



Munich Personal RePEc Archive

Improving biodiversity monitoring by modeling relative abundance from "presence only" data

Jingwa A, Brian

Hasselt University, Belgium

12 September 2011

Online at <https://mpra.ub.uni-muenchen.de/35232/>
MPRA Paper No. 35232, posted 06 Dec 2011 14:05 UTC

2010

2011

FACULTY OF SCIENCES

*Master of Statistics: Epidemiology & Public Health
Methodology*

Masterproef

*Improving biodiversity monitoring by modeling relative
abundance from "presence only" data*

Promotor :
Prof. dr. Geert MOLENBERGHS

Promotor :
Dr. MARC HERREMANS

Brian Jingwa Awungnja

*Master Thesis nominated to obtain the degree of Master of Statistics , specialization
Epidemiology & Public Health Methodology*

De transnationale Universiteit Limburg is een uniek samenwerkingsverband van twee universiteiten in twee landen:
de Universiteit Hasselt en Maastricht University

universiteit
hasselt

UNIVERSITEIT VAN DE TOEKOMST

 Maastricht University

Universiteit Hasselt | Campus Diepenbeek | Agoralaan Gebouw D | BE-3590 Diepenbeek
Universiteit Hasselt | Campus Hasselt | Martelarenlaan 42 | BE-3500 Hasselt

 Maastricht University

universiteit
hasselt
UNIVERSITEIT VAN DE TOEKOMST

2010

2011

FACULTY OF SCIENCES

*Master of Statistics: Epidemiology & Public Health
Methodology*

Masterproef

*Improving biodiversity monitoring by modeling relative
abundance from "presence only" data*

Promotor :

Prof. dr. Geert MOLENBERGHS

Promotor :

Dr. MARC HERREMANS

Brian Jingwa Awungnja

*Master Thesis nominated to obtain the degree of Master of Statistics , specialization
Epidemiology & Public Health Methodology*

Improving biodiversity monitoring by modelling relative abundance from “presence-only” data

**A thesis submitted to the Centre for Statistics in partial fulfilment
of the requirements for the degree of Masters of sciences**

By

Jingwa Awungnjia Brian

Supervised by

Prof. Dr. Geert Molenberghs, Internal supervisor

Marc Herremans, External Supervisor

12/09/2011

ACKNOWLEDGEMENTS

Thousands of volunteer naturalists contributed the data I analysed in this project. I am grateful for their input and I thank Natuurpunt Studie for providing me the data.

I am also grateful to Censtat professors for the immense work they did throughout the program. Their determination to inspire and lead us into statistical research was never-ending, in particular my internal supervisor Prof. Geert Molenberghs and my external supervisor Marc Herremans.

I would like to thank Hasselt University for providing access to valuable resources such as publications, software and pedagogic support necessary for the Masters of statistics program.

Many people contributed indirectly by providing moral support. I am particularly indebted to my wife Esther Vermote, my parents; Tambo George and Tambo Stella whose steadfast prayers and encouragement kept my spirits high. I heartily thank Edward and Maggy Vermote for inspiring me with patience, and faith.

ABSTRACT

One of the greatest concerns of this age is finding definitive ways to curb the loss of biodiversity. Although there have been growing calls on biodiversity protection and conservation by governments and institutions, the challenge lies on fitting models that best explain the trends of target species over space and time.

The objective of this project was to determine the probability of detection of four species and how it changes with time during the season. Also, the occupancy of each species and its relationship with abundance was obtained.

Site-occupancy models were fitted to the data, with each model having four components; occupancy, colonization, extinction and detection probability. Model selection was done on the basis of the AIC criterion.

The results show different trends for each species over time. The occupancies of each species do reflect the abundances in the various stages of their life cycles. The colonization, extinction and detection probability estimates were also obtained at each time point.

TABLE OF CONTENTS

1	INTRODUCTION.....	4
2	OBJECTIVES.....	6
3	PRESENCE-ONLY DATA.....	6
3.1	REASONS FOR USING PRESENCE-ONLY DATA.....	7
3.2	INHERENT PROBLEMS OF USING PRESENCE-ONLY DATA.....	7
3.3	STRATEGIES FOR ANALYSIS.....	7
3.4	HIERACHICAL MODELS FOR ANALYSING PRESENCE-ONLY DATA.....	8
4	DATA.....	9
4.1	STUDY AREA AND SPECIES.....	9
4.2	DATA STRUCTURE.....	10
4.3	SELECTION BIAS.....	10
4.4	SAMPLING SCHEME.....	11
5	STATISTICAL METHODS.....	12
5.1	ABUNDANCE.....	12
5.2	SITE OCCUPANCY MODELS.....	13
5.3	SITE OCCUPANCY MODELLING FRAMEWORK.....	13
5.4	PROBABILITY OF OBSERVATION (DETECTION).....	14
5.5	DETECTION HISTORIES OF BUTTERFLIES.....	16
5.6	DETECTION HISTORIES OF BIRDS.....	18
5.7	NAÏVE OCCUPANCY AND MAXIMUM LIKELIHOOD.....	19
5.8	MODEL SELECTION.....	19
6	APPLICATION TO THE DATA.....	21
6.1	BUTTERFLY SPECIES.....	21
6.2	BIRD SPECIES.....	26
7	DICUSSION.....	34
8	CONCLUSION.....	36
9	RECOMMENDATIONS.....	37
10	REFERENCES.....	38

1. INTRODUCTION

Since time immemorial, the earth has been endowed with millions of species of plants as well as animals. Biodiversity has been used to describe the variation of these life forms. The balance of these species in the planet is essential for our survival. Nazeri *et al* (2010) suggests that one of the greatest threats the planet now faces is that of species going into extinction. The World Wide Fund for Nature (WWF) uses biodiversity to express the desire for us to save the planet. The fund estimates that biodiversity has decreased by more than a quarter in the past 35 years.

Stork and Samways (1995) defines biodiversity as the variability among living organisms from all sources, including inter alia, terrestrial, marine and other aquatic ecosystems, and the ecological complexes of which they are a part.

The challenge today for conservation and biodiversity organisations is to develop strategies to monitor species. This is important for organisations, governments, NGOs and individuals involved in decision making. The World Resource Institute (WRI) explains how scientists were startled in 1980 by the discovery of a tremendous diversity of insects in tropical forests. A stunning discovery in the Panama of 80% of unknown beetle species out of 1200 found on a small area shows how much species richness could be present in an ecosystem. At least 6 million to 9 million species of arthropods and possibly more than 30 million are now thought to dwell in the tropics with only a small fraction currently described.

A report, commissioned by the European Union in 2008 estimated that 50 billion Euros worth of biodiversity-related services is lost every year. Also, the cumulative annual loss in ecosystem services will be worth 14 trillion Euros by the year 2050. This amount is will constitute about 7% the global GDP.

Many assessments of biodiversity abound. Conservation International and other organisations have underscored the importance of biodiversity conservation. By any measure, the conservation of biodiversity provides substantial benefits to meet immediate human needs, such as those for clean and consistent water supplies, protection from floods and storms and a stable climate.

Despite advances made in biodiversity conservation, the commitment of world leaders remain limited and the greatest question we now face is how to ensure that the benefits of biodiversity conservation remain available for future generations, whether or not we will be able to find new ways to benefit from nature and whether biodiversity continues to provide solutions to our most pressing problems, even as these problems change over time.

Many organisations, governments and institutions have made it a priority to reduce significantly the loss of biodiversity. In the year 2002, world leaders met during the Convention of Biological Diversity (CBD) and made a commitment to reduce the loss of biodiversity by 2010. However, they have failed to achieve this goal and today the rate of biodiversity loss is even higher. The vision of the EU commission is the halting of biodiversity loss and ecosystem degradation in the EU by 2020, making sure that resources

are used in a sustainable way by enhancing the positive contribution of agriculture and forestry as well as increasing contribution to global biodiversity.

One way of addressing this situation is by improving the monitoring of biodiversity which involves developing a strategic framework for policy making. Niemela (2000) suggests that modelling biodiversity constitutes an integral part of monitoring and hence a means to an end even though not an end in itself. The advantage of modelling biodiversity is that it determines the status of species at one or more ecological levels. In this respect, researchers are increasingly looking at modelling strategies for the relative abundance of species. The first challenge here is finding an effective way to measure biodiversity. Most ecologists and researchers identify 2 aspects of biodiversity that are important in quantification. These are species richness and relative abundance. Species richness refers to the number of species in an area. This is often the only information available.

2. OBJECTIVES

The objectives of the project include the determination of detection histories for each species and in each site, then using this information to obtain detection probabilities. The interest is to determine how species detection varies with time during the seasons. Another objective was to obtain the occupancies of each species as a determinant of abundance using site occupancy models.

3. PRESENCE - ONLY DATA

In biodiversity monitoring, it has become increasingly important to use presence-only data due to the high cost of collecting count data with a strict protocol. In many monitoring programs involving modelling species of animals, one of the most difficult steps to overcome is obtaining the data. Most of the data obtained in biodiversity monitoring are provided by volunteers who may either have taken prior engagement in the process or accidentally come across and observe a species of interest. Therefore the time and location of the species observed in this manner is hardly consistent. Also these volunteers are most often amateurs, but the data they provide is very informative and this is the only realistic way of obtaining information about species' presence.

Data in this form is sometimes referred to by many as presence-absence data. More appropriately, Kery *et al.* (2009) describes it as detection-nondetection data. Dorazio *et al.* (2011) shows that it involves the recording of the presence or apparent absence of species of interest. Here, counts of individuals can be used to estimate the abundance of the species. Also important in presence-only data are samples of locations with known presences and samples of locations with unknown presences.

Many studies have used presence-only and presence-absence data to analyse abundance of species. Royle and Nichols (2003) used repeated observations of presence-absence data in the estimation of detection probability. In Nazeri *et al.* (2010), the distribution of some tropical endangered species was analysed geospatially. In studies where the communities involve highly mobile species, Dorazio, Gotelli and Ellison (2011) recommends using presence-absence data rather than abundance surveys.

Many organisations support this kind of data collection through online data-portals such as the United States National Science Association (NEON), the Long Term Ecological Network (LTER), the US National Phenology Network, The Galaxy Zoo, the REEF Environmental Education Foundation, eBird program and many others. In Europe more specifically, several projects have been launched for collecting observations of birds and other species of interest. These include Natuurpunt in Belgium with the portal Waarneming.be in Belgium, the Stichting Natuurinformatie with Waarneming.nl in the Netherlands, and a world version (observado.org), as the country portal in Portugal and Spain, BirdTrack in UK, and the Ornitho group of dataportals in Switzerland, Italy and Germany, among others.

3.1. REASONS FOR USING PRESENCE-ONLY DATA

There are many reasons for using this type of data. Firstly, presence-only data constitutes the vast majority of data that is currently available on species occurrence (Zaniewski *et al*, 2002). Furthermore, Reliable information on absence data is not easily obtained leading to ambiguity. Also, Ottaviani *et al*, (2004) underscores that most available data obtained and used were collected without any specific design and without information of species absent hence more cost efficient.

3.2. INHERENT PROBLEMS OF USING PRESENCE-ONLY DATA

Even though presence-only data is widely available, there are many inherent problems involved in its interpretation. Perhaps the most obvious is the lack of absence data which is reliable and accurate. Most statistical models require both presence and absence information. Another problem as shown in Zaniewski *et al*, (2002) is the presence of bias.

Since there is no consistency in the design or collection methods, there is bound to be differential methods of obtaining data due to type of environment and accessibility, the distance to the cover and the amount of time volunteers are willing to commit. Another drawback, pointed out by Kery *et al*. (2009) is the lack of information about detection probability when presence-only data is obtained.

Perhaps one of the most important issues that researchers face when dealing with this data is the fact that non-detections are not synonymous to absences, in which case conventional binomial modelling can be used to obtain probabilities. If a species is not detected by the observer, it does not imply absence. Therefore, different methodologies are needed to model presence-only data.

3.3. STRATEGIES FOR ANALYSIS

Many authors and researchers have applied special technique to study species distributions. In the past, simple models were used. The simplest approaches proposed by Chakraborty *et al* (2011) involve summarizing the environmental attributes of the site of species occurrence and extrapolating the presence information to other areas with similar attributes. In this scenario, the spatial aspect of the data is not taken into account.

