PRELIMINARY RESULTS OF ORANGUTAN DISTRIBUTION USING OCCUPANCY MODELLING AT THE SEDILU-SEBUYAU-LESONG LANDSCAPE IN SARAWAK.

*JOSHUA PANDONG¹, MARY BALANG², YONNY ANYIE³, MELYEN MANGGAT⁴, SYLVIA NG¹, MELVIN GUMAL⁵ AND D. MARK RAYAN¹

¹ Wildlife Conservation Society (WCS) Malaysia
No 7 Jalan Ridgeway 93200 Kuching, Sarawak, Malaysia

Previously with WCS Malaysia, currently affiliated with:
² Pengurusan Lebuhraya Pan Borneo Sarawak Sdn Bhd
³ Department of Wildlife and National Parks (DWNP) Peninsular Malaysia
⁴ Rumah Manggat, Menyang Taib Gaharu Tea
⁵ Biodiversity Conservation and Research, Sarawak Forestry Corporation

*Corresponding author’s email: jPandong@wcs.org

Abstract

The knowledge on orangutan distribution is currently imperfect and investigation into methods such as occupancy modelling could be used to map orangutan occurrence over time and space. For this pilot study, we spatially indexed 303 hexagonal tiles, each 1 km² in size, to represent plots over the Sedilu-Sebuyau-Lesong (SSL) landscape. We integrated rapid assessments of orangutan nest and occupancy modelling to generate proxy orangutan distribution for the survey period between July 2018 to November 2019. Preliminary results showed an estimate of 11%-22% probability of orangutans occupying 13 out of the 303 hexagonal tiles. These predicted occupied hexagonal tiles represent areas that are around the Ulu Sebuyau National Park and its proposed extension. Our study not only aids in refining monitoring methodology but also guides where management interventions are most needed to ensure the long-term survival of orangutans within the landscape.

Keywords: Bornean orangutan, hexagonal tiles, occupancy modelling, rapid assessment, Sarawak.
Introduction

The orangutans (Pongo spp) are the only great ape in Asia and found on the islands of Sumatra and Borneo (Rijksen and Meijaard, 1999; Goossens et al., 2009). The three species are namely the Sumatran (Pongo abelii), Bornean (P. pygmaeus) and Tapanuli orangutans (P. tapanuliensis) (Ancrenaz et al., 2016; Nowak et al., 2017; Singleton et al., 2017). The three orangutan species are listed as Critically Endangered under the IUCN Red List of Threatened Species.

In the past 60 years, threats to orangutan viability and conservation have escalated at an alarming level. Sumatra lost approximately 60% of its key orangutan forest habitat (11,570 km²) between 1985 and 2007 (Wich et al., 2011). Borneo lost approximately 24% (70,000 km²) of core orangutan range between the 1950s and 2010 (Struebig et al., 2015). The drastic decline of orangutan habitats can be attributed to several causes, mainly forest conversion into plantations, encroachment, fire, and shrinking fragmented habitats (Wich et al., 2012; Bryan et al., 2013; Gaveau et al., 2013; Hansen et al., 2013; Pandong et al., 2019).

Continued monitoring of orangutan population and distribution across various land use types have been crucial to assist researchers and managers in making informed decisions on conservation and enforcement (Santika et al., 2017; Utami-Atmoko et al., 2019). In Borneo, this is important as 24% of orangutan distribution occurs outside of protected areas (36,542 km²) and a further 49% in concessions expected to be converted to plantations (74,373 km²) or other commercial land uses (Wich et al., 2012). These mixed landscapes include oil palm plantations used by orangutans to move between forest patches (Ancrenaz et al., 2015), proposed agroforestry sites and communal lands without legal protection status but are viable habitats for orangutans (Wich et al., 2012; Pandong et al., 2018). As part of the continued monitoring effort, researchers incorporated various sampling techniques and technology to map prediction of orangutan distribution across various habitat units and land use (Husson et al., 2009; Struebig et al., 2015; Wich et al., 2016; Santika et al., 2017).

One of the existing techniques is a basic occupancy study design to assess spatial distributions of wide-ranging and elusive species (Magoun et al., 2007; Johnson et al., 2013). This involves surveying several sampling units or habitat patches within the larger study area to record signs of species presence or to determine resource use by the target species (Mackenzie and Royle, 2005; Johnson et al., 2013). Each sampling unit is visited by one or more trained observers to
detect animal signs and in more than one visit to help account for false negatives. The modelling approach then deals with imperfect detection of animal signs at neighbouring sampling units (Magoun et al., 2007), which are subsequently used for prediction and inference of species occupancy at the given sampling unit over the larger landscape.

We adapted many of the survey principles from Magoun et al. (2007) and Johnson et al. (2013) in preparation for this study. We did proceed with caution as based on past reconnaissance (recce) surveys, we found that portions of the study sites were without signs of old or new orangutan nest despite being identified as suitable orangutan habitat. This needs to be mitigated as in the past, it was alleged that orangutan population and their projected decline numbers were overestimated due to extrapolation of orangutan density in areas perceived to be ‘suitable habitats’ but potentially not occupied by orangutans such as in Maludam National Park.

