

Scaling Pattern of Occupancy Model for Estimating Species Occupancy

Evans Otieno Ochiaga¹ & Frederic Ntirenganya²

¹Phastar Limited Company, Nairobi, Kenya

²University of Lay Adventists of Kigali, Rwanda

Corresponding Authors: evansochiaga@aims.ac.za, fredo@aims.ac.za

Abstract

The positive relationship in occupancy and area is one of the fundamental aspects of conservation biology. The relationship has been described using different models ranging from spatially explicit to spatially implicit ones. In this paper, we present Individual Based Model for estimating occupancy area relationship numerically. The paper also compares performance of 9 occupancy area relationship parametric models. The performances of the models were examined using akaike information criterion (AIC) and residual standard error (RSE) procedures. Among all the models, Hui model which is spatially implicit and derived based on global and local densities of species distribution performs the best. Though the models we developed lacked generalized formulae due to complexity involved in coming up with it, its numerically solutions prove that occupancy increases with increase in the area, and this pattern is observed on other models.

Keywords: Occupancy area relationship, Scaling Pattern of Occupancy

INTRODUCTION

Recently changes in occupancy across spatial scales is one of the main research interests in spatial ecological (Kunin, 1998). Occupancy is obtained by dividing study area into small grids or cells of equal sizes then compute the fraction of cells occupied by focal species

(Barwell, Azaele, Kunin, & Isaac, 2014). The correlation between occupancy and resolution has been modelled for long, however, little attention has been given to spatial characteristic of the species distribution in coming up with these models. In this paper, the word resolution is the unit spatial size of any unit cell of a regular grid

superimposed on the entire study region and scale refers to the underlying characteristic length of the unit cells (Hui, 2009).

The increase in the resolution enables one to study occupancies at a relatively small spatial scale and vice versa (Hui, McGeoch, & Warren, 2006), and this is well documented using occupancy area curve which is also called scale area curve (herein referred as occupancy area curve). The occupancy area curve helps in analysing the species patchiness at different resolution (i.e. in sparsely distributed species, species individuals occupy several superimposed grids and highly aggregated species, species individuals are presented in some of those grids of the study area), and thus it addresses one of the crucial areas in spatial ecology.

At a relatively coarse resolution, occupancy is easily analysed since it depends on the occurrence of the species. However, this procedure cannot be used in documenting occupancy at a finer resolution because coarse resolution masked some spatial pattern that can only be identified when the cells are small. That is species with the same occupancy at coarse resolution could dramatically change in their spatial pattern at a finer resolution. This is because occupancy changes across the resolution and is strongly

correlated to spatial pattern (Hui & McGeoch, 2007).

As the cell resolution become finer and finer such that only one individual occupies a cell, alongside occupancy, abundance of the species can be easily estimated. These estimations are only possible by using downscaling models since conducting a survey in large area for estimating occupancy is quite expensive and time consuming (Hui, Boonzaaier, & Boyer, 2012).

For sometimes, different downscaling occupancy area relationship models have been proposed for both extrapolating and interpolating occupancy at different spatial scale (Barwell, Azaele, Kunin, & Isaac, 2014). However, most of these models, with exception of Hui and Thomas models which are spatially explicit, the rest are spatially implicit, that is, they do not accommodate spatial features in explicit manner and derivation of species occupancy curve is quite empirical (Holt, Gaston, & He, 2002).

In contrast, I propose a model that explicitly incorporates space called scaling pattern of occupancy (SPO). This model is based on the bisection framework by taking into consideration Harte's self-similarity assumption in the distribution of the species. The Harte assumption states that, species

distribution is self-similar if it's uniformly distributed across the assemblage such that the probability of the species occupying any given sub habitat of the area is constant and independent of the area, this probability is given by a (Harte, Kinzig, & Green, 1999).

