

**This item is the archived peer-reviewed author-version of:**

Discrete choice modelling of natal dispersal : choosing where to breed from a finite set of available areas

**Reference:**

Vardakis Michail, Goos Peter, Adriaensen Frank, Matthysen Erik.- Discrete choice modelling of natal dispersal : choosing where to breed from a finite set of available areas

Methods in Ecology and Evolution - ISSN 2041-210X - 6:9(2015), p. 997-1006

Full text (Publishers DOI): <http://dx.doi.org/doi:10.1111/2041-210X.12404>

To cite this reference: <http://hdl.handle.net/10067/1252490151162165141>

# 1 Discrete choice modelling of natal dispersal: “Choosing” where 2 to breed from a finite set of available areas

3 Running head: Discrete choice models for Dispersal

4 Michalis Vardakis<sup>1</sup>, Peter Goos<sup>2,3</sup>, Frank Adriaensen<sup>1</sup> and Erik Matthysen<sup>1</sup>

5 <sup>1</sup>Evolutionary Ecology Group, Universiteit Antwerpen, Antwerpen, Belgium.

6 <sup>2</sup>Department Engineering Management, Universiteit Antwerpen, Antwerpen, Belgium.

7 <sup>3</sup>Faculty of Bioscience Engineering, University of Leuven.

8 Corresponding author:

9 Michalis Vardakis

10 Address: Groenenborgerlaan 171, B-2020 Antwerpen, Belgium

11 Telephone number: +32 3 265 35 06

12 Email address: [Michail.Vardakis@uantwerpen.be](mailto:Michail.Vardakis@uantwerpen.be)

## 13 Summary

14 1. Classic natal dispersal studies focus mainly on distance travelled. Although distances capture  
15 some of the main selective pressures related to dispersal, this approach cannot easily  
16 incorporate the properties of the actual destination versus the available alternatives. Recently,  
17 movement ecology studies have addressed questions on movement decisions in relation to  
18 availability of resources and/or availability of suitable habitats through the use of discrete choice

1 models (DCMs), a widely used type of models within econometrics, which explains individual  
2 choices as a function of the properties of a finite number of alternatives.

3 2. In this contribution, we show how the dispersal discrete choice model (DDCM) can be used for  
4 analysing natal dispersal data in patchy environments given that the natal and the breeding area  
5 of the disperser are observed. We test this method using a case study on Great Tits (*Parus major*)  
6 in an archipelago of small woodlots.

7 3. Our results show that DDCMs are able to capture the results of classic distance-based  
8 approaches, and simultaneously allow testing hypotheses on how departure and settlement are  
9 affected by variables that characterize the disperser, the natal patch and the breeding area, as  
10 well as their interactions.

11 4. DDCMs can be applied to any other species and system that uses some form of discrete breeding  
12 location or a certain degree of discretization can be applied.

13

14 Keywords: Multinomial regressions, conditional & mixed logit model, random effects, dispersal.

15

## 16 **Introduction**

17 Natal dispersal is the movement from birth place to the site of first reproduction and is a key mechanism  
18 for spatial population dynamics (Clobert *et al.* 2012). It is often represented as a threefold decision  
19 process, which includes leaving the natal area, an intermediate (prospecting) stage and settling in a  
20 breeding area (Ronce 2007; Clobert *et al.* 2009). The causation of natal dispersal is complex because it  
21 often involves interactions between individual characteristics and environmental variables, different  
22 costs and benefits that may vary among individuals, causal factors acting at different spatial scales and

1 different dispersal mechanisms (Lambin *et al.* 2001; Bowler & Benton 2005; Nathan *et al.* 2008; Benard &  
2 McCauley 2008; Clobert *et al.* 2009; Bonte *et al.* 2012). The complex and context dependent nature of  
3 these interactions highly confounds our ability to elucidate which mechanisms are important in shaping  
4 patterns of natal dispersal.

5 To unravel the mechanisms behind individual dispersal decisions, the majority of dispersal studies have  
6 focused on variation in distances, implicitly or explicitly assuming that distance captures the most  
7 important selective pressures associated with dispersal (Crespi & Taylor 1990; Nathan & Muller-Landau  
8 2000; Levin *et al.* 2003; Nathan *et al.* 2008). For example, inbreeding risk is likely to decrease  
9 monotonically with distance, whereas travel cost is likely to increase with distance, leading to testable  
10 predictions about variation in dispersal distance (e.g. Baker & Rao 2004; Bowler & Benton 2005). An  
11 additional advantage of using distances is that these can be easily adopted in behavioural and population  
12 models (Ronce 2007; Korsten *et al.* 2013). Yet, dispersal distance analyses as such are unlikely to fully  
13 capture the complexity of the mechanisms underlying dispersal, since they cannot explain why  
14 individuals settle in a particular location in contrast to what is available to them (Davis & Stamps 2004).  
15 In a similar manner distance studies may underestimate the influence of different landscape properties,  
16 social and behavioural cues and individual characteristics (morphology, condition) affecting settling  
17 decisions (Clobert *et al.* 2012).

18 By default, natal dispersal is a risky decision associated with imperfect information that is biased towards  
19 the natal area. Most actively dispersing organisms try to minimize this risk to some extent by gathering  
20 information on the quality of natal and surrounding areas, including distances and conspecific densities  
21 (Clobert *et al.* 2009). This sampling takes place in all three stages of the dispersal process (Schjørring  
22 2002; Davis & Stamps 2004; Stamps & Krishnan 2005). Thus, if we want to model natal dispersal as a

1 choice process, we have to accommodate variables describing both the natal environment as well as the  
2 potential destination environments.

3 So far, mainly two approaches have been used for statistical modelling of dispersal patterns. One  
4 approach models distances using linear regression (Greenwood *et al.* 1979; Garrard *et al.* 2012) or more  
5 generally by quantifying dispersal kernels (Nathan *et al.* 2012), while the second approach models the  
6 probability of departing as a binomial response variable (Doligez *et al.* 2002). These approaches have  
7 rarely taken habitat availability directly into account (but see: Davis & Stamps 2004; Stamps & Krishnan  
8 2005). An alternative approach is to model dispersal using a multinomial response variable, where the  
9 choice from a set of available breeding locations is predicted from a range of explanatory variables.

10 Outside the dispersal context, the multinomial approach has been used extensively in studies analysing  
11 movement decisions in a foraging or resource use context, where an animal chooses its next move based  
12 on habitat preferences, conspecifics density or other variables (Manly *et al.* 1993; Cooper & Millspaugh  
13 1999; Johnson & Nielsen 2006; Kesler & Walters 2012). However, multinomial models have rarely been  
14 applied in a dispersal context (Richard & Armstrong 2010; Fernández-Chacón *et al.* 2013) and to the best  
15 of our knowledge, there have been no formal applications to explain observed choices from a number of  
16 available areas, including the natal area. One study modelled dispersal during the prospecting stage with  
17 a similar approach adopted in resource use studies, implying that staying in the natal area was not an  
18 available option (Richard & Armstrong 2010). Another study used mark-recapture multi-event modelling,  
19 involving a multinomial approach, but here again departure and settling choices were separated  
20 (Fernández-Chacón *et al.* 2013). Moreover, instead of incorporating explanatory variables directly in the  
21 multinomial model, the latter authors related them post-hoc to the probability of departure and  
22 settlement.

1 In this paper, we introduce a new general framework for modelling natal dispersal among a finite set of  
2 alternatives. We use discrete choice models (DCMs, also referred to as multinomial regression models), a  
3 widely used technique within econometrics (McFadden 1974; Ben-Akiva & Lerman 1985; Train 2009) and  
4 previously used in resource use modelling (Cooper & Millspaugh 1999). The major advantage of DCMs in  
5 studying natal dispersal is that they allow to model the choice for each available alternative including the  
6 natal area, thereby considering simultaneously the reasons shaping departure and settlement decisions.  
7 Furthermore, DCMs can relate preferences for a particular alternative to individual characteristics as well  
8 as variables describing the alternatives. The most commonly used DCM, the conditional logit model, can  
9 be extended with random effects to allow for heterogeneity of preferences between individuals and/or  
10 to capture similarities between siblings or individuals originating from the same natal area (Daniels &  
11 Gatsonis 1997; Congdon 2006; Kuss & McLerran 2007; Train 2009). The resulting models are named  
12 mixed logit models or multi-level choice models. We first describe how the DCM can be used to model  
13 individual dispersal destinations in general. Then we demonstrate how the model can be applied to a  
14 case study on Great Tits (*Parus major*) breeding and dispersing in an archipelago of small woodlots as a  
15 proof of concept.