Another strategy involves the modelling of the probability of presence given the environmental conditions instead of just counts given the environmental conditions. Since it is not possible to obtain probabilities from presence data alone, the data is upgraded to presence-absence data. Pearce and Boyce (2006) and many other researchers have shown that data from background environment can be used to obtain the pseudo-absence data. These pseudo-absences are therefore assumed to be absence data. This methodology leads to the calculation of presence probabilities.

A review of some approaches is described in Peace and Boyce (2006). These include describing distribution of species using presence-only data, making use of presence-only as

well as pseudo-absence data, contrasting presence data and available locations and modelling abundance.

Depending on the research question, many different models have been used to analyse presence-only data. The number of different statistical methods is growing. Multiple regression models, neural networks, logistic regression models, classification trees have all been used in literature as in Guisan and Zimmermann (2000) with the application of each dependent on the trade-off between accuracy and bias. David and Shepherd (2010) show some weaknesses associated with logistic models such as model specifications and hence propose poisson point processes models instead. Gelfand *et al.* (2003) have used Bayesian hierarchical modelling to model presence-only data and obtain estimates for the parameters. Hierarchical models have also been used in by Chakraborty *et al* (2011), whereby the presence-only data are treated as realisations of a spatial process.

3.4. HIERACHICAL MODELS FOR ANALYSING PRESENCE-ONLY DATA

Different methodologies have been used in order to model presence-only data. An elaborate study carried out by Elith *et al.* (2006) shows comparisons of many models. In the study, 16 models were fitted over 226 species from 6 regions of the world. The central idea in these models is to obtain the probability of the true presence at a location y given the covariates $P(y = 1|x)$ whereby the presence-only data are obtained from the conditional probability of observed presences. A widely used method to model this conditional probability is the Logistic regression model. However, Keating and Cherry (2004) highlights the somewhat misinterpretations associated with use of logistic regression models by wildlife ecologists.

More generally Elith *et al.* (2006) distinguishes two broad categories based on the type of data involved. These include models that use presence-only records and those that characterise the background. Those that use presence-only records are known as envelop style methods and involve sites (locations) within an environmental hyper-space in which the species are located. Those that characterise the background are regression models such as Generalised Linear Models (GLMs) and Generalised additive models (GAMs).

Kery *et al* (2005) applied Poisson and negative binomial mixture distributions to model abundance. These models are used to model both abundance and detection. Since count data are biased due to heterogeneity, these mixture models have been applied because of their ability to estimate unbiased abundance.

Another formulation of models is shown in McKenzie and Kendall (2002), McKenzie et al (2003) and McKenzie (2005). These are unbiased methods of estimating proportion of area occupied by the species. Parameter estimates are estimated by maximizing the likelihood. Also, Bayesian methods have been applied to estimate parameters by sampling from the posterior distribution.

4. DATA

4.1. STUDY AREA AND SPECIES

The study area was divided into kilometre-squares (1km x 1km). There are four species of interest in this study. These include two butterfly species and two bird species. The butterfly species are Swallowtail (*Papilio machaon*) and Speckled wood (*Pararge aegeria*). The bird species are Buzzard (*Buteo buteo*) and the Black woodpecker (*Dryoscopus martius*).

The datasets are of two types. The presence-only data are the roving records available on the online portal www.waarnemingen.be. Benchmarking data (also known as Reference data) is obtained from counting programs in which strict protocols for recording are respected.

Roving records (presence only) were obtained from the portal www.waarnemingen.be for the year 2009 and 2010. Here, missing data implies no records of the species of interest, which means true absence or not recorded (apparent absence). Swallowtail is widespread but sparse and hence well reported while Speckled wood is commonly widespread and therefore underreported in the presence only dataset from www.waarnemingen.be.

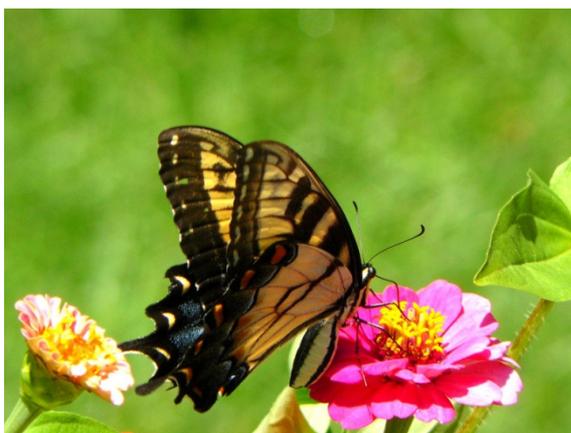


Figure 1: Swallowtail (*Papilio machaon*)

Courtesy of Words and Pictures



Figure 2: Speckled wood (*Pararge aegeria*).

The Woodpecker is local, sparse and well reported whereas the Buzzard, which is common and widespread is underreported in the presence-only data.



Figure 3: Buzzard (*Buteo buteo*)



Figure 4: Black woodpecker (*Dryoscopus martius*)

Benchmarking count data for the bird species was obtained from the Common Bird Census in Flanders. This is a standardized count program in which birds are counted in a selection of km-squares. For each visit, the birds are counted during 5 minutes at six fixed points within the km-square. The counting procedure is then repeated 3 times between March and July.

Benchmarking data for the butterfly species involve count of butterflies during 1 hour search in a km-square. Absence data at this point implies no butterflies of the species of interest were encountered (apparent absence).

4.2. DATA STRUCTURE

Of particular importance is the fact that not all animals present are detected in any field study. Hence each species is therefore “recordable” or “not recordable” during a field study. In the roving records, the km-squares that corresponded to those in the count survey were selected. In this situation, data was combined for each species to obtain a “proven present” set of km-squares.

The dataset for the Common Buzzards and Black Woodpeckers consists of 119017 visits (observations), most of which are repeated observations of the same km-squares on different dates. These dates span over the year 2009 and 2010. For the Speckled Wood and Swallowtail, there are 887 observations and were done in one year (2009). For each km-square, the observations for each species are recorded for the date observed. This results in a series of profiles (dataset strings) for each species.

4.3. SELECTION BIAS

Since the data is obtained without a statistical sampling design, there is a degree of bias involved. One way of limiting selection bias, was the introduction of the 1km-squares. Each square is then considered as the potential breeding range of the species of interest, on which the analysis was conditioned.

4.4. SAMPLING SCHEME

The design of the study is a fairly simple one. For each species, a number of sites are selected (s sites) and a single species occupancy model for multiple seasons is built. This can be represented schematically as shown in figure 5 below.

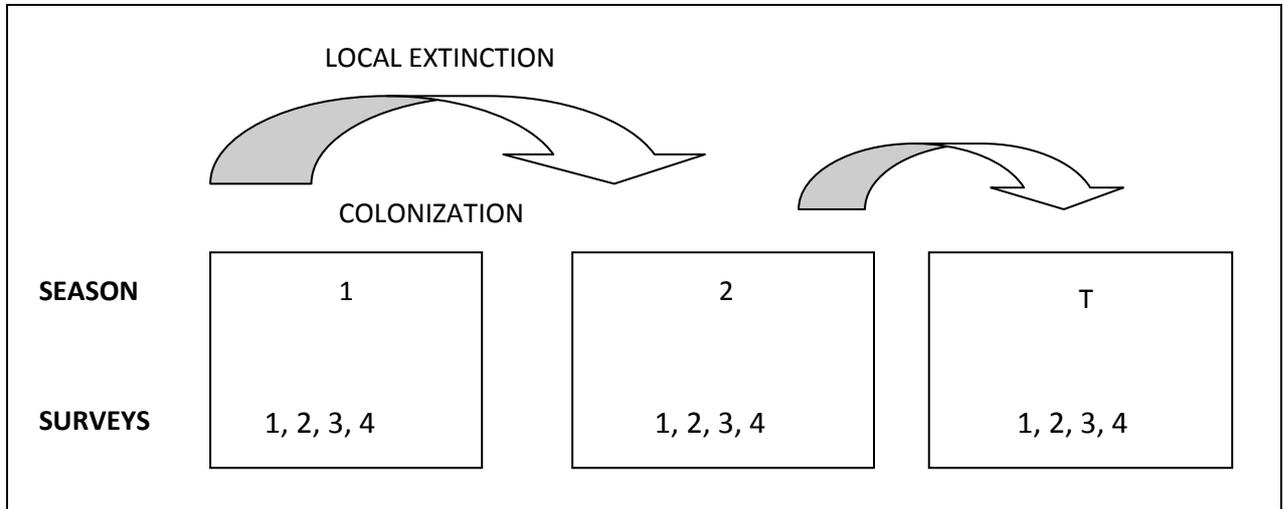


Figure 5: Sampling scheme of the study

The square box represents a km-square of the study area. The changes that occur within the km-square over time are represented as season 1, 2 ... T. In each season, there could be up to K surveys. In the study, there were multiple surveys within a week in some km-squares than others. In order to achieve the hierarchical design with an eye on reducing bias of detections, each month (season) was divided into 4 weeks. If the species was seen during the week, it was considered as detected. MacKenzie *et al* (2006) makes an important assumption in the derivation of the occupancy of the species.

The assumption here is that each site is closed to changes in occupancy within seasons, but changes may occur between seasons through the processes of local extinction and colonization.

The hierarchical sampling scheme is evident as there design shows two levels.

1. The primary sampling periods or seasons which are in this case the months.
2. The secondary sampling periods consist of short time periods within each primary sampling period. These are the weeks in this case.

The design is therefore a series of single species single-seasons put together. For each site, the detection of the species is recorded across all surveys within each season. This gives the detection history of the species in the sites.

It is possible, from any given survey, to calculate detectability and model the occupancy of a species. However, this cannot be reliable since it only gives a snapshot of the population at a single point in time. To reliably observe and detect species distributions, the process has to be repeated many times. Therefore, looking at data from many seasons to estimate occupancy

and detectability would give more stable results. The constraint of using this sampling scheme is that it costs more since more than one sampling occasion is required.

5. STATISTICAL METHODS

To analyse the data, it is imperative to understand the characteristics of the data type involved. Field surveys and counts result in recorded and not recorded data. In the instance where a species is seen by the observer, it is recorded as present hence considered as “proven present”. When the observer fails to find the species of interest, it is not recorded. This however does not imply the species is absent. The presence (proven presence) of a species gives the impression that it occupies that place at that moment (Occupancy). The subject of Occupancy has been explored by many researchers, but one of the most recent methodologies by McKenzie, (2005) describes it as the proportion of area, patches or sample units that is occupied by the species. This however does not stay constant as they move from place to place and migrate from season to season. Therefore, field surveys only result in “apparent occupancy”.

Apparent occupancy has many drawbacks which make it not a valid method. These include different detection probabilities for each species. Since birds may migrate with changing seasons, and butterflies fly in short generations, the probability to find them during a visit changes with time.

Also, the survey method used can affect the outcomes. Most people find it difficult to consecrate enough time in their search effort.

Another drawback is that some species may be difficult to find due to difficulty in accessing the habitat. These involve dense forests and rugged terrain. Species found on more open places are easier to find.

Since not all observations are recorded, the reported rate is $< 100\%$.

True occupancy results from apparent occupancy whereby the latter is repeated a number of times to overcome the difficulties mentioned.

5.1. ABUNDANCE

Abundance is used to describe the number of species of interest present across the occupied areas. Estimation of abundance is much more difficult and expensive than occupancy. Species are most often detected imperfectly with detection probabilities less than 1. When this occurs, Royle et al (2005) has shown that observed counts are biased estimates of true abundance. In many recent studies, occupancy has been used to replace abundance.

5.2. SITE OCCUPANCY MODELS

In most surveys, observation of occurrence doesn't occur without errors due to non detection of species that may otherwise be present. Kéry and Schmid (2006) highlighted the problem. Recent studies have attempted to solve the problem of non detection, occurrence probability, and site occupancy by widely adopting site occupancy models.

MacKenzie *et al.* (2002) proposed that by repeated surveying of the sites, the probability of detecting the species can be estimated which then enables unbiased estimation of proportion of area occupied and the probability a site is occupied by a species. Site occupancy takes into account two aspects:

1. A single point in time which involves surveys of the geographic locations, in this case the km-squares. Also, habitat relationships and incidence functions (a popular approach of estimating extinction and colonization probabilities) are taken into account.
2. Another aspect is change over time. This involves migrations, colonisation and extinction, and monitoring change over time.

The motivation for using site occupancy models is that the within-season (weeks, months) replications allows modelling of true species distribution corrected for detection probability. These models are hierarchical and well described by Royle and Kéry (2007) and Kéry *et al.* (2009).