The framework of this model was first laid by Hui in (Hui & McGeoch, 2007) by dividing one occupied cell into two sub cells having saturation and local densities as $p = (2a - 1)$ and $q = (1 - a)$ respectively. Similarly,

REVIEW OF OAR MODELS

in my model I used the same rational as Hui's model, but I divided occupied cell into four sub cells which is more complex with saturation probability p and local densities q , r , and s . This work is organized in a way that, in the methods section I reviewed already existing species occupancy area curve models that have been used for distributional downscaling and their related details. The performance of the model is presented in result and discussions sections

Quite number of occupancy area curve models have been proposed by different ecologists

Model Name	Formulae	Number of parameters	Reference
Poisson	$p = 1 - e^{-da}$	1	(Holt, Gaston, & He, , 2002)
Negative Binomial	$p = 1 - \left(1 + \frac{da}{k}\right)^{-k}$	2	(He, Gaston, & Wu, 2002)
Finite negative binomial	$p = 1 - \frac{\left(N + \frac{A_0k}{A} - k\right) \left(\frac{A_0k}{A}\right)}{\left(N + \frac{A_0k}{A}\right) \left(\frac{A_0k}{A} - k\right)}$	3	(Zillio & He, 2010)
Improved negative binomial	$p = 1 - \left(\frac{da}{\alpha(da)^\beta}\right)^{\frac{(da)^2}{\alpha(da)^\beta - (da)}}$	3	(He & Gaston , 2003)
Power	$p = ca^z$	2	(Kunin, 1998)
Nechman	$p = 1 - e^{-ca^z}$	2	(Nachman, 1981)
Logistic	$p = \frac{ca^z}{1 + ca^z}$	2	(Hanski& Gyllenberg, 1997)
Generalized negative binomial	$p = 1 - \left(1 + \frac{ca^z}{k}\right)^{-k}$	3	(He, Gaston, & Wu, 2002)
Thomas	$p = 1 - e^y$	3	(Azaele, Cornell, & Kunin, 2012)
Hui	$p = 1 - \left(\theta X \beta^{2a^2} \delta^a\right)$	3	(Hui, 2009)

Table 1: Table showing summary of OAR models, p and a are occupancy and grain respectively

SPO MODEL

Scaling Pattern of Occupancy (SPO) model is a model which is spatially explicit and derived based on Harte’s bisection framework. The bisection in SPO is done by considering self-similarity assumption of species distribution. In this model assemblage is divided into four 4 similar rectangles by connecting sides such that ration of length to width of sub-habitat is $\sqrt{2}:1$. In doing so, number of sub-habitat after i^{th} subdivision is 4^i with area of each one being $A_i = \frac{A_0}{4^i}$, where A_0 is area of original habitat. Therefore, based on the Harte’s assumption, if species is present in sub-habitat H_i , after i^{th} subdivision we have 4 mutually exclusive events;

- Species will be present in one subdivided sub-habitat unit of H_i with probability q.
- Species will be present in two subdivided sub-habitats unit of H_i with probability r.
- Species will be present in three subdivided sub-habitats unit of H_i with probability s.
- Species will be present in all subdivided sub-habitats unit of H_i with probability p.

Since the above 4 events are mutually exclusive, then $q + r + s + p = 1$. P is the saturation coefficient and given by $p = 1 - (q + r + s)$.

If there is species in the original habitat, then after first subdivision we have following occupancies (number of sub-habitats occupied by focal species divided by total number of sub-habitat) and condition probabilities as;

$$\frac{1}{4}(q_1) \frac{2}{4}r_1 \frac{3}{4}s_1 \frac{4}{4}p_1 \tag{1}$$

If the occupancy is $\frac{1}{4}$ with conditional probability of q_1 after first subdivision, then after second subdivision we have

$$\overbrace{\frac{1}{16}(q_1(q_2)) \frac{2}{16}(q_1(r_2)) \frac{3}{16}(q_1(s_2)) \frac{4}{16}(q_1(p_2))} \quad (2)$$