16

## 17 Methods

### 18 Discrete Choice Models

19 Multinomial DCMs describe choices of individuals from a finite number of mutually exclusive  
20 alternatives, the choice set. The basic underlying assumption is that an individual selects the alternative  
21 with the maximum “perceived” utility. This utility is known to the individual but not the observer. The  
22 observer can identify attributes of the alternatives available to the individual as well as attributes of the

1 individual itself. Using these, the observer can build a regression model to describe the utility of the  
2 various alternatives. These models are sometimes referred to as the alternatives' "systematic" utilities.  
3 The difference between the latent "perceived" utility for each alternative and the corresponding  
4 "systematic" utility is captured by a random error term. In the following section, we define utility and  
5 explain in detail the use of DCMs based on the conditional logit model and its generalization, the mixed  
6 logit model. Then we explain how DCMs can be used to model dispersal choices. Note that use of the  
7 term "utility" does not imply any specific cognitive processes or fitness outcomes of the dispersal  
8 process; "utility" refers to the combination of all properties that may increase the likelihood of settling in  
9 a specific destination.

## 10 Conditional logit model

11 In the conditional logit model, the most commonly used discrete choice model, an individual  $j$  is  
12 assumed to select the most attractive alternative from a set of alternatives,  $i=1,\dots,I$ . Each alternative is  
13 associated with a utility  $U_{i,j}$  for each individual  $j$ . The model assumes that an individual chooses the  
14 alternative  $n$  that provides the maximum utility  $U_{n,j} > U_{i,j} \forall n \neq i$ . The utility  $U_{i,j}$  is modelled as

$$15 \quad U_{i,j} = V_{i,j} + \varepsilon_{i,j} ,$$

16 where,

$$17 \quad V_{i,j} = x'_{i,j} \beta ,$$

18  $V_{i,j}$  is the systematic utility with  $x'_{i,j}$  a vector of observed covariates related to alternative  $i$  and specific  
19 to individual  $j$  and  $\beta$  the vector of their coefficients. The utility of option  $i$  for individual  $j$  is  
20 complemented with the unobserved random error  $\varepsilon_{i,j}$ . The random errors are assumed to be

1 independently and identically standard Gumbel distributed (Train 2009; ch2 p 34-35). As a result, the  
2 probability of individual  $j$  choosing alternative  $i$  is

$$3 \quad P_{i,j} = \frac{e^{V_{i,j}}}{\sum_i e^{V_{i,j}}}$$

4 The conditional logit model assumes that there is no heterogeneity in individual preferences and that all  
5 utilities have the same variance (Congdon 2006; Train 2009).

## 6 Mixed logit model

7 The mixed logit model is an extension of the conditional logit model, allowing for correlations among the  
8 error terms in the alternatives' utilities as well as for preference heterogeneity between individuals.  
9 Heterogeneity can be defined either by using random coefficients for the covariates or by using common  
10 random effects for groups of individuals (Daniels & Gatsonis 1997; Congdon 2006; Kuss & McLerran  
11 2007; Train 2009). In general, the systematic utility for individual  $j$  and alternative  $i$  can then be written  
12 as

$$13 \quad V_{i,j}(\eta_j) = \beta' x_{i,j} + \eta_j' w_{i,j},$$

14 with  $\eta_j$  the vector of random effects for individual  $j$  and  $w_{i,j}$  the vector of corresponding covariates. The  
15 logit choice likelihood conditional on the random effects in the vector  $\eta_j$  is then given by

$$16 \quad L_{i,j}(\eta_j) = \frac{e^{V_{i,j}(\eta_j)}}{\sum_i e^{V_{i,j}(\eta_j)}} .$$

17 The unconditional choice probabilities in the mixed logit model are obtained by integrating the random  
18 effects out of the conditional choice probabilities.

$$19 \quad P_{i,j} = \int L_{i,j}(\eta_j) f(\eta_j) d\eta_j,$$

1 where  $f(\eta_j)$  is the population distribution of the random effects.

2 The mixed logit choice probability thus is a weighted average of the conditional logit probability  
3 evaluated at different values of the random effects vector  $\eta_j$  with weights given by the distribution of  
4 the random effects. This distribution is usually assumed to be a multivariate normal, but it is possible to  
5 use other distributions as well. The parameters of the distribution can be either estimated with the  
6 maximum likelihood estimator or with hierarchical Bayes (HB) procedures. Both parameterization  
7 techniques have their advantages and disadvantages, but in most cases, they result in remarkably similar  
8 estimates (Bolduc *et al.* 1997; Carlin *et al.* 2001).

## 9 **Application to natal dispersal**

10 Dispersal can be modelled as a choice process where the possible breeding locations are the alternatives  
11 in the choice set and the corresponding utilities represent the “perceived” attractiveness of every  
12 possible dispersal destination (including the natal patch). In order to model dispersal decisions, DCMs  
13 have to incorporate two inter-related decision processes, which are departure and settlement. Firstly, an  
14 individual faces the decision whether or not to leave the natal patch and, in case it leaves, where to  
15 settle for breeding. Three kinds of variables impact these decisions: (i) individual-specific variables  
16 characterizing the decision-maker (e.g., sex), (ii) alternative-specific variables characterizing every  
17 possible destination (e.g., size of the patch), and (iii) variables specific to the individual-alternative  
18 combination (e.g., distance to available patches). In this section, we present a general structure for  
19 dispersal discrete choice models (DDCMs) using these three kinds of explanatory variables. In the next  
20 section, we apply the model to a case study.

21 Consider first an individual  $j = 1$  born in patch 1, and suppose, for notational simplicity, that the  
22 individual can either stay in patch 1, or disperse to patches 2 or 3. We assume that the individual’s

1 choice is influenced by the patch sizes, the distances between them, and the competitor densities in the  
2 three patches. To further illustrate the potential of the model, we assume that departure is only  
3 influenced by density in the natal patch, while settlement is also influenced by density in the destination  
4 patches.

5 A simple model for the utility  $V_{j,i}$  of a given alternative  $i$  perceived by individual  $j$  would include only the  
6 patch size ( $Q_i$ , specific to each patch  $i$ ) and its distance from the natal patch ( $D_{j,i}$  specific for the  
7 combination of patch  $i$  and individual  $j$ ). The systematic utility corresponding to individual 1 then  
8 becomes

$$V_{1,1} = \beta_2 Q_1 + \beta_3 D_{1,1}$$

9  $V_{1,2} = \beta_2 Q_2 + \beta_3 D_{1,2}$

10  $V_{1,3} = \beta_2 Q_3 + \beta_3 D_{1,3},$

11 In these expressions,  $\beta_2$  and  $\beta_3$  are the effects of patch size and distance on utility. This basic model can  
12 already be used to predict both the probability of staying (by comparing utility  $V_{1,1}$  to that of the other  
13 alternatives) and the probabilities of settling in patches 2 or 3 (by comparing utilities  $V_{1,2}$  versus  $V_{1,3}$ ).

14 Note that by introducing distance as a variable for each destination, including a zero distance for the  
15 natal patch, the probability of staying versus leaving is directly coupled to the probability of settling in  
16 patches at varying distances. In order to allow more flexibility in capturing a possible intrinsic inclination  
17 to stay in or leave the natal patch, independent of distances to other patches, we include a “home  
18 advantage” dummy variable  $\alpha_i$  in the utility model for patch  $i$ . This dummy variable takes the value 1 for  
19 each destination that equals the natal patch and zero otherwise. The utilities then become

$$V_{1,1} = \beta_1 \alpha_1 + \beta_2 Q_1 + \beta_3 D_{1,1}$$

1  $V_{1,2} = \beta_1 a_2 + \beta_2 Q_2 + \beta_3 D_{1,2}$

$$V_{1,3} = \beta_1 a_3 + \beta_2 Q_3 + \beta_3 D_{1,3}$$

2 Note that the home advantage coefficient  $\beta_1$  cannot be interpreted independently from the distance  
3 effect  $\beta_3$  (see Figure 5 from the case study for further illustration). This is because of the unavoidable  
4 nonzero correlation between the explanatory variables  $\alpha_i$  and  $D_{j,i}$ , (if  $D_{j,i}$  is zero, then  $\alpha_i$  is one).

5 Likewise, we can increase the complexity of the model by incorporating additional individual- and  
6 alternative-specific variables. For example, we include a variable  $M_i$  which is the competitor density in  
7 the natal patch (note that this variable is zero for all non-natal patches) and variable  $G_i$  as the competitor  
8 density in each patch  $i$  at the time of settling.

9 By definition, individual-specific variables remain constant across all alternatives within an individual's  
10 choice set. Therefore, interaction terms between individual and alternative-specific variables should be  
11 included to capture the fact that the former can only affect individual choices in combination with the  
12 latter.