This hierarchical nature is highlighted in the nested structure. Here, surveys are repeated several times both within the breathing season of the species and over several years. Another terminology applied by many authors is the "state-space" dynamic model. Site occupancy models are able to estimate true (rather than apparent) species distribution. They also do differentiate between species distribution and species detectability. Hence they are best suitable in this study.

5.3. SITE OCCUPANCY MODELLING FRAMEWORK

The modelling framework is based on the dynamic model of MacKenzie *et al.* (2002). Royle and Kéry (2007) employs the term *dynamic* because these models explicitly incorporate the meta-population's dynamic processes of colonization and extinction.

The model for the observed detection/non-detection data can be expressed as a product of two component models. These include a model for the data conditioned on the unobserved (or partially observed) process and also a model for the occupancy process.

In developing the model takes into account three assumptions.

1. The detection process is independent at each site
2. Heterogeneity is accounted for by covariates if present

3. There are no changes in the occupancy of the sites between sampling occasions. This is possible within a short time such as a year, but MacKenzie *et al.* (2003) warns that it may not hold in the long term such as many years.

The models of MacKenzie *et al.* (2002) allow the estimation of site occupancy, colonization and local extinction probabilities. This model works only when the species is not detected with absolute certainty, hence when species detection probability is less than 1. These models have been further extended in MacKenzie *et al.* (2003) to directly permit the estimation of the parameters.

The data obtained from the surveys consists of repeated presence/absence of four species under study. In each case, the spatial units (km-squares) involved can be represented as i where $i = 1, 2, 3, \dots, R$. Each spatial unit is surveyed j times where $j = 1, 2, 3, \dots, J$ within each season or primary sample period t where $t = 1, 2, \dots, T$.

The occupancy status of site i and survey j for season t can be expressed as $y_j(i, t)$. Royle and Kéry (2007) assumed occupancy status to be independent and identically distributed over the different survey times j . The true occupancy is given by $z(i, t)$ where $z(i, t) = 1$ implies that the unit (i) at season (t) is occupied by the species and $z(i, t) = 0$ implies that the unit (i) at season (t) is not occupied by the species. This can be used as the basis of developing the concept of detectability.

5.4. PROBABILITY OF OBSERVATION (DETECTION)

MacKenzie *et al.* (2002) uses straight forward probabilistic arguments. The probability of occurrence for any defined time period t is given as

$$\phi_t = \Pr (z(i, t) = 1)$$

This is also the probability that the site remains occupied. In order to understand what happens over time, the parameterization can be obtained by extending ϕ_t so that colonization and extinction of the species can be taken into account. Considering time $t + 1$ given time t , we have that

$$\phi_t = \Pr(z(i, t + 1) = 1 / z(i, t) = 1)$$

From probability theory, we deduce that the local extinction probability is $\varepsilon_t = 1 - \phi_t$. This parameterization can be extended to involve other time points and hence the detection history.

Generally, the probability of observing a full detection history is given by the expression

$$\Pr(\mathbf{X}_i) = \phi_0 \prod_{t=1}^{T-1} D(\mathbf{P}_{X,t}) \phi_t \mathbf{P}_{X,T}$$

Where ϕ_t is a 2x2 matrix of probabilities between occupancy states when t goes to $t + 1$ at primary sample period $t = 1$ to $T - 1$. Here, rows of ϕ_t is the occupancy status at of the site at t and the columns represent occupancy state at $t + 1$. In matrix notation, it is represented as follows

$$\boldsymbol{\phi}_i = \begin{bmatrix} 1 - \varepsilon_t & \varepsilon_t \\ \gamma_t & 1 - \gamma_t \end{bmatrix}$$

$\boldsymbol{\phi}_0$ is a vector $\boldsymbol{\phi}_0 = [\psi_1 \quad 1 - \psi_1]$ where ψ_1 is the probability that the site is occupied in the first primary period.

$\mathbf{P}_{X,t}$ is a column vector whereby the elements denote the probability of observing the detection history $\mathbf{X}_{i,t}$ in primary period t given occupancy state.

$D(\mathbf{P}_{X,t})$ is a diagonal matrix containing the elements of $\mathbf{P}_{X,t}$

The probability of observing the seasonal detection history is given by the relationship

$$\psi_t = \psi_{t-1}(1 - \varepsilon_{t-1}) + (1 - \psi_{t-1})\gamma_{t-1}$$

To obtain the parameter values $\psi_1, \varepsilon, \gamma$ and p Betts *et al* (2008) proposed maximizing the likelihood.

$$L(\psi_1, \varepsilon, \gamma, \mathbf{p}/\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_N) = \prod_{i=1}^N \Pr(\mathbf{X}_i)$$

Where ψ is the occupancy, γ is the colonization, and ε is the local extinction

The maximum likelihood modelling procedure relies on data from detection histories in order to estimate occupancy.

A crucial assumption here in the estimation of the model parameters is all parameters are constant across sites and at any given time. This implies that all sites are assumed to be homogeneous. If this assumption is neglected, the implications would be far reaching and parameter estimates would be biased. In that case, parameter estimates would be conditioned on the site.

The representation of the model components is shown in figure 6 below. Occupancy, colonization and extinction probabilities are of interest. The figure shows that a species is detected with a probability equal to ψ_1 and if it's not detected, the probability of $1 - \psi_1$ is associated to it. If a km-square is occupied (shaded squares), two things can happen as time goes on. Local extinction represented by ε_1 hence making the km-square to become empty. On the other hand, if the species does not become locally extinct, it's occupancy is $\psi_1(1 - \varepsilon_1)$ and also $(1 - \psi_1)\gamma_1$. Hence generally sites that will be occupied in the next season is a combination of two things: sites occupies this season where species doesn't become locally extinct ($\psi_t(1 - \varepsilon_t)$) and sites which are not occupied now but become colonized

$$\psi_{t+1} = \psi_t(1 - \varepsilon_t) + (1 - \psi_t)\gamma_t$$

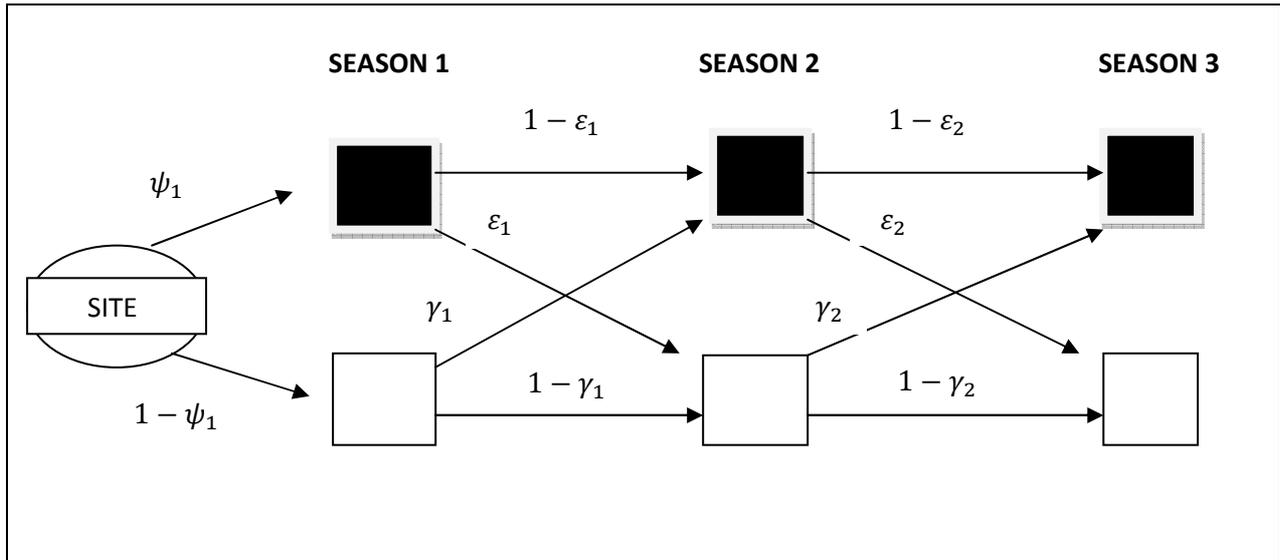


Figure 6: Schematic representation of occupancy state. courtesy of Mackenzie *et al* (2006)

5.5. DETECTION HISTORIES OF BUTTERFLIES

Detection histories are obtained from the detection/non-detection data of the species. In each survey, a species can be detected or not detected when it is present at the site. However, a species cannot be detected if it is not present at the site. The detection history can be represented by a vector of values representing presence and absence. If the target species is detected at least once, then the presence information is given the value “1” and if the species is not detected, it is given the value “0”. When the site was not surveyed, it was given “-”

The resulting detection histories for site i in season t is denoted $h_{t,i}$ while that of site i across all seasons is h_i .

The full detection histories and associated probabilities are given as follows;

For the speckled wood butterfly, the full detection history for the first km-square (DS7261) is given as 0110 101- --10 -- 1- ---1 - 0 --. There are altogether 24 weeks (6 months). Each represents the detection or non detection of the species in the km-square DS7261 at a particular time (week). We can interpret the history as follows:

In the first month (April) the speckled wood butterfly was detected during the second and third weeks. There were no detections in the first as well as the fourth week.

In May, there were detections in the first and third weeks but no detection in the second week. There was no survey in the fourth week

Interpretation of the detection histories of the rest of the months can be done likewise. For the second km-square, the detection history is given as -1-- 0010 ----- 010- 0-0- - 1--

$$\mathbf{X}_{DS7261} = 0110 \ 101- \ --10 \ -- 1- \ ---1 \ - 0 \ --$$

$$\mathbf{X}_{\text{DS7360}} = -1-- \quad 0010 \quad - - - - \quad 010- \quad 0-0- \quad -1--$$

The corresponding probabilities of observing the full detection histories for the first two km-squares are given as follows. The seasonal occupancy probabilities are also shown in table 1.

Table 1: Seasonal occupancy detection histories and associated probabilities of butterflies

\mathbf{X}_t	$\Pr(\mathbf{X}_t) = \boldsymbol{\phi}_0 \prod_{t=1}^{T-1} D(\mathbf{P}_{X,t}) \boldsymbol{\phi}_t \mathbf{P}_{X,T}$
$\mathbf{X}_{\text{DS7261}} = 0110 \quad 101- \quad --10 \quad --1- \quad ---1 \quad -0--$	
April (Season 1) = 0010	$\psi_1(1 - p_{1,1})(1 - p_{1,2})p_{1,3}(1 - p_{1,4})$
May (Season 2) = 101-	$\psi_2 p_{2,1}(1 - p_{2,2})p_{2,3}$
June (Season 3) = --1-	$\psi_3 p_{3,3}$
July (Season 4) = --10	$\psi_4 p_{4,3}(1 - p_{4,4})$
Aug (Season 5) = ---1	$\psi_5 p_{5,5}$
Sept (Season 6) = -0--	$\psi_6(1 - p_{6,2})$
$\mathbf{X}_{\text{DS7360}} = -1-- \quad 0010 \quad - - - - \quad 010- \quad 0-0- \quad -1--$	
April (Season 1) = -1--	$\psi_1 p_{1,2}$
May (Season 2) = 0010	$\psi_2(1 - p_{2,1})(1 - p_{2,2})p_{2,3}(1 - p_{2,4})$
June (Season 3) = - - - -	ψ_3
July (Season 4) = 010-	$\psi_4(1 - p_{4,1})p_{4,2}(1 - p_{4,3})$
Aug (Season 5) = 0-0-	$\psi_5(1 - p_{5,1})(1 - p_{5,3})$
Sept (Season 6) = -1--	$\psi_6 p_{6,2}$

The same procedure was repeated for all the km-squares in the two butterfly species as well as the two bird species.

5.6. DETECTION HISTORY OF BIRDS

The data for the birds was divided into two parts; the detection of the two species in 2009 and 2010. The same hierarchical design as for the butterfly species was implemented. However, the number of months (seasons) in this case is 5 (March, April, May, June and July) with each season having 4 surveys (weeks). Again, the point was to find out whether or not the species was detected in that any given week. Since the precise dates of detection were not consistent over the km-squares, it was necessary to revise the survey period from a day to a week. In this design, only three outcomes are possible. If the species was seen in a given week, it was coded as detected (“1”) for the km-square in which it was seen. If the species was not seen, it was coded as not detected or present (“0”). If the km-square was not surveyed in a given week, it was coded not surveyed (“-“). Also important to note is the fact that non detection is not equivalent to absent because the species may be present but not seen by the observer. For the first two km-squares of the year 2009 the probabilities of observing the detection histories are shown in table 2.