For this study, we chose to survey the Sedilu-Sebuyau-Leṣong (SSL) Landscape, where a remnant orangutan population is still found outside the core habitats of Batang Ai-Lanjak-Entimau (BALE) Complex. However, the habitats at the SSL Landscape are fragmented and vulnerable due to large-scale land-use development in the surrounding area. The research questions were: a) do orangutans occupy or use the entire SSL Landscape?; and b) what is the probability of occupancy for new orangutan nests at the landscape?

We applied a spatial autologistic model on a tessellation of hexagonal tiles to analyse new nest data based on rapid assessments of new orangutan nest using the pilot field protocol developed for this paper. The model uses occupancy with imperfect detection of new nests in neighbouring tiles to predict the occupancy probability at the given tile with a measure of precision (coefficient of variation or CV) (Hoeting et al., 2000; Royle and Dorazio, 2008).

The advantages of using hexagonal tiles over common square grids include: a) suitable connectivity for orthogonal movements; b) equal length sides and identical nearest neighbours; c) lowest perimeter to area ratio (after a circle) to form a grid; and d) better fit for curved surfaces (Oom et al., 2004; Birch et al., 2007). We also adopted a Bayesian framework and used Markov chain Monte Carlo (MCMC) for inference (Mackenzie, 2006; Johnson et al., 2013).

The aims of this study were to: a) determine the occupancy probability of orangutans (subspecies P. p. pygmaeus), b) assess the degree of variability or uncertainty of the results, and c) generate proxy orangutan distribution maps for conservation planning. We further
discussed improvements of the study design for future surveys, limitation, and advantages of using the occupancy modelling, and the conservation implications of our findings.

Materials and Methods

4.1. Study area

Our study site consists of six contiguous areas with different land statuses: 1) areas with legal protection status, namely Sedilu National Park (NP), Ulu Sebuyau NP and Gunung Lesong NP; 2) non-protected areas proposed as connectors or extension of the NPs, namely the proposed Ulu Sebuyau-Sedilu connector, proposed Ulu Sebuyau extension and proposed Ulu Sebuyau-Gunung Lesong connector (Figure 1). These protected and non-protected areas are collectively known as the Sedilu-Sebuyau-Lesong (SSL) Landscape. The area sizes range from 5.95 km\(^2\) to 182.87 km\(^2\), and the total area combined is 306.91 km\(^2\) (Table 1).

Historically, the SSL Landscape was once known worldwide to be crucial orangutan habitats through the works of prominent naturalists such as Barbara Harrisson and George Schaller from the 1960s to 1970s. The former Sarawak Government under the British Crown Colony prior to the formation of Malaysia in 1963, assigned the Maias Protection Commission* in 1960 to determine orangutan distribution for protection in the region. (*Note: ‘Maias’ and the now standard ‘Mayas’ is the word for ‘Orangutan’ in the Iban language).

The Maias Protection Commission proposed the protection of the peat swamps of Sedilu, Sebuyau and Simunjan; these areas are within the Sadong and Batang Lupar river networks, which Schaller (1961) estimated to have had approximately 350 orangutans in the 1960s. Since then, updated protection statuses included as ‘Permanent Forest’, ‘Protected Forest’, and ‘Forest Reserve’ (1960s to 2000s). The Sarawak Government then gazetted the ‘National Park’ status to Sedilu and Ulu Sebuyau in 2010, and Gunung Lesong (also known as Bukit Lingga) in 2013. To this day, the SSL Landscape remains the only known remnant orangutan population outside (or 120 km to the west) of the core habitats of Batang Ai-Lanjak-Entimau (BALE) Landscape in Sarawak (Pandong et al., 2019).

Sedilu NP and Ulu Sebuyau NP consists of mainly peat swamp and kerangas forests, whilst Gunung Lesong NP consists of sandstone mountain (400-900 asl), mixed dipterocarp and semi-kerangas forests. The SSL Landscape is also important habitats for flying foxes (Gumal, 2004), at least four hornbill species in Sarawak (D. Kong, pers. comms.), and other endangered wildlife species. The Landscape is located within the administrative districts of Sebangan, Sebuyau, Simunjan, Pantu and Lingga. It is surrounded by more than 80 longhouses within 5 km buffer of the Landscape, oil palm plantations and smallholders’ agriculture lands.

4.2. Sampling scheme
We divided the study area into a tessellation of hexagonal tiles in ArcMap 10.5.1 (www.esri.com) based on Johnson et al. (2013). Each individual tile was 1 km$^2$ in size and there were 303 tiles overlaid on the SSL Landscape. The hexagonal design allows the tiles to have six identical and evenly spaced neighbouring units, with distance between all neighbours and the centroids to be the same (Oom et al., 2004; Birch et al., 2007).