13 Let us assume that both sex and birthdate (relative to other individuals in the population) affect the  
14 tendency to settle in distant versus nearby patches. For this reason, we include the interaction involving  
15 sex  $S_j$  and distance  $D_{j,i}$  and birthdate  $F_j$  and distance  $D_{j,i}$ , respectively. The systematic utilities then  
16 become

$$V_{1,1} = \beta_1 a_1 + \beta_2 Q_1 + \beta_3 D_{1,1} + \beta_4 G_1 + \beta_5 M_1 + \beta_6 S_1 \times D_{1,1} + \beta_7 F_1 \times D_{1,1}$$

17  $V_{1,2} = \beta_1 a_2 + \beta_2 Q_2 + \beta_3 D_{1,2} + \beta_4 G_2 + \beta_5 M_2 + \beta_6 S_1 \times D_{1,2} + \beta_7 F_1 \times D_{1,2}$

$$V_{1,3} = \beta_1 a_3 + \beta_2 Q_3 + \beta_3 D_{1,3} + \beta_4 G_3 + \beta_5 M_3 + \beta_6 S_1 \times D_{1,3} + \beta_7 F_1 \times D_{1,3}$$

1 An alternative way to model individual variation would be to include random preference variation by  
 2 adding a random coefficient on a specific predictor for every individual (Train 2009). This would be  
 3 appropriate in case there is an unobserved source of individual variation that cannot be captured by  
 4 fixed individual-specific variables. However, this approach can only be used if an individual is exposed to  
 5 the same decision more than once which by definition cannot be the case in natal dispersal, contrary to  
 6 most other movement data.

7 When there is preference heterogeneity at a group level the common random effects are applicable. For  
 8 example, siblings may tend to disperse over similar distances, due to genetic resemblance, common  
 9 environment effects or simply as a consequence of travelling together. This effect can be captured by  
 10 including a random slope for distance  $D_{j,i}$  which differs among families. We denote the difference  
 11 between the slope for family  $C_j$  of individual  $j$  and the average slope for the entire population by  $\eta_{C_j}$ ,  
 12 and assume that  $\eta_{C_j}$  is a normally distributed random variable with zero mean and variance  $\sigma_D^2$ . The  
 13 utility models for patches 1, 2 and 3 then become

$$14 \quad V_{1,1} = \beta_1 a_1 + \beta_2 Q_1 + (\beta_3 + \eta(C_1))D_{1,1} + \beta_4 G_1 + \beta_5 M_1 + \beta_6 S_1 \times D_{1,1} + \beta_7 F_1 \times D_{1,1}$$

$$15 \quad V_{1,2} = \beta_1 a_2 + \beta_2 Q_2 + (\beta_3 + \eta(C_1))D_{1,2} + \beta_4 G_2 + \beta_5 M_2 + \beta_6 S_1 \times D_{1,2} + \beta_7 F_1 \times D_{1,2}$$

$$V_{1,3} = \beta_1 a_3 + \beta_2 Q_3 + (\beta_3 + \eta(C_1))D_{1,3} + \beta_4 G_3 + \beta_5 M_3 + \beta_6 S_1 \times D_{1,3} + \beta_7 F_1 \times D_{1,3}$$

16 In the next section, a mixed logit model is applied to analyse natal dispersal of Great Tits breeding in an  
 17 archipelago of small woodlots.

## 18 Case Study Materials and Methods

19 The study area consists of 10 woodlots from 0.25 ha up to 12 ha in the Boshhoek study area close to  
 20 Antwerp, Belgium (Figure 1; see Supporting information Appendix 1). In these woodlots (henceforth

1 “patches”), breeding data were collected from 1996 to 2012 and we identified all birds born in year  $y$   
2 and breeding in  $y + 1$ . Note that many birds do not recruit inside the study area (estimated at about  
3 50%; Matthysen *et al.* 2001). Therefore our model analyses breeding choices conditional on the decision  
4 to stay within the study area. We will address the consequences of this in the Discussion.

5 Every individual is a unique data entry described by the following individual-specific variables: patch of  
6 birth, patch of breeding, brood identity, sex, and relative fledging date (Matthysen *et al.* 2011) and  
7 standardized within years by subtracting the overall median (Supporting information Appendix 1). The  
8 analysis includes 1022 natal dispersal records with on average 63.9 ( $\pm 16.2$ ) individuals per year, and an  
9 average distance of 332.3 ( $\pm 78.4$ ) meters with a maximum of 2.2 km. Of these, 52% settled outside the  
10 natal patch but within the study area. Every patch was described by the following alternative-specific  
11 variables: 1) size, 2) home advantage (see above for explanation), 3) competitor density prior to  
12 departure, measured as the mean number of fledglings per nest box, 4) competitor density at the time of  
13 settlement, measured as the proportion of nest boxes occupied by roosting Great Tits in November (see  
14 Supporting information; Appendix 1 for explanation). Finally, we included a variable specific to every  
15 individual-patch combination, namely the minimum distance from the nest of birth to the closest border  
16 of that patch (set at zero for the natal patch). We scaled all the continuous variables to zero mean and  
17 unity variance. We considered candidate DDCMs using the aforementioned predicting variables and their  
18 combinations provided that no collinearity occurred. All models included patch size, distance and home  
19 advantage, assuming that these would be the minimal attributes to model the basic characteristics of  
20 any dispersal pattern (see above).

21 In order to account for similarities in siblings' choices, we used three types of random effects. First, we  
22 allowed for family variation in the tendency to stay or leave, modelled by a random effect on the home  
23 advantage dummy variable  $a_j$ . Second, we allowed for family variation in the effect of the distance

1 variable  $D_{i,j}$ , as explained above. Finally we allowed for family variation in the actual choice  $J$  for all but  
2 one choices (reference level). We assumed that all random effects are normally distributed with zero  
3 mean, and variance  $\sigma_{\alpha}^2$ ,  $\sigma_D^2$ , and  $\sigma_j^2$  respectively. We constructed a set of candidate models as outlined in  
4 the *Application to Dispersal* section, i.e. starting from a baseline conditional logit model and adding fixed  
5 effects as well as random effects ending up with a mixed logit model (Table 1; 1<sup>st</sup> column) (for the  
6 algebraic form of all the effects tested: Supporting information; Appendix 1).

7 The parameters as well as the variances of the three types of random effects were calculated with  
8 Bayesian procedures using Markov Chain Monte Carlo (MCMC) methods (Lunn *et al.* 2009) in WinBUGS  
9 (WinBUGS14) via the R package R2WinBUGS (Sturtz *et al.* 2005). We used independent vague normal  
10 prior distributions  $N(0, 0.0001)$  for the fixed model parameters  $\beta_1$  to  $\beta_7$ . For the variances of the  
11 random effects, we used independent gamma prior  $Gamma(0.5, 0.5)$ . The R and WinBUGS code  
12 corresponding to our approach is given in the Supporting information; Appendix 2. To ensure  
13 convergence of the Bayesian estimates, two MCMC chains of 200 000 samples were generated; the first  
14 100 000 samples formed the burn-in sample and were discarded when computing the final estimates.  
15 We subsequently checked for convergence using Gelman's  $\hat{R}$  (Brooks & Gelman 1998).

## 16 **Model selection**

17 We assessed the importance of fixed variables by checking the 95% credible intervals of the posterior  
18 densities. Moreover we assessed model parsimony with the Deviance Information Criterion (DIC) for  
19 conditional logit models. However, use of the DIC is not appropriate for selecting among mixed-effects  
20 models (Gelman *et al.* 2014). We therefore decided to disregard random effects only when their variance  
21 was close to zero or negative, and kept them in the model otherwise.

## 22 **Model validation**

1 Model performance was assessed by the proportion of cases where the model correctly predicted the  
2 observed destination (classification accuracy). These predictions were derived using estimated mean  
3 parameters omitting random effects and partitioning the dataset into a Training-set and a Test-set  
4 (Supporting information; Appendix 2). Although point estimates result in more accurate predictions than  
5 mean estimates obtained directly from WinBUGS, this is a computationally daunting task. As this overall  
6 measure may be strongly influenced by the model’s success in assigning individuals born in the largest  
7 patches, we made an additional evaluation at patch-level, based on frequencies of individuals assigned  
8 to particular patches. For each of the ten patches we calculated the observed and predicted frequency of  
9 individuals assigned to each of the ten choices, including the natal patch. The model fit was tested by  
10 fitting a linear regression model to both the Training-set and the Test-set data with the observed choice  
11 frequencies for every patch combination as a response variable and the predicted frequencies as  
12 explanatory variable.

$$13 \text{ Observed}_i = \text{Predicted}_i * \beta + \alpha,$$

14 where  $i = 1, \dots, 100$ , (10 natal patches times 10 destinations).

15 While the two previous approaches are based on comparing observed and predicted choices, they ignore  
16 the actual strength of the prediction at the individual level, i.e. how much the observed choice differs  
17 from the predicted choice in terms of utility. Thus, observed choices may be classified as wrongly  
18 assigned even if their utility is only marginally lower than the most likely option. Therefore we calculated  
19 for each individual the probability assigned by the model to the observed choice (“observed  
20 probability”), and examined the distribution of these probabilities for all the individuals. Because  
21 probabilities sum up to 1 for each individual across the ten choices, a completely random choice of  
22 breeding locations without any information should result in an average “observed probability” of 0.1.