If occupancy is $\frac{2}{4}$ with conditional probability of r_1 after first subdivision, then we have the following occupancies and conditional probabilities after second subdivision;

$$\overbrace{\frac{2}{16}(r_1(q_2^2)) \frac{3}{16}(r_1(2r_2q_2)) \frac{4}{16}(r_1(s_2q_2 + r_2^2)) \frac{5}{16}(r_1(p_2q_2 + 2s_2r_2))} \\ \overbrace{\frac{6}{16}(r_1(p_2r_2 + s_2^2)) \frac{7}{16}(r_1(2p_2s_2)) \frac{8}{16}(r_1(p_2^2))} \quad (3)$$

Again, if occupancy is $\frac{3}{4}$ with conditional probability s_1 first subdivision, then we have the following occupancies and their conditional probabilities;

$$\overbrace{\frac{3}{16}(s_1(q_2^3)) \frac{4}{16}(s_1(3r_2q_2^2)) \frac{5}{16}(s_1(3s_2q_2^2 + 3q_2r_2^2))} \\ \overbrace{\frac{6}{16}(s_1(3p_2q_2^2 + 6s_2r_2q_2 + r_2^3)) \frac{7}{16}(s_1(6p_2r_2q_2 + 3s_2^2q_2 + 3s_2r_2^2))} \\ \overbrace{\frac{8}{16}(s_1(6p_2s_2q_2 + 3p_2r_2^2 + 3r_2s_2^2)) \frac{9}{16}(s_1(3p_2^2q_2 + 6p_2s_2r_2 + 3s_2^3))} \\ \overbrace{\frac{10}{16}(s_1(3p_2^2r_2 + 3p_2s_2^2)) \frac{11}{16}(s_1(3p_2^2)) \frac{12}{16}(s_1(3p_2^3))} \quad (4)$$

Also, if occupancy is $\frac{4}{4}$ with conditional probability p_1 after first subdivision, then we have the following occupancies and conditional probabilities for second subdivision

$$\overbrace{\frac{4}{16}(p_1(q_2^4)) \frac{5}{16}(p_1(4r_2q_2^3)) \frac{6}{16}(p_1(4s_2q_2^2 + 6q_2^2r_2^2))} \\ \overbrace{\frac{7}{16}(p_1(4p_2q_2^3 + 12s_2r_2q_2^2 + 4q_2r_2^3))}$$

$$\begin{aligned}
 & \frac{8}{16} \left(p_1 (12p_2 r_2 q_2^2 + 12q_2 s_2 r_2^2 + 6s_2^2 q_2^2 + r_2^4) \right) \\
 & \frac{9}{16} \left(p_1 (12p_2 s_2 q_2^2 + 12p_2 q_2 r_2^2 + 12q_2 r_2 s_2^2 + 4r_2^3 s_2) \right) \\
 & \frac{10}{16} \left(p_1 (24q_2 r_2 s_2 p_2 + 6q_2^2 p_2^2 + 6s_2^2 r_2^2 + 4s_2^2 q_2 + 4r_2^2 p_2) \right) \\
 & \frac{11}{16} \left(p_1 (12p_2^2 q_2 r_2 + 12p_2 q_2 s_2^2 + 12p_2 s_2 r_2^2 + 4s_2^3 r_2) \right) \\
 & \frac{12}{16} \left(p_1 (12p_2^2 s_2 q_2 + 12p_2 r_2 s_2^2 + 6p_2^2 r_2^2 + s_2^4) \right) \\
 & \frac{13}{16} \left(p_1 (4p_2^3 q_2 + 12p_2^2 s_2 r_2 + 4p_2 s_2^3) \right) \frac{14}{16} \left(p_1 (4p_2^3 r_2 + 6p_2^2 s_2^2) \right) \\
 & \frac{15}{16} \left(p_1 (p_2^3 s_2) \right) \frac{16}{16} \left(p_1 (p_2^4) \right) \tag{5}
 \end{aligned}$$

Where values in the bracket are the conditional probabilities and fractions are occupancies. By induction, after analyzing all bisection up to third level, it's realized that there are only 4^i occupancies after i^{th} bisection and they include; $1/4^i, 2/4^i, 3/4^i, \dots$ 1. Other occupancies can not be obtained based on this assumption.