Ideally, the survey routes would follow the method by Magoun et al. (2007) where a survey route was plotted through a tile centre, exited another side (with all sides had the potential to be included), and then entered the next tile in the direction of the tile centre. For our survey design, we further divided each selected hexagonal tile into six triangles and marked the six centre points. Where possible, the teams visited all six points either in a clockwise or anti-clockwise manner until they returned to their point of origin and completed a tile survey within a day (Figure 2).

We employed a systematic sampling design to select hexagonal tiles for the occupancy surveys. The advantage of this method is that it allows researchers to make the most of their travelling time to look for orangutan nests over a wide area coverage (Figure 2). We used ArcMap 10.5.1 to randomly select a point on the hexagonal grid, which subsequently select a cluster of four to seven tiles. A total of 80 hexagonal tiles were selected to be surveyed (Figure 3).

At this preliminary stage, we conducted this study to acquire a spatial overview of the SSL Landscape with a simplifying assumption that all surveys were conducted during the same survey window.

4.3. Model framework

We used the spatial autologistic model to map occupancy probability of new orangutan nest at each hexagonal tile and accounted for imperfect detection (Royle and Dorazio, 2008). The occupancy process ($\psi$) is the primary interest in this analysis. Occupancy probability for tile $i$ ($\psi_i$) is modelled as a logistic function of the proportion of neighbouring tiles occupied.

The probability of detection ($p$) allowed us to assess false zeros due to imperfect detection due to the detection process and variability at the tiles sampled. We summarize the Bayesian analysis for the spatial autologistic models of this study into three steps:

**Step 1: Identify the number of occupied neighbours**
We overlaid a neighbourhood of non-overlapping hexagonal tiles across the study sites. Each tile unit shares boundary with its neighbours and has between one to six neighbours. We identified $x_i$, the number of occupied neighbours of tile $i$ ($i = 1, 2, 3… G$), and modelled this relationship using a model by Royle and Dorazio (2008) as a reference and is shown in Equation (1) below:
The notation $n_i$ refers to the number of neighbours each tile has. $N_i$ is a matrix of $G \times \text{max}(N_i)$, where: a) $G$ refers to 303 rows of hexagonal tiles, and b) $\text{max}(N_i)$ indicates which of tiles $j$ are neighbours to tile $i$ based on $n_i$ in each row. The latent variable $z$ denotes the occupancy status of new orangutan nest at the tile during the surveys.

**Step 2: Describe the latent and observation processes**

To differentiate between variation in detectability and occupancy, we first introduce the distinction between latent process (underlying state variable observed imperfectly) and observation process (Royle and Dorazio, 2008). This is denoted by the latent $z$ and observation $y$. To accommodate imperfect sampling, we recognized that $y$ is equal to $z$ only sometimes, and at other times, we may falsely observe $y = 0$.

There are two possible mechanisms for non-occupancy observations. Firstly, ‘sampling zero’ means new orangutan nest was present but not detected. Second, ‘structural zero’ refers to no new nest was built by some random chance at a suitable habitat (Mackenzie, 2006). However, we have no way of knowing which of the two occur when $y = 0$. Thus, probability of detection ($p$) is the probability of observing new orangutan nest over the six spatial points in a hexagonal tile, given that it is present. Occupancy probability is the probability of tile occupied when $z = 1$. Details of the relationship between $p$ and $\psi$ are described below:

### 2.3.1. Latent process:

We implied that occupancy status at the landscape does not change for detection probability and occupancy probability estimation, which is the closure assumption. This meant that surveys were not conducted and separated by a long break (up to one year). In our survey design, we assumed that the occupancy states of $z_1$, $z_2$, $z_3$, ..., $z_G$ were spatially dependent within the landscape structure. The standard framework for modelling spatial dependence in a binary state variable is the use of the spatial autologistic model. The set of possibilities for the latent $z$ has two values: $z = 1$ for ‘occupied’ or $z = 0$ for ‘not occupied’. For tiles that were not surveyed, we needed to obtain a prediction on how many of the unsurveyed tiles were occupied. The relationship for the model depends on a Bernoulli process ($z_i$, detected or not detected) and varies among tiles based on the occupancy probability, $\psi$. The relationship is shown in Equation (2):

$$z_i \sim \text{Bernoulli} (\psi_i) \quad \ldots \quad \text{Equation (2)}$$

We specified a functional relationship between $\psi$ and the spatial auto-covariate $x_i$ in a linear relationship form using the logit link function in Equation (3):

$$\text{logit} (\psi_i) = \alpha + \beta x_i \quad \ldots \quad \text{Equation (3)}$$
where $\alpha$ and $\beta$ are parameters to be estimated. We chose the logit link function to model the probability of ‘success’ occupancy as a function of covariates (Mackenzie, 2006). The purpose of the logit link is to take a linear combination of covariate values and transform those values to the scale of probability, i.e. between 0 and 1. $x_i$ were expected to be constant over the six spatial points.

