



# Measuring dispersal as distance-dependent recruitment rates: testing the performance of DDRR on simulated data

Arie J. Noordwijk

## ► To cite this version:

Arie J. Noordwijk. Measuring dispersal as distance-dependent recruitment rates: testing the performance of DDRR on simulated data. *Journal für Ornithologie = Journal of Ornithology*, 2011, 152 (S1), pp.239-249. 10.1007/s10336-010-0637-2 . hal-00664098

**HAL Id: hal-00664098**

**<https://hal.science/hal-00664098>**

Submitted on 29 Jan 2012

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Measuring dispersal as Distance Dependent Recruitment Rates: Testing the performance of DDRR on simulated data.

Arie J. van Noordwijk

Netherlands Institute of Ecology, Boterhoeksestraat 48, 6666GA Heteren, The Netherlands

(Note, the institute will move at the end of 2010)

Email [a.vannoordwijk@nioo.knaw.nl](mailto:a.vannoordwijk@nioo.knaw.nl)

Phone: ++31 26 4791258

Fax ++31 26 4723227

Keywords: methods, spatial ecology, breeding, population dynamics, observer effects, behaviour, movement analysis

Suggested running title: Testing the measurement of dispersal as DDRR

Word counts

Title 110 chars. Running title 44 chars.

Abstract 198

Main text 2864

Fig legends 298

## **Abstract**

Dispersal is an important process in ecology, but its measurement is difficult.

In particular natal dispersal, the net movement between site of birth and site of first reproduction is important, since it determines population structure. Using simulated data, I study the claim that measuring dispersal in terms of distance dependent recruitment rates filters out many problems. Using several dispersal rules and several spatial distributions of breeding sites, it is shown that distance dependent recruitment rate (DDRR) estimates are independent of the spatial distribution of breeding sites and are sensitive to differences in dispersal rules. These simulations were carried out with sample sizes of 200 individuals, which is a number exceeded in many studies. Variation in clumping of breeding sites (colony sizes) also has little effect on the resulting DDRR estimates. The effects of individuals entering and leaving the study area was simulated by assuming that only half the area was observed. Comparing the ‘observed’ movements with the total distribution of distances dispersed shows that the shape of the DDRR is not affected, although the absolute values are, of course, lower. Thus, DDRR estimates will allow us to start studying dispersal behaviour independent of the peculiarities of the study area and independent of the distribution of observer effort.

## Introduction

Dispersal is an important process in ecology (Johnson and Gaines 1990; Tinbergen 2005). In particular natal dispersal, the net displacement between site of birth and site of first reproduction determines the spatial scale of population dynamics and gene flow. In birds, large amounts of data have been collected where individuals have been ringed in the nest and later been identified as breeding birds. Yet the measurement of natal dispersal is quite problematic, because nearly always there are important differences in the opportunity to observe movements over different distances (Korner-Nievergelt et al. 2010). This heterogeneity in observability quickly leads to discrepancies between the observed dispersal and the true dispersal distance distribution (Baker et al. 1995; Kendall and Nichols 2004; Koenig et al. 1996; Nichols and Kaiser 1999).

Two of the main problems in this respect are individuals that move out of the study area and are thereby lost from observation (Barrowclough 1978) and the fact that within a study the distribution of all possibly observed distances varies between locations (van Noordwijk 1984; van Noordwijk 1995; Winkler et al. 2005). For example the maximum distances are longer at the periphery than in the centre and the number of sites at short distances are higher in the centre. It has been suggested that this latter problem can largely be circumvented by expressing dispersal as distance dependent recruitment rates DDRR In which observed dispersal movements are expressed relative to the numbers ringed in that distance class (van Noordwijk 2006).

Here the performance of DDRR as a measure of dispersal is studied using simulated data. There are many aspects of measuring dispersal that can be investigated by means of simulations. Birds breed either in more or less evenly spaced

territories or in colonies. Colonial breeding adds complications due to variation in colony sizes. To include these problems the simulations are made in terms of breeding colonies. The results presented here are by no means exhaustive, but they illustrate several aspects of the method:

- 1) There should be a substantial reduction in the variation in results obtained from replicate studies with different distributions of breeding colonies.
- 2) Effects of variation in colony size should largely be eliminated.
- 3) The method should be sensitive to differences in dispersal rules used.
- 4) Effects of incomplete knowledge due to animals moving into and out of the study area should be largely eliminated from the results.

These simulations also show that the resulting DDDR are easy to interpret, because they show the dispersal rules used in the simulation in a direct way.