## 23 Case Study Results

1 The best conditional logit model we obtained included *size*, *distance*, the combination of *sex* \*  
2 *distance*, the combination of *fledging date* \* *distance* and *home\_advantage* (model 3; Table 1)  
3 based on the 95% credible intervals (Supporting information Appendix 1, Table S1-S7) and DIC . All  
4 random effects had nonzero positive estimates and were kept in the model resulting in model 9  
5 (Supporting information Appendix 1, Table S1-S7), which includes the same fixed effects as model 3.  
6 Concerning classification accuracy model 9 correctly classified 53% and 51% of the observed choices for  
7 the Test-set and Training-set, respectively (Table 1). In terms of predictive accuracy, a linear model of  
8 predicted against observed frequencies showed that all models performed similarly with model 9 and  
9 model 1 performing slightly better.

10 For model 9, Figure 2 shows the predicted and observed frequencies for each patch and destination, and  
11 Figure 3 shows the relationship between the two. The “observed probability” was 0.34 for the Test-set  
12 (Table 1), which is higher than the 0.1 expected if choices were completely unpredictable. Figure 4  
13 illustrates the distribution of “observed probabilities” across the different possible natal and destination  
14 patches for the Test-set. For some natal patches many of the observed choices were predicted with high  
15 probability; this is for example the case in patch ZZ, which is relatively large and distant from other  
16 patches. Therefore the model correctly predicts high probabilities to stay in this patch (albeit with some  
17 variation based on individual covariates, see further). For other natal patches such as VS, a small patch  
18 with several larger patches nearby, the predicted probabilities were low for most individuals, because  
19 for most birds this patch would be one of many options with similar (or higher) utilities.

20 If we look into the actual estimates, Model 9 showed a strong effect of inter-patch distance (-2.51, SD=  
21 0.16) indicating that birds were less likely to settle in more distant patches (Table 2). This distance effect  
22 was weaker for females than males (*sex*\**distance* interaction term = 1.17, SD=0.16), implying that  
23 females were more likely to settle in distant patches. Similarly, the distance effect was weaker for early-

1 fledged individuals (0.04, SD= 0.01). Birds were more likely to settle in larger patches (0.60, SD=0.07). No  
2 significant effects of competitor density were found. The estimate of the home advantage coefficient  
3 was significantly negative (Table 2). This does not necessarily mean that birds are less likely to settle in  
4 the natal patch, but that the probability of staying is less than would be predicted from the overall slope  
5 of settlement against distance, which is by definition zero for the natal patch. The combined effect of  
6 home advantage and distance in the two sexes is illustrated in Figure 5. Finally, the variance estimates of  
7 the random family effects were clearly positive, showing that siblings tend to make similar dispersal  
8 choices with respect to staying at home and dispersal distance, but also at the level of actual settlement  
9 in patches regardless of other covariates such as distance (Table 2).

## 10 Discussion

11 To the best of our knowledge, this paper presents the first application of Discrete Choice models to  
12 analyse variation in dispersal choices. The main novelty of the Dispersal Discrete Choice Model (DDCM) is  
13 the utilization of three different types of explanatory variables, namely individual-specific variables,  
14 alternative-specific variables, and variables describing combinations of individual and alternatives.  
15 Additionally, compared to the multinomial models used so far for modelling dispersal choices (Richard &  
16 Armstrong 2010; Fernández-Chacón *et al.* 2013), our DDCM is the first that simultaneously incorporates  
17 the decision to stay in the natal patch versus leaving, and the settlement choice. Therefore, the DDCM  
18 explicitly accounts for the fact that the decision whether or not to breed in the natal patch is also  
19 impacted by the features of the alternative breeding locations. Finally, our DDCM involves random  
20 effects that can describe similarities in dispersal behaviour due to unexplained variation such as family  
21 effects.

22 The results of our case study on Great Tits show that the DDCM is able to capture the known main  
23 drivers of variation in natal dispersal distance such as sex and fledging date, but in addition provides us

1 with additional insights in dispersal decisions based on individual and patch characteristics. Distance and  
2 patch size have previously been identified as the two main variables that structure dispersal movements  
3 for this and other populations (Matthysen *et al.* 2001; Fernández-Chacón *et al.* 2013). Longer dispersal  
4 distances for females are found in many bird species including Great Tits (Clarke *et al.* 1997; Verhulst *et*  
5 *al.* 1997) while evidence for effects of fledgling date varies among studies (Dhondt 1979; Verhulst *et al.*  
6 1997). A novel element in our case study was to test whether competitor density affects either the  
7 settling and/or the leaving phase. Even though no significant effects were found, it does illustrates the  
8 potential of the modelling approach. Note that our case study is not intended as a final test of these  
9 hypotheses, but rather as a proof of concept, and should be viewed as such.

10 The DDCM performed well at an overall level, predicting more than half of individuals' breeding sites  
11 correctly using simple predictors such as distance, density, patch size and sex. However, at the individual  
12 level, when we examined the predicted probability for the observed breeding site, the model performs  
13 less well. This suggests that in order to improve our understanding of individual dispersal choices, it will  
14 be necessary to add more explanatory variables. These can be additional individual-level variables such  
15 as previous experience or personality of the disperser, but it can also be additional patch-level variables  
16 such as habitat quality or landscape connectivity. The importance of additional individual-level variables  
17 is supported by the large variances we obtained for the random family effects, indicating that members  
18 of the same brood make similar choices. Previous studies have provided evidence for at least two  
19 mechanisms leading to sibling similarity, i.e. heritable variation in dispersal distance linked to personality  
20 (Korsten *et al.* 2013) and shared early experiences through post-fledging family movements (Matthysen  
21 *et al.* 2010). A particularly promising component of our model is to model the interaction between  
22 individual-level and alternative-specific characteristics on the breeding patch choice, which allows to test  
23 hypotheses on why certain individuals end up in particular patches. One example that was not yet  
24 explored in this case study would be to test the classic hypothesis that late-born offspring have a

1 competitive disadvantage in settling (Wilkin *et al.* 2006), by incorporating the interaction between  
2 conspecific density at patch level, and individual fledging date. We would also be able to examine these  
3 effects separately for staying in the natal patch and for choosing among other patches. Yet another  
4 option would be to include the effect of previous experience with a patch (prospecting) in a similar  
5 fashion.

6 Another topic for future research is to allow for correlations between utilities of different patches. Such  
7 correlations might be useful when choices for certain patches are similar due to factors not incorporated  
8 in the model, for example if they are connected by corridors. Another avenue where DDCMs can be  
9 applied is to investigate the replacement of the Euclidean distance with alternative metrics (Richard &  
10 Armstrong 2010). Finally, in our DDCM, we assume that the error terms for the different alternatives  
11 within a choice set are homoscedastic. This may not necessarily be realistic: individuals might possess  
12 less precise information concerning patches located farther from the natal patch. A heteroscedastic  
13 DDCM can therefore be useful for future studies (Congdon 2006).

14 An obvious limitation of the DDCM approach is that dispersal options have to be categorized in a finite  
15 number of available options. Thus DDCMs can be applied to any other species and system that uses  
16 some form of discrete breeding location. Examples are woodlots in an agricultural mosaic (Pärt &  
17 Gustafsson 1989; Doligez *et al.* 2002, 2004), naturally fragmented habitats (Coulon *et al.* 2010), breeding  
18 colonies of seabirds (Fernández-Chacón *et al.* 2013) and experimental systems with artificial patches  
19 (Dahirel *et al.* 2014). DDCMs might also be useful for continuous systems, as long as a certain degree of  
20 discretization can be applied. An alternative to discretization is sampling for control locations which is  
21 frequently found in the resource selection function literature (Forester *et al.* 2009), yet the main  
22 difference is that in this approach the choice set for every movement step of an individual is unique.  
23 Another limitation which is apparent from our case study is that DDCMs can only predict choices among

1 the set of alternatives included in the study. However, the same is true for most distance-based  
2 approaches where distances are truncated and results might be biased. Such limitations would not apply  
3 to systems where all breeding areas are included as in some colonial species (Fernández-Chacón *et al.*  
4 2013), and/or where detection is not constrained by distance such as in large-scale remote-tracking or  
5 ringing studies. Furthermore, while some of our estimates may be biased by not including the full range  
6 of alternative options, we are convinced that this is mainly true for the basic structural parameters such  
7 as distance and home advantage, but much less so for more interesting behavioural and ecological  
8 variables such as phenotypic traits or competitor densities. In conclusion, we believe that the discrete  
9 choice modelling approach offers the potential to provide additional insight into the mechanisms and  
10 the causality of natal dispersal patterns beyond traditional distance-based methodologies.

11

## 12 Acknowledgements

13 We thank many researchers and field assistants who contributed to collection of the data used in this  
14 study, in particular Frans Fierens and Joris Elst. This study was supported by a University of Antwerp  
15 BOF/TOP grant to EM.