2.3.2. Observation process:

It is vital that imperfect detection of new orangutan nest be considered to infer about occupancy probability (Mackenzie and Royle, 2005). Ignoring imperfect detection would have understated the occupancy probability and distribution.

The observation $y$ refers to the number of spatial points on which new orangutan nest was detected at the tile; ‘NA’ if the tile was not surveyed. This model is based on a Binomial argument that assumes detection probability ($p$) is independent and identical for all expected parameter $p \times z_i$. Thus, if the tile is occupied ($z_i = 1$), the observations are binomial with parameter $p$. Conversely, if the tile is unoccupied ($z_i = 0$), then observations are binomial with probability 0 (i.e., the observations must be zero). This relationship is modelled in Equation (4):

$$y_i \sim \text{Binomial}(K, p \times z_i) \quad \text{... Equation (4)}$$

where $K$ is the total number of spatial points at each tile (six for each tile). However, this is irrelevant if the tile was not surveyed, but it must not be assigned as ‘NA’ in the analysis.

We could run Steps 1 and 2 using WinBUGS, that is specifications for an autologistic model with observation of the state variable $z$ subject to imperfect detection (Royle and Dorazio, 2008).

Step 3: Assess the variability or uncertainty of the estimates

The coefficient of variation (CV) is a measure of relative variability. It is the ratio of the standard deviation (SD) to the mean. The CV is useful to compare results between tiles that have different measures or values (Johnson et al., 2013). Thus, the posterior CV is a useful indicator of how ‘good’ or ‘weak’ the point estimate for occupancy probability is at each tile. ‘Good’ here refers to the spread of the posterior around the point estimate. Thus, higher CV means more variability or uncertainty; strong reliability for lower CV. From our results, there were more variability or uncertainty in areas that were spatially far from the surveyed tiles and at tiles with unobserved occupancy or sparse sampling effort. The equation for posterior CV in percentage is shown in Equation (5):

$$%CV = \left( \frac{SD_{\psi}}{\text{Mean}_{\psi}} \right) \times 100\% \quad \text{... Equation (5)}$$

4.4. Bayesian analysis using WinBUGS and implementation in R

Kruschke (2013) defines Bayesian analysis as the process of reallocating prior credibility consistent with the new data observed. Possibilities that were consistent
with the data gain more credibility; possibilities that were not, lose credibility. The Bayesian framework is the structure where the reallocation takes place. All the possibilities were spread out as a probability distribution; thus, the total area under the histogram (or curve) is equal to 1. The most credible range of possibilities which covers 95% of the posterior distribution is the highest density interval (HDI).

We adopted a Bayesian perspective and used Markov chain Monte Carlo (MCMC) (Plummer, 2003) for inference on the occupancy probabilities and imperfect detection estimates of this study. We ran the occupancy model in WinBUGS (Wintle and Bardos, 2006) using the R2WinBUGS interface (Sturtz et al., 2005) in R (version 4.0.3) (R_Core_Team, 2020). The analysis converged quickly, and 25,000 iterations were used of which 5,000 were discarded as burn-in. The mean and 95% HDI of the MCMC samples were used to summarize the posterior probabilities. The complete R scripts for these analyses are shown in the Supplementary Appendix section.

**Results**

The map of observed occupancy (Figure 4) shows that there were 6 out of 67 hexagonal tiles where new orangutan nests were recorded (Surveyed = 1). The number of tiles surveyed were short of the ideal 80 tiles targeted, which would have been systematically spread across the SSL Landscape. We surveyed 61 tiles without recording any new orangutan nest (Surveyed = 0) and did not survey 236 tiles. For this pilot study, our preliminary results of proxy orangutan distribution based on the rapid assessments between July 2018 to September 2020 showed that there were 21 tiles with higher occupancy ($\psi$, 18%-35%) (Figure 5). These were located within the Ulu Sebuyau NP and its extension area.

[Insert ‘Figure 4’ and caption here]

[Insert ‘Figure 5’ and caption here]

[Insert ‘Figure 6’ and caption here]

[Insert ‘Figure 7’ and caption here]

At this point, the maps of posterior occupancy probability (Figure 5) and its posterior CV (Figure 6) when visually overlapped, provided the following spatial narrative: the degree of variability or uncertainty becomes higher for tiles without detection of new orangutan nest or at unsurveyed tiles that were spatially far from the surveyed tiles (Figure 7). We then used the posterior CV as an indicator of how strong (CV < 1) or weak (CV > 1) the reliability of a $\psi$ estimate generated was for each tile.
Two other notable results generated were probability of detection and spatial autologistic parameters. Probability of detection ($p$) was at 0.9744 with 0.9188 to 1.0000 within the 95% HDI. However, this is unreasonably high and would have meant that teams conducting surveys detected new orangutan nests 97.44% of the times the nests were present. The probability of non-detection would then be too low for practical purposes, especially for unsurveyed tiles. For the spatial autologistic parameters, the intercept $\alpha$ was -2.9 and the coefficient $\beta$ was 3.3. This verifies that tiles were less likely to be occupied if the number of occupied neighbours were zero, and vice versa.

