



HAL
open science

Modelling unbiased dispersal kernels over continuous space by accounting for spatial heterogeneity in marking and observation efforts

Joel J. Chadoeuf, Alexandre Millon, Jean-luc Bourrioux, Thierry Printemps, Benoît van Hecke, Vincent Lecoustre, Vincent V. Bretagnolle

► To cite this version:

Joel J. Chadoeuf, Alexandre Millon, Jean-luc Bourrioux, Thierry Printemps, Benoît van Hecke, et al.. Modelling unbiased dispersal kernels over continuous space by accounting for spatial heterogeneity in marking and observation efforts. *Methods in Ecology and Evolution*, 2018, 9 (2), pp.331-339. 10.1111/2041-210X.12872 . hal-01581789

HAL Id: hal-01581789

<https://hal.science/hal-01581789v1>

Submitted on 26 Jun 2018

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.

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/319163524>

Modelling unbiased dispersal kernels over continuous space by accounting for spatial heterogeneity in marking and observation efforts

Article in *Methods in Ecology and Evolution* · August 2017

DOI: 10.1111/2041-210x.12872

CITATIONS

0

READS

145

7 authors, including:



Joel Chadoeuf

French National Institute for Agricultural Research

71 PUBLICATIONS **200** CITATIONS

[SEE PROFILE](#)



Alexandre Millon

Aix-Marseille Université

40 PUBLICATIONS **541** CITATIONS

[SEE PROFILE](#)



Vincent Bretagnolle

French National Centre for Scientific Research

372 PUBLICATIONS **6,810** CITATIONS

[SEE PROFILE](#)

Some of the authors of this publication are also working on these related projects:



CESAB Disco-Weed - Disentangling the role of anthropic DISTurbances and eCOlogical processes on weed community assembly [View project](#)



Ecology and conservation of Montagu's harrier (*Circus pygargus*) in France, Europe and worldwide. [View project](#)

Modelling unbiased dispersal kernels over continuous space by accounting for spatial heterogeneity in marking and observation efforts

Joël Chadœuf^{1*}, Alexandre Millon^{2*†}, Jean-Luc Bourrioux³, Thierry Printemps³,
5 Benoît van Hecke³, Vincent Lecoustre⁴ & Vincent Bretagnolle^{4,5}

*these two authors equally contributed to this work

¹ INRA-Avignon, Statistics, UR 1052, Site de Saint Maurice, 84140 Montfavet, France

² Aix Marseille Univ, CNRS, IRD, Avignon Université, IMBE, Institut Méditerranéen de Biodiversité et d'Ecologie marine et continentale, Technopôle Arbois-Méditerranée, Bât.
10 Villemin – BP 80, F-13545 Aix-en-Provence cedex 04, France

³ Groupe d'Etudes et de Protection des Busards (GepB), Maison forestière de Blinfey, 52110 Beurville, France

⁴ Centre d'Etudes Biologiques de Chizé (CEBC), UMR CNRS-Université La Rochelle, 79360 Beauvoir-sur-Niort, France

15 ⁵LTSER Zone Atelier « Plaine & Val de Sèvre », CNRS, F-79360 Villiers-en-Bois, France

Keywords: Montagu's harrier, *Circus pygargus*, citizen science, imperfect detection, bias correction

Running title: Estimating unbiased dispersal kernels,

20 † author for correspondence : Alexandre Millon, Aix Marseille Univ, email address :

