



## Modelling large-scale dispersal distances

Emmanuel Paradis, Stephen R Baillie, William J Sutherland

### ► To cite this version:

Emmanuel Paradis, Stephen R Baillie, William J Sutherland. Modelling large-scale dispersal distances. Ecological Modelling, 2002, 151 (2-3), pp.279-292. 10.1016/S0304-3800(01)00487-2 . ird-02063958

**HAL Id: ird-02063958**

**<https://ird.hal.science/ird-02063958>**

Submitted on 11 Mar 2019

**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.

# Modelling large-scale dispersal distances

Emmanuel Paradis <sup>a</sup>, Stephen R. Baillie <sup>b</sup>, William J. Sutherland <sup>c</sup>

<sup>a</sup> *Laboratoire de Paléontologie, Institut des Sciences de l'Évolution, Université Montpellier  
II, 34095 Montpellier cédex 5, France*

<sup>b</sup> *British Trust for Ornithology, The Nunnery, Thetford, Norfolk, IP24 2PU, UK*

<sup>c</sup> *School of Biological Sciences, University of East Anglia, Norwich, NR4 7TJ, UK*

*Address for correspondence:*

Emmanuel Paradis  
Laboratoire de Paléontologie  
Institut des Sciences de l'Évolution  
Université Montpellier II  
34095 Montpellier cédex 5  
France

*phone:* +33 4 67 14 39 64

*fax:* +33 4 67 14 36 10

*e-mail:* paradis@isem.univ-montp2.fr

## Abstract

We present an approach to analyse dispersal distance data. This approach allows one to take into account accuracy of the recorded dispersal distances. Three distributions were used, all assuming continuous space; a maximum likelihood approach was used for parameter estimation and model selection. Numerical simulations showed that our method is statistically consistent since it selected the correct model with increasing frequency when sample size increased. Ringing data on two species of tits (*Parus caeruleus* and *P. major*) in Britain and Ireland were used to illustrate the potentialities of our method. In both species, adults dispersed significantly further than juveniles. The differences between species within an age-class were not statistically significant. In all species and age-classes, the model finally selected was the one assuming a heavy-tailed half-Cauchy distribution where long-distance dispersers are predicted to be more frequent than in the exponential model. The proposed methodology can potentially be applied to any organisms, and the model selection procedure can be used with any model of the distribution of dispersal distances. Several extensions are presented in the discussion, such as generalized linear modelling of the dispersal parameters, or interfacing with capture-recapture models.

**Keywords:** Dispersal; Estimation; Hypothesis testing; Likelihood; Model; Spatial scale.

## 1. Introduction

The last few years have seen an increasing interest in dispersal from ecologists and population biologists (Kot et al., 1996; Fagan, 1997; McCarthy, 1997). At least four reasons justify this focus of attention. First, the development of spatially-explicit population models which integrate the spatial structure of the landscape where plants and animals live underlines the importance of how they move in their natural habitat for the dynamics of their populations (Conroy et al., 1995). Second, theoretical studies have shown that the frequency distribution of dispersal distances affect invasion patterns and processes (Shaw, 1995; Clark, 1998; Clark et al., 1998; Higgins and Richardson, 1999). Third, dispersal is the vector of gene flow, and the relationship between dispersal and genetic differentiation among local populations was recognized a long time ago (Mayr, 1963); this also has impacts for considering the likely spread of genes from genetically modified crops. Fourth, it has been realized recently that dispersal has an impact on the way we estimate fitness in local populations: taking into account dispersal is necessary to better understand the processes underlying adaptation (Lambrechts et al., 1999).

There are two main approaches to modelling dispersal in biological populations. The first one considers the rates of movement among discrete populations. The second approach considers the frequency distribution of dispersal distances in a continuous space. In the first approach, the dispersal parameters can be estimated from capture-recapture or mark-resight data (Brownie et al., 1993). These parameters have a straightforward biological meaning as they are the probabilities of movement from a given population to

another. Their estimation requires a large quantity of data (Spendelov et al., 1995; Lindberg and Sedinger, 1998). In the second approach, the observed distribution of dispersal distances (DDD) is compared to an expected distribution derived from a model; we will distinguish two categories of such models. The first one can be called mechanistic because the expected DDD is derived from biological relationships involving parameters (e.g., seed weight and shape, wind direction) which have intuitive interpretations (Okubo and Levin, 1989; Nurminiemi et al., 1998). However, it seems that these models can be developed only for organisms with passive dispersal since similar models for actively dispersing animals are likely to include too many parameters to be tractable. In the second category of models, the DDD is described by a frequency curve involving few parameters; the shape of the curve depends on the value(s) of the parameter(s). In both categories of models, the parameters can be estimated from dispersal distance data.

The purpose of this paper is to present an approach for the modelling of DDD. Three simple models, derived from some probability density functions, with one or two parameters were used. A maximum likelihood method was used to estimate the parameters, to test which model best describes the data, and to test biological hypotheses. A probabilistic approach was used to account for uncertainty in the observations. The developed approach is general and applies to other models of dispersal. Some simulations to assess the performances of our approach were run. Finally, we analysed some data collected on two bird species.

This paper gives some emphasis on hypothesis testing. This may seem odd with respect to recent trends in statistical modelling where the emphasis seems to be on

parameter estimation. However, tests of hypotheses are particularly valuable in the present context since the biology of dispersal is still poorly known and testing biological hypotheses on dispersal is likely to be valuable to quantify this trait.

## **2. Statistical modelling of dispersal distances**

Like other field measurements, dispersal distances may be observed with errors. Two reasons why a dispersal distance is not known accurately are considered here. First, this may arise because of the accuracy of the data recording procedure. For instance, in most studies on dispersal of insects, distances are recorded in discrete sampling units that catch individuals (e.g., Mayer and Atzeni, 1993; Smith and Wall, 1998). In studies on plants, data on dispersal are usually recorded as the quantity of pollen or seeds dispersing a distance-class (Nathan and Muller-Landau, 2000). Second, a dispersal distance may be unknown but it may be known that it is greater than a given distance; this is likely to occur when a disperser has moved out of the study area, which is typical for birds, for instance.

Now define for dispersal data three kinds of observation. When the dispersal distance is precisely known, the observation is said to be *accurate*. When the observation falls within an interval, it is said to be *interval-censored*. When the distance dispersed is known to be greater than a distance given, the observation is said to be *censored* (by analogy with censoring in survival studies, Cox and Oakes, 1984).