## **Methods**

### **Calculation of Distance Dependent Recruitment Rates**

The basic data consist of observations on individuals that were born at a known location and initiated reproduction at a known location. For each location, where a new breeding bird settled, it can be calculated how many individuals were ringed at each distance in the birth year of the recruits. These data are summarised in distance classes. These distributions of distances to ringing locations are then averaged over all recruits. The resulting distribution describes the average numbers ringed in each distance class, which gives a complete description of what could possibly be observed. Dividing the frequencies of the actually observed number of recruits per distance class by the average numbers ringed per distance class gives the number of recruits observed per nestling ringed, that is a recruitment rate per distance

class, hence DDRR. Formal definitions are given in appendix 1 and a step by step manual for calculating DDRR on real data is given in appendix 2.

### **Basic simulations**

In each run, 25 points (studied colonies) were generated with coordinates drawn from uniform distributions. Next, 200 individuals were generated that started at one colony and moved to a second colony, according to one of the sets of dispersal rules (see below). The distance between natal and breeding colony of each individual was then calculated and these distances were grouped in a frequency distribution to be presented as numbers observed. The same data were also analysed to generate DDRR values. For each recruited individual, all distances to the starting colonies for all individuals in that run were calculated to generate the average number of individuals marked in each distance class. DDRR values are the number of individuals observed per distance class, divided by the average number of individuals marked in that distance class. All simulations were performed in PASCAL programs.

### **Dispersal rules**

In all cases, the starting colony for each individual was drawn randomly from the set of colonies. The following dispersal rules were used:

- 1) *Random redistribution*. One point was drawn randomly from the total set of colonies. The probability that an individual moved to any colony (including the colony of origin) is thus equal to one over the number of colonies.
- 2) *Favouring short distances*. Two colonies were drawn at random. The distances from the starting colony were calculated and the colony with the smallest distance was selected as destination.

3) *Strongly favouring short distances*. Similar to 2, but now the minimum distance from 5 randomly picked colonies was used instead of the minimum from 2.

4) *Favouring medium distances*. The median distance from 3 randomly picked colonies was used to determine the destination colony.

#### **Extensions to the simulations.**

In the real world colonies are unlikely to have the same size, the biggest colonies may be evenly spaced and moreover study sites are limited, so that individuals will be lost from sight by moving over the border of the study site. These aspects were included in the analysis by three further extensions to the simulations. First, colony sizes were made unequal. Instead of 25 colonies with equal sizes, 5 colonies were created with relative size 10, 5 with relative size 5, 10 with relative size 2 and 5 with relative size 1. This was achieved by using a list of 100 colonies, but replicating the coordinates as many times as the relative size.

A second extension consisted of fixing the coordinates of the five biggest colonies at [200,200],[200,800],[800,200],[800,800] and [500,500] in a field of 1000 \* 1000 creating an excess of movements of about 430 units and 600 units. The other colonies were still located at coordinates drawn randomly from uniform distributions and differed among runs.

The final complication added was that two quarters of the total area were considered to be unobserved. Animals starting and/or finishing in these unobserved areas were included in the 'total' dataset but excluded from the 'observed' dataset.

Numbers of replicate runs.

142           In all but the final analysis only five runs of the model are presented. This low  
143   number was chosen, because in real datasets it is often possible to create a number of  
144   sub-sets of the data in this order of magnitude. Since standard errors depend to a large  
145   extent on the number of replicates, the standard errors presented here are indicative of  
146   what could be obtained with real data. Thus, when differences between dispersal rules  
147   are highly significant with these numbers, one can also expect them to be visible in  
148   real data. The only exception is the final evaluation of how much better DDRR  
149   performs when only partial data are available. Here 100 runs are presented, which just  
150   still allows to show the individual datapoints.

151

152



## Results.

The results for four different dispersal rules are presented in Fig 1 in the form of means and standard errors over 5 replicate runs. For random redistribution in the study area, Fig 1a shows the numbers observed and Fig 1b shows the corresponding DDRR estimates. Fig 1a illustrates how difficult it is to interpret raw data on dispersal. The numbers observed in the 2<sup>nd</sup>, 3<sup>rd</sup> and 4<sup>th</sup> distance classes are considerably lower than in the next classes and numbers rapidly become lower after distance class 13. The first of these two aspects is due to the relation between distance and area. The area within a distance band increases linearly with the distance and thus the number of possible destinations increases with distance (see van Noordwijk 1995). The second aspect is due to the fact that at larger distances an increasing proportion of the total area at that distance falls outside the study and thus there are few observations of individuals moving over longer distances because there are few opportunities to move long distances and be observed (only moving from one edge to the opposite edge).