Bisecting occupied cell from 1 to 2 gives a good picture of how occupancies is inversely related to number of bisections and easy to compute. However, using Harte's assumption by doing bisection from 1 occupied cell to 4 subdivision is more complicated and observing the pattern in occupancies as bisection increases is

complicated as observed in scaling pattern occupancy model. However, some simulations can be done to describes occupancy and pattern based this model. Through, simulation, positive relationship between occupancy and area is still explainable by developed SPO model.

CURVE FITTING AND ANALYSES

For the analysis all the species occupancy area relationship models in table1 were considered together with UK species binary data. The models were fitted using non linear least square regression method contained in min pack. Lm library. In additions, despite the lack of pattern in scaling pattern of occupancy model making its' performance to

be hard to evaluate, performance of other models used in describing the relationship in area and occupancy is possible. The models

were evaluated using Akaike information criterion (AIC) and residual standard error values as summarized in the table 2.

Model	AIC	RSE
Hui	-103.5032	0.008299
Nachman	-89.22244	0.0149215
Generalized nbd	-87.26891	0.015474
Logistic	-81.66481	0.01908867
Improved nbd	-81.74041	0.0196255
Nbd	-80.43574	0.02001117
Fnbd	-71.27764	0.022855
Power	-59.16873	0.03382333
Poission	-50.98314	0.04339633

Table 2: Table showing residual standard error values (RSE), akaike information criterion (AIC) for fitted species occupancy area models.

RESULTS

Occupancy frequency distribution (OFD) of bisection framework from one occupied cell

to four occupied cells indicate that as bisection increases, the limit of expected occupancies tend to zero. This is demonstrated using pattern of OFD which is more and more right skewed as bisection increase.

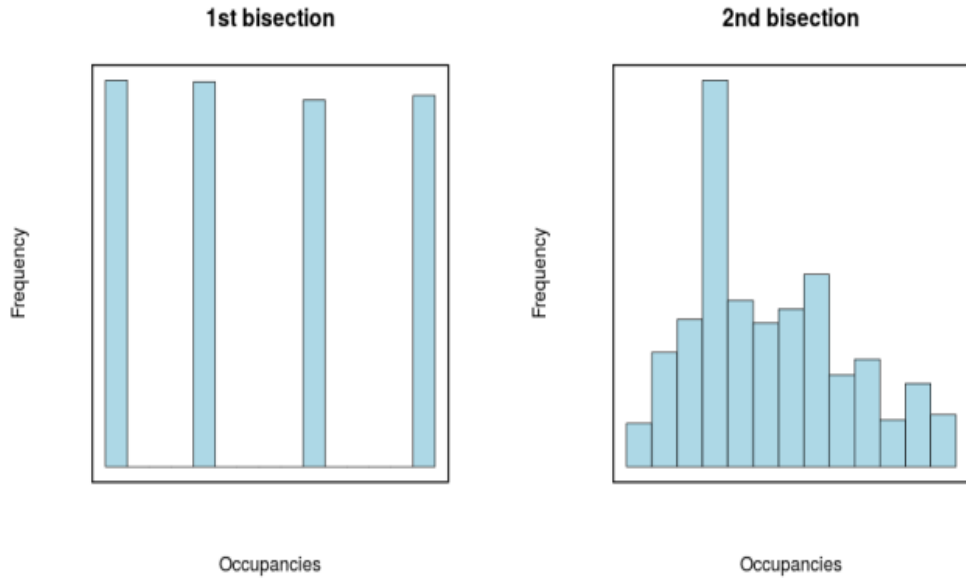


Figure 1: Histograms showing occupancy frequency distribution (OFD) for 1st and 2nd bisection simulated 1000 times, q,r,s, and p

are 0.1, 0.2, 0.5 and 0.2 respectively for both bisections.