A probability density function is appropriate to describe the theoretical DDD since, by definition, such a function sums to one when integrated over space. If we have a

probability density function  $f_{\theta}(d)$  which describes the expected DDD with some parameter(s)  $\theta$ , then it is possible to compute the probabilities of the events corresponding to the three kinds of precision just described. Note that these probabilities are not strictly equal but proportional to those given by the probability density function, but since they are subsequently used in a likelihood function, only their definition up to a constant is needed (Edwards, 1992). The probability of a dispersal distance  $D$  to be equal to  $d$  is given by:

$$Pr(D = d) = f_{\theta}(d).$$

The probability of  $D$  to fall between  $d'$  and  $d''$  is given by the integration on this interval:

$$Pr(d' \leq D \leq d'') = \int_{d'}^{d''} f_{\theta}(u) du.$$

Similarly, the probability of  $D$  to be greater than  $c$  is given by the integration between  $c$  and infinity:

$$Pr(D \geq c) = \int_c^{+\infty} f_{\theta}(u) du.$$

For a sample of  $n$  observations of dispersal distances, we have for the  $i$ th observation one of the followings: (i) an exact distance denoted  $d_i$ , (ii) an interval denoted  $(d'_i, d''_i)$ , or (iii) a censored distance denoted  $c_i$ . Let  $y_1$ ,  $y_2$ , and  $y_3$  be three indicator variables such that for the  $i$ th observation  $y_{1i} = 1$  if it is accurate ( $y_{1i} = 0$  otherwise),  $y_{2i} = 1$  if it is interval-censored ( $y_{2i} = 0$  otherwise), and  $y_{3i} = 1$  if it is censored ( $y_{3i} = 0$  otherwise). The likelihood of the sample for the model of DDD defined by  $f_{\theta}$  is:

$$L = \prod_{i=1}^n [y_{1i} f_{\theta}(d_i) + y_{2i} \int_{d_i'}^{d_i''} f_{\theta}(u) du + y_{3i} \int_{c_i}^{+\infty} f_{\theta}(u) du].$$

If the cumulative density function  $F_{\theta}$  corresponding to  $f_{\theta}$  is known, it is possible to re-write the likelihood as:

$$L = \prod_{i=1}^n \{y_{1i} f_{\theta}(d_i) + y_{2i} [F_{\theta}(d_i'') - F_{\theta}(d_i')] + y_{3i} [1 - F_{\theta}(c_i)]\}. \quad (1)$$

This paper considers three distributions: exponential, half-Cauchy, and Weibull.

This restriction is mainly due to the fact that we studied birds for which mechanistic models are likely to have too many parameters. However, as mentioned in the Introduction, our approach is general, and the algorithms and computer programs (see below) were written in a way that they can easily be extended to other models of DDD.

The three distributions have one (exponential, half-Cauchy) or two parameters (Weibull). The exponential distribution has a biological interpretation in the present context. Consider an animal or a seed moving in a landscape: if it has a constant probability of settling then the dispersal distances will follow an exponential distribution with rate parameter  $\lambda$  equal to this probability. The Cauchy distribution is unimodal and symmetric (Kotz and Johnson, 1982). The half-Cauchy distribution is derived from the standard Cauchy distribution by mirroring the curve on the origin so that only positive values can be observed. As a heavy-tailed distribution, the half-Cauchy has been used as an alternative to the exponential to model theoretical DDD (Shaw, 1995), since the former predicts more frequent long-distance dispersal events than the latter (Fig. 1). The Weibull distribution is a generalization of the exponential distribution (Antle and Bain, 1988). It has an additional

parameter (the shape  $\beta$ ) determining how the probability of settling varies: this probability increases (decreases) with increasing distance for  $\beta > (<) 1$ . If  $\beta = 1$ , then the Weibull distribution reduces to the exponential one. The probability density function  $f$  and the cumulative density function  $F$  for each distribution are given in Table 1.

<Table 1 near here>

All computations were programmed in the statistical language R (Ihaka and Gentleman, 1996); the codes are available upon request from the first author. R is an interpreted language derived mainly from the S language (Becker et al., 1988) which has many built-in functions for statistical analyses and computing (CRAN, 2001). The likelihood (Eq. 1) was transformed as the deviance:  $\text{Dev} = -2\ln L$ . Thus, finding the maximum likelihood was similar to minimizing the deviance: this was done with the nonlinear minimization function of R which uses a Newton-type algorithm for unconstrained minimization (Schnabel et al., 1985). This method allows one to find the minimum of a function even when its partial and second derivatives are unknown, which is particularly useful for complicated expressions such as the likelihood function in Eq. 1. However, these derivatives can be computed numerically. The values of the parameters which maximize the likelihood are the maximum likelihood estimates (MLEs). The standard-errors of the MLEs can be computed using the second derivatives of the likelihood function at its maximum under the assumption that this function is normal. However, it is known that the likelihood function may be asymmetric when there are censored data (Cox

and Oakes, 1984). An alternative is to compute directly the confidence interval of the MLE on the likelihood function (Hudson, 1971). Since twice the difference between two log-likelihoods must be at least 3.84 to be significant with  $P = 0.05$ , then the area below 1.92 units of the observed maximum  $\ln L$  defines a 95% confidence interval of the MLEs. A third method is to use a bootstrap approach: this was implemented giving equal probabilities of resampling for all observations whatever their precision (see Efron, 1981).

The deviance residuals and the interval lengths can be used as diagnostics of model adequacy and contributions of the individual observations (Farrington, 2000). The deviance residual  $r_i$  is the square root of the contribution of the  $i$ th observation to the deviance function. Unlike in a regression model, the deviance residuals are here all positive and thus not distributed around an expected value. However, residuals after fitting the data to two different distributions can be plotted against each others to see which observations are most influential to the difference of fit between both distributions (see application below). The interval length  $I_i$  is defined as: (i)  $I_i = 0$  if the  $i$ th observation is accurate, (ii)  $I_i = d_i'' - d_i'$  if the  $i$ th observation is interval-censored, (iii)  $I_i = -1$  if the  $i$ th observation is censored. If the interval lengths vary among observations, a plot of these against the deviance residuals may reveal some possibly influential observations.

Using distributions such as those we have described above, it is possible to build models that can be fitted to dispersal data. For instance, if one is interested in testing for sex-biased dispersal (Greenwood, 1980), we can consider two models: the first one with different parameters for the two sexes, and the second with the same parameter for both sexes. The second model is nested in the first one, and both models can be compared with a

likelihood ratio test (LRT) which is twice the difference in  $\ln L$  of both models. Under the null hypothesis of no difference in both categories, this test follows a  $\chi^2$  distribution with a number of degrees of freedom  $df = \Delta p$ , where  $\Delta p$  is the difference in numbers of parameters of both models. For example, to test for a difference in dispersal between sexes assuming an exponential distribution, one will build two models: the first one with a common parameter for both sexes estimated by fitting the exponential distribution to the pooled data, and the second one with two parameters estimated by fitting the distribution separately for both sexes. Both models will be compared with a LRT with  $df = 1$ : if the test is significant, then one will reject the null hypothesis that both sexes have the same DDD.