Table 2: Seasonal occupancy detection histories and associated probabilities of birds

\mathbf{X}_i	$\Pr(\mathbf{X}_i) = \boldsymbol{\phi}_0 \prod_{t=1}^{T-1} D(\mathbf{P}_{X,t}) \boldsymbol{\phi}_t \mathbf{P}_{X,T}$
$\mathbf{X}_{DS6858} = \text{---}0 \text{ -}0\text{--} \text{ ---}0 \text{ -}0\text{--} \text{ 0--}$ --	
March (Season 1) = ---0	$\psi_1(1 - p_{1,4})$
April (Season 2) = -0--	$\psi_2(1 - p_{2,2})$
May (Season 3) = ---0	$\psi_3(1 - p_{3,3})$
June (Season 4) = -0--	$\psi_4(1 - p_{4,2})$
July (Season 5) = 0---	$\psi_5(1 - p_{5,1})$
$\mathbf{X}_{DS6859} = \text{--}0\text{-} \text{ 010-} \text{ 1000} \text{ 000-} \text{ 00--}$	
March (Season 1) = --0-	$\psi_1(1 - p_{1,3})$
April (Season 2) = 010-	$\psi_2(1 - p_{2,1})p_{2,2}(1 - p_{2,3})$
May (Season 3) = 1000	$\psi_3p_{3,1}(1 - p_{3,2})(1 - p_{3,3})(1 - p_{3,4})$
June (Season 4) = 000-	$\psi_4(1 - p_{4,1})(1 - p_{4,2})(1 - p_{4,3})$
July (Season 5) = 00--	$\psi_5(1 - p_{5,1})(1 - p_{5,2})$

5.7. NAÏVE OCCUPANCY AND MAXIMUM LIKELIHOOD

The naïve occupancy is the proportion of sites out of the total number of sites where the species is detected. This is given as

$$\text{Occupancy} = \frac{\text{Number of sites where species is detected}}{\text{Total number of sites surveyed}}$$

This estimate is biased and therefore the survey needs to be repeated several times.

To obtain unbiased estimates of occupancy and other parameters, we proceed by maximum likelihood. If there was a single season involved, we would divide the frequencies of each detection history by the total number of sites to obtain the detection probability of that history. These probabilities correspond to the maximum likelihood estimates of the data. The log likelihood is obtained by multiplying the likelihood by the natural log and summed across all histories. When this is multiplied by -2, it becomes -2LogL.

In the case of the multiple seasons, we follow another approach. Here, there are 4 important players (Occupancy, colonization, extinction and detection probability). Here, we make use of a multinomial likelihood function of the parameters.

In theory, each detection history occurs a certain number of times (frequency). The objective is to derive estimates of the parameters that maximize the likelihood of observing the frequencies of individual sites.

Given a detection history, we can multiply the number of times it occurs by the natural log of the probability of observing the history. The estimates are then obtained after maximization of the log likelihood.

The advantage of maximizing likelihood is that the estimates are asymptotically unbiased, normally distributed and the estimators do not vary much with one another.

5.8. MODEL SELECTION

The parameters of the models are the occupancy probability (ψ), the extinction probability (ϵ_t) which is also given as $Pr(\text{absence at time } t + 1 / \text{presence at time } t)$, the colonization probability (γ_t) which is also given as $Pr(\text{presence at time } t + 1 / \text{absence at time } t)$ and the detection probability at a location or km-square P_i .

From the parameters just described, many models can be conceived. A period is used to represent a time-constant function.

1. $\psi(.)\epsilon(.)\gamma(.)p(.)$. This is a model where occupancy, extinction, colonization and detection probabilities are all constant over time.

2. $\psi(\cdot)\varepsilon(\cdot)\gamma(\cdot)p(t)$. In this model, only occupancy, extinction and colonization are constant. Detection probabilities are allowed to change over time.
3. $\psi(\cdot)\varepsilon(\cdot)\gamma(t)p(\cdot)$. Here, all parameters are constant except for colonization which can change over time.
4. $\psi(\cdot)\varepsilon(t)\gamma(\cdot)p(\cdot)$. In this case, only the extinction is allowed to change over time. The other parameters are kept constant.
5. $\psi(t)\varepsilon(\cdot)\gamma(\cdot)p(\cdot)$. Occupancy depends on time while extinction, colonization and detection probability are constant.
6. $\psi(t)\varepsilon(t)\gamma(t)p(t)$. In this model, all parameters are a function of time. Hence they can all change with time.
7. $\psi(\cdot)\varepsilon(t)\gamma(t)p(t)$. Only Occupancy is time independent. All other parameters depend on time.

Each of the models listed above can be parameterized in four different ways. These however do not have an effect on the fit of the model or the parameter estimates. It is therefore at the discretion of the researcher to apply the desired parameterization. These are as follows

1. Initial occupation (ψ_1), local colonization (γ_t), extinction (ε_t) and detection. Hence ψ_1 , γ_t and ε_t are estimated and then used to derive ψ_2 , ψ_3 ... as shown in the expression

$$\psi_{t+1} = \psi_t(1 - \varepsilon_t) + (1 - \psi_t)\gamma_t$$

Thus we have that

$$\psi_2 = \psi_1(1 - \varepsilon_1) + (1 - \psi_1)\gamma_1$$

Since Occupancy is determined by the state of the first primary period, ψ_2, ψ_3 ... do not need to be estimated.

2. Seasonal occupancy (ψ_t) and colonization (γ_t). In this option, ψ_t is estimated directly and we derive ε_t as follows

$$\begin{aligned} \psi_{t+1} &= \psi_t(1 - \varepsilon_t) + (1 - \psi_t)\gamma_t \\ \varepsilon_t &= 1 - \frac{\psi_{t+1} - (1 - \psi_t)\gamma_t}{\psi_t} \end{aligned}$$

3. Seasonal occupancy (ψ_t) and local extinction (ε_t). In this option, ψ_t and ε_t are estimated directly and then γ_t is derived. This is done as follows

$$\begin{aligned} \psi_{t+1} &= \psi_t(1 - \varepsilon_t) + (1 - \psi_t)\gamma_t \\ \gamma_t &= \frac{\psi_{t+1} - (1 - \varepsilon_t)\psi_t}{(1 - \psi_t)} \end{aligned}$$

4. Season occupancy whereby only ψ_t is modelled and local detection is forced to be

$$\varepsilon_t = 1 - \gamma_t$$

In this study we used the first parameterization scheme.

These models are fitted in PRESENCE 2.0 and the best model selected.

Model selection is an integral part of this study. The objective of this exercise is to find a model that best fits the data. The models are fitted based on the species in question. An increasingly applied approach also presented in MacKenzie and Bailey (2004) for analysing models in ecological studies is to fit a range of competing models and then selecting the best fitting model(s) using a suitable technique such as the Akaike's information criterion (AIC). However, it is important to note that the best fitting model does not imply a good model. Importantly, the method is flexible enough to incorporate potential model covariates that may vary across sites. Models will be developed for each species and then compared for best fit using their AICs.

6. APPLICATION TO THE DATA

6.1. BUTTERFLY SPECIES

I. SPECKLED WOOD BUTTERFLY

The analysis was done on 149 sites during 24 sampling occasions (weeks). There were 2814 missing observations (missing in terms of not surveyed). There were no site covariates. The link function used to fit the model is the logistic link function.

The naïve occupancy estimate is 0.8435. This is the proportion of sites where the species (Speckledwood) was detected at least once.

Table 3: Fit statistics of competing occupancy models for Speckled wood Butterfly

Model	AIC	Δ AIC	L	#Par	-2Log L
$\psi, \gamma (\cdot) \varepsilon (\cdot) p(t)$	910.07	0.00	1.0000	27	856.07
$\psi, \gamma (\cdot) \varepsilon(t) p(t)$	913.44	3.37	0.1854	31	851.44
$\psi, \gamma (t) \varepsilon(t) p(\cdot)$	917.89	7.82	0.0163	31	855.89
$\psi, \gamma (t) \varepsilon(\cdot) p(t)$	917.89	7.82	0.0163	31	855.89
$\psi, \gamma (t) \varepsilon(t) p(t)$	921.45	11.38	0.0034	35	851.45
$\psi, \gamma (\cdot) \varepsilon(t) p(t)$	940.38	30.31	0.0000	8	924.38
$\psi, \gamma (\cdot) \varepsilon(\cdot) p(\cdot)$	952.29	42.22	0.0000	4	944.29
$\psi, \gamma (t) \varepsilon(\cdot) p(\cdot)$	960.29	50.22	0.0000	8	944.29

The best fitting model is the $\psi, \gamma (\cdot) \varepsilon (\cdot) p(t)$ whereby the probability of detection is allowed to change over time. This model has 6 primary and 27 sampling periods and 27 sampling parameters.

The colonization remains constant from one season to the next. The extinction rate also stayed constant at 1.72% throughout all seasons. Detection probability depended on the time.

The parameter estimates and standard errors are shown in table 3 below.

Table 4: Parameter estimates and standard errors of the model $\psi, \gamma(\cdot), \epsilon(\cdot), p(t)$

PARAMETER	ESTIMATE	SE(ESTIMATE)	95% CI
$\psi(1)$	0.9638	0.0368	0.8105 - 0.9940
$\psi(2)$	0.9472	0.0287	0.8910 - 1.0034
$\psi(3)$	0.9309	0.03	0.8720 - 0.9897
$\psi(4)$	0.9148	0.0356	0.8451 - 0.9846
$\psi(5)$	0.8991	0.0434	0.8140 - 0.9842
$\psi(6)$	0.8836	0.0523	0.7811 - 0.9861
$\gamma(\cdot)$	0	-	-
$\epsilon(\cdot)$	0.0172	0.013	0.0039 - 0.0730
P[wk1]	0.0909	0.0614	0.0228 - 0.2996
P[wk2]	0.2227	0.0746	0.1098 - 0.3994
P[wk3]	0.5467	0.0925	0.3696 - 0.7128
P[wk4]	0.7445	0.0848	0.5495 - 0.8744
P[wk5]	0.5868	0.1002	0.3881 - 0.7607
P[wk6]	0.4378	0.0866	0.2821 - 0.6067
P[wk7]	0.3901	0.0722	0.2615 - 0.5360
P[wk8]	0.4121	0.0678	0.2894 - 0.5468
P[wk9]	0.1044	0.099	0.0144 - 0.4816
P[wk10]	0.377	0.152	0.1454 - 0.6827
P[wk11]	0.1129	0.0755	0.0282 - 0.3579
P[wk12]	0.0769	0.0523	0.0193 - 0.2610
P[wk13]	0.4057	0.0847	0.2556 - 0.5757
P[wk14]	0.4103	0.0977	0.2407 - 0.6044
P[wk15]	0.4541	0.0835	0.3014 - 0.6159
P[wk16]	0.582	0.0777	0.4276 - 0.7219
P[wk17]	0.7161	0.0789	0.5446 - 0.8417
P[wk18]	0.445	0.1014	0.2641 - 0.6418
P[wk19]	0.3027	0.1153	0.1303 - 0.5572
P[wk20]	0.3835	0.0934	0.2239 - 0.5730
P[wk21]	0.1634	0.1068	0.0406 - 0.4744
P[wk22]	0.567	0.1579	0.2721 - 0.8211
P[wk23]	0.3362	0.1156	0.1551 - 0.5828
P[wk24]	0.2826	0.1218	0.1085 - 0.5603

The derived parameters represent the probability that a site is occupied given its detection history. The derived parameters are shown in the table below

Since there were no covariates in the model or under investigation, all sites have a $\psi(1)$ estimate of 0.9638. Likewise all sites have a $\psi(2)$ estimate of 0.9472 and $\psi(3)$ equal to 0.9309.

The detection probabilities across the sampling periods vary as shown in the figure below. We see that the first week has a very low detection probability (0.09). This rapidly increases to 0.74 at the end of the month (April). The month of May experiences a steady decline and the probabilities remain low in the month of June. Another peak season is July where the detection probabilities increase before falling again in August.