16

## 1 References

- 2 Baker, M. & Rao, S. (2004). Incremental costs and benefits shape natal dispersal: theory and example  
3 with *Hemilepistus reaumuri*. *Ecology*, **85**, 1039–1051.
- 4 Ben-Akiva, M.E. & Lerman, S.R. (1985). *Discrete Choice Analysis: Theory and Application to Travel*  
5 *Demand*.
- 6 Benard, M.F. & McCauley, S.J. (2008). Integrating across life-history stages: consequences of natal  
7 habitat effects on dispersal. *The American Naturalist*, **171**, 553–67.
- 8 Bolduc, D., Fortin, B. & Gordon, S. (1997). Multinomial probit estimation of spatially interdependent  
9 choices: an empirical comparison of two new techniques. *International Regional Science Review*,  
10 **20**, 77–101.
- 11 Bonte, D., Van Dyck, H., Bullock, J.M., Coulon, A., Delgado, M., Gibbs, M., Lehouck, V., Matthysen, E.,  
12 Mustin, K., Saastamoinen, M., Schtickzelle, N., Stevens, V.M., Vandewoestijne, S., Baguette, M.,  
13 Barton, K., Benton, T.G., Chaput-Bardy, A., Clobert, J., Dytham, C., Hovestadt, T., Meier, C.M.,  
14 Palmer, S.C.F., Turlure, C. & Travis, J.M.J. (2012). Costs of dispersal. *Biological reviews of the*  
15 *Cambridge Philosophical Society*, **87**, 290–312.
- 16 Bowler, D.E. & Benton, T.G. (2005). Causes and consequences of animal dispersal strategies: relating  
17 individual behaviour to spatial dynamics. *Biological Reviews of the Cambridge Philosophical Society*,  
18 **80**, 205–25.
- 19 Brooks, S. & Gelman, A. (1998). General methods for monitoring convergence of iterative simulations.  
20 *Journal of computational and graphical statistics*, 37–41.
- 21 Carlin, J.B., Wolfe, R., Brown, C.H. & Gelman, a. (2001). A case study on the choice, interpretation and  
22 checking of multilevel models for longitudinal binary outcomes. *Biostatistics (Oxford, England)*, **2**,  
23 397–416.
- 24 Clarke, A.L., Sæther, B.-E. & Røskft, E. (1997). Sex biases in avian dispersal: a reappraisal. *Oikos*, **79**,  
25 429–438.
- 26 Clobert, J., Baguette, M., Benton, T. & Bullock, J. (2012). *Dispersal ecology and evolution*, Firstn. Oxford  
27 University Press.
- 28 Clobert, J., Le Galliard, J.-F., Cote, J., Meylan, S. & Massot, M. (2009). Informed dispersal, heterogeneity  
29 in animal dispersal syndromes and the dynamics of spatially structured populations. *Ecology*  
30 *Letters*, **12**, 197–209.
- 31 Congdon, P. (2006). *Bayesian statistical modelling*, Secondn. Wiley Series in Probability and Statistics.
- 32 Cooper, A. & Millsbaugh, J. (1999). The application of discrete choice models to wildlife resource  
33 selection studies. *Ecology*, **80**, 566–575.

- 1 Coulon, A., Fitzpatrick, J.W., Bowman, R. & Lovette, I.J. (2010). Effects of habitat fragmentation on  
2 effective dispersal of Florida scrub-jays. *Conservation biology : the journal of the Society for*  
3 *Conservation Biology*, **24**, 1080–8.
- 4 Crespi, B. & Taylor, P. (1990). Dispersal rates under variable patch density. *American Naturalist*, **135**, 48–  
5 62.
- 6 Dahirel, M., Ansart, A. & Madec, L. (2014). Stage- and weather-dependent dispersal in the brown garden  
7 snail *Cornu aspersum*. *Population Ecology*, **56**, 227–237.
- 8 Daniels, M.J. & Gatsonis, C. (1997). Hierarchical polytomous regression models with applications to  
9 health services research. *Statistics in medicine*, **16**, 2311–25.
- 10 Davis, J.M. & Stamps, J. a. (2004). The effect of natal experience on habitat preferences. *Trends in*  
11 *ecology & evolution*, **19**, 411–6.
- 12 Dhondt, A. (1979). Summer dispersal and survival of juvenile great tits in southern Sweden. *Oecologia*,  
13 **157**, 139–157.
- 14 Doligez, B., Danchin, E. & Clobert, J. (2002). Public information and breeding habitat selection in a wild  
15 bird population. *Science (New York, N.Y.)*, **297**, 1168–70.
- 16 Doligez, B., Pärt, T., Danchin, E., Clobert, J. & Gustafsson, L. (2004). Availability and use of public  
17 information and conspecific density for settlement decisions in the collared flycatcher. *Journal of*  
18 *Animal Ecology*, **73**, 75–87.
- 19 Fernández-Chacón, A., Genovart, M., Pradel, R., Tavecchia, G., Bertolero, A., Piccardo, J., Forero, M.G.,  
20 Afán, I., Muntaner, J. & Oro, D. (2013). When to stay, when to disperse and where to go: survival  
21 and dispersal patterns in a spatially structured seabird population. *Ecography*, **36**, 1117–1126.
- 22 Forester, J., Im, H. & Rathouz, P. (2009). Accounting for animal movement in estimation of resource  
23 selection functions: sampling and data analysis. *Ecology*, **90**, 3554–3565.
- 24 Garrard, G.E., McCarthy, M. a, Vesk, P. a, Radford, J.Q. & Bennett, A.F. (2012). A predictive model of  
25 avian natal dispersal distance provides prior information for investigating response to landscape  
26 change. *The Journal of animal ecology*, **81**, 14–23.
- 27 Gelman, A., Carlin, J.B., Stern, H., Dunson, D.B., Vehtari, A. & Rubin, D.B. (2014). *Bayesian Data Analysis*,  
28 Thirdn. CRC Press Taylor & Francis Group.
- 29 Greenwood, P., Harvey, P. & Perrins, C. (1979). The role of dispersal in the great tit (*Parus major*): the  
30 causes, consequences and heritability of natal dispersal. *The Journal of Animal Ecology*, **48**, 123–  
31 142.
- 32 Johnson, C. & Nielsen, S. (2006). Resource selection functions based on use-availability data: theoretical  
33 motivation and evaluation methods. *Journal of Wildlife Management*, **70**, 347–357.

- 1 Kesler, D.C. & Walters, J.R. (2012). Social composition of destination territories and matrix habitat affect  
2 red-cockaded woodpecker dispersal. *The Journal of Wildlife Management*, **76**, 1028–1035.
- 3 Korsten, P., van Overveld, T., Adriaensen, F. & Matthysen, E. (2013). Genetic integration of local dispersal  
4 and exploratory behaviour in a wild bird. *Nature communications*, **4**, 2362.
- 5 Kuss, O. & McLerran, D. (2007). A note on the estimation of the multinomial logistic model with  
6 correlated responses in SAS. *Computer methods and programs in biomedicine*, **87**, 262–9.
- 7 Lambin, X., Aars, J. & Piernet, S. (2001). Dispersal, intraspecific competition, kin competition and kin  
8 facilitation: a review of the empirical evidence. *Dispersal* (eds J. Clobert, E. Danchin & A. Dhondt),  
9 pp. 261–272. Oxford University Press, New York.
- 10 Levin, S., Helene, C., Muller-Landau, H., Nathan, R. & Chave, J. (2003). The ecology and evolution of seed  
11 dispersal: a theoretical perspective. *Annual Review of Ecology, Evolution, and Systematics*, 575–  
12 604.
- 13 Lunn, D., Spiegelhalter, D., Thomas, A. & Best, N. (2009). The BUGS project: Evolution, critique and future  
14 directions. *Statistics in Medicine*, 3049–3067.
- 15 Manly, B., McDonald, L. & Thomas, D. (1993). *Resource selection by Animals - Statistical design and  
16 analysis for field studies*. Chapman & Hall, London-Glasgow-New York-Tokyo-Melbourne-Madras.
- 17 Matthysen, E., Adriaensen, F. & Dhondt, A. a. (2001). Local recruitment of great and blue tits (*Parus  
18 major*, *P. caeruleus*) in relation to study plot size and degree of isolation. *Ecography*, **24**, 33–42.
- 19 Matthysen, E., Adriaensen, F. & Dhondt, A. a. (2011). Multiple responses to increasing spring  
20 temperatures in the breeding cycle of blue and great tits (*Cyanistes caeruleus*, *Parus major*). *Global  
21 Change Biology*, **17**, 1–16.
- 22 Matthysen, E., Van Overveld, T., Van de Castele, T. & Adriaensen, F. (2010). Family movements before  
23 independence influence natal dispersal in a territorial songbird. *Oecologia*, **162**, 591–7.
- 24 McFadden, D. (1974). Conditional logit analysis of qualitative choice behavior. *Frontiers in Econometrics*  
25 (ed P. Zarembka), pp. 105–142. Academic Press, New York.
- 26 Nathan, R., Getz, W., Revilla, E., Holyoak, M., Kadmon, R., Saltz, D. & Smouse, P. (2008). A movement  
27 ecology paradigm for unifying organismal movement research. *Proceedings of the National  
28 Academy of Sciences of the United States of America*, **105**, 19052–19059.
- 29 Nathan, R., Klein, E., Robledo-Arnuncio, J.J. & Revilla, E. (2012). Dispersal kernels: a review. *Dispersal  
30 Ecology and Evolution* (eds J. Clobert, M. Baguette, T.G. Benton & J.M. Bullock), pp. 187–210.  
31 Oxford University Press.
- 32 Nathan, R. & Muller-Landau, H. (2000). Spatial patterns of seed dispersal, their determinants and  
33 consequences for recruitment. *Trends in Ecology & Evolution*, **15**, 278–285.