In contrast, the corresponding DDRR estimates (Fig 1b) are close (within 2 s.e.) to 1.0 for all distance classes. The dip in the distance classes 2, 3 and 4 is absent and although the standard errors increase dramatically for the last few distance classes, the results for distance classes 14 to 20 look interpretable in the DDRR, whereas they are strongly affected by the limitations of the study area in the observed numbers. One final aspect is that in the observed numbers, the standard errors obtained from replicate simulation runs are high when the numbers observed are high and low in the higher distances classes, whereas in the DDRR the standard errors are high in the last distance classes. This latter pattern corresponds much better to the

greater imprecision of the estimates in the higher distance classes, which are based on small numbers.

The easiest way to create a dispersal pattern that is biased towards smaller distances is to draw two random destinations and each time choose the shortest distance (see van Noordwijk 1984). Results are presented in Fig 1c and 1d. The raw data are again difficult to interpret. Although the relative dip in distance classes 2 and 3 is smaller than in Fig 1a, it is still present. In contrast, Fig 1d shows a gradual decline in DDRR with increasing distance. In this series of runs, the 2<sup>nd</sup> distance class has a lower value than the first and third, but the difference is only about two s.e., instead of five in the corresponding raw data. In Fig 1e-h, two more dispersal rules are shown, strongly favouring small distances and favouring intermediate distances. In all cases the DDRR is easily interpretable and peculiar aspects in the raw data have been eliminated.

### **Unequal colony sizes**

Randomly distributed colonies of equal size is a rather artificial situation, in practice, it is more likely that colony sizes are unequal. With unequal colony sizes (see methods) the standard errors increased, both in the raw numbers and in the DDRR, but the dispersal rule used is still easily seen in the DDRR (Fig 2). In a next step, the five big colonies were given fixed regularly distributed coordinates (see Fig 3) that were the same in replicate runs. This has the effect that the distances between these colonies will be overrepresented in the raw data, (Fig 4a & 4c). This effect was not present in the DDRR estimates derived from the same data (Fig 4b & 4d).

## **Evaluation.**

How can we measure how much better the DDRR performs than the raw data?

There are two things that we want. First, the same dispersal rules in different settings should give us similar results and second, different dispersal rules in the same setting should give us different results. Here we compare the raw data and the corresponding DDRR values for two dispersal rules (random redistribution and favouring short distances) in two settings (randomly distributed colonies of equal size versus big colonies at fixed locations plus randomly distributed small colonies).

For each distance class, we have a value and a standard error over 5 runs, which allows us to do a t-test for each distance class. This gives us the probability that the two values for the same distance class obtained for the different configurations of the study area come from the same distribution. We can then combine the probabilities for each point using Fisher's combination test ( $\chi^2_{[2n]} = -2\sum \ln P$ ) to give an overall statement on the similarity of the curves (for the raw data: Random redistribution:  $\chi^2_{[40]} = 72.52$ ;  $P = 0.00125$ , Favouring small distances:  $\chi^2_{[40]} = 88.91$ ;  $P = 0.000014$ ). We can do the same for the DDRR estimates (Random redistribution:  $\chi^2_{[38]} = 28.75$ ;  $P = 0.86$ , Favouring small distances:  $\chi^2_{[38]} = 28.74$ ;  $P = 0.86$ ). Thus the raw numbers are quite different between the two different colony configurations, but the DDRR measures are very similar for both dispersal rules used. At the same time, in both configurations the DDRR results were very different for the two dispersal rules (both  $P < 10^{-8}$ ). Thus DDRR performed as required.

## **The effects of partial observations.**

In practice, study areas are nearly always limited and thus individuals will move into and out of the study area. The extent to which conclusions are affected by

these movements is another aspect to be investigated. This was simulated by limiting the observations to two quarters of the total. Thus the simulations were carried out as before, but only when both the starting and end point were within the observed part, was the individual added to the observations.

One can now compare the total data-set with the observed sub-set (Fig 5). Using the setting with the big colonies with fixed co-ordinates (two of which are now hidden) and favouring short distances as the dispersal rule, the DDRR estimates are proportional, while the raw data have quite differently shaped distributions. The effects of the fixed big colonies are clearly visible in the raw data and these irregularities have disappeared from the DDRR estimates. It is obvious that the recruitment rates for the observed dataset are lower than the full data-set. DDRR estimates are relative measures and not absolute measures. In this case, half the individuals marked at birth disappear out of sight, so that the observed recruitment is half as high (see discussion).