If several models are not nested, they can be compared with the Akaike Information Criterion (AIC, Akaike, 1973). It is computed for each model as  $AIC = -2\ln L + 2p$ , where  $p$  is the number of parameters in the model. The model with the smallest AIC is accepted as the one that best describes the data.

### **3. Simulation study**

#### *3.1 Methods*

<Fig. 1 near here>

An important current issue in dispersal theory is whether long-distance dispersers are more frequent than expected under the exponential distribution. If true, the half-Cauchy

distribution with its heavy tail would be an appropriate model to describe DDDs. Fig. 1 shows the theoretical curves for the exponential distribution with rate  $\lambda = 0.5$ , and for the half-Cauchy distribution with scale  $\sigma = 1.8$ . The similarity of the two curves is obvious. However, the half-Cauchy distribution predicts more frequent long-distance dispersers than the exponential one. Thus a central question was: can the AIC-based approach find the distribution that generated a sample of dispersal distances in a realistic sampling frame? We considered the curves on Fig. 1. Two sampling frames were simulated with respect to spatial scale. In both cases, random samples were generated from one of both distributions. On the small spatial scale (0–5 km), if the generated observation was less than or equal to 5 km, then it was considered as accurate; if it was greater than 5 km, then it was considered as censored (thus, even the individuals that dispersed outside the study area were taken into account in the analysis). On the large spatial scale (0–500 km), if the generated observation was less than or equal to 500 km, then it was considered as interval-censored using the two nearest integer values; if it was greater than 500 km, then it was considered as censored. We considered four sample sizes (50, 100, 200 and 400), and replicated the simulations for each set of parameters 1000 times. The exponential and the half-Cauchy distributions were fitted to each sample: the two AIC values were stored for comparisons.

<Fig. 2 near here>

The problem of heterogeneity in dispersal may be illustrated with a hypothetical population composed of two types of individuals dispersing according to an exponential

distribution but with different rates (see Rees, 1993). Then, if heterogeneity is ignored, the population mean rate will appear as decreasing with distance (because the individuals with the lowest rate are settling first), even though the parameters are constant. The question was: can our method detect such heterogeneity if it is present in the population? A population with two types of dispersers was simulated using two Weibull distributions: the first type had a decreasing probability of settling with distance ( $\beta < 1$ ) and were called philopatric, and the second type had an increasing probability of settling with distance ( $\beta > 1$ ) and were called dispersers. We considered three pairs of Weibull distributions with a high (0.5, 2), medium (0.65, 1.5), or low (0.8, 1.2) difference in the shape  $\beta$ . In all cases, we took the scale  $\delta = 1$ . Fig. 2 shows the contrast for the three pairs of distributions on the range 0–5 km; all curves look very similar when plotted on the range 0–500 km. It should be noted that the tail of the curve for a Weibull distribution becomes heavier with smaller  $\beta$ . We considered the same sampling scales (0–5 km, and 0–500 km), sample sizes (50, 100, 200, and 400), and number of replications (1000) than above. We assumed that both types of individuals (philopatric and disperser) could be identified (e.g., sex), and that each sample was composed of half of each type. Four models were fitted to the simulated data sets: (i) assuming an exponential distribution, (ii) assuming a half-Cauchy distribution, (iii) assuming a Weibull distribution (ignoring heterogeneity in all these three models), and (iv) assuming a Weibull distribution with different parameters for both categories (i.e. the true model). The fitted models had one, one, two, and four parameters, respectively. The AIC values were computed and compared.

### 3.2 Results

<Table 2 near here>

On the range 0–5 km, when the data were simulated with the half-Cauchy distribution, the correct model was detected in 85% of the cases with the smallest sample size (Table 2). This figure increased with increasing sample size to reach 100% when  $n = 400$ . On the other hand, when the simulated data were from the exponential distribution, the performance were less satisfying, though still statistically consistent as the percentage of correct model selection increased with increasing sample size (Table 2). On the range 0–500 km, the performance of model selection were very good reaching more than 95% when  $n = 50$ , and 100% for  $n \geq 200$ .

<Table 3 near here>

For the simulations where heterogeneity was present, the efficiency of model selection generally increased with increasing sample size and/or increased contrast between the shape parameters (Table 3). The “exponential” model provided a good fit when the contrast between the shapes was low which was due to the fact that both shapes were close to one, and thus both Weibull distributions were well approximated by an exponential one. On the other hand, the “half-Cauchy” model fitted well to the data when the contrast between the shapes was high, particularly on the range 0–500 km. This is due to the fact

that the Weibull distribution with  $\beta = 0.5$  has a relatively heavy tail which is well fitted by the half-Cauchy distribution; however, this artifact disappears progressively with increasing sample size. Note that the Weibull model assuming homogeneity provided the best fit (averaged over the replications) to the data in no cases.

#### **4. Application: dispersal in tits**

There is considerable interest in the dynamics of bird populations at a large spatial scale where long distance dispersal is likely to have a critical rôle in population dynamics (Baillie et al., 2000). Most field studies are spatially limited to quantify dispersal distances, we thus analysed data from a large scale ringing programme. Since such data are likely to be recorded with some inaccuracies, it was essential to take them into account in the analyses.

##### *4.1 Methods*

The Ringing Scheme of the British Trust for Ornithology (BTO) has been run since 1909: birds have been ringed by volunteers in Britain and Ireland since this time (Clark et al., 1996). Recoveries are largely by members of the general public, often when they find dead birds. Distances moved by recovered birds between ringing and recovery sites are stored in computerized databases to the nearest kilometer. Unlike most other studies of dispersal, this data set has the advantage that the relocation is not restricted to a limited

number of sites. We considered here two species: the blue tit (*Parus caeruleus*) and the great tit (*P. major*). This choice was directed by the quantity of data for these two species which are very abundant in the British Isles (Gibbons et al., 1993), but also because they are extremely sedentary so that winter movements are unlikely to contaminate significantly the dispersal data. For both species, we considered ring recoveries from 1948 to 1994. Only birds ringed and recovered during breeding season (between April and July), and recovered at least one year after ringing were used (see Paradis et al., 1998 for details on data selection). Two age-classes were defined with respect to the age of the birds when ringed: juvenile for birds ringed in their year of birth (either as nestlings or as fledglings), and adult for birds ringed later.