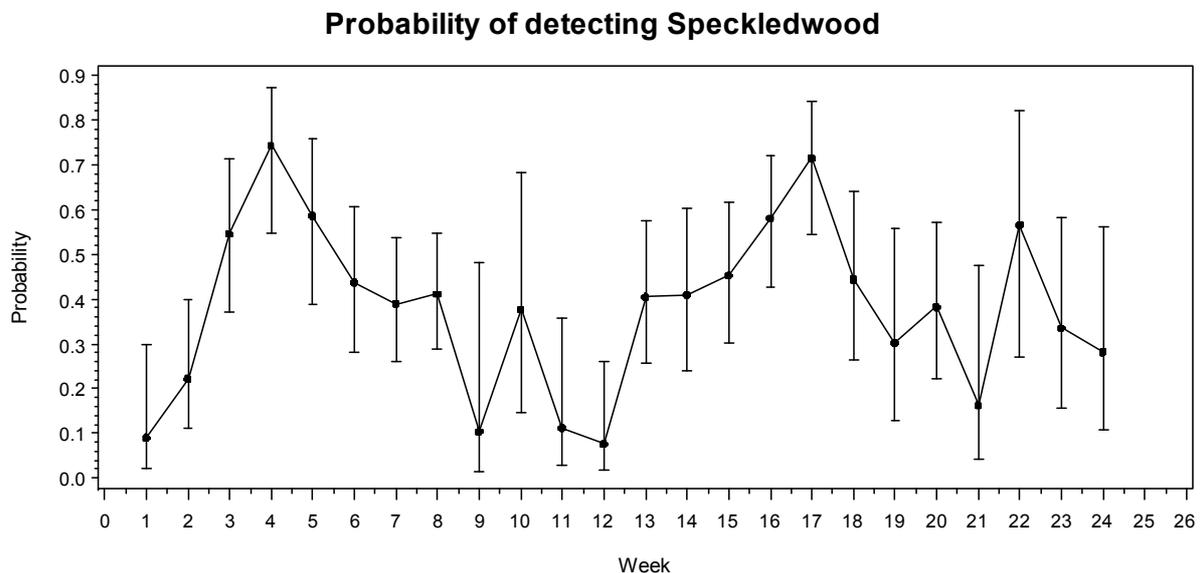


Figure 7: Probability of detecting Speckled wood butterfly (April-August 2009)

This relates closely to the two generations of butterfly this species has: a first generation emerges from the pupae early April, lays eggs and dies off in May. The eggs hatch, caterpillars grow and the second generation of butterflies emerges primarily in July. There is also a partial third generation late August – September.

II. SWALLOWTAIL BUTTERFLY

The analysis was done on 150 sites each with 24 sampling occasions (weeks) and 6 seasons with each season having 4 sampling occasions. 3863 observations were missing (missing because no survey was carried out). The naïve occupancy estimate is 0.2400.

Table 5: Fit statistics of competing occupancy models for Swallow tail butterfly

Model	AIC	Δ AIC	L	# Par	-2Log L
$\psi, \gamma(t)\varepsilon(\cdot)p(\cdot)$	367.43	0.00	1.0000	8	351.43
$\psi, \gamma(\cdot)\varepsilon(t)p(\cdot)$	367.99	0.56	0.7558	8	351.99
$\psi, \gamma(\cdot)\varepsilon(\cdot)p(\cdot)$	368.52	1.09	0.5798	4	360.52
$\psi, \gamma(t)\varepsilon(t)p(\cdot)$	370.63	3.20	0.2019	12	346.63
$\psi, \gamma(\cdot)\varepsilon(\cdot)p(t)$	380.97	13.54	0.0011	27	326.97
$\psi, \gamma(\cdot)\varepsilon(t)p(t)$	385.24	17.81	0.0001	31	323.24
$\psi, \gamma(t)\varepsilon(\cdot)p(t)$	386.59	19.16	0.0001	31	324.59
$\psi, \gamma(t)\varepsilon(t)p(t)$	390.84	23.41	0.0000	35	320.84

Even though the first model has the least AIC (367.43), it was not selected as the best model because the parameter estimates are very unstable with very large standard errors. The second model has approximately the same AIC as the first with much lower standard errors. Hence the model with the best fit for the swallowtail butterfly is $\psi, \gamma(\cdot)\varepsilon(t)p(\cdot)$. The colonization rate and the detection rate are constant while the extinction rate was depended on time.

Table 6: Parameter estimates of the model $\psi, \gamma(\cdot)\varepsilon(t)p(\cdot)$.

Parameter	Estimate	SE(estimate)	95% CI
$\psi(1)$	0.2564	0.1188	0.0923 - 0.5390
$\psi(2)$	0.3180	0.1034	0.1152 - 0.5207
$\psi(3)$	0.3596	0.1295	0.1057 - 0.6136
$\psi(4)$	0.4500	0.1201	0.2146 - 0.6853
$\psi(5)$	0.3910	0.1443	0.1082 - 0.6738
$\psi(6)$	0.0859	0.0449	-0.0022 - 0.1740
$\gamma(\cdot)$	0.1411	0.0768	0.0455 - 0.3612
$\varepsilon(1)$	0.169	0.2716	0.0053 - 0.8858
$\varepsilon(2)$	0.1715	0.4161	0.0007 - 0.9844
$\varepsilon(3)$	0	-	-
$\varepsilon(4)$	0.3035	0.2761	0.0346 - 0.8413
$\varepsilon(5)$	1	0.0557	0.0000 - 1.0000
P[wk1-wk24]	0.203	0.0524	0.1197 - 0.3230

The occupancy rates change over the seasons. Instead of a naïve and constant rate of 24%, we see that the occupancy ranges from 25.64% in the first season (April) to 8.59% in the last season (August). The derived parameters $\psi(2) - \psi(6)$ are given in table 6 above

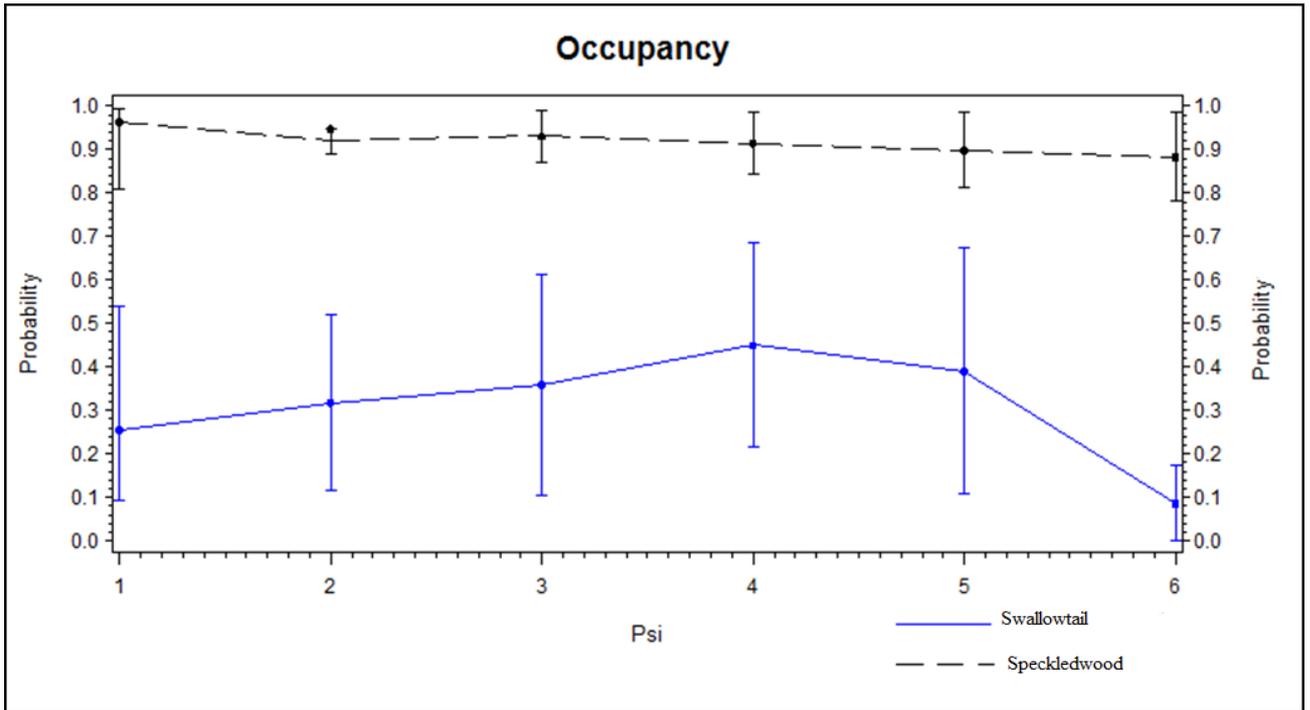


Figure 8: Occupancy of speckled wood and swallowtail butterflies

From the figure above, the occupancy of the speckled wood butterfly in each season is relatively higher than that of the swallowtail butterfly. That of the speckled wood butterfly falls very slightly from 0.96 in the first season to 0.88 in the last season. For the Swallowtail species, the occupancy increases from 0.26 to a maximum point at 0.45 (season 4) and then falls to 0.08 in the last season.

There is a constant probability of detection throughout the seasons. Even though Swallowtail also has two generations, the difference in between them is very small, possibly because of the lower occupancy rates which could not be captured in the modelling.

6.2. BIRD SPECIES

I. COMMON BUZARDS

For the year 2009, the following models were fitted.

Table 7: Fit statistics of plausible models in 2009

Model	AIC	Δ AIC	L	# Par	-2LogL
$\psi, \gamma (\cdot) \varepsilon (\cdot) p(t)$	20628.35	0.00	1.0000	23	20582.35
$\psi, \gamma (\cdot) \varepsilon(t) p(t)$	20630.22	1.87	0.7558	26	20578.22
$\psi, \gamma (t) \varepsilon(\cdot) p(t)$	20631.87	3.52	0.5798	26	20579.87
$\psi, \gamma (t) \varepsilon(t) p(t)$	20632.13	3.78	0.2019	29	20574.13
$\psi, \gamma (\cdot) \varepsilon(\cdot) p(\cdot)$	20659.71	31.36	0.0011	4	20651.71
$\psi, \gamma (t) \varepsilon(\cdot) p(\cdot)$	20661.66	33.31	0.0001	7	20647.66
$\psi, \gamma (\cdot) \varepsilon(t) p(\cdot)$	20662.35	34.00	0.0001	7	20648.35
$\psi, \gamma (t) \varepsilon(t) p(\cdot)$	20713.34	84.99	0.0000	10	20693.34

The best model from the table above is $\psi, \gamma (\cdot) \varepsilon (\cdot) p(t)$. In this model, the colonization probability and probability of detection depends on the time point. Extinction probability is constant.

There were 8274 sites involved with each having 20 sampling occasions. There were 5 seasons each with 4 sapling periods. The naïve occupancy was 0.2244.

Table 8: Parameter estimates of the model $\psi, \gamma (\cdot) \varepsilon (\cdot) p(t)$.

Parameter	Estimate	SE	95% CI
$\psi(1)$	0.4332	0.0206	0.3934 - 0.4739
$\psi(2)$	0.3665	0.0121	0.3428 - 0.3902
$\psi(3)$	0.3246	0.0121	0.3009 - 0.3484
$\psi(4)$	0.2984	0.0143	0.2703 - 0.3265
$\psi(5)$	0.282	0.0162	0.2502 - 0.3138
$\gamma(\cdot)$	0.0948	0.0101	0.0769 - 0.1165
$\varepsilon(\cdot)$	0.2781	0.0211	0.2386 - 0.3214
P(wk1)	1	-	-
P(wk2)	0.2943	0.0252	0.2474 - 0.3461
P(wk3)	0.3504	0.0209	0.3107 - 0.3924
P(wk4)	0.3399	0.0208	0.3005 - 0.3817
P(wk5)	0.3214	0.0172	0.2887 - 0.3560
P(wk6)	0.315	0.0171	0.2826 - 0.3493
P(wk7)	0.2625	0.0165	0.2315 - 0.2962
P(wk8)	0.2474	0.0155	0.2183 - 0.2790

Table 8 Cont.