In this case, the performance of DDRR can be evaluated by considering the correlation between the numbers observed per distance class with those in the total data and to do the same for the DDRR estimates. In Fig 6 the results are presented for 100 runs. If one excludes DDRR estimates which are based on less than 5 individuals ringed in that distance class, the correlations between the DDRR estimates of the observed sub-set and the total are quite high, (mean 0.90, median 0.93), and higher than for the raw numbers (mean 0.78, median 0.79). We are particularly interested in the quantity  $1-r^2$  as a measure of the unexplained variance. Over the 100 runs, this was smaller for the DDRR than for the raw numbers in 93 % of the runs (on average 51 % (of the  $1-r^2$  in the numbers) smaller). In the few cases where DDRR did not perform better, either the correlation was high for both the raw numbers and the

253 DDRR estimates or the number of datapoints in the observed set was very low. Thus,  
254 DDRR allows us to draw better conclusions about the dispersal behaviour when (a  
255 substantial) part of the movements are unobserved then the raw numbers.  
256

## Discussion

The simulations described here demonstrate that DDRR estimates are easier to interpret than raw data. Moreover, changes in the lay-out of the study have no effect on the resulting DDRR estimates, while different dispersal rules used can easily be distinguished. Furthermore, hiding data from observation has little effect on the shape of the obtained DDRR estimates. DDRR estimates are not absolute numbers, they tell us how recruitment varies with distance. They clearly reach their limits at distances near the size of the area studied.

In calculating DDRR estimates, no assumptions are made about any sort of underlying distribution. The only two assumptions made are that dispersal can be summarised in terms of distances (see below) and that averages can be made over the data. This is equivalent to the assumption that there no heterogeneities in the dispersal rules use by the birds in space and time. Whenever there are sufficient data available, this last assumption can be checked by subdividing the data and checking whether the resulting DDRR values are different.

In situations where it is reasonable to make assumptions about equality of immigration and emigration, one could estimate the proportion of the dispersal process that has been lost from view from the proportion of immigrant first breeders. There are other advantages in studying dispersal in terms of recruitment. Whereas it is not possible to tell where emigrants went to, it is sometimes possible to obtain some information on where immigrants came from, based on isotope ratios in their feathers(Clark et al. 2004; Hobson et al. 2004).

It is an open question whether summarising dispersal in terms of physical distances is biologically the most relevant. For a forest bird, a distance of 1 km over

open landscape or over water is probably quite different from the same distance through forest or along hedgerows. In principle, there is no limitation to the different distance measures that could be used when calculating DDRR (Heinz et al. 2005). At present, too little is known about for example the relation between dispersal and density (Kim et al. 2009; Matthysen 2005) to evaluate whether expressing distance in terms of number of territories moved is biologically relevant.

Among the dispersal rules tested in these simulations, it is easy to interpret the resulting DDRR measures. This should facilitate the connection between dispersal patterns observed and the behaviour of the individuals moving (Bowler and Benton 2005; Dingemanse et al. 2003; Greenwood 1980; Hawkes 2009; Russell and Rowley 1993).

A first step to take the limitations on dispersal observations into account has been to compare the observed movements with a random redistribution of the animals over the observed natal and observed breeding sites (van Noordwijk 1995; Winkler et al. 2005). DDRR estimates are different in a number of ways. Whereas the reference distribution under random redistribution changes if e.g. the study area is enlarged, DDRR estimates should not change, unless the dispersal behaviour or the density is different in the added area. The unit of movements observed per nestling ringed at that distance also does not imply any null assumption about dispersal.

The dispersal rules used in these simulations, such as favouring short distances are not formulated in terms of the actual behaviour. In terms of the behaviour, two processes can be distinguished. First there is the location, size and shape of the area that an individual is familiar with, which depends on how individuals learn about the world. Second, there is the decision to settle somewhere within this area. The dispersal rules used in the simulations with random

redistribution or favouring short distances are equivalent to familiar areas centred around the site of birth but differing in size. The rule of random redistribution implies that individuals move up to the borders of the study area. This could happen either when the study area is smaller than the area with which individuals are familiar, or when the study area is a (habitat) island with reflective boundaries. The dispersal rule of favouring intermediate distances could come about by individuals first moving away from the natal site and then having a limited home range. In some cases, this move away from the natal site or part of it may come about before independence, which then leads to correlations of dispersal distances by siblings (Massot et al. 1994; Matthysen et al. 2010). Thus, although the dispersal rules used were formulated in terms of the resulting pattern, there are plausible mechanisms underlying them.

The unit used to describe dispersal is observed recruit per nestling ringed as a function of distance. This unit is easy to understand and should facilitate the incorporation of dispersal in models of populations or metapopulations (Reed and Levine 2005). The interpretability of results in terms of the dispersal rules, the filtering out of specific properties of the study area and the resulting robustness should therefore make DDRR a very attractive way to describe dispersal.

The simulations reported here are by no means exhaustive. They show that DDRR is a step forward in isolating the dispersal behaviour from peculiarities of the study area. This should allow us to start analysing variation in dispersal that is due to biologically interesting processes.