All data were considered as interval-censored: we took the value stored in the databases as the lower bound of the interval ( $d'$ ), and this value plus one as the upper bound of the interval ( $d''$ ).

The four data sets were first analysed separately, then different possibilities of pooling (among species, or age-classes) were considered. We had some data on sex for the great tits ringed as adults, and tested for a possible difference between males and females in breeding dispersal.

## 4.2 Results

<Fig. 3, Fig. 4, and Table 4 near here>

Figure 3 shows the frequencies of the observed dispersal distances for all four data sets. In all four data sets, the half-Cauchy distribution was selected by the AIC (Table 4). The difference in AIC values with the Weibull distribution was slight, however, particularly for juveniles. Plots of deviance residuals of these two distributions show that the observations that were most influential on this difference of fit were those with the longer distances (and thus the greater residuals, Fig. 4). Since all interval lengths were equal, it was possible to compute for each observed distance the expected frequency at this particular distance under the fitted half-Cauchy distribution: the plots of these expected frequencies against the observed ones suggest a good agreement between them (Fig. 5).

<Fig. 5, and Table 5 near here>

The estimates of the scale parameters show some similarities across species for both age-classes (Table 5). The 95% confidence intervals were globally congruent, though the normal approximation yielded the smallest intervals, and the bootstraps the larger ones. The profile likelihood and bootstrap intervals were slightly asymmetric. We pooled the data of both species separately for juveniles and adults, and fitted the half-Cauchy distribution: the LRTs were computed to test for difference between species within each age-class. The differences were not significant, either for juveniles ( $\chi^2 = 0.17$ ,  $df = 1$ ,  $P = 0.677$ ), or for adults ( $\chi^2 = 0.01$ ,  $df = 1$ ,  $P = 0.920$ ). Within species, the difference between age-classes was statistically significant:  $\chi^2 = 80.19$ ,  $df = 1$ ,  $P < 0.0001$ , for blue tits, and  $\chi^2 = 62.69$ ,  $df = 1$ ,  $P < 0.0001$ , for Great Tits.

Identification of sex was done on 152 great tits (83 males and 69 females). The estimated scale was slightly greater for males ( $= 0.640$ ,  $SE = 0.076$ , 95% profile likelihood CI: 0.459-0.885), than for females ( $= 0.460$ ,  $SE = 0.060$ , 95% profile likelihood CI: 0.343-0.577), but the difference was not significant ( $\chi^2 = 1.766$ ,  $df = 1$ ,  $P = 0.183$ ).

## **5. Discussion**

The likelihood approach of model fitting is, certainly in the present context, superior to the least squares method because the latter is scale-dependent and yields different results when the scale of the data is changed (Edwards, 1992). Hence, fitting a model to dispersal data by least squares is likely to depend on the spatial scale of the study. No such abnormalities are expected with the likelihood method. Colbach and Satche (2001) fitted a Weibull distribution to seed dispersal data, but they used a least squares approach. Baker et al. (1995) presented a Bayesian method that corrects mean dispersal distance using data collected on a limited study area, but this method actually ignores long-distance dispersers.

Wu et al. (2000) studied a two dimensional random walk model of animal movement for which they derived the expected mean dispersal distance. Comparing this expected mean with a population observed mean is a more approximative approach than computing the likelihood. However, it seems possible to derive the probability distributions of dispersal distances from Wu et al.'s (2000) model, making thus possible the likelihood approach presented here.

Simulations in our study were focused on model selection, since the most important

problem in the study of dispersal is that of hypothesis testing. The simulation study showed that the model selection procedure was statistically consistent (the proportion of correctly identified model increased with increasing sample size), and had power to detect long-distance dispersal, and power to detect heterogeneity. When the latter was present, statistical consistency was also observed as the proportion of correctly selected models was positively associated with the contrast between both categories of dispersers. It should be noted that good performance (see Tables 2 and 3) was observed when a large spatial scale was simulated even though the data were considered as only accurate to the nearest kilometer. The results from these simulations served to illustrate some properties of the method. From a practical point of view, they cannot serve directly as a guide to choose sampling sizes to optimize statistical power for a particular study; however, these can be very easily assessed in R using our functions and some very simple programming.

There are two straightforward extensions of our approach. The first one is in using other distributions to describe the theoretical DDD. These distributions can be either simple ones (such as those developed here: e.g. gamma, normal, or lognormal), or mechanistic such as those developed for plants (see references in the Introduction). The statistical approach presented here can be used provided the theoretical DDD can be described in probabilistic terms. Such distributions can be obtained from Monte Carlo simulations.

The second extension is a generalized linear modelling of the parameters of an exponential or a half-Cauchy distribution in order to test biological hypotheses (McCullagh and Nelder, 1989). For instance, it is simple to fit an exponential distribution where the rate parameter depends, using a logarithmic link, on covariate(s)  $X$  such as:  $\ln\lambda = \gamma X + \zeta$ , where

$\gamma$  is a coefficient and  $\zeta$  an intercept, both to be estimated. If  $\gamma = 0$ , then this model reduces to the exponential distribution used in this paper with rate  $\lambda = \exp(\zeta)$ ; so the two models are nested and can be compared with a LRT. The same approach is apparently possible with the half-Cauchy distribution, however, there are theoretical considerations that GLMs cannot be fitted to data which are Cauchy distributed (McCullagh and Nelder, 1989, p.20, see also Lindsey, 1996). On the other hand, there are statistical approaches to the regression analysis of data when the dispersion of residuals is higher than expected with the normal distribution, which is the case with the Cauchy distribution (Ronchetti, 1997). Further investigation in this area of research is definitely needed.

Another extension, which seems less easy, would be to combine our approach on modelling dispersal distances with capture-recapture (Pollock et al., 1990) and ringing-recovery (Brownie et al., 1985) probabilistic models. The aim of these models is to estimate survival probabilities taking into account the recapture and/or resighting probabilities. Pollock (1982) developed a sampling design that can estimate separately local recruitment from immigration, but this requires some assumptions which may be difficult to check (Pollock et al., 1993). Generally, capture-recapture models cannot estimate separately local survival and emigration. It may seem possible to include the DDD in the recapture parameters of a capture-recapture model, and so constrain the recapture probabilities with respect to the frequency of dispersal movements. Further studies are needed to make this approach feasible. A more straightforward approach may be to combine a model of DDD with a multi-strata capture-recapture model (Brownie et al., 1993) to include constraints on transition probabilities between strata with respect to the distance separating them.