Parameter	Estimate	SE	95% CI
P(wk9)	0.2321	0.0166	0.2012 - 0.2662
P(wk10)	0.2397	0.0173	0.2074 - 0.2752
P(wk11)	0.2521	0.0181	0.2182 - 0.2892
P(wk12)	0.2656	0.0178	0.2322 - 0.3019
P(wk13)	0.2694	0.0221	0.2284 - 0.3147
P(wk14)	0.2728	0.0236	0.2290 - 0.3215
P(wk15)	0.2469	0.022	0.2063 - 0.2925
P(wk16)	0.2065	0.0192	0.1713 - 0.2468
P(wk17)	0.2096	0.0242	0.1661 - 0.2608
P(wk18)	0.3153	0.03	0.2597 - 0.3768
P(wk19)	0.5	7905694150	0.0000 - 1.0000
P(wk20)	0.5	7905694150	0.0000 - 1.0000

For the year 2010, the following models were fitted

Table 9: Fit statistics of plausible models in 2010

Model	AIC	Δ AIC	L	# Par	-2LogL
$\psi, \gamma (.)\varepsilon(.)p(t)$	26659.79	0.00	1.0000	23	26613.79
$\psi, \gamma (t)\varepsilon(.)p(t)$	26662.00	2.21	0.3312	26	26610.00
$\psi, \gamma (.)\varepsilon(t)p(t)$	26662.21	2.42	0.2982	26	26610.21
$\psi, \gamma (t)\varepsilon(t)p(t)$	26665.90	6.11	0.0471	29	26607.90
$\psi, \gamma (t)\varepsilon(t)p(.)$	26728.35	68.56	0.0000	10	26708.35
$\psi, \gamma (.)\varepsilon(t)p(.)$	27124.57	464.78	0.0000	7	27110.57
$\psi, \gamma (.)\varepsilon(.)p(.)$	27174.67	514.88	0.0000	4	27166.67
$\psi, \gamma (t)\varepsilon(.)p(.)$	27252.25	592.46	0.0000	7	27238.25

The best model for the common buzzards in 2010 was the $\psi, \gamma (.)\varepsilon(.)p(t)$ whereby only the detection probability depended on time. The total number sites included is 9164 and the naïve occupancy estimate is 0.2548

Table 90: Parameter estimates of the model $\psi, \gamma(\cdot)\varepsilon(\cdot)p(t)$

Parameter	Estimate	SE	95% CI
$\psi(1)$	0.4883	0.0202	0.4489 - 0.5279
$\psi(2)$	0.4167	0.0120	0.3932 - 0.4402
$\psi(3)$	0.3694	0.0114	0.3471 - 0.3917
$\psi(4)$	0.3382	0.0135	0.3116 - 0.3647
$\psi(5)$	0.3176	0.0157	0.2867 - 0.3484
$\gamma(\cdot)$	0.0943	0.0103	0.0760 - 0.1164
$\varepsilon(\cdot)$	0.2455	0.0171	0.2135 - 0.2805
P(wk1)	0	0.0001	0.0000 - 1.0000
P(wk2)	0.2526	0.0251	0.2066 - 0.3049
P(wk3)	0.3158	0.0168	0.2839 - 0.3495
P(wk4)	0.2889	0.0154	0.2597 - 0.3201
P(wk5)	0.2968	0.0142	0.2696 - 0.3254
P(wk6)	0.293	0.0144	0.2656 - 0.3219
P(wk7)	0.2534	0.0135	0.2278 - 0.2809
P(wk8)	0.2174	0.0121	0.1947 - 0.2420
P(wk9)	0.2032	0.0136	0.1778 - 0.2311
P(wk10)	0.2214	0.0143	0.1947 - 0.2507
P(wk11)	0.2705	0.0152	0.2418 - 0.3012
P(wk12)	0.2708	0.0148	0.2429 - 0.3007
P(wk13)	0.2605	0.0176	0.2276 - 0.2964
P(wk14)	0.2117	0.0172	0.1800 - 0.2474
P(wk15)	0.2366	0.0183	0.2027 - 0.2742
P(wk16)	0.2472	0.0179	0.2139 - 0.2838
P(wk17)	0.2692	0.0224	0.2276 - 0.3153
P(wk18)	0.2599	0.0234	0.2168 - 0.3083
P(wk19)	0.5	7905694150	0.0000 - 1.0000
P(wk20)	0.5	7905694150	0.0000 - 1.0000

The plot below shows that the occupancy estimates in 2009 are consistently lower than those in 2010. Here, Occupancy probabilities in during each year are highest in the first seasons and least at the last seasons.

Occupancy of Common Buzzards

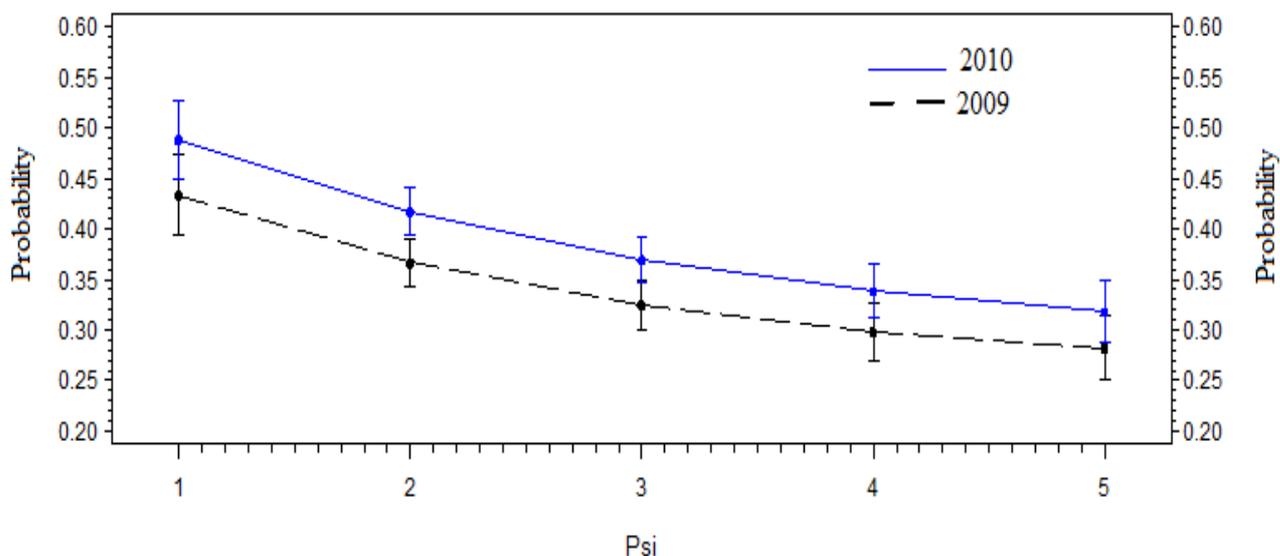


Figure 9: Occupancy probabilities of Common Buzzards in 2009 and 2010

Probability of detecting Common Buzzards

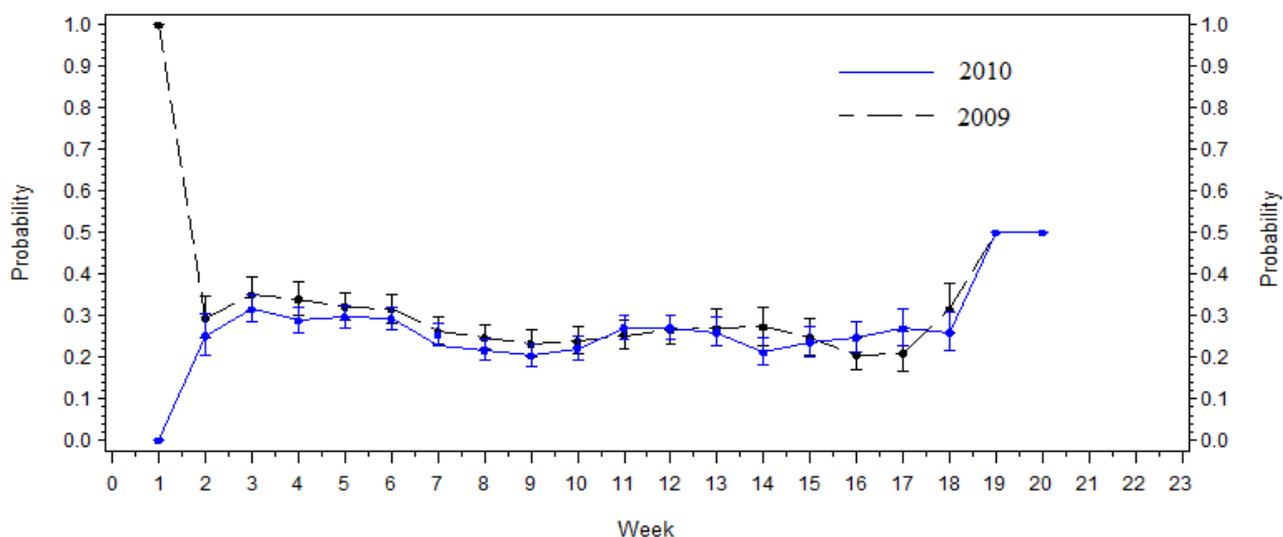


Figure 10: Detection probabilities of Common Buzzards in 2009 and 2010

The difference between the years is an artefact of roving records. More and more people are realising the value of the portal data and more and more enter more records of common species (Herremans 2009), while initially they would only report rare observations. This also applies to raptors like Common buzzard (Herremans 2010).

The seasonal fluctuation refers primarily to behavioural changes: in early spring buzzards defend territories and have a conspicuous aerial display. Later, during breeding they become more silent and retiring and they hide more efficiently because the deciduous trees are in leaf.

When the young leave the nest they loudly beg for food easily drawing attention of observers. And at the end of the season migration starts, with more birds becoming more obvious in the sky.

II. BLACK WOOD PECKERS

The models fitted for the 2009 data are shown in table 11 below

Table 11: Fit statistics of plausible models of Black woodpeckers in 2009

Model	AIC	Δ AIC	L	# Par	-2Log L
$\psi, \gamma (t)\varepsilon(t)p(\cdot)$	5814.45	0.00	1.0000	10	5794.45
$\psi, \gamma (t)\varepsilon(\cdot)p(t)$	5817.80	3.35	0.1873	26	5765.80
$\psi, \gamma (t)\varepsilon(t)p(t)$	5819.10	4.65	0.0978	29	5761.10
$\psi, \gamma (\cdot)\varepsilon(\cdot)p(t)$	5820.72	6.27	0.0435	23	5774.72
$\psi, \gamma (\cdot)\varepsilon(t)p(t)$	5820.75	6.30	0.0429	26	5768.75
$\psi, \gamma (t)\varepsilon(\cdot)p(\cdot)$	5822.55	8.10	0.0174	7	5808.55
$\psi, \gamma (\cdot)\varepsilon(t)p(\cdot)$	5841.81	27.36	0.0000	7	5827.81
$\psi, \gamma (\cdot)\varepsilon(\cdot)p(\cdot)$	5844.17	29.72	0.0000	4	5836.17

The best fitting model is $\psi, \gamma (t)\varepsilon(t)p(\cdot)$. There are 8274 sites in total involved. Also there are 5 primary periods with 4 sampling periods in each. The model estimates 10 parameters. In the table below, we see that there are 14 parameters. This is because $\psi(2) - \psi(5)$ are derived parameters. The naïve occupancy estimate is 0.0518

Table 12: Parameter estimates of the model $\psi, \gamma(t)\varepsilon(t)p(\cdot)$

Parameter	Estimate	SE(estimate)	95% CI
$\psi(1)$	0.1291	0.0115	0.1082-0.1534
$\psi(2)$	0.1139	0.0104	0.0935 - 0.1344
$\psi(3)$	0.1021	0.0098	0.0828 - 0.1213
$\psi(4)$	0.0609	0.0079	0.0453 - 0.0765
$\psi(5)$	0.0601	0.0117	0.0371 - 0.0831
$\gamma(1)$	0	-	-
$\gamma(2)$	0.0411	0.0089	0.0268 - 0.0624
$\gamma(3)$	0.0048	0.0079	0.0002 - 0.1096
$\gamma(4)$	0.0285	0.0116	0.0127 - 0.0625
$\varepsilon(1)$	0.1177	0.0726	0.0328 - 0.3443
$\varepsilon(2)$	0.4234	0.0681	0.2988 - 0.5586
$\varepsilon(3)$	0.4454	0.0827	0.2944 - 0.6073
$\varepsilon(4)$	0.4523	0.1766	0.1696 - 0.7695
P[Wk1] - P[Wk20]	0.187	0.0134	0.1628 - 0.2138

The following models were fitted for the Black Woodpeckers in 2010

Table 13: Fit statistics of plausible models for Black woodpeckers in 2010

Model	AIC	Δ AIC	L	# Par	-2Log L
$\psi, \gamma(t)\varepsilon(t)p(t)$	6962.94	0.00	1.0000	29	6904.94
$\psi, \gamma(\cdot)\varepsilon(t)p(t)$	6966.28	3.34	0.1882	26	6914.28
$\psi, \gamma(t)\varepsilon(t)p(\cdot)$	6968.30	5.36	0.0686	10	6948.30
$\psi, \gamma(\cdot)\varepsilon(\cdot)p(\cdot)$	6974.52	11.58	0.0031	4	6966.52
$\psi, \gamma(\cdot)\varepsilon(t)p(\cdot)$	6998.29	35.35	0.0000	7	6984.29
$\psi, \gamma(\cdot)\varepsilon(\cdot)p(t)$	700838	45.44	0.0000	23	6962.38
$\psi, \gamma(t)\varepsilon(\cdot)p(t)$	7481.35	518.41	0.0000	26	7429.35
$\psi, \gamma(t)\varepsilon(\cdot)p(\cdot)$	7889.18	928.24	0.0000	7	7875.18

The best model is the $\psi, \gamma(t)\varepsilon(t)p(t)$. The model estimates 29 parameters and is not over parameterized because the number of possible histories is much bigger than 29. Again, $\psi(2)$ - $\psi(5)$ are derived parameters. There are 9164 sites involved, each of which has 20 sampling occasions. 143311 observations were recorded as missing (not surveyed) and the naïve occupancy estimate is 0.0512

Table 14: Parameter estimates of the model $\psi, \gamma(t)\varepsilon(t)p(t)$.