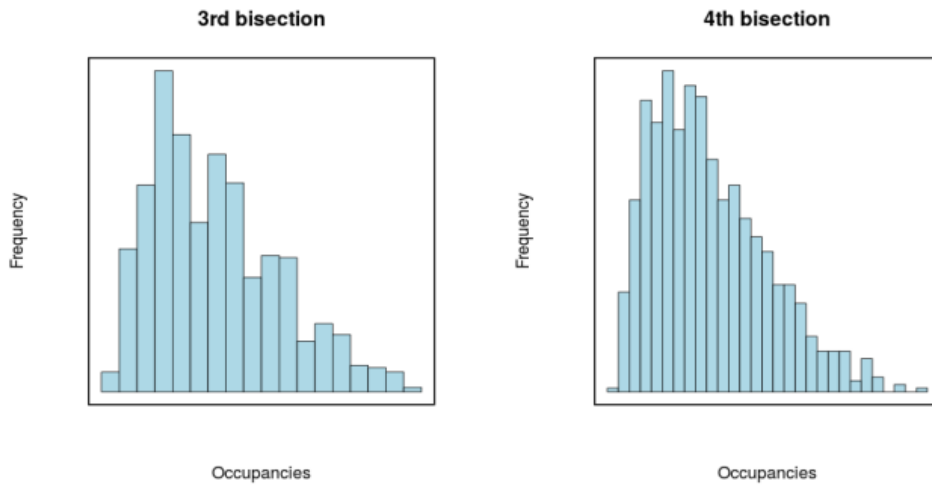


Figure 2: Histograms showing occupancy frequency distribution (OFD) for 3rd and 4th bisection simulated 1000 times, using q,r,s and p to as 0.1, 0.2, 0.5 and 0.2 respectively for all bisections.

occupancy becomes more and more right skewed. This is because occupancy is inversely related to bisection. The right skewness of expected occupancy implies that as spatial scale of study area increased occupancy increases as well and this behavior is also described in OAR models as presented in following figures;

From figures 1 and 2 it is observed that as number of bisection increases, expected

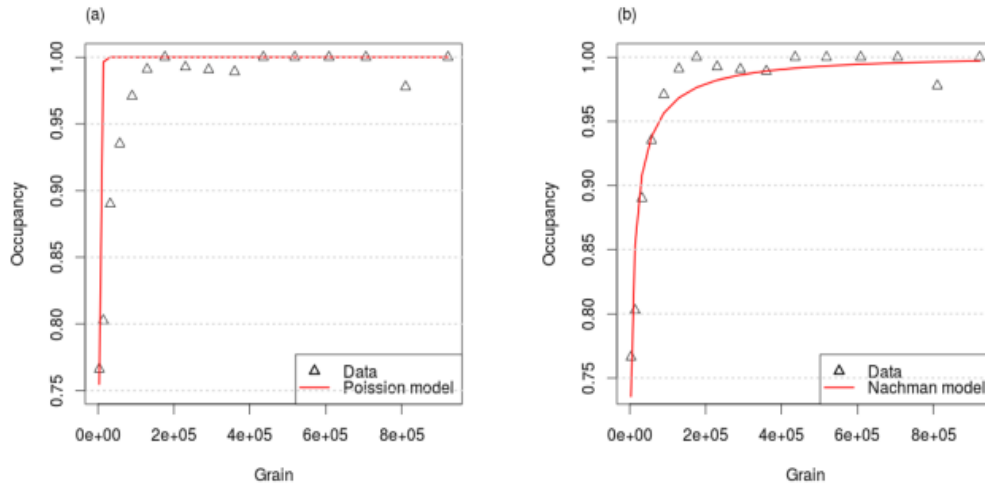


Figure 3: Figures showing fitted (a): poisson and (b): nachman models to occupancy area curve