**Discussion**

Our preliminary results show that the occupancy modelling allowed us to use new orangutan nests as proxies to map orangutan distribution across surveyed tiles and their neighbouring unsurveyed tiles for the period of the surveys. However, the glaring issues of unreasonably high probability of detection as well as the high variability or uncertainty in the posterior CV show that many adjustments are needed to fine-tune the sampling scheme. In this section, we discuss ways to improve the study design for an occupancy analysis, the limitation, and advantages of using the occupancy modelling, as well as the conservation implications of our paper for orangutan conservation and management.

4.1. **Study design improvements**

4.1.1. **Probability of detection**

For this study, we estimated the probability of detection at the hexagonal tiles ($p$) based on the observation at the six spatial points within each tile. High $p$ was expected given detections were high during the surveys in tiles where new nests were observed. At present, the low probability of false negatives in our study shows that more sampling units or hexagonal tiles are needed to be surveyed across the SSL Landscape.

If resources are sufficient, we recommend the consideration to deploy multiple observers to conduct independent surveys in more tile clusters as well as to consider a second team to conduct surveys in reverse, to address the low counts of new orangutan nest. However, we caution that decisions should be based on project objectives. Mackenzie (2006) recommends shorter survey or minimal
sampling periods in a season. If a sampling season is too long, orangutans may appear present in all or appear to use all sampling units within the study area. But if it is too short, a sampling season may not provide sufficient opportunity for researchers to encounter new nests. Even so, survey results from previous sampling seasons can be used as ‘prior knowledge’ to generate trend analysis, if the same field methods are applied.

4.1.2. **Coefficient of variation (CV)**

The high variability or uncertainty in the posterior CV was due to the high number of unsurveyed tiles that were spatially far from the surveyed tiles as well as tiles without detection of new orangutan nest included in the analysis. Currently, there is a lack of surveys at the Sedilu NP, the proposed Ulu Sebuyau-Sedilu connector and Gunung Lesong NP. The map showing the overlap between occupancy probability and its posterior CV provided a visual guide for researchers and conservation practitioners at the SSL Landscape to plan future surveys at tiles with lower $\psi$ and weak reliability (CV > 1) if an investigation is needed. Otherwise, more tile clusters could be surveyed as part of a larger-scale occupancy surveys and monitoring across the landscape in the future (Hines et al., 2010).

4.2. **Limitation and advantages of using occupancy modelling**

The inference about orangutan distribution made using this study design heavily depends on the actual field survey method used to collect new orangutan nest data and the measure of imperfect detection estimated from them. We were unable to survey all the initially targeted 80 tiles systematically spread across the SSL Landscape due to seasonal flooding at the peat swamp areas as well as travel restrictions due to the Covid-19 pandemic. Below, we discuss two lessons learned from this study for a potential follow up research.

Firstly, the optimal number of tiles must be surveyed to reduce high variability of CV. Mackenzie and Royle (2005) suggested that the “optimal strategy for rare species is to conduct fewer surveys at more sites, while for a common species … conduct more surveys at fewer sites.” This means survey design for rare species with low occurrence such as orangutans should emphasize surveys at greater number of tiles and for surveys
done in a shorter timeframe or sampling season. We also intend to continue having systematic selection of tiles and with replacements, to avoid biasing them to easy-to-access or favoured sites.

Secondly, multiple teams should survey adjacent tiles on the same days whenever possible to allow detection under similar conditions across the study area (Magoun et al., 2007). The change in occupancy probabilities between multiple sampling seasons and across the landscape could infer trends about orangutan use of the landscape and possibly about dispersal, colonization at the tiles by orangutans, or perhaps temporal or local extinctions at the tiles if orangutans had not been observed in the area for decades.

4.3. Conservation implications for orangutan conservation and management

This study is useful to help assist researchers and managers in learning about occurrence and habitat use of orangutans across the SSL Landscape. The map showing the overlap between occupancy probabilities and their posterior CV provided a current distribution of orangutans at the study site. Hexagonal tiles with higher occupancy were an indication of orangutans using the habitats during the sampling season (Magoun et al., 2007). In our study, we identified 21 out of 303 tiles with $\psi$ between 18% and 35% as priority sites for protection and monitoring. Using this information, researchers and managers can examine the characteristics of the occupied and unoccupied tiles as part of investigating orangutan and habitat associations there.

If integrated into population studies, the SSL Landscape could then be stratified to determine different population estimates at various habitats during a sampling season. This is critical for reliable inferences for conservation and enforcement, amid rapid land development around the study site over time. We designed this study to use new nest data only and excluded old nests to provide distribution of orangutans during a sampling season. A different study design using new and old nest data could potentially be developed using this occupancy approach for conservation practitioners collecting both datasets.