- 1 Pärt, T. & Gustafsson, L. (1989). Breeding dispersal in the collared flycatcher (*Ficedula albicollis*): possible  
2 causes and reproductive consequences. *The Journal of Animal Ecology*, **58**, 305–320.
- 3 Richard, Y. & Armstrong, D.P. (2010). Cost distance modelling of landscape connectivity and gap-crossing  
4 ability using radio-tracking data. *Journal of Applied Ecology*, **47**, 603–610.
- 5 Ronce, O. (2007). How Does It Feel to Be Like a Rolling Stone? Ten Questions About Dispersal Evolution.  
6 *Annual Review of Ecology, Evolution, and Systematics*, **38**, 231–253.
- 7 Schjørring, S. (2002). The evolution of informed natal dispersal: inherent versus acquired information.  
8 *Evolutionary Ecology Research*, 227–238.
- 9 Stamps, J. & Krishnan, V. (2005). Nonintuitive cue use in habitat selection. *Ecology*, **86**, 2860–2867.
- 10 Sturtz, S., Ligges, U. & Gelman, A. (2005). R2WinBUGS: a package for running WinBUGS from R. *Journal of*  
11 *Statistical Software*, **12**, 1–16.
- 12 Train, K. (2009). *Discrete choice methods with simulation*. Cambridge University Press, New York.
- 13 Verhulst, S., Perrins, C. & Riddington, R. (1997). Natal dispersal of great tits in a patchy environment.  
14 *Ecology*, **78**, 864–872.
- 15 Wilkin, T. a, Garant, D., Gosler, A.G. & Sheldon, B.C. (2006). Density effects on life-history traits in a wild  
16 population of the great tit *Parus major*: analyses of long-term data with GIS techniques. *The Journal*  
17 *of animal ecology*, **75**, 604–15.

18

19

## 20 **Tables**

1 Table 1. Performance of different candidate DDCMs in predicting Great Tit dispersal choices. Deviance Information Criterion (DIC) is reported for every model.  
 2 The most parsimonious model is highlighted in bold.  $\eta(\text{family})$  refers to the random effect of family. “Prior density” and “Settle density” refer to competitor  
 3 density prior to departure and at the time of settlement. ‘Correctly Classified’ refers to the one to one predicted versus observed breeding choices for the  
 4 Training-set. ‘Training-set fit’ and ‘Test-set fit’ refer to a linear regression of predicted versus observed frequencies of choices for the Training-set and the Test-  
 5 set (Intercept, Slope,  $R^2$ ) respectively.  $R^2$  is the proportion of the explained variation. “Observed probability” is the total average of the predicted probabilities  
 6 for the observed Great Tits’ choices in the Test-set.

7

Model ( $U = \text{utility}$ )	DIC	Correctly Classified (Training/Test-set)	Training-set fit (Intercept, Slope, $R^2$ )	Test-set fit (Intercept, Slope, $R^2$ )	“Observed probability”
<b>Conditional logit model</b>					
1. $V = \text{size} + \text{distance} + \text{home\_advantage}$	2205	0.51/0.49	0.09, 0.99, 0.97	0.05, 0.97, 0.92	0.33
2. $V = \text{size} + \text{distance} + \text{home\_advantage} + \text{sex\_females} * \text{distance}$	2141	0.52/0.51	-0.29, 1.04, 0.97	0.08, 0.96, 0.93	0.34
3. $V = \text{size} + \text{distance} + \text{home\_advantage} + \text{sex\_females} * \text{distance} + \text{fledging\_date} * \text{distance}$	2131	0.52/0.52	-0.36, 1.04, 0.98	0.07, 0.97, 0.92	0.34
4. $V = \text{size} + \text{distance} + \text{home\_advantage} + \text{sex\_females} * \text{distance} + \text{fledging\_date} * \text{distance} + \text{prior\_comp\_density}$	2132	0.52/0.51	-0.36, 1.05, 0.97	0.05, 0.98, 0.92	0.34
5. $V = \text{size} + \text{distance} + \text{home\_advantage} + \text{sex\_females} * \text{distance} + \text{fledging\_date} * \text{distance} + \text{settle\_comp\_density}$	2132	0.52/0.51	-0.4, 1.05, 0.98	0.07, 0.97, 0.92	0.34

6. $V = size + distance + home\_advantage + sex\_females * distance + fledging\_date * distance + \mathbf{distance * size}$	2133	0.52/0.51	-0.38, 1.05, 0.98	0.06, 0.98, 0.92	0.34
<b>Mixed logit model</b>					
7. $V = size + distance + home\_advantage + sex\_females * distance + fledging\_date * distance + \eta(\mathbf{family}) * \mathbf{home\_advantage}$	2115	0.53/0.50	-0.2, 1.03, 0.97	0.49, 0.50, 0.86	0.30
8. $V = size + distance + home\_advantage + sex\_females * distance + fledging\_date * distance + \eta(\mathbf{family}) * \mathbf{home\_advantage} + \eta(\mathbf{family}) * \mathbf{distance}$	2111	0.53/0.50	-0.5, 1.06, 0.96	0.4, 0.84, 0.90	0.31
9. $V = size + distance + home\_advantage + sex\_females * distance + fledging\_date * distance + \eta(\mathbf{family}) * \mathbf{home\_advantage} + \eta(\mathbf{family}) * \mathbf{distance} + \eta_j(\mathbf{family})$	<b>2068</b>	<b>0.53/0.51</b>	<b>-1.22, 1.16, 0.96</b>	<b>0.11, 0.95, 0.92</b>	<b>0.34</b>

1 Table 2. Posterior parameter estimates (mean, standard deviation and Credible Intervals at 95%) for  
 2 model 9 of Table 1. Random effects variances related to within-family variation on the actual choice, on  
 3 the home advantage and on the distance effect are denoted with  $\sigma_J^2$ ,  $\sigma_\alpha^2$  and  $\sigma_D^2$ , respectively. Estimates  
 4 are taken from 200 000 MCMC samples after discarding 100 000 samples as burn in. All the chains are  
 5 converged with Gelman's  $\hat{R} < 1.1$ .