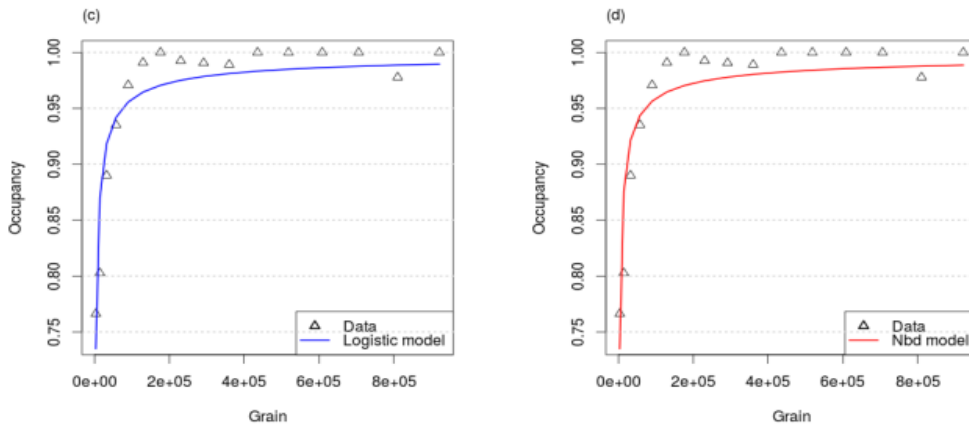


Figure 4: Figures showing fitted (c): logistic and (d): nbd models to occupancy area curve.

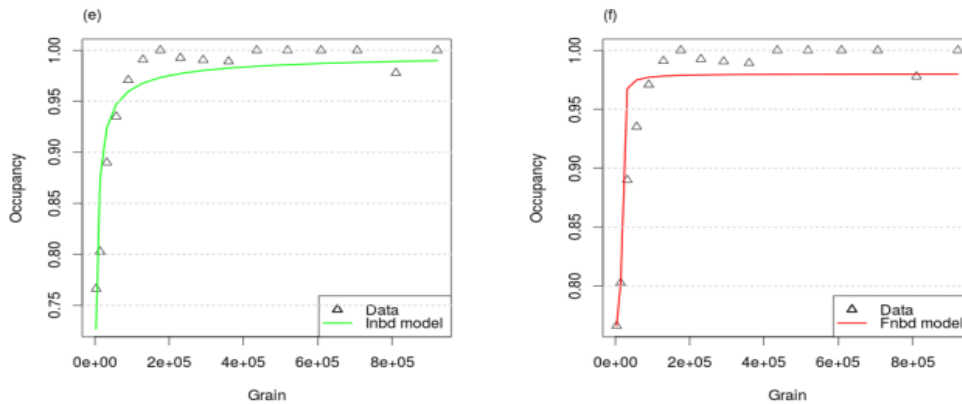


Figure 5: Figures showing fitted (e): inbd and (f): fnbd models to occupancy area curve