For managing such rare species, where population estimates are difficult to obtain, occupancy probabilities could be used as a proxy to indicate the ‘persistence’ of the
species within a landscape. An increased occupancy probability could indicate greater travel for the existing individuals and/or an increase in population. A stable occupancy probability over different seasons could also be interpreted as a potential stable population whereas a dramatic decrease in occupancy could mean reduced use of habitat resources or a decrease in population. As these are long-lived species, much care is needed to interpret the results of the changes in occupancy probabilities.

4.4. Conclusion

In this paper, we showed that the occupancy estimation and modelling based on detection and non-detection of new nest data could be used as proxies to map current orangutan distribution across the SSL Landscape during a closed project period. The aim of this study was to use a tessellation of hexagonal tiles to determine occupancy probability of orangutan across the landscape. However, the unreasonably high probability of detection and variability or uncertainty in the posterior CV show that many adjustments are needed to fine-tune the sampling scheme.

We believe that occupancy modelling could be further developed as an effective way to explore change in distribution across time and space especially for orangutans as the critically endangered species is not always detected with certainty. We recommend more independent sampling units or tiles conducted in shorter timeframe for larger-scale occupancy surveys and monitoring in the future.

The ideal use of occupancy modelling at the SSL Landscape in our opinion is to use it as an integrative approach with population studies. With an integrated approach, it is possible to map the orangutan population and habitat use across various habitats over multiple short sampling seasons and over multiple occasions. In this way, we can avoid overestimation of population estimate based on suitable but unoccupied habitats.
Acknowledgements

We acknowledge the support of the Arcus Foundation, Borneo Adventure, Fondation Segré, Hock Seng Lee Berhad, Josephine Peiser Charitable Foundation, Orangutan Project Sdn Bhd, The Enlyst Fund, United Nations Development Programme (UNDP) Malaysia, Whitley Fund for Nature, Whitley-Segré Conservation Fund, Department of Survey and Mapping Malaysia, District Offices of Sri Aman, Pantu, Simunjan and Sebuyau, Forest Department Sarawak, Ministry of Finance Malaysia, Sarawak Forestry Corporation, and the Sarawak State Government. We also acknowledge the support of Mike Meredith, the statistician who developed the study design and R scripts of this study. Finally, we wish to acknowledge the tireless contribution of former WCS researchers and field guides who laboured tough field conditions to collect orangutan data over the years, including for this paper.
References


orangutan (Pongo abelii). *Science Advances, 2*(3), e1500789-e1500789. doi: 10.1126/sciadv.1500789

Appendices

List of figures.

Figure 1. Map showing the study sites of Sedilu-Sebuyau-Lesong (SSL) Landscape. The area consists of protected areas and non-protected proposed connectors/extension. Coordinate system: WGS84.
**Box 1.** The area size of each hexagonal tile is 1 km². There are six triangles with their centres marked in each tile. The position of the triangles are indexed as shown on the left.

**Box 2.** Ideally, a team of researchers will survey a hexagonal tile by visiting all six triangle centres. The team first head towards the centroid of the hexagon from the side closest to the hexagon in the middle of the cluster (see Box 3). They subsequently walk towards the centres of the triangles (either clockwise or anti-clockwise is fine) until they return to their point of origin. The tile size is designed for each team to complete surveys within a day.

**Box 3.** This study design uses systematic sampling to select hexagonal tiles for the occupancy surveys. This means not all tiles have equal opportunity of being selected for the surveys. A black tile (shown right) will be selected at random, and subsequent cluster of up to seven tiles from fixed distances (gray tiles) will be selected for our surveys.* The advantage of this method is that it allows researchers to make the most of their travelling time to look for orangutan nests over a wider area/coverage. *Note: The ideal distance between clusters of gray tiles is currently under review.

Figure 2. Infographic to summarize field methodology and systematic sampling.
Figure 3. Map of the study sites overlaid with 1 km² hexagonal grids (303 tiles in total). We adapted Johnson et al. (2013)’s method by integrating rapid assessments of orangutan nest with occupancy modelling into the study design. Coordinate system: WGS84.
Figure 4. Map of observed occupancy based on the rapid assessments of new orangutan nest conducted during the study period. There were six hexagon tiles with records of new orangutan nest out of 67 tiles surveyed. Coordinate system: WGS84.
Figure 5. Map of posterior mean of occupancy probability of new orangutan nest based on the rapid assessments. Tiles with higher occupancy (18%-35%) were at the Ulu Sebuyau NP and its extension area. Coordinate system: WGS84.
Figure 6. Map showing the posterior coefficient of variation (CV) of occupancy probability of new orangutan nest at the SSL Landscape. We used posterior CV to measure the reliability of the $\psi$ estimates: strong (CV < 1) or weak (CV > 1). Coordinate system: WGS84.
Figure 7. Map showing the overlap between the occupancy probability and its posterior CV.