The results from our analyses on dispersal distances of two species of tits revealed similar patterns of natal and breeding dispersal. The half-Cauchy distribution was preferred in all cases. There was a marked contrast between adults and juveniles, meaning that the well-documented difference between natal and breeding dispersal on small-scale study areas (e.g., Greenwood and Harvey, 1982) has some large-scale consequences for the dispersal of individuals. We note the observed difference, though not significant, with respect to sex in the DDD of adult great tits: males had a greater estimated scale than females. This is in contrast with local scale studies (Greenwood, 1980) which recorded greater dispersal distances in females compared to males for this species. It could be that the mechanisms acting on dispersal at a local scale (such as inbreeding avoidance, Greenwood and Harvey, 1982) are not relevant at a large-scale, thus explaining the lack of significant difference between sexes with our data.

If it appears that the half-Cauchy distribution adequately describes DDD in real situations (as suggested by our analyses with the British tits data), then this suggests a reconsideration of comparisons of mean dispersal distances estimated from different studies (as usually done in comparative analyses, e.g. Cain et al., 1998) if there are some discrepancies in the areas sampled.

## **Acknowledgements**

The authors are grateful to the many volunteers who contributed to the BTO ringing data, to the Ringing Unit of the BTO for their work in managing these data, and to Steven

Freeman, Marcel Lambrechts, and two anonymous referees for very helpful comments on an earlier draft of the manuscript. Special thanks to David Vaughan, Kamal Ibrahim, and Rob Freckleton for discussions on the issue. Financial support for the present study was provided by NERC grant GST/02/1197 as part of the NERC/SOAFED special topic on Large Scale Processes in Ecology and Hydrology. The post held by Stephen Baillie is supported by a contract from the Joint Nature Conservation Committee on behalf of English Nature, the Countryside Council for Wales, Scottish Natural Heritage and the Environment and Heritage Service in Northern Ireland. This is publication 01-131 of the Institut des Sciences de l'Évolution (Unité Mixte de Recherche 5554 du Centre National de la Recherche Scientifique).

## References

- Akaike, H., 1973. Information theory and an extension of the maximum likelihood principle. In: B.N. Petrov and F. Csaki (Editors), Proceedings of the second international symposium on information theory. Akadémia Kiado, Budapest, pp. 267–281.
- Antle, C.E. and Bain, J.L., 1988. Weibull distribution. In: S. Kotz and N.L. Johnson (Editors), Encyclopedia of statistical sciences. Volume 9. John Wiley & Sons, New York, pp. 549–556.
- Baillie, S.R., Sutherland, W.J., Freeman, S.N., Gregory, R.D. and Paradis, E., 2000. Consequences of large-scale processes for the conservation of bird populations.

- Journal of Applied Ecology, 37: 88–102.
- Baker, M., Nur, N. and Geupel, G.R., 1995. Correcting biased estimates of dispersal and survival due to limited study area: theory and an application using wrentits. *Condor*, 97: 663–674.
- Becker, R.A., Chambers, J.M. and Wilks, A.R., 1988. *The New S Language*. Chapman & Hall, London.
- Brownie, C., Anderson, D.R., Burnham, K.P. and Robson, D.S., 1985. Statistical inference from band recovery data—a handbook (second edition). U.S. Fish and Wildlife Service Resource Publication, 156: 1–305.
- Brownie, C., Hines, J.E., Nichols, J.D., Pollock, K.H. and Hestbeck, J.B., 1993. Capture-recapture studies for multiple strata including non-markovian transitions. *Biometrics*, 49: 1173–1187.
- Cain, M.L., Damman, H. and Muir, A., 1998. Seed dispersal and the Holocene migration of woodland herbs. *Ecological Monographs*, 68: 325–347.
- Clark, J.A., Adams, S.Y., Peach, W.J. and Simons, J.R., 1996. Report on bird ringing in Britain and Ireland for 1994. *Ringling & Migration*, 17: 36–79.
- Clark, J.S., 1998. Why trees migrate so fast: confronting theory with dispersal biology and the paleorecord. *American Naturalist*, 152: 204–224.
- Clark, J.S. et al., 1998. Reid’s paradox of rapid plant migration. Dispersal theory and interpretation of paleoecological records. *BioScience*, 48: 13–24.
- Colbach, N. and Sache, I., 2001. Blackgrass (*Alopecurus myosuroides* Huds.) seed dispersal from a single plant and its consequences on weed infestation. *Ecological modelling*,

139: 201–219.

- Conroy, M.J., Cohen, Y., James, F.C., Matsinos, Y.G. and Maurer, B.A., 1995. Parameter estimation, reliability, and model improvement for spatially explicit models of animal populations. *Ecological Applications*, 5: 17–19.
- Cox, D.R. and Oakes, D., 1984. *Analysis of survival data*. Chapman and Hall, London.
- CRAN, 2001. Comprehensive R Archive Network. URL: <http://www.R-project.org/>,
- Edwards, A.W.F., 1992. *Likelihood* (expanded edition). Johns Hopkins University Press, Baltimore.
- Efron, B., 1981. Censored data and the bootstrap. *Journal of the American Statistical Association*, 76: 312–319.
- Fagan, W.F., 1997. Introducing a “boundary-flux” approach to quantifying insect diffusion rates. *Ecology*, 78: 579–587.
- Farrington, C.P., 2000. Residuals for proportional hazards models with interval-censored survival data. *Biometrics*, 56: 473–482.
- Gibbons, D.W., Reid, J.B. and Chapman, R.A., 1993. *The new atlas of breeding birds in Britain and Ireland: 1988–1991*. T. & A.D. Poyser, London.
- Greenwood, P.J., 1980. Mating systems, philopatry and dispersal in birds and mammals. *Animal Behaviour*, 28: 1140–1162.
- Greenwood, P.J. and Harvey, P.H., 1982. The natal and breeding dispersal of birds. *Annual Review of Ecology and Systematics*, 13: 1–21.
- Higgins, S.I. and Richardson, D.M., 1999. Predicting plant migration rates in a changing world: the role of long-distance dispersal. *American Naturalist*, 153: 464–475.