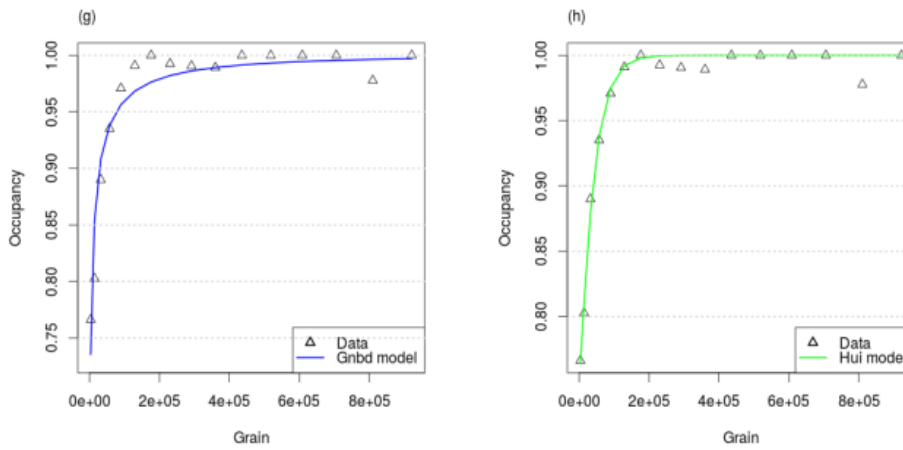


Figure 6: Figures showing fitted (g): gnbd and (h): hui models to occupancy area curve

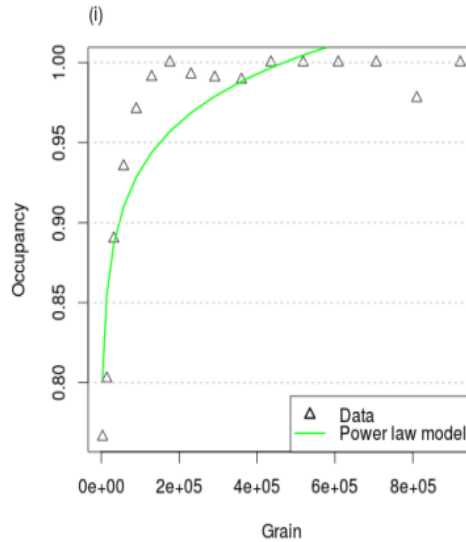


Figure 7: Figure showing fitted (i): power model to occupancy area curve

DISCUSSION

Relationship in spatial change and occupancy is a key area in ecology. For quite sometimes occupancy and area patterns had been analyzed using some occupancy area relationship models (Magurran & McGill,

2011). In ecological literature, with exception of Hui (Hui, 2009) and Thomas (Azaele, Cornell, & Kunin, 2012) models, no other model is derived based on the spatial characteristics of the species distribution . However, I developed a model that is spatially explicit and numerically describes

positive correlation between occupancy and area.

Despite the lack of generalized formulae of SPO model, its' numerical simulation gives a good picture of how occupancy is affected by spatial scale of the study area which is inversely related to bisection. From figures 1 and 2 the shift in the occupancy is small but it still shows the positive relationship in occupancy and area, this small difference in occupancy is due to self-similarity assumption.

In ecological literature, different methods have been used in evaluating models. Though those other methods also give a good evaluation procedure, in this work, the use of AIC and RSE in evaluating the models give one main observation; variation in performance of the model across the spatial scale which has not been reported before in any literature. It was realized that performance of models changes across different scales and in general after several simulations across different scales, Hui model was found to be the best.

The performance of species occupancy area models can be increased by taking into consideration approaches used in modelling species distributions (i.e. climatic condition) (Araujo, Thuiller, Williams, & Reginster,

2005). Although this aspect has not been fully utilized, Hui and SPO model has considered this to some extent since they use spatial information and a change in climate will affect occupancy of a species in a given spatial scale and hence a change in spatial information (Barwell, Azaele, Kunin, & Isaac, 2014). Therefore, for a better modelling of species occupancy, this factor should be considered in future.

Unlike other downscaling models, finite negative binomial models performs good for finite landscape (Zillio & He, 2010), although the model still converges for infinite area, its convergence do not reflect the real pattern of occupancy area relationship plot as shown in figure 5. On the other hand, power law model also exceeds unity for large spatial grain and this implies that for large area of studies, the two models are not appropriate.