**Acknowledgements.** I gratefully acknowledge discussions with many people and in particular Stein Saether over the last decade that led to the formulation of



331 DDRR in its present form and I thank Pascaline le Gouar for comments on the ms.

332 This is paper XXXX from the Netherlands Institute of Ecology.

333

334

- Baker M, Nur N, Geupel GR (1995) Correcting Biased Estimates of Dispersal and Survival Due to Limited Study Area - Theory and an Application Using Wrentits. *Condor* 97:663-674
- Barrowclough GF (1978) Sampling bias in dispersal studies based on a finite area. *Bird-Band.* 49:333-341
- Bowler DE, Benton TG (2005) Causes and consequences of animal dispersal strategies: relating individual behaviour to spatial dynamics. *Biological Reviews* 80:205-225
- Clark RG, Hobson KA, Nichols JD, Bearhop S (2004) Avian dispersal and demography: Scaling up to the landscape and beyond. *Condor* 106:717-719
- Dingemanse NJ, Both C, van Noordwijk AJ, Rutten AL, Drent PJ (2003) Natal dispersal and personalities in great tits (*Parus major*). *Proceedings of the Royal Society of London Series B-Biological Sciences* 270:741-747
- Greenwood PJ (1980) Mating systems, philopatry and dispersal in birds and mammals. *Animal Behaviour* 28:1140-1162
- Hawkes C (2009) Linking movement behaviour, dispersal and population processes: is individual variation a key? *Journal of Animal Ecology* 78:894-906
- Heinz SK, Conradt L, Wissel C, Frank K (2005) Dispersal behaviour in fragmented landscapes: Deriving a practical formula for patch accessibility. *Landscape Ecology* 20:83-99
- Hobson KA, Wassenaar LI, Bayne E (2004) Using isotopic variance to detect long-distance dispersal and philopatry in birds: An example with Ovenbirds and American Redstarts. *Condor* 106:732-743
- Johnson ML, Gaines MS (1990) Evolution of Dispersal - Theoretical-Models and Empirical Tests Using Birds and Mammals. *Annual Review of Ecology and Systematics* 21:449-480
- Kendall WL, Nichols JD (2004) On the estimation of dispersal and movement of birds. *Condor* 106:720-731
- Kim SY, Torres R, Drummond H (2009) Simultaneous positive and negative density-dependent dispersal in a colonial bird species. *Ecology* 90:230-239
- Koenig WD, VanVuren D, Hooze PN (1996) Detectability, philopatry, and the distribution of dispersal distances in vertebrates. *Trends in Ecology & Evolution* 11:514-517
- Korner-Nievergelt F et al. (2010) Improving the analysis of movement data from marked individuals through explicit estimation of observer heterogeneity. *Journal of Avian Biology* 41:8-17
- Massot M, Clobert J, Chambon A, Michalakis Y (1994) Vertebrate Natal Dispersal - the Problem of Nonindependence of Siblings. *Oikos* 70:172-176
- Matthysen E (2005) Density-dependent dispersal in birds and mammals. *Ecography* 28:403-416
- Matthysen E, Van Overveld T, Van de Castele T, Adriaensen F (2010) Family movements before independence influence natal dispersal in a territorial songbird. *Oecologia* 162:591-597
- Nichols JD, Kaiser A (1999) Quantitative studies of bird movement: a methodological review. *Bird Study* 46:289-298
- Reed JM, Levine SH (2005) A model for behavioral regulation of metapopulation dynamics. *Ecological Modelling* 183:411-423

383 Russell EM, Rowley I (1993) Philopatry or Dispersal - Competition for Territory  
 384 Vacancies in the Splendid Fairy-Wren, *Malurus-Splendens*. *Animal Behaviour*  
 385 45:519-539  
 386 Tinbergen JM (2005) Biased estimates of fitness consequences of brood size  
 387 manipulation through correlated effects on natal dispersal. *Journal of Animal*  
 388 *Ecology* 74:1112-1120  
 389 van Noordwijk AJ (1984) Problems in the Analysis of Dispersal and a Critique on Its  
 390 Heritability in the Great Tit. *Journal of Animal Ecology* 53:533-544  
 391 van Noordwijk AJ (1995) On bias due to observer distribution in the analysis of data  
 392 on natal dispersal in birds. *Journal of Applied Statistics* 22:683-694  
 393 van Noordwijk AJ (2006) Measuring natal dispersal as distance-related recruitment  
 394 rates. *Journal of Ornithology* 147:45-45  
 395 Winkler DW et al. (2005) The natal dispersal of tree swallows in a continuous  
 396 mainland environment. *Journal of Animal Ecology* 74:1080-1090  
 397  
 398  
 399

**Figure legends:**

**Fig 1.** A comparison of raw data and resulting DDRR estimates for four dispersal rules. The means and se per distance class over five runs are given. In each run 25 randomly distributed colonies and 200 individuals were generated. a) raw data and b) DDRR estimates for random redistribution, c) raw data and d) DDRR estimates for favouring short distances, e) raw data and f) DDRR estimates for strong favouring short distances and g) raw data and h) DDRR estimates for favouring intermediate distances.