Parameter	Estimate	SE	95% CI
$\psi(1)$	0.0932	0.0079	0.0788 - 0.1098
$\psi(2)$	0.0761	0.0068	0.0628 - 0.0894
$\psi(3)$	0.0954	0.0093	0.0772 - 0.1136
$\psi(4)$	0.0882	0.0146	0.0597 - 0.1168
$\psi(5)$	0.0547	0.0259	0.0039 - 0.1055
$\gamma(1)$	0.0000	0.0000	0.0000 - 1.0000
$\gamma(2)$	0.0347	0.0076	0.0225 - 0.0531
$\gamma(3)$	0.0189	0.0084	0.0079 - 0.0448
$\gamma(4)$	0.0008	0.0072	0.0000 - 1.0000
$\varepsilon(1)$	0.1831	0.0656	0.0866 - 0.3463
$\varepsilon(2)$	0.1676	0.0771	0.0638 - 0.3731
$\varepsilon(3)$	0.2544	0.1262	0.0847 - 0.5570
$\varepsilon(4)$	0.388	0.2994	0.0508 - 0.8824
P(1)wk1	1	0	0.0000 - 1.0000
P(2)wk2	0.0935	0.0373	0.0418 - 0.1963
P(3)wk3	0.2633	0.0322	0.2051 - 0.3311
P(4)wk4	0.2393	0.0287	0.1876 - 0.2999
P(5)wk5	0.2586	0.0337	0.1981 - 0.3299
P(6)wk6	0.2092	0.0298	0.1567 - 0.2736
P(7)wk7	0.2292	0.0308	0.1744 - 0.2950
P(8)wk8	0.2738	0.0335	0.2133 - 0.3440
P(9)wk9	0.2376	0.0334	0.1784 - 0.3090
P(10)wk10	0.2207	0.0315	0.1651 - 0.2885
P(11)wk11	0.2302	0.0309	0.1753 - 0.2961
P(12)wk12	0.2021	0.0283	0.1523 - 0.2631
P(13)wk13	0.2059	0.0395	0.1390 - 0.2939
P(14)wk14	0.1393	0.0327	0.0866 - 0.2166
P(15)wk15	0.1005	0.0266	0.0590 - 0.1658
P(16)wk16	0.1487	0.0331	0.0947 - 0.2258
P(17)wk17	0.2831	0.136	0.0960 - 0.5950
P(18)wk18	0.1698	0.0873	0.0573 - 0.4078
P(19)wk19	0.5	7905694150	0.0000 - 1.0000
P(20)wk20	0.5	7905694150	0.0000 - 1.0000

In order to understand the occupancy trends in 2009 and 2010, the probabilities have been plotted against the occupancy status. Occupancy in 2009 from $\psi(1)$ to $\psi(3)$ appears higher than in 2010. Then the trend reverses. Generally, we observe a reduction in both years of the occupancy estimates over the seasons.

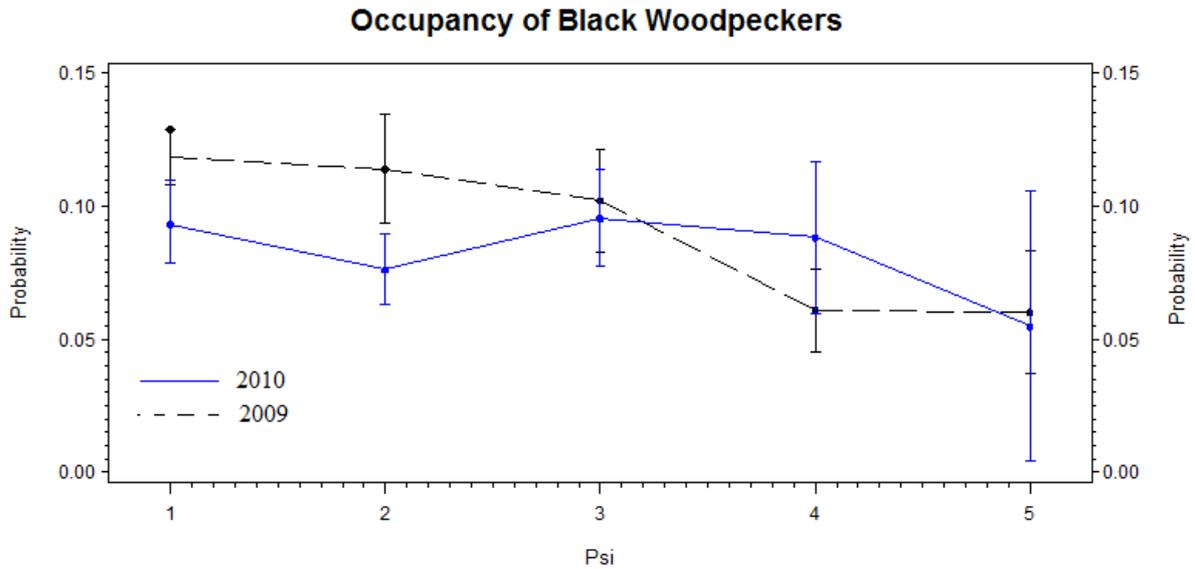


Figure 11: Occupancy probabilities of Black woodpeckers in 2009 and 2010

The species is somewhat sensitive to severe winters, which may explain the lower numbers early 2010. Based on the number of records relative to search effort, Herremans (2009) indicated a decline of some 15% from 2009 to 2010.

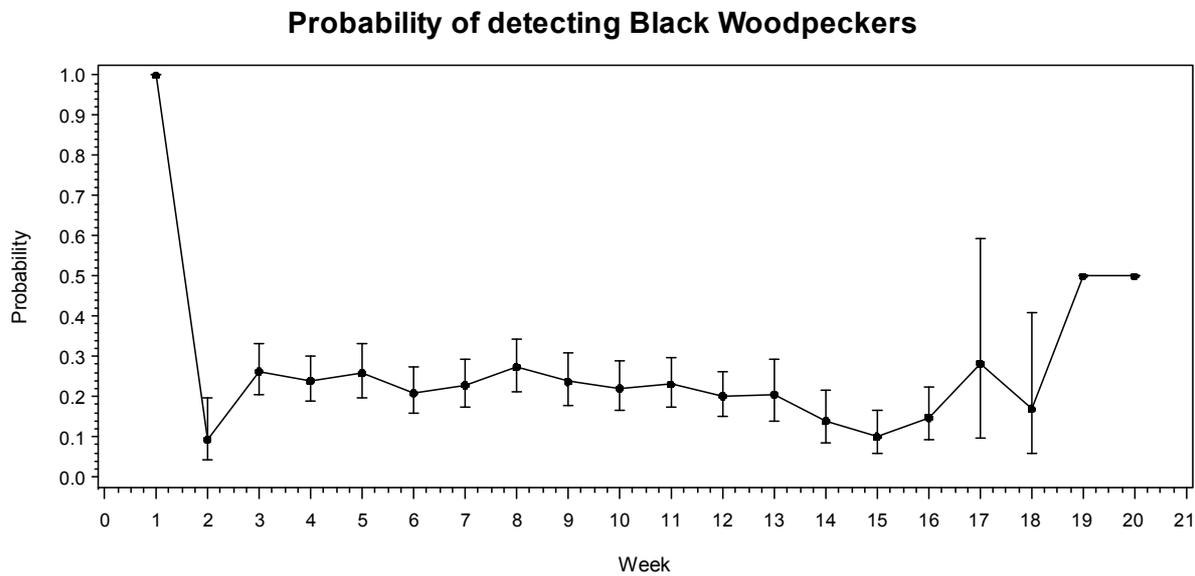


Figure 12: Detection probabilities of Black woodpeckers in 2010

Black woodpeckers defend territories with far-carrying vocalisations, particularly in early spring. They have fairly long breeding seasons, but the vocalisations gradually decline. In midsummer the species is quieter. The detecting probabilities follow this pattern.

7. DISCUSSION

The problems we face nowadays are growing and concern is also mounting about the rate of loss of biodiversity if nothing is done. The realisation of this rapid decline and loss of biodiversity has prompted policy makers to commit to different measures to reduce or stop the trend. This commitment is translated into strategies for quantifying the amount of biodiversity and modelling the life cycles of species.

One of the major hurdles in estimating of the number of species is that data is more available as presence-only data. This involves only information about the species that are seen or detected by the observer. Species that are not detected could be genuinely absent or present but not detected. Detectability is therefore imperfect and detection probabilities are most often less than 1.

A natural strategy would be to estimate the abundance of each species and over time. However, this is too costly and difficult to perform. Recent methodologies of MacKenzie *et al.* (2000), MacKenzie *et al.* (2003), MacKenzie *et al.* (2004), Royle *et al.* (2005), Royle *et al.* (2007) use occupancy as a measure of true abundance. Abundance –occupancy relationships are usually positive. Species declining in abundance also, in most cases, show declines in the number of sites they occupy. This relationship is well explored in Gaston *et al.* (2000).

Each species was monitored for detected/non-detected information. Each site was visited repeatedly and the number of the target species recorded. For the butterfly species (Speckled wood and Swallowtail) the period of observation ranged from April to September 2009 while for the two bird species (Common Buzzards and Black Woodpeckers) the period was between March and July of 2009 and 2010. This trend is also reflected in the statistics of these species waarneming.be.

As a bias reducing measure, a hierarchical statistical design was used. This involved grouping the data by weeks as primary sampling units and then by months (also used as seasons) as secondary sampling units. This method is very efficient the imposed hierarchy reduces variance and increases precision of parameter estimates. This is the multi-seasonal representation.

Site occupancy models were fitted for each species taking into account the four dynamic processes of occupancy, colonization, extinction and detectability. For each species, the best fitting model was selected from a variety of competing models using the AIC criterion.

For the Speckled wood butterfly, best model was $\psi, \gamma(.)\varepsilon(.)p(t)$. This means that the detection probability depends on time, meanwhile colonization and extinction are time constant. For the Swallowtail butterfly, the best model was $\psi, \gamma(.)\varepsilon(t)p(.)$ in which only the extinction probability depended on time. The occupancy of Speckled wood starts out high in the first season and stays high throughout the seasons. That of Swallow tail is lower comparatively. This is probably due to the fact that the swallow tail species is rare.

In the case of the Common Buzzards, the models were fitted based on the year. The best model for 2009 was $\psi, \gamma(.)\varepsilon(.)p(t)$ and that for 2010 was $\psi, \gamma(.)\varepsilon(.)p(t)$. We observe,

in both cases, that probability of detection varies with time. There is therefore a slight increase in detection probability for both years. The occupancy increased from 2009 to 2010. The observation shows that more detection cases are being reported.

For the Black woodpeckers, the best model was $\psi, \gamma(t)\varepsilon(t)p(\cdot)$ while for 2010, the best model was $\psi, \gamma(t)\varepsilon(t)p(t)$. We see that in 2010, colonization, extinction and detectability all depended on time whereas in 2009, only colonization and extinction were time independent. Detectability was time constant.