We use this map to show the preliminary result of proxy orangutan distribution with a measure of reliability at the SSL Landscape. Coordinate system: WGS84.
## List of tables.

Table 1. List of study sites and their area sizes at the Sedilu-Sebuyau-Lesong (SSL) Landscape.

<table>
<thead>
<tr>
<th>No.</th>
<th>Study site</th>
<th>Gazette year</th>
<th>Protection status</th>
<th>Size (km²)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>Sedilu NP</td>
<td>2010</td>
<td>Totally protected area</td>
<td>63.11</td>
</tr>
<tr>
<td>2.</td>
<td>Ulu Sebuyau NP</td>
<td>2010</td>
<td>Totally protected area</td>
<td>182.87</td>
</tr>
<tr>
<td>3.</td>
<td>Gunung Lesong NP</td>
<td>2013</td>
<td>Totally protected area</td>
<td>5.95</td>
</tr>
<tr>
<td>4.</td>
<td>Ulu Sebuyau-Sedilu</td>
<td>NA</td>
<td>Proposed connector</td>
<td>~16.18</td>
</tr>
<tr>
<td>5.</td>
<td>Ulu Sebuyau extension</td>
<td>NA</td>
<td>Proposed extension</td>
<td>~31.69</td>
</tr>
<tr>
<td>6.</td>
<td>Ulu Sebuyau-Gunung Lesong</td>
<td>NA</td>
<td>Proposed connector</td>
<td>~7.11</td>
</tr>
<tr>
<td></td>
<td><strong>TOTAL</strong></td>
<td></td>
<td></td>
<td><strong>306.91</strong></td>
</tr>
</tbody>
</table>
Supplementary Appendix S1. R script for occupancy estimation for individual tiles adapted from Panel 9.4 of Royle & Dorazio (2008). This R script was saved as ‘SS_Spatial nests_new.txt’ and sourced into the model to run in WinBUGS (Supplementary Appendix S3).

```
# Title: WinBUGS model specification for an auto-logistic model with observation
# of the state variable z subject to imperfect detection.
# Adapted from: Panel 9.4 of Royle & Dorazio (2008)
# Probability of occupancy for each site,
# psi[i], is modelled as a logistic function of the proportion of
# neighbouring sites occupied.
# The data consist of:
# = nsite : total number of sites.
# = y (vector) : the number of occasions on which new nest was detected
# at the tile; NA if the tile was not surveyed.
# = K (vector) : the total number of survey occasions at each tile
# (irrelevant if the tile was not surveyed, but must not be NA).
# = NumNbrs (vector) : the number of neighbours for each tile.
# = Nbrs (matrix) : tiles in rows; the first NumNbrs entries in each
# row indicate which tiles neighbour the tile in questions.

model{

alpha ~ dnorm(0,.01)
beta ~ dnorm(0,.01)
p ~ dunif(0,1)

for(i in 1:nsite){
    x[i,1]<-0
    for(j in 1:NumNbrs[i]){
```
\[
x[i,j+1] \leftarrow x[i,j] + z[Nbrs[i,j]]
\]

logit(\psi[i]) \leftarrow \alpha + \beta \times (x[i, \text{NumNbrs}[i]+1] / \text{NumNbrs}[i])

\(z[i] \sim \text{dbern}(\psi[i])\)

\(\mu[i] \leftarrow z[i] \times p\)

\(y[i] \sim \text{dbin}(\mu[i], K[i])\)
Supplementary Appendix S2. R script to plot hexagonal tessellation. This R script was saved as ‘hexTess.R’ and sourced into Step 3 and 4 of the Supplementary Appendix S3. Mike Meredith (the main statistician of this study) wrote the original script.

```r
# Title: Plot hexagonal tessellation.
# Written by: Mike Meredith

hexTess <- function(data, colors, horizontal=TRUE, border=NULL, ...) {
  # data : a data frame with the coordinates of the centres
  # of the hexagons in the columns 1 and 2, and the values to plot in
  # column 3.
  # colors : a vector of colour specifications
  # horizontal : if TRUE, the top and bottom sides of the hexagon are
  # horizontal
  # border : colour of hexagon borders, NA = no border, NULL = default
  # colour (black)
  # ... additional arguments for plotting functions (untested!)

  # Get together some basic info
  npix <- nrow(data)
  dist <- as.matrix(dist(data[, 1:2]))
  diag(dist) <- Inf
  rad <- min(dist) / (2 * cos(pi/6)) # distance from centre of hexagon to
  # vertices.
  rad <- rad * 1.01 # increase slightly to avoid gaps between hexagons
    # when plotted.
  # Get bounding box
  bbox <- data.frame(x = c(min(data[,1]) - rad, max(data[,1]) + rad),
                     y = c(min(data[,2]) - rad, max(data[,2]) + rad))

  # Determine what to add to centre coods to get vertices
  v1 <- c(-rad, -rad*cos(pi/3), rad*cos(pi/3), rad, rad*cos(pi/3),
          -rad*cos(pi/3))
```
v2 <- c(0, rad*sin(pi/3), rad*sin(pi/3), 0, -rad*sin(pi/3), -rad*sin(pi/3))

if(horizontal) {
  vX <- v1 ; vY <- v2
} else {
  vX <- v2 ; vY <- v1
}

# Create array with coordinates of vertices: 6 x 2 x npix
verts <- array(NA, dim=c(6, 2, npix))
verts[, 1, ] <- outer(vX, data[,1], FUN="+")
verts[, 2, ] <- outer(vY, data[,2], FUN="+")