- Hudson, D.J., 1971. Interval estimation from the likelihood function. *Journal of the Royal Statistical Society. Series B. Methodological*, 33: 256–262.
- Ihaka, R. and Gentleman, R., 1996. R: a language for data analysis and graphics. *Journal of Computational and Graphical Statistics*, 5: 299–314.
- Kot, M., Lewis, M.A. and van den Driessche, P., 1996. Dispersal data and the spread of invading organisms. *Ecology*, 77: 2027–2042.
- Kotz, S. and Johnson, N.L., 1982. Cauchy distribution. In: S. Kotz and N.L. Johnson (Editors), *Encyclopedia of statistical sciences*. Volume 1. John Wiley & Sons, New York, p. 386.
- Lambrechts, M.M. et al., 1999. Will estimates of lifetime recruitment of breeding offspring on small-scale study plots help us to quantify processes underlying adaptation? *Oikos*, 86: 147–151.
- Lindberg, M.S. and Sedinger, J.S., 1998. Ecological significance of brood-site fidelity in Black Brant: spatial, annual, and age-related variation. *Auk*, 115: 436–446.
- Lindsey, J.K., 1996. *Parametric statistical inference*. Clarendon Press, Oxford.
- Mayer, D.G. and Atzeni, M.G., 1993. Estimation of dispersal distances for *Cochliomyia hominivorax* (Diptera, Calliphoridae). *Environmental Entomology*, 22: 368–374.
- Mayr, E., 1963. *Animal species and evolution*. Belknap Press of Harvard University Press, Cambridge, Massachusetts.
- McCarthy, M.A., 1997. Competition and dispersal from multiple nests. *Ecology*, 78: 873–883.
- McCullagh, P. and Nelder, J.A., 1989. *Generalized linear models* (second edition).

Chapman & Hall, London.

- Nathan, R. and Muller-Landau, H.C., 2000. Spatial patterns of seed dispersal, their determinants and consequences for recruitment. *Trends in Ecology & Evolution*, 15: 278–285.
- Nurminiemi, M., Tufto, J., Nilsson, N.-O. and Rognli, O.A., 1998. Spatial models of pollen dispersal in the forage grass meadow fescue. *Evolutionary Ecology*, 12: 487–502.
- Okubo, A. and Levin, S.A., 1989. A theoretical framework for data-analysis of wind dispersal of seeds and pollen. *Ecology*, 70: 329–338.
- Paradis, E., Baillie, S.R., Sutherland, W.J. and Gregory, R.D., 1998. Patterns of natal and breeding dispersal in birds. *Journal of Animal Ecology*, 67: 518–536.
- Pollock, K.H., 1982. A capture-recapture design robust to unequal probability of capture. *Journal of Wildlife Management*, 46: 752–757.
- Pollock, K.H., Kendall, W.L. and Nichols, J.D., 1993. The “robust” capture-recapture design allows components of recruitment to be estimated. In: J.-D. Lebreton and P.M. North (Editors), *Marked individuals in the study of bird population*. Birkhäuser Verlag, Basel, pp. 245–252.
- Pollock, K.H., Nichols, J.D., Brownie, C. and Hines, J.E., 1990. Statistical inference for capture-recapture experiments. *Wildlife Monographs*, 107: 1–97.
- Rees, M., 1993. Null models and dispersal distributions: a comment on an article by Caley. *American Naturalist*, 141: 812–815.
- Ronchetti, E., 1997. Robustness aspects of model choice. *Statistica Sinica*, 7: 327–338.
- Schnabel, R.B., Koontz, J.E. and Weiss, B.E., 1985. A modular system of algorithms for

- unconstrained minimization. *ACM Transactions on Mathematical Software*, 11: 419-440.
- Shaw, M.W., 1995. Simulation of population expansion and spatial pattern when individual dispersal distributions do not decline exponentially with distance. *Proceedings of the Royal Society of London. Series B. Biological Sciences*, 259: 243–248.
- Smith, K.E. and Wall, R., 1998. Estimates of population density and dispersal in the blowfly *Lucilia sericata* (Diptera: Calliphoridae). *Bulletin of Entomological Research*, 88: 65–73.
- Spendelov, J.A. et al., 1995. Estimating annual survival and movement rates of adults within a metapopulation of Roseate terns. *Ecology*, 76: 2415–2428.
- Wu, H., Li, B.-L., Springer, T.A. and Neill, W.H., 2000. Modelling animal movement as a persistent random walk in two dimensions: expected magnitude of net displacement. *Ecological Modelling*, 132: 115–124.

Table 1

The three distributions used in this paper to model dispersal distances

Distribution	Probability density function	Cumulative density function	Parameter(s) (name)
Exponential	$\lambda e^{-\lambda d}$	$1 - e^{-\lambda d}$	$\lambda$ (rate)
Half-Cauchy	$2\sigma/\pi(\sigma^2 + d^2)$	$(2/\pi)\arctan(d/\sigma)$	$\sigma$ (scale)
Weibull	$\beta/\delta(d/\delta)^{\beta-1}\exp[-(d/\delta)^\beta]$	$1 - \exp[-(d/\delta)^\beta]$	$\beta$ (shape), $\delta$ (scale)

Table 2

Simulation results. CAU and EXP give the number of simulations where the half-Cauchy and exponential distributions, respectively, were accepted out of the 1000 replicates;  $n$ : sample size. The distance ranges (0–5 km and 0–500 km) are the simulated sampling areas (see more details on the simulation procedure in the text).

Simulated distributions	$n$	0–5 km		0–500 km	
		CAU	EXP	CAU	EXP
Half-Cauchy	50	853	147	968	32
	100	921	79	997	3
	200	983	17	1000	0
	400	1000	0	1000	0
Exponential	50	267	733	43	957
	100	165	835	16	984
	200	124	876	0	1000
	400	65	935	0	1000

Table 3

Simulation results when heterogeneity was present in the data. CAU, EXP, W(1) and W(2) give the number of simulations where the half-Cauchy, exponential, Weibull (assuming homogeneity), and Weibull (assuming heterogeneity) “models”, respectively, were accepted out of the 1000 replicates; shapes: values of  $\beta$  used to generate the two Weibull samples;  $n$ : sample size. The distance ranges (0–5 km and 0–500 km) are the simulated sampling areas (see more details on the simulation procedure in the text).

Shapes	$n$	0–5 km				0–500 km			
		CAU	EXP	W(1)	W(2)	CAU	EXP	W(1)	W(2)
0.5, 2	50	0	0	0	1000	67	19	8	906
	100	0	0	0	1000	105	0	0	895
	200	0	0	0	1000	195	0	0	805
	400	0	0	0	1000	319	0	0	681
0.65, 1.5	50	6	39	1	954	16	191	49	744
	100	0	0	0	1000	2	45	3	950
	200	0	0	0	1000	3	0	0	996
	400	0	0	0	1000	5	0	0	995
0.8, 1.2	50	30	472	65	433	10	619	115	256
	100	3	268	29	700	0	401	90	509
	200	0	47	7	946	0	191	37	772
	400	0	1	0	999	0	27	5	968

Table 4

Model fitting results for two species of birds in Britain and Ireland.