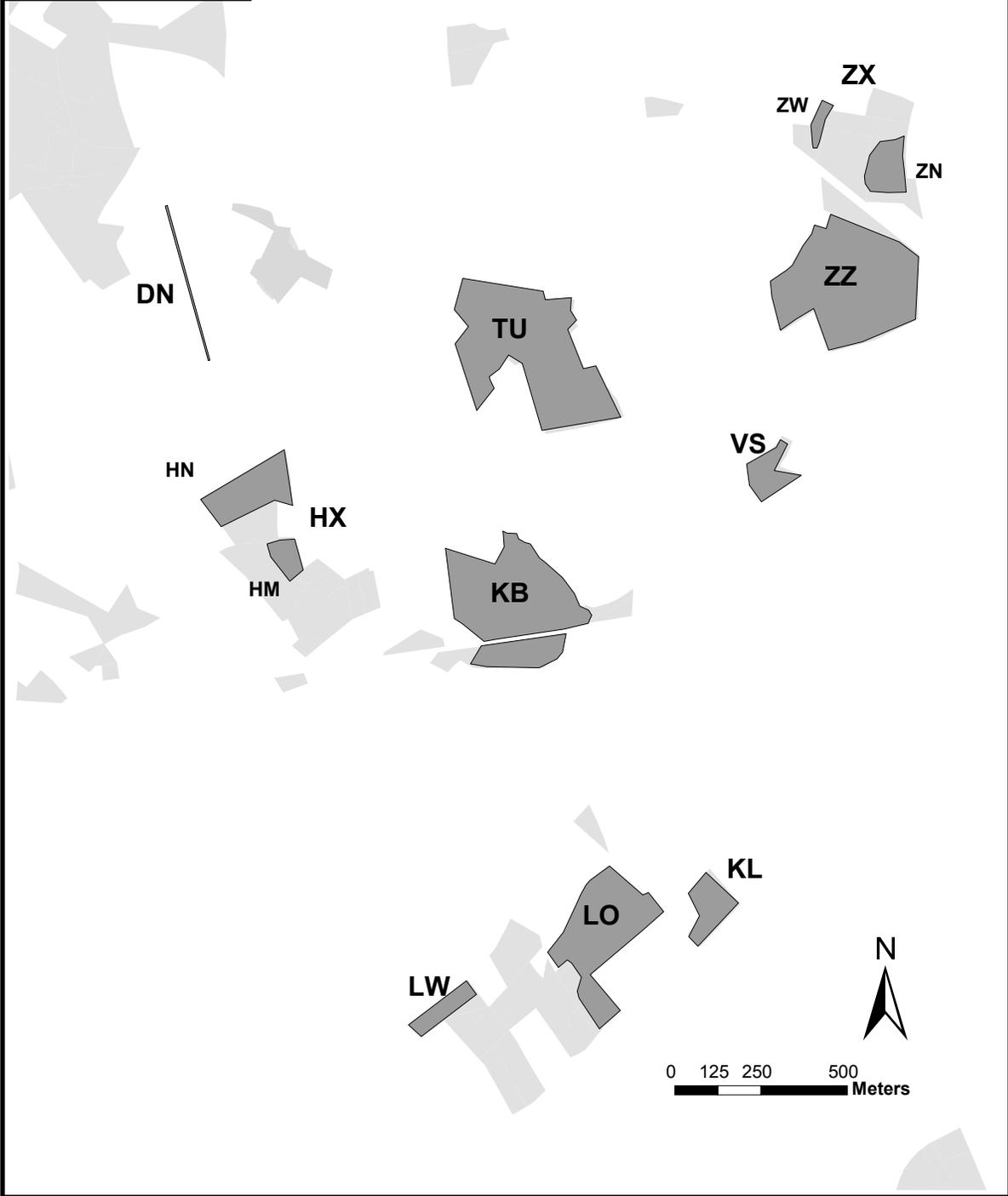
Parameter	mean	sd	2.50%	97.50%	$\hat{R}$
<i>size</i>	0.598	0.072	0.457	0.743	1.00
<i>distance</i>	-2.509	0.161	-2.844	-2.212	1.00
<i>home_advantage</i>	-0.653	0.207	-1.086	-0.269	1.00
<i>sex_females * distance</i>	1.170	0.159	0.863	1.491	1.00
<i>fledging_date * distance</i>	0.039	0.011	0.016	0.061	1.00
$\sigma_J^2$ HX	0.641	0.400	0.149	1.646	1.00
$\sigma_J^2$ KB	1.482	0.663	0.443	2.995	1.00
$\sigma_J^2$ KL	1.158	0.674	0.223	2.736	1.00
$\sigma_J^2$ ] LO	1.501	0.725	0.360	3.132	1.00
$\sigma_J^2$ LW	1.137	0.736	0.211	2.945	1.00
$\sigma_J^2$ TU	0.594	0.411	0.135	1.671	1.00
$\sigma_J^2$ VS	0.661	0.400	0.153	1.665	1.00
$\sigma_J^2$ ZX	0.693	0.432	0.157	1.775	1.00
$\sigma_J^2$ ZZ	0.962	0.582	0.202	2.399	1.00
$\sigma_J^2$ HX	0.641	0.400	0.149	1.646	1.00
$\sigma_\alpha^2$	0.316	0.139	0.114	0.649	1.00
$\sigma_D^2$	0.603	0.411	0.136	1.682	1.00

6

## 7 **Figures**

8

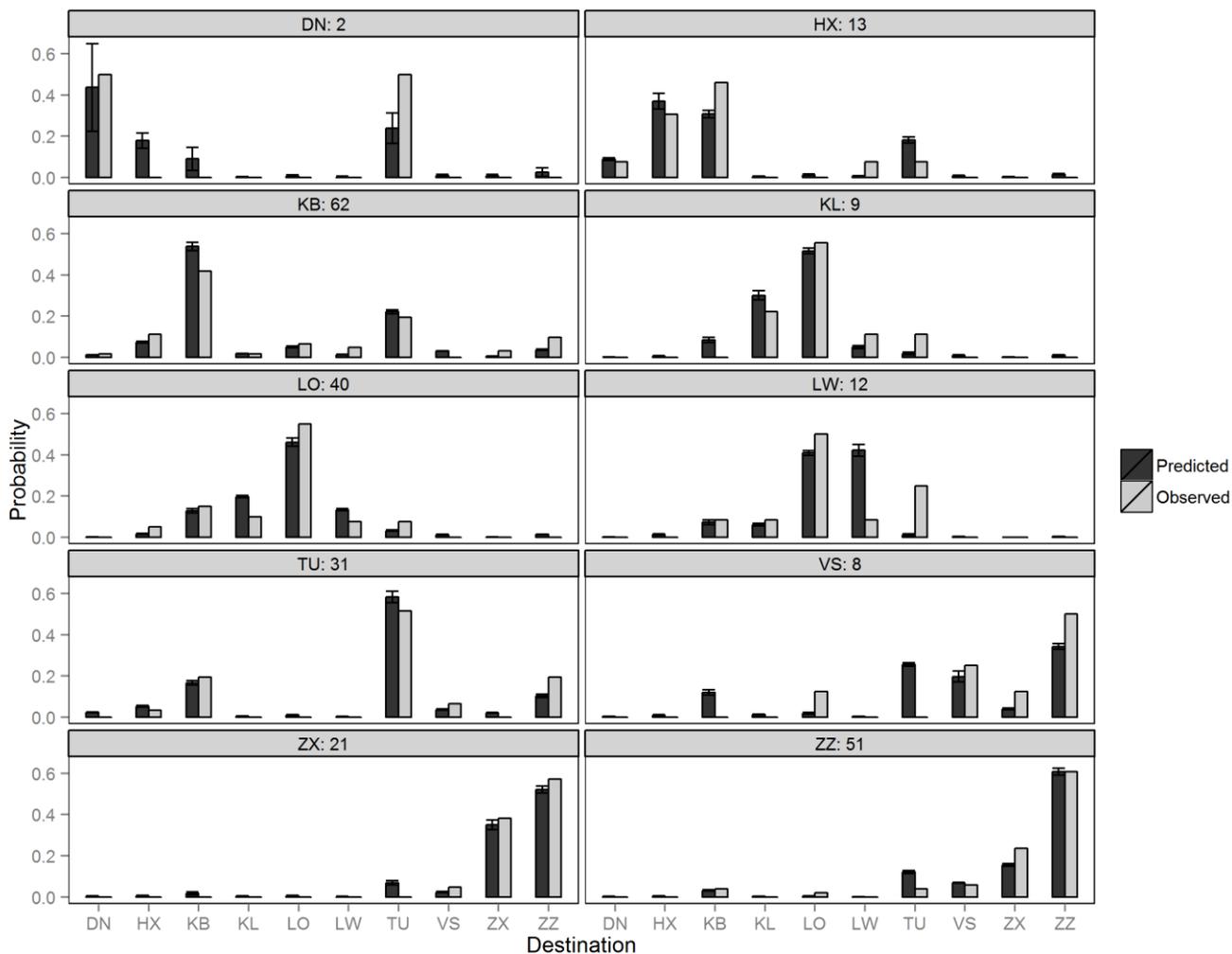
# Boshoek



1

1 Figure 1 Map of study area Boshhoek, Belgium. Wooded areas are shown in grey, with woodlots  
 2 containing nest boxes used in the case study in dark grey. Note that woodlots HN, HM and ZN, ZW were  
 3 lumped for the analysis into HX and ZX respectively.

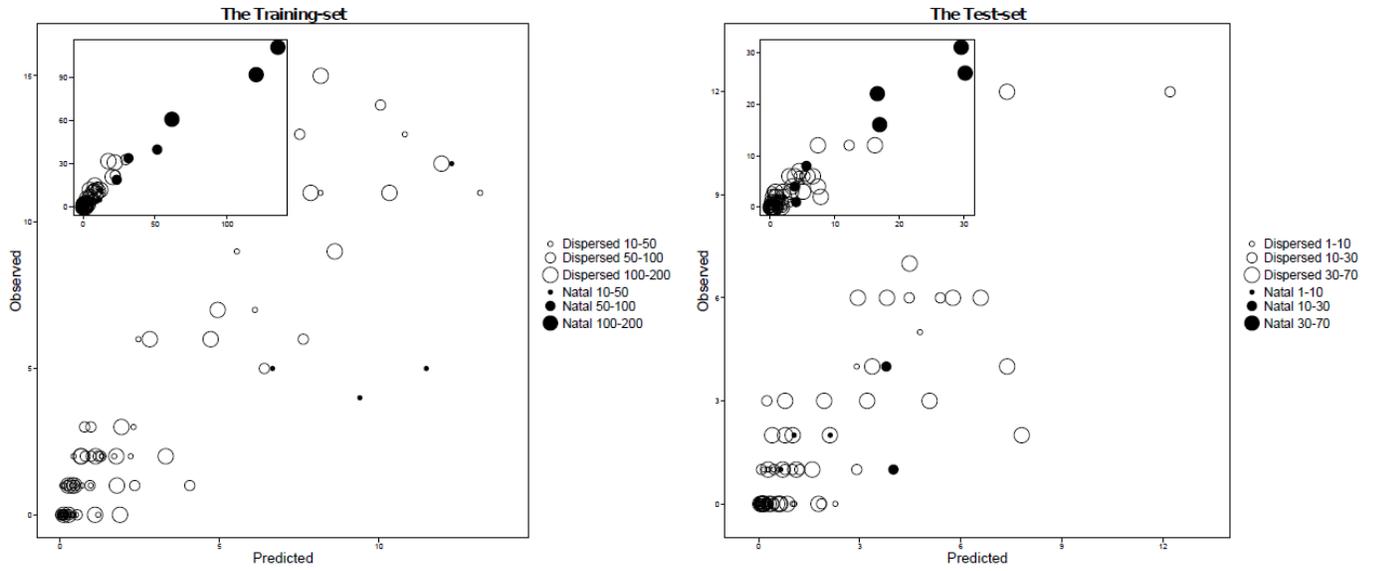
4



5

6 Figure 2. Average predicted (dark grey; mean  $\pm$  standard errors) and observed probabilities of choice  
 7 (light grey) for every natal patch and different destinations, based on model 4 applied to the Test-set  
 8 data. Panel headers indicate the number of individuals born in each patch.