Describing species spatial pattern is important in understanding and linking biodiversity pattern and epicenter for conservative biology. In this paper I reviewed all the downscaling models and due to increased interested in modelling species occupancy I developed downscaling model which numerically describes occupancy and area pattern. All these models were implemented in SPO package and after their

evaluation Hui model was found to be the best. Limitations of the models were also identified for their improvement in future applications in theoretical and conservation biology.

CONCLUSION

Conservation biology is a field that plays a role in proper management and planning of our community. In this paper, one main concepts in ecology has been discussed; species occupancy. Species occupancy helps in knowing distribution pattern of species in the community, for example, aggregated species are expected not to reach their saturation occupancy (i.e. occupancy 1) faster as compared to sparsely or randomly distributed species whose occupancy converges to 1 at a faster rate. Knowing species occupancy helps conservation ecologists in proper planning for the species since their extent would be known at hand. Due to increased demand on modelling of species occupancy, I also proposed a model that numerically describes the positive relationship in occupancy and area. Despite the fact this model lacks its generalized formulae due to the complexity involved, its' simulation gives a good picture of occupancy area relationship. The model is considered for future research and its variance will presented in future work.

REFERENCES

- Araujo, M. B., Thuiller, W., Williams, P. H., & Reginster, I. (2005). *Global Ecology and Biogeography*, 14(1), 17-30.
- Azaele, S., Cornell, S. J., & Kunin, W. E. (2012). Downscaling species occupancy from coarse spatial scales. *Ecological Applications*, 22(3), 1004-1014.
- Barwell, L. J., Azaele, S., Kunin, W. E., & Isaac, N. J. (2014). Can coarse grain pattern in insect atlas data predict local occupancy? *Diversity and Distribution*, 20(8), 895-907.
- Hanski, I., & Gyllenberg, M. (1997). Uniting two general patterns in the distribution of species. *Science*, 275(5298), 397-400.
- Harte, J., Kinzig, A., & Green, J. (1999). Self-similarity in the distribution and abundance of species. *Science*, 284(5412), 334-336.
- He, F., & Gaston, K. J. (2003). Occupancy, spatial variance, and the abundance of species. *The American Naturalist*, 162(3), 366-375.
- He, F., Gaston, K. J., & Wu, J. (2002). On species occupancy-abundance models. *Ecoscience*, 119-126.
- Holt, A. R., Gaston, K. J., & He, F. (2002). Occupancy-abundance relationships and spatial distribution. *Basic and Applied Ecology*, 3(1), 1-13.
- Holt, A. R., Gaston, K. J., & He, F. (2002). On species occupancy abundance models. *Ecoscience*, 119-126.

- Hui, C. (2009). A bayesian solution to the modifiable areal unit problem. *Foundations of Computational Intelligence*, 2, 175-196.
- Hui, C. (2009). On the scaling patterns of species spatial distribution and association. *Theoretical Biology*, 261(4), 481-487.
- Hui, C., & McGeoch, M. A. (2007). A self similarity model for the occupancy frequency distribution. *Theoretical population biology*, 71(1), 61-67.
- Hui, C., Boonzaaier, C., & Boyer, L. (2012). Estimating changes in species abundance from occupancy and aggregation. *Basic and Applied Ecology*, 13(2), 169-177.
- Hui, C., McGeoch, M. A., & Warren, M. (2006). A spatially explicit approach to estimating species occupancy and spatial correlation. *Journal of Animal Ecology*, 75(1), 140-147.
- Kunin, W. E. (1998). Extrapolating species abundance across spatial scales. *Science*, 281, 1513-1515.
- Magurran, A. E., & McGill, B. J. (2011). *Biological diversity: frontiers in measurement and assessment* (Vol. 12). Oxford: Oxford University Press.
- Nachman, G. (1981). A mathematical model of the functional relationship between density and spatial distribution of a population. *The Journal of Animal Ecology*, 453-460.
- Zillio, T., & He, F. (2010). Modeling spatial aggregation of nite populations. *Ecology*, 91(12), 3698-3706.