<b>Species</b>	<b>Age-class</b>	<b><i>n</i></b>	<b>Model</b>	<b>AIC</b>
Blue Tit	adult	201	exponential	856
			half-Cauchy	583
			Weibull	692
	juvenile	703	exponential	3965
			half-Cauchy	3558
			Weibull	3585
Great Tit	adult	173	exponential	761
			half-Cauchy	508
			Weibull	602
	juvenile	560	exponential	3179
			half-Cauchy	2757
			Weibull	2819

Table 5

Maximum likelihood estimates of the scale parameters of the half-Cauchy distribution for two species of birds in Britain and Ireland.

Species	Age-class	$\hat{\sigma}$	95% confidence intervals		
			Normal approximation	Profile likelihood	Bootstrap
Blue Tit	adult	0.525	0.446–0.604	0.423–0.649	0.451–0.623
	juvenile	1.604	1.462–1.746	1.415–1.818	1.361–1.900
Great Tit	adult	0.561	0.471–0.652	0.445–0.704	0.474–0.670
	juvenile	1.619	1.466–1.772	1.415–1.851	1.380–1.923

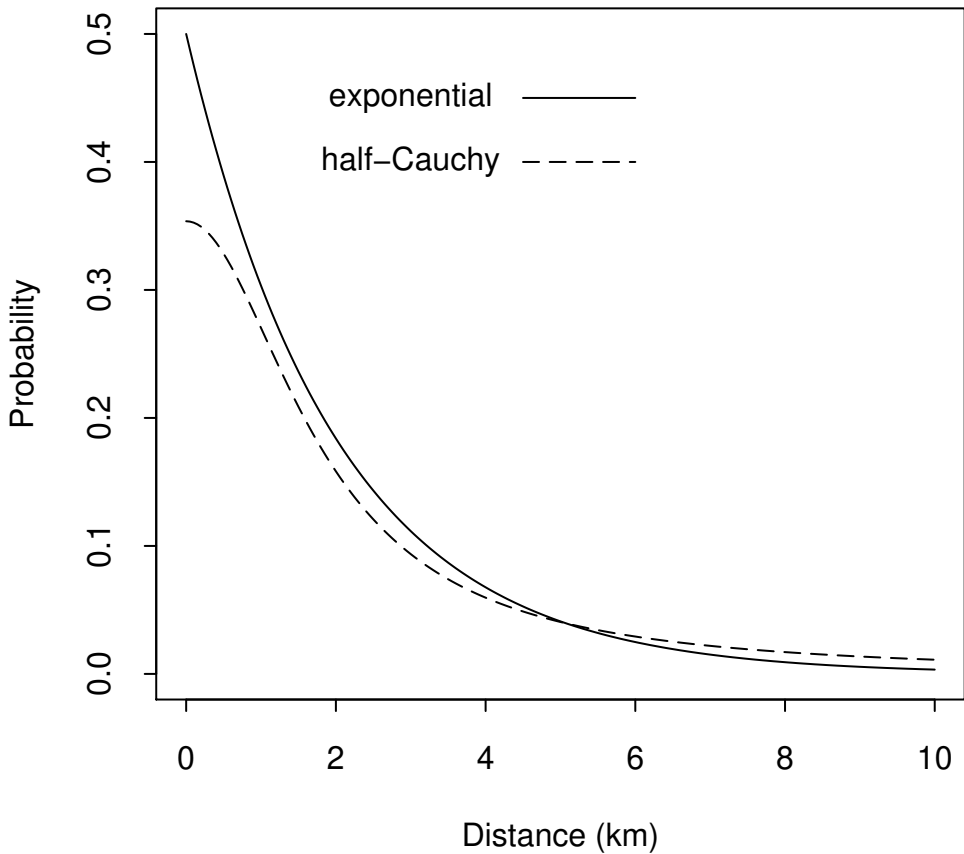
Fig. 1. Half-Cauchy and exponential distributions with scale  $\sigma = 1.8$  and rate  $\lambda = 0.5$ , respectively.

Fig. 2. Three pairs of Weibull distributions with different shapes  $\beta$ , the scale is the same in all cases ( $\delta = 1$ ).

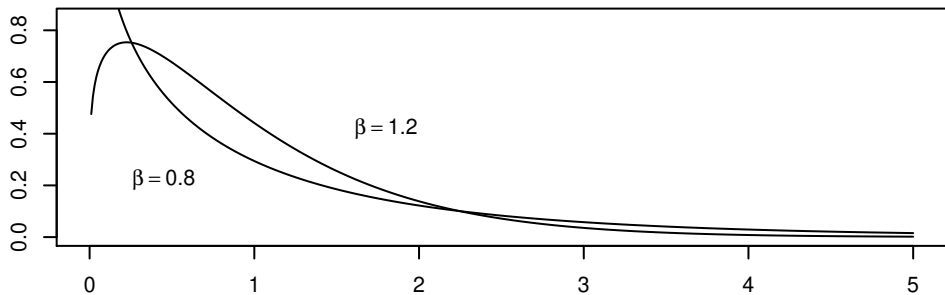
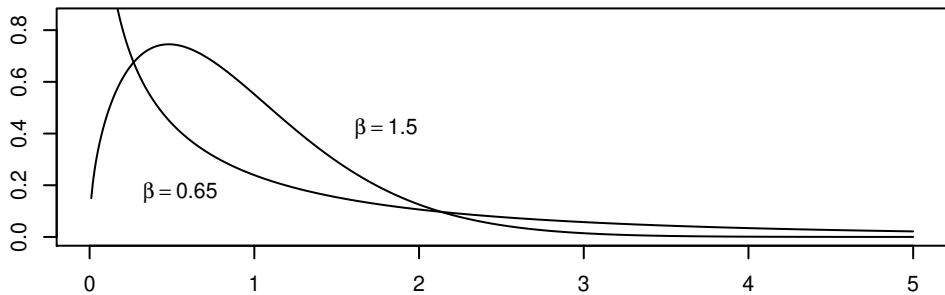
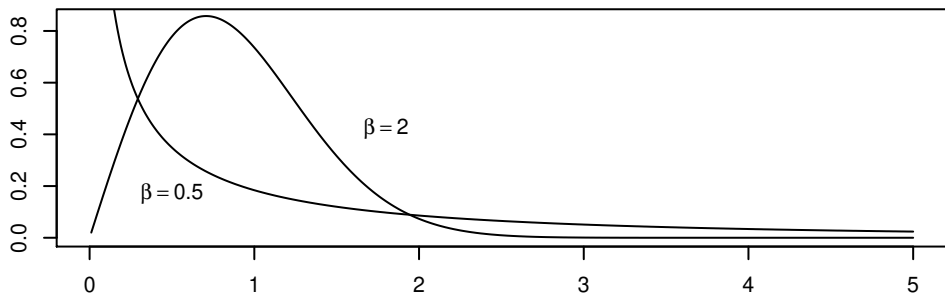
Fig. 3. Observed frequencies (numbers) of dispersal distances for two species of birds in Britain and Ireland. A: adult blue tits, B: juvenile blue tits, C: adult great tits, D: juvenile great tits.