**Fig 2.** A comparison of raw data and resulting DDRR estimates for colonies of unequal size. a) raw data and b) DDRR estimates for random redistribution, c) raw data and d) DDRR estimates for favouring short distances. Otherwise similar to fig 1.

**Fig 3.** Set-up of “Regular” and incomplete observations. The five big colonies are in fixed positions as indicated and the co-ordinates of the other colonies are drawn independently from uniform distributions separately in each run. In the simulations with incomplete observations, all movements starting or ending in the stippled areas are included in the total, but not in the observed dataset.

**Fig 4.** A comparison of two dispersal rules (random redistribution a,b) and favouring short distances (c,d) in two different colony lay-outs (randomly distributed colonies of equal size (continuous blue line) and fixed big colonies with random small colonies (dashed red line). Numbers observed (mean and se over 5 runs) are given in a and c and the resulting DDRR estimates in b and c.

**Fig 5.** Comparison of the ‘observed’ with the ‘total’ dataset when half the area is hidden from view (see fig 3). Means and se over 5 runs using the dispersal rule favouring short distances. a) numbers observed, b) corresponding DDRR values.

**Fig 6.** A scatterplot of the correlation between the 'total' and 'observed' dataset in raw numbers per distance class (horizontal axis) and as DDRR estimates per distance class (verticalaxis). The diagonal indicates equality of the correlations (see text).

433 Appendix 1. Formal definitions.

434

435 Let  $A_i$  be the  $i$ th (potential) breeding location with co-ordinates  $x_i$  and  $y_i$ .

436 Let  $B_j$  be the  $j$ th breeding location where  $N_j$  individuals were marked in year  $t-1$ .  $B$  is  
437 a sub-set of  $A$ .

438 Let  $C_k$  be the  $k$ th breeding location where a recruit was observed in year  $t$ .  $C$  is a sub-  
439 set of  $A$  and  $k$  is the total number of recruits observed.

440 
$$D_{jk} = \sqrt{(x_j - x_k)^2 + (y_j - y_k)^2}$$

441 The distribution of all possible observations is then given by

442 
$$\{ N_j * D_{jk} \} \text{ for all } j \text{ and all } k.$$

443

444 Let  $N_p = \text{Count}(\{ N_j * D_{jk} \mid D_{p-1} \leq D_{jk} < D_p \})$  where  $D_p$  is the maximum distance of  
445 the  $p$ th distance class and  $D_0 = 0$ .

446 Then  $N_p / k$  gives the average number of nestlings ringed at distances between  $D_{p-1}$   
447 and  $D_p$ , measured from the locations where recruits were observed in year  $t$ .

448 Let  $\{D\}$  be the set of distances between site of birth and site of recruitment for all  
449 recruits with known birth locations.

450 
$$M_p = \text{Count}(\{D \mid D_{p-1} \leq D < D_p\})$$

451 
$$DDRR_p = M_p / (N_p / k).$$

452

453 Note 1.

454 It is assumed here that individuals recruit one year after birth. The set  $B$  should  
455 always be taken from the appropriate year

456 Note 2.

457 There is a choice to either use the set C over all recruits irrespective of whether these  
458 recruits have a known origin or not, or alternatively to restrict the set to recruits with  
459 known origin. This will make a difference (only) when immigrants (recruits with  
460 unknown origin) settle in different places from local recruits. This could happen for  
461 example when the study area is very large relative to dispersal distances and more  
462 immigrants are expected in the periphery.

463 Note 3.

464 When data are collected over several years, it is recommended to calculate an average  
465 distribution weighted by the number of recruits per year:

466  $N_p = ( \sum N_{pt} ) / \sum k_t$  where  $N_{pt}$  and  $k_t$  are the quantities for year t.

467

## **Appendix 2. A step-by-step guide to calculate distance dependent recruitment rates.**

### **Step 1. Ingredients.**

We need the following data:

- a) A list of all nests where nestlings were ringed with their co-ordinates and the numbers ringed at each location. Using UTM co-ordinates or national grid co-ordinates makes the calculation of distances a bit easier, but this is not essential.
- b) A list of all locations where recruits (first-time breeders) were observed with their co-ordinates. Here one has a choice of either using only recruits whose birth location is known or also including recruits with unknown birth locations. This will make a difference if immigrants settle at different locations compared to local recruits.
- c) A list with the distances for the observed recruits, i.e. individuals for which both the birth and the recruitment location are known.