# Get the colours to plot
ncolors <- length(colors)
# scale z to range [0, 1]
tmp <- data[, 3] - min(data[, 3])
z <- tmp / max(tmp)
zcols <- floor(z * (ncolors-1)) + 1
# zcols[zcols > ncolors] <- ncolors # doesn't seem to be needed

op <- par(mar=c(5,4,4,6)+0.1) ; on.exit(par(op))
MASS::eqscplot(bbox, type="n", bty="n", xlab="", ylab="", ...) 

for(i in 1:npix)
  polygon(verts[, i], border=border, col=colors[zcols[i]], ...)

AHMbook::image_scale(data[,3], colors, digits=2)

return(invisible(verts))
Supplementary Appendix S3. Main R script to run the analysis and source codes from Supplementary Appendix S1 and S2. Mike Meredith (the main statistician of this study) wrote the original script.

```r
# Title: Main R script to run orangutan occupancy probability paper.
# Written by: Mike Meredith

library(R2WinBUGS)
raw <- read.csv("new_nests_spatial.csv", row.names=1)
attach(raw)

# = raw : data frame of new nest data imported via read.csv
# The variables in the data frame are:
# = Tile : original code for main hexagonal tiles
# = East : easting, Coordinate System: Borneo RSO Timbalai 1948
# = North : northing, Coordinate System: Borneo RSO Timbalai 1948
# = K : number of visits to each plot (9 if not surveyed)
# = Surveyed : 1 = surveyed, 0 = not surveyed
# = Nestnew : number of visits when nest was recorded,
# blank (NA) if not surveyed.

# Step 1. Do a matrix with pair-wise distances between tiles
# =========================================================
coords <- raw[, 1:2]
dist <- as.matrix(dist(coords))
# Put Inf on the diagonal (otherwise tile is its own neighbour)
diag(dist) <- Inf
# Check - look at neighbours for point #1
pt1 <- which(dist[1, ] < 3160)
# function to get neighbours for 1 column of dist
getNeigh <- function(x) {
  tmp <- which(x < 3160)
}
return(c(tmp, rep(NA, 6 - length(tmp))))

Nbrs <- t(apply(dist, 2, getNeigh))
dim(Nbrs)
head(Nbrs)

# See how many neighbours each sub-tile has:
NumNbrs <- rowSums(!is.na(Nbrs))
min(NumNbrs) # None less than 2

# Step 2. Run model in WinBUGS
# ........................
ntiles <- nrow(raw)
# Set up data for WinBUGS:
bugdat <- list(
y  = Nestnew,
nsite = ntiles,
K  = K,
NumNbrs = NumNbrs,
Nbrs = Nbrs)
str(bugdat)
pars <- c("p", "alpha", "beta", "psi", "z")

# The main run (takes ca. 10 mins):
spat1 <- bugs(bugdat, NULL, pars, "SS_Spatial nests_new.txt",
               n.chain=2, n.burn=5000, n.iter=25000, debug=TRUE)
# bugs.run.time()

spat1$mean$p # Mean probability of detection
# Put mean values for occupancy into a separate vector:
z <- spat1$mean$z
# Step 3. Generate map of posterior mean occupancy probability of orangutans
# ---------------------------------------------------------------------------
psi <- spat1$mean.psi
psicol <- round(psi*10)+1
colorz <- rev(grey.colors(5))

source("hexTess.R")
hexTess(cbind(raw[, 1:2], psi), colorz, border="grey")
title(main="Posterior mean of occupancy probability of new
orangutan nest at the greater SSL Landscape",
 xlab="East", ylab="North", cex=1.5, las=1,
 sub="Coordinate systems: Borneo RSO Timbalai 1948 (EPSG: 29873)")

# Step 4. Generate map of posterior CV of occupancy probability
# ---------------------------------------------------------------------------
psi.sd <- spat1$sd.psi
psi.CV <- spat1$sd.psi / spat1$mean.psi
colorz.se <- rev(grey.colors(5)) # new colours
hexTess(cbind(raw[, 1:2], psi.CV), colorz.se, border="grey") # with borders
title(main="Posterior CV of occupancy probability
of new orangutan nest at the greater SSL Landscape",
 xlab="East", ylab="North", cex=1.5, las=1,
 sub="Coordinate systems: Borneo RSO Timbalai 1948 (EPSG: 29873)"