K. L., Jr., D. J. Z., Lutmerding, M., Aponte, V., & Hudson, M. (2019). North american breeding bird survey dataset 1966-2018, version 2018.0. *US Geological Survey*. <https://doi.org/https://doi.org/10.5066/P9J6QUF6>