### **Step 2. Calculating distances between all locations.**

This is the first step in describing all possible observations. For most data sets one will want to make these calculations separately for each year and the amount of data will then be manageable within a spreadsheet programme. In the example in Table 1, the numbers ringed and the co-ordinates of the nestboxes where birds fledged are given in columns **A**, **B** and **C**, and the co-ordinates of recruited birds in the next year are given in rows **1** and **2**, starting in column **E**. The distances between each combination of fledging and recruiting are given in the cells **E4:I15**. The formulas for the first three elements on the diagonal are given below the table. Through the use of the dollar sign, a single formula can be copied for the whole table. In cell **A1** is the



492 number of recruits. Finally, in cell **A16** is given the sum of all nestlings ringed, to be  
 493 used for checking for calculation errors.

494

495 **Table 1. Calculation of distances, see text.**

|           | <b>A</b>  | <b>B</b> | <b>C</b> | <b>D</b> | <b>E</b>    | <b>F</b>    | <b>G</b>   | <b>H</b> | <b>I</b> |
|-----------|-----------|----------|----------|----------|-------------|-------------|------------|----------|----------|
| <b>1</b>  | <b>5</b>  |          |          |          | 5983        | 6217        | 7044       | 7082     | 5891     |
| <b>2</b>  |           |          |          |          | 7803        | 7624        | 7843       | 8041     | 7794     |
| <b>3</b>  |           |          |          |          |             |             |            |          |          |
| <b>4</b>  | 6         | 7371     | 7941     |          | <b>1395</b> | 1197        | 341        | 306      | 1487     |
| <b>5</b>  | 6         | 7180     | 8011     |          | 1215        | <b>1038</b> | 216        | 102      | 1307     |
| <b>6</b>  | 6         | 6958     | 7571     |          | 1002        | 743         | <b>285</b> | 486      | 1090     |
| <b>7</b>  | 7         | 7212     | 7636     |          | 1240        | 995         | 267        | 425      | 1330     |
| <b>8</b>  | 7         | 6043     | 7922     |          | 133         | 345         | 1004       | 1046     | 199      |
| <b>9</b>  | 7         | 5891     | 7794     |          | 92          | 368         | 1154       | 1216     | 0        |
| <b>10</b> | 6         | 5935     | 7770     |          | 58          | 318         | 1111       | 1179     | 50       |
| <b>11</b> | 7         | 5922     | 7737     |          | 90          | 316         | 1127       | 1199     | 65       |
| <b>12</b> | 4         | 5887     | 7268     |          | 544         | 485         | 1292       | 1423     | 526      |
| <b>13</b> | 8         | 5963     | 7321     |          | 482         | 395         | 1200       | 1331     | 478      |
| <b>14</b> | 4         | 5940     | 7620     |          | 188         | 277         | 1126       | 1217     | 181      |
| <b>15</b> | 6         | 6317     | 7557     |          | 415         | 120         | 781        | 905      | 487      |
| <b>16</b> | <b>74</b> |          |          |          |             |             |            |          |          |

**E4** **=SQRT((E\$1-\$B4)^2+(E\$2-\$C4)^2)**

**F5** **=SQRT((F\$1-\$B5)^2+(F\$2-\$C5)^2)**

**G6**     **=SQRT((G\$1-\$B6)^2+(G\$2-\$C6)^2)**

496

**Step 3. Aggregating distances into frequencies per class.**

Given the distances between all points calculated in the previous step, we should now calculate the frequencies of distances within each distance class. The easiest way of doing this is first to calculate the cumulative numbers, because this requires only one criterion at a time. In table 2, the criterion applied for each column is given in row **1**. For example in row **4** in table 1, we find two distances between 300 and 400 m. and three distances between 1000 and 1500 m. In each case we multiply the frequency by the number ringed from that box (See formulas for cells **L4** and **M5** below the table). Row **16** gives us the sums in each column. In row **17** we transform the cumulative distribution into numbers per class, by subtracting the sum of the previous classes (see example formula for cell **M17**) below the table. In row **18**, the last step is to divide these total numbers of possible observations by the number of recruits, which gives us the average numbers of nestlings ringed in each distance class, averaged over the points where recruits were observed. One important check is that the sum of these nestling densities (in cell **V18**) must be equal to the total number of nestlings ringed from cell **A16** in Table 1. Errors in copying formulas are easily made, in particular when the size of the table has changed. One advantage of specifying the criteria in row **1** is that it becomes very easy to redo the calculations with different class boundaries.