8. CONCLUSION

For the speckled wood butterfly, the proportion of sites occupied reduces slightly from 0.96 in the first season to 0.88 in the last season. This corresponds to an 8% decrease in number of sites occupied. For the swallow tail butterfly, there is first of all an increase of 28% of the number of sites occupied, then decreases rapidly from July to September. There is an overall decrease of 18% in occupancy. At any given month, the occupancy of speckled wood is significantly greater than that of swallow tail. For the common buzzards 43% of sites were occupied in March of 2009. By July of the same year, this proportion dropped to 28% hence a reduction of 15%. In the following year, the proportion ranges from 49% in March to 32% in July hence an overall reduction of 17%. At any given month, the occupancy of common buzzards is higher in 2010 than in 2009. For the black woodpeckers, occupancy ranges from 13% in March to 6% in July in 2009 and from 9% in March to 5% in July of 2010. These correspond to a decrease of 7% in 2009 and 4% in 2010.

Considering colonization, virtually no sites were colonized by the speckled wood species. On the contrary, 14% of the sites were colonized by swallowtail species every month. About 9% of the sites were colonized by the common buzzards every month from March to July in 2009 and 2010. The proportion of sites occupied by black woodpeckers depended on the month in both 2009 and 2010. It ranges from 0% to 3% in 2009 and from 0% to 0.08% in 2010.

In terms of extinction, the proportion of sites in which the speckled wood species became locally extinct was 0.017. This value was constant from April to September. The swallow tail species was locally extinct in 17% of the sites between April and May. The extinction spread to almost all the sites between August and September. In 28% of the sites, the common buzzards became locally extinct and this remained constant throughout March to July of 2009. This reduced to 25 % in 2010. Black woodpeckers showed in different trend from common buzzards. In 2009, extinction increased from 12% at the beginning of the season to 45% at the end. In the year 2010, the range was between 18% at the beginning and 39% at the end.

The detectability of speckled wood increased from 9% at the beginning of April to 74% by the end of the month. It then reduced to 9% at the end of June. Another steady increase leads to 72% detectability in August. Another small peak appears in September. For the swallow tail species, there was a constant detectability of 20%. For the common buzzards, the detectability was most of the time around 30% in 2009 as well as 2010. For the black woodpeckers, the detectability was constant in 2009 at 19%. In 2010, the detectability varied slightly but most of the time was around 20% but increased to 50% in September.

9. RECOMMENDATIONS

In this study, we have been able to understand occupancy, colonization, extinction and detection probabilities of four species common to the European fauna. In order to get a full scale understanding of these factors, it would be necessary to extend the time frame and by implication more data over many years.

Also, this study assumes homogeneity of sites to reduce the bias of the estimates. Even though models have been developed to account for heterogeneity of sites, application remains a major challenge. Future studies on the subject should be able to include estimation of these parameters under heterogeneity.

10. REFERENCES

1. Betts, M., Rodenhouse, N., Sillett, S., Doran, P. and Holmes, R. (2008) Dynamic occupancy models reveal within-breeding season movement up a habitat quality gradient by migratory songbird. *Ecography* **31**, 592-600.
2. Chakraborty, A., Gelfand, A., Wilson, A., Latimer, A. and Silander J. (2011) Poing pattern modelling for degraded presence-only data over large regions to appear. *Journal of Royal Society, Series C*.
3. Elith, J., Graham, C., Anderson, R., Dudk, M., Ferrier, S., Guisan, A., Hijmans, R., Huettmann, F., Leathwick, J., Lehmann, A., Li, J., Lohmann, L., Loiselle, B., Manion, G., Moritz, C., Nakamura, M., Nakazawa, Y., Overton, J., Peterson, A., Phillips, S., Richardson, K., Scachetti-Pereira, R., Schapire, R., Sober'on, J., Williams, S., Wisz, M., and Zimmermann, N. (2006) Novel methods improve prediction of species' distributions from occurrence data. *Ecography*, **29**, 129–151.
4. European Commission. (2011) communication from the commission to the European parliament, the council, the economic and social committee and the committee of the regions our life insurance, our natural capital: an EU biodiversity strategy to 2020. *Com 244*.
5. Gaston, K., Blackburn, T., Greenwood, J., Gregory, R., Quinn, R. and Lawton, J. (2000) Abundance-occupancy relationships. *Journal of Applied Ecology*, **37**, 39-59.
6. Gelfand, A., Schmidt, A., Wu, S. and Rebelo, A. (2003) Modelling species diversity through level hierarchical modelling. *Appl. Statist*, **54**, 1-20
7. Guisan, A. and Zimmermann, N. (2000) predictive habitat distribution models in ecology. *Ecological Modelling*, **135**, 147-186.
8. Herremans, M. (2009) Monitoring using waarnemingen.be. *Natuur.oriolus* **76(3)**, 94-108.
9. Herremans, M. (2010) Is the Kestrel hovering on the brink ? *Natuur.oriolus* **77(2)**, 60-67.
10. Keating K., and Cherry S. (2004) Use and interpretation of logistic regression in habitat-selection studies. *Journal of Wildlife Management*, **68(4)**, 774-789.
11. Kéry, M., Royle, A. and Schmid, H. (2005) Modelling avian abundance from replicated counts using binomial mixture models. *Ecological Applications*, **15(4)**, 1450-1461.
12. Kéry, M., Royle, A. and Schmid, H. (2006) Estimating species richness: calibrating a large avian monitoring program. *Journal of applied ecology*, **43**, 101-110.
13. Kéry, M., Royle, A., Schmid, H., Volet, B., Hafliger, G. and Zbinden N. (2009) Site-occupancy distribution to correct population-trend derived from opportunistic observations. *Conservational Biology*, 1388-1397.
14. MacKenzie, D and Bailey, L. (2004) Assessing the fit of site-occupancy models. *Journal of Agricultural, biological and Environmental Statistics*, **9(3)**, 300-318.
15. MacKenzie, D. (2005) Was it there? Dealing with imperfect detection for species presence/absence data. *Aust. N.Z.J. Stat.* **47(1)**, 65-74.
16. MacKenzie, D., Kendall, W. (2002) How should detection probability be incorporated into estimates of relative abundance. *Ecology*, **83(9)**, 2387-2393.

17. MacKenzie, D., Nicholas, J., Hines, J., Knutson, M., Franklin, A. (2003) Estimating site occupancy, colonization and local extinction when a species is detection imperfectly. *Ecology*, **84(8)**, 2200-2207.
18. MacKenzie, D., Nichols, J., Lachman, G., Droege, S., Royle, A. and Langtimm, C. (2002) Estimating site occupancy when detection probabilities are less than one. *Ecology*, **83**, 2248-2255.
19. Nazeri, M., Jusoff, K., Bahaman, A. and Madani, N. (2010) Modelling the potential distribution of wildlife species in the tropics. *World Journal of Zoology*, **5(3)**, 225-231.
20. Niemela, J. (2000) Biodiversity monitoring for decision making. *Ann. Zool. Fennici*, **37**, 307-317.
21. Ottaviani, D., Lasinio, G. and Boitani, L. (2004) Two statistical methods to validate habitat suitability models using presence-only data. *Ecological Modelling*, **179**, 417-443.
22. Pearce, J. and Boyce M. (2003) Modelling distribution and abundance with presence-only data. *Journal of Applied Ecology*, **43**, 405-412.
23. Royle, A. and Nichols, J. (2003) Estimating abundance from repeated presence-absence data or point counts. *The Ecological Society of America*, **84(3)**, 777-790.
24. Royle, A. and Kéry, M. (2007) A bayesian state-space formulation of dynamic occupancy models. *Ecology*, **88(7)**, 1813-1823.
25. Royle, A., Nicholas, D. and Kéry, M. (2005) Modelling occurrence and abundance of species when detection is imperfect. *Oikos*, **110**, 353-359.
26. Stork, N. and Samways, M. (1995) Inventory and monitoring. *Global Biodiversity Assessment*, 453-543.
27. Warton, D. and Shepherd, L. (2010) Poisson point process models solve the “pseudo-absence problem” for presence-only data in ecology. *The Annals of Applied Statistics*, **4(3)**, 1383-1402
28. Wintle, B., McCarthy, M., Parris, K. and Burgman, M. (2004) Precision and bias of methods for estimating point survey detection probabilities. *Ecological Applications* **14(3)**, 703-712
29. Zaniwski, E., Lehmann, A., and Overton, J. (2002) Predicting species spatial distributions using presence-only data: a case study of native New Zealand ferns. *Ecological Modelling*, **157**, 261-280.

30. Dorazio, R., Gotelli, N. and Ellison, A. (2011) *Modern Methods of Estimating Biodiversity from Presence-Absence Surveys*. Biodiversity / Book 4, O. Grillo (ed.)
31. MacKenzie, D., Nichols J., Royle A., Pollock, K., Bailey L., and Hines J. (2006) *Occupancy Estimation and Modelling: Inferring Patterns and Dynamics of Species Occurrence*. Elsevier Inc.

32. Donovan, T. and J. Hines. (2007) Exercises in occupancy modelling and estimation. <http://www.uvm.edu/envnr/vtcfwru/spreadsheets/occupancy.htm>

- (Accessed on 19/07/2011)
33. 27. World Resource Institute
http://archive.wri.org/page.cfm?id=535&page=pubs_content_text
(Accessed on 19/08/2011)
 34. World Resource Institute
 35. <http://www.conservation.org/learn/biodiversity/Pages/overview.aspx>
(Accessed on 25/07/2011)
 36. European Commission 2008: The Global Cost of Biodiversity Loss: 14 Trillion Euros? *Science for Environmental Policy*
 37. World Wide Fund for Nature
http://wwf.panda.org/about_our_earth/biodiversity/threatsto_biodiversity/
(Accessed on 20/08/2011)
 38. Convention on Biological Diversity 2010
<http://www.conservation.org/newsroom/pressreleases/Pages/World-Governments-Fail-Deliver-2010-Biodiversity-Target.aspx>
(Assessed on 25/07/2011)
 39. Oiseau Rare. A French organisation for rare birds.
<http://www.ornitho.fr>
(Accessed on 10/07/2011)
 40. Natuur Punt (Belgian nature point)
www.natuurpunt.be
(Assessed on 15/07/2011)
 41. Perception website with records of species and perception times
www.waarneming.be
(Assessed on 05/07/2011)
 42. US international bird tracking program EBIRD
<http://ebird.org>
(Assessed on 7/7/2011)
 43. PRESENCE 2.0
<http://www.mbr-pwrc.usgs.gov/software/presence.html>
 44. Words and Pictures
<http://www.wordsandpictures.me.uk/insects1a.html>
(Accessed on 13/07/2011)

Auteursrechtelijke overeenkomst

Ik/wij verlenen het wereldwijde auteursrecht voor de ingediende eindverhandeling:

Improving biodiversity monitoring by modeling relative abundance from "presence only" data

Richting: **master of Statistics-Epidemiology & Public Health Methodology**

Jaar: **2011**

in alle mogelijke mediaformaten, - bestaande en in de toekomst te ontwikkelen - , aan de Universiteit Hasselt.

Niet tegenstaand deze toekenning van het auteursrecht aan de Universiteit Hasselt behoud ik als auteur het recht om de eindverhandeling, - in zijn geheel of gedeeltelijk -, vrij te reproduceren, (her)publiceren of distribueren zonder de toelating te moeten verkrijgen van de Universiteit Hasselt.

Ik bevestig dat de eindverhandeling mijn origineel werk is, en dat ik het recht heb om de rechten te verlenen die in deze overeenkomst worden beschreven. Ik verklaar tevens dat de eindverhandeling, naar mijn weten, het auteursrecht van anderen niet overtreedt.

Ik verklaar tevens dat ik voor het materiaal in de eindverhandeling dat beschermd wordt door het auteursrecht, de nodige toelatingen heb verkregen zodat ik deze ook aan de Universiteit Hasselt kan overdragen en dat dit duidelijk in de tekst en inhoud van de eindverhandeling werd genotificeerd.

Universiteit Hasselt zal mij als auteur(s) van de eindverhandeling identificeren en zal geen wijzigingen aanbrengen aan de eindverhandeling, uitgezonderd deze toegelaten door deze overeenkomst.

Voor akkoord,

Jingwa Awungnjia, Brian

Datum: **12/09/2011**