9



1

2 Figure 3. Observed versus predicted frequencies of choices for every combination of natal patch and

3 destination in both the Training and the Test-set (data from Figure 2). The larger plot is the details of the

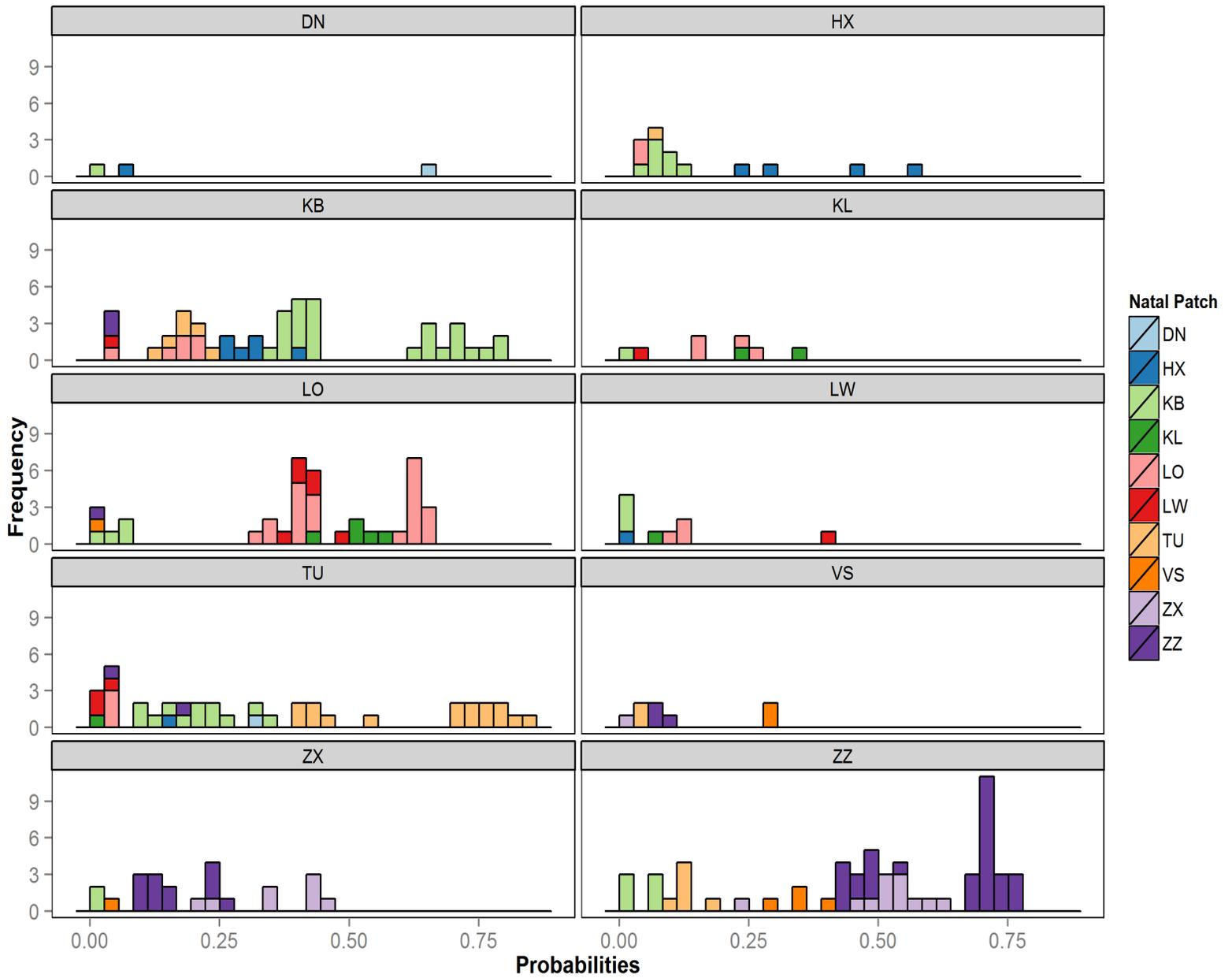
4 low-value region of the smaller plot (the whole dataset). Filled bubbles represent the no dispersal,

5 whereas empty bubbles are the dispersed individuals. Size reflects sample size from every natal patch.

6 The slope and the intercept are from model 9 (Table 1).

7

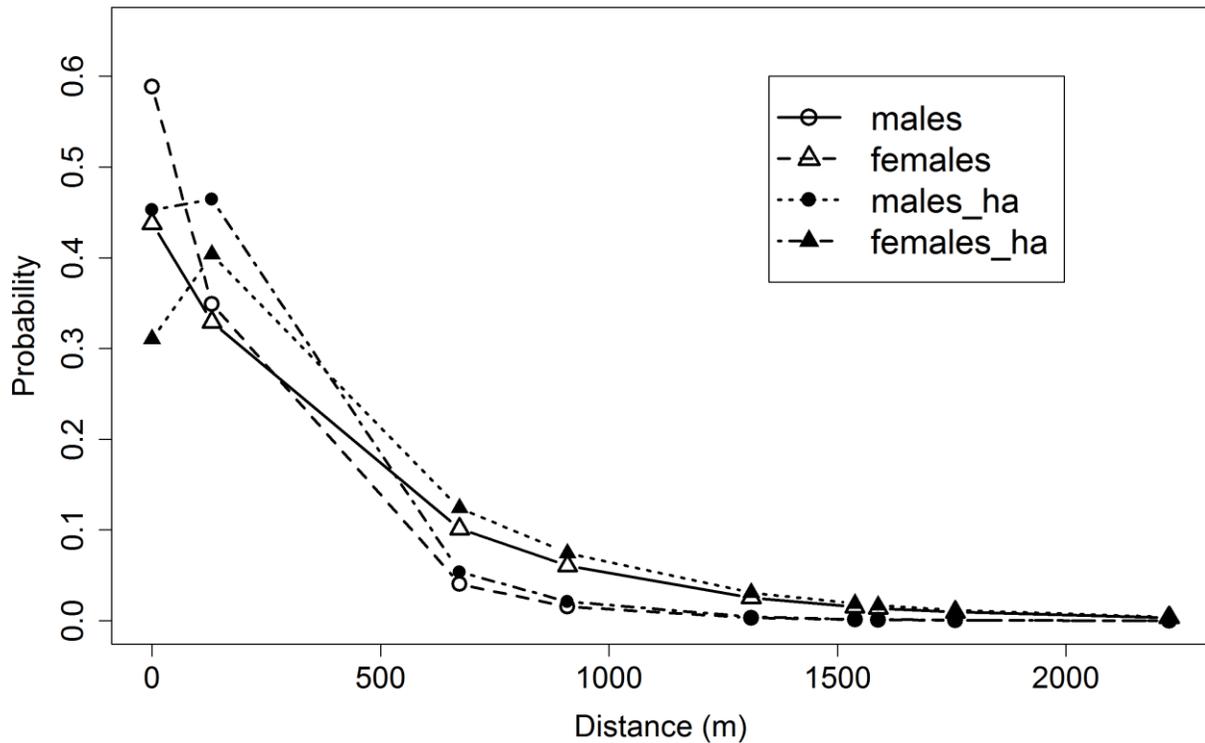
8



1 Figure 4. Frequencies of “observed probabilities” in the Test-set. Each panel shows the choices for a  
 2 particular destination patch, while colours reflect the natal patch.

3

4



1  
 2 Figure 5. Probability plot of choices in relation to distance, illustrating the combined effect of home  
 3 advantage and distance. Symbols represent patches at different distances, including the natal patch at  
 4 the left. Solid and dashed lines show how choices change with distance in a model without and with the  
 5 home advantage parameter, respectively. Symbols show predictions for males and females, respectively.  
 6 Note that for this plot patches are assumed to be of uniform size corresponding to the overall average  
 7 (5.2 ha). Outcomes are based on a model using scaled distance and back-transformed to actual distances  
 8 for clarity.

9  
 10  
 11