518 **Table 2. Aggregating distances and calculating densities of nestlings ringed.**

|     | K                           | L    | M    | N    | O    | P    | Q    | R     | S     | T     | U      | V   |
|-----|-----------------------------|------|------|------|------|------|------|-------|-------|-------|--------|-----|
| 1   |                             | <100 | <200 | <300 | <400 | <600 | <800 | <1000 | <1500 | <2000 | <20000 |     |
| 2   |                             |      |      |      |      |      |      |       |       |       |        |     |
| 3   |                             |      |      |      |      |      |      |       |       |       |        |     |
| 4   |                             | 0    | 0    | 0    | 12   | 12   | 12   | 12    | 30    | 30    | 30     |     |
| 5   |                             | 0    | 6    | 12   | 12   | 12   | 12   | 12    | 30    | 30    | 30     |     |
| 6   |                             | 0    | 0    | 6    | 6    | 12   | 18   | 18    | 30    | 30    | 30     |     |
| 7   |                             | 0    | 0    | 7    | 7    | 14   | 14   | 21    | 35    | 35    | 35     |     |
| 8   |                             | 0    | 14   | 14   | 21   | 21   | 21   | 21    | 35    | 35    | 35     |     |
| 9   |                             | 14   | 14   | 14   | 21   | 21   | 21   | 21    | 35    | 35    | 35     |     |
| 10  |                             | 12   | 12   | 12   | 18   | 18   | 18   | 18    | 30    | 30    | 30     |     |
| 11  |                             | 14   | 14   | 14   | 21   | 21   | 21   | 21    | 35    | 35    | 35     |     |
| 12  |                             | 0    | 0    | 0    | 0    | 12   | 12   | 12    | 20    | 20    | 20     |     |
| 13  |                             | 0    | 0    | 0    | 8    | 24   | 24   | 24    | 40    | 40    | 40     |     |
| 14  |                             | 0    | 8    | 12   | 12   | 12   | 12   | 12    | 20    | 20    | 20     |     |
| 15  |                             | 0    | 6    | 6    | 6    | 18   | 24   | 30    | 30    | 30    | 30     |     |
| 16  | sums 1.                     | 40   | 74   | 97   | 144  | 197  | 209  | 222   | 370   | 370   | 370    |     |
| 17  | sums 2.                     | 40   | 34   | 23   | 47   | 53   | 12   | 13    | 148   | 0     | 0      | 370 |
|     | densitie                    |      |      |      |      |      |      |       |       |       |        |     |
| 18  | s                           | 8    | 6.8  | 4.6  | 9.4  | 10.6 | 2.4  | 2.6   | 29.6  | 0     | 0      | 74  |
| 19  |                             |      |      |      |      |      |      |       |       |       |        |     |
| 20  |                             |      |      |      |      |      |      |       |       |       |        |     |
| L4  | =A4*COUNTIF(\$E4:\$J4,L\$1) |      |      |      |      |      |      |       |       |       |        |     |
| M5  | =A5*COUNTIF(\$E5:\$J5,M\$1) |      |      |      |      |      |      |       |       |       |        |     |
| M17 | =M16-L16                    |      |      |      |      |      |      |       |       |       |        |     |
| P18 | =P17/\$A\$1                 |      |      |      |      |      |      |       |       |       |        |     |

**Step 4. Calculating averages over years.**

In step 3, we calculated the average densities of nestlings ringed in each distance class relative to the positions where new breeding birds were recruited for one single year, or rather one specific combination with a year of birth and a year of recruiting. We will normally have such data for a number of years and it is straightforward to calculate the average per distance class over the years. However, there are a number of choices to be made. One choice is the number of different (replicate) estimates we can make, another choice is whether or not to weight the average densities of nestlings ringed by the numbers of recruits whose dispersal distances were observed. When there is little variation in the densities and or the numbers recruited per year, weighted and unweighted averages will give the same results. When there is substantial variation among years, then weighting is a good idea, since the limitations on what can be observed should be related to the observations made.

**Step 5. Evaluating the number of classes.**

If it is possible to make replicate estimates, keeping the variation in densities within each estimate low is a secondary criterion. Most important are the numbers observed in each distance/time period class. Given the number of observations, we can either choose a finer resolution in space by choosing more distance classes, or in time by calculating more replicate estimates. As a rule of thumb, I would recommend that the average number of actual observations per class is on average at least 5 and that not more than 25 % of the cells has less than 5 observations. Given a spatial scale, the optimal number of replicates can also be determined by searching for the minimum in the standard errors of the recruitment estimates per distance class.

545    **Step 6. Calculating the recruitment rates.**

546    The recruitment rates are obtained by dividing the actual number of observations of  
547    individuals that have moved a particular distance by the average density of nestlings  
548    ringed in that class. The unit is thus observed recruit per nestling ringed.

549

550