Fig. 4. Deviance residuals for two fitted distributions (half-Cauchy and Weibull). A: adult blue tits, B: juvenile blue tits, C: adult great tits, D: juvenile great tits. Dotted lines are  $x = y$ .

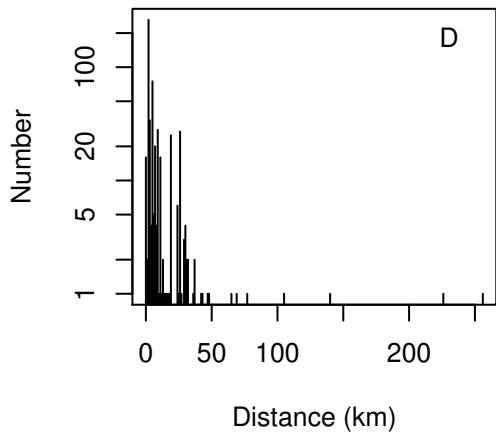
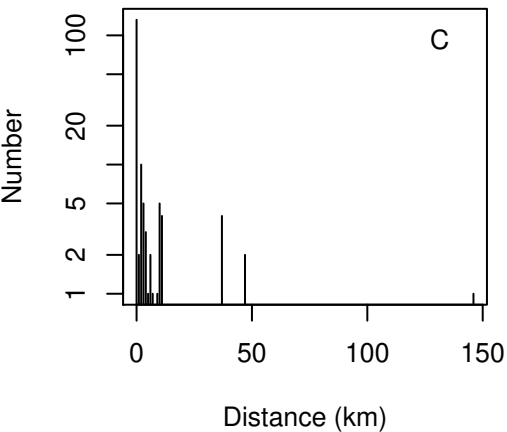
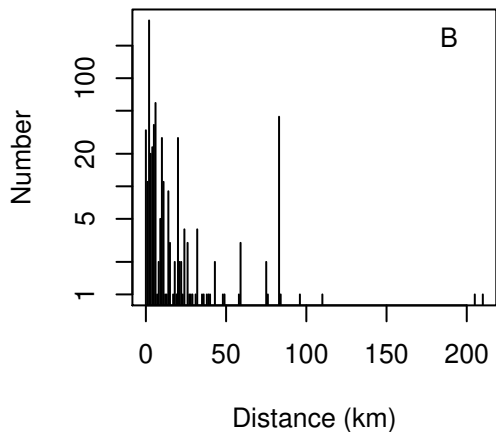
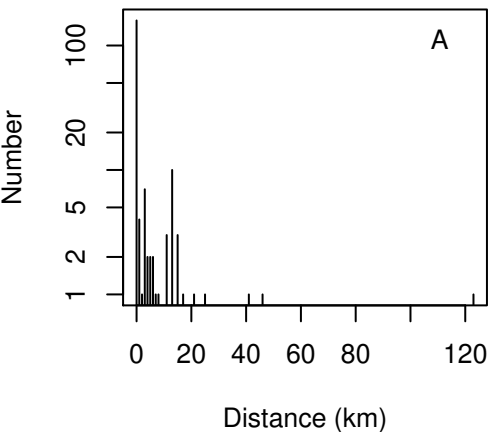
Fig. 5. Observed and expected frequencies of dispersal distances under the fitted half-Cauchy distribution. A: adult blue tits, B: juvenile blue tits, C: adult great tits, D: juvenile great tits. Dotted lines are  $x = y$ .



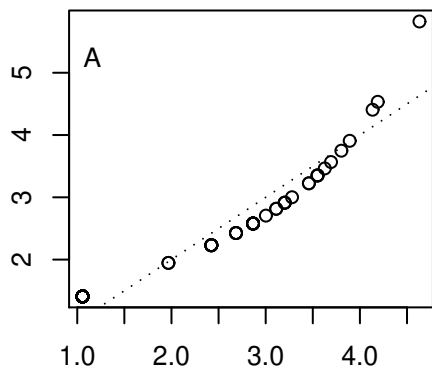
Probability



Distance (km)

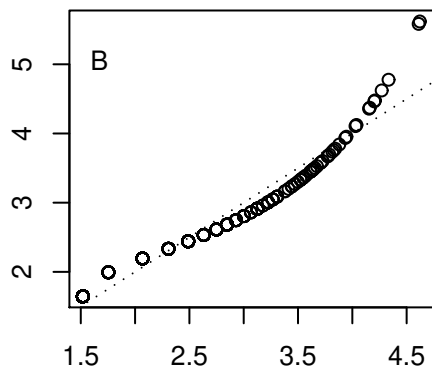


Weibull residuals



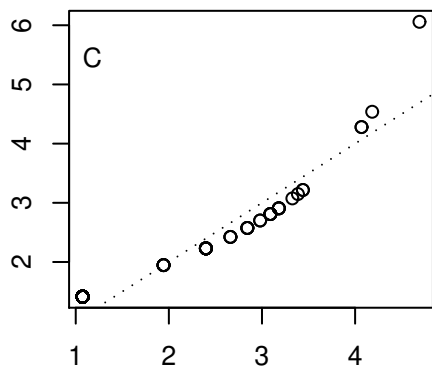
half-Cauchy residuals

Weibull residuals



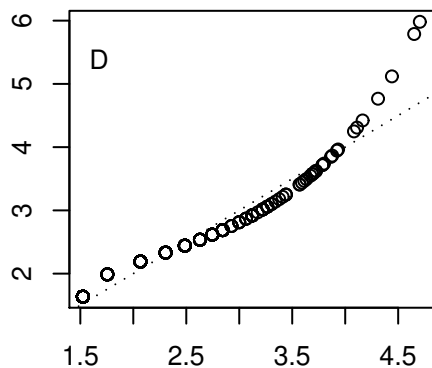
half-Cauchy residuals

Weibull residuals



half-Cauchy residuals

Weibull residuals



half-Cauchy residuals