alexandre.millon@imbe.fr

Abstract

1. Although a key demographic trait determining the spatial dynamics of wild populations, dispersal is notoriously difficult to estimate in the field. Indeed, dispersal distances obtained from the monitoring of marked individuals typically lead to biased estimations of dispersal kernels as a consequence of i) restricted spatial scale of the study areas compared to species potential dispersal and ii) heterogeneity in marking and observation efforts and therefore in detection probability across space.
2. Here we propose a novel method to circumvent these issues that does not require data on observation effort per se, to correct for the variability in detection of marked individuals across space. Observed dispersal events were weighted by the distribution of departure points and an eroded spatial window approach was applied so as to deal with border effect. We conducted a set of simulations which indicated that our method was successful in correcting the effect of spatially heterogeneous detectability and produce unbiased dispersal kernels.
3. We applied this method to a real dataset on Montagu's harrier (>5000 chicks tagged), providing ca. 6000 resightings collected in entire France by a network of 1200 volunteers within a citizen-science program. The median dispersal distance observed was 32 km (range: 0.1-627 km). Once corrected for spatial heterogeneity in marking and observation efforts and border effect, the modelled dispersal kernel indicated a median dispersal distance of 78-123 km depending on the spatial scale considered (constrained within French borders or not, respectively).
4. *Synthesis and applications*: The current rise of citizen-science programs is likely to stretch our estimate of the ecologically-relevant spatial scale at which dispersal takes place for many taxa. Our method is particularly suited for such large scale data that typically suffer from high spatial heterogeneity in marking and observation efforts and offers the possibility to derive

unbiased dispersal kernels, a key component for modelling population dynamics and species distribution in a context of environmental change. Currently, our method assumes homogeneity in both habitat and dispersal behaviour across individuals. We discuss however
50 how to relax these hypotheses to further investigate the effect of e.g. local conspecific density or habitat quality on dispersal propensity.

INTRODUCTION

Dispersal is defined either as the movement of an individual from its birthplace to the location
55 where it will reproduce for the first time (natal dispersal), or the movement between
successive reproduction events (adult dispersal). Dispersal is a pivotal process underpinning
spatial population dynamics (review in Clobert *et al.* 2001). Indeed, movements of
individuals, both within and among populations, sustain gene flow across space and time, thus
driving genetic diversity as well as adaptation to local conditions (Cain, Milligan & Strand
60 2000; Postma & van Noordwijk 2005). Dispersal is a spatial demographic trait that affects
extinction and colonisation rates of habitat patches in a metapopulation context (Hanski 1999;
Sutherland, Elston & Lambin 2012), and also the size of more continuously spread
populations (Doncaster *et al.* 1997; Schaub, Jakober & Stauber 2013; Lieury *et al.* 2016).
Ignoring spatial demographic processes when investigating population dynamics may indeed
65 lead to spurious interpretations regarding e.g. the conservation status of a population (Schaub
et al. 2010), the efficiency of management actions (Lieury *et al.* 2015), the estimation of
fitness (Dhondt 1979; Tinbergen 2005; Doligez & Pärt 2008) and more generally the
evolution of populations in response to environmental changes (Kokko & López-Sepulcre
2006; Ronce 2007). In addition, the way dispersal is accounted for greatly affects the results
70 of species distribution models predicting range shifts in response to climate change (Guisan &
Thuiller 2005; Bellard *et al.* 2012).

However, estimating dispersal is notoriously difficult and has been the subject of intense
methodological developments over the last two decades (van Noordwijk 1995, 2011; Koenig,
Van Vuren & Hooge 1996; Paradis, Baillie & Sutherland 2002; Schaub & Royle 2014). The
75 classical way for investigating the dispersal process involves the longitudinal monitoring of
marked individuals over space. A major issue of such studies, however, arises from the finite
nature of the study area, as the closer a departure point is from the border, the higher the

probability of the corresponding arrival point to fall outside the study area. This results in the so-called border effect, which produces an inevitable reduction in the detectability of long-
80 distance dispersal (Barrowclough 1978). Indeed, in most instances, the spatial scale of the study area is defined according to logistical constraints so as to ensure an exhaustive marking of offspring and maximising recapture/resighting probability of previously marked individuals. However, there is growing evidence for the dispersal ability of individuals to largely overcome the limits of 'standard' study areas (Van Houtan *et al.* 2007). In birds
85 particularly, data collected from individuals fitted with VHF, Argos or GPS devices have revealed much larger dispersal distances than those previously estimated using passive marks (Van Vuren & Armitage 1994; Koenig, Van Vuren & Hooge 1996; Hénaux *et al.* 2011). Weight and economic costs of such electronic devices currently limit the number of species on which such equipment can be deployed and the number of equipped individuals, therefore
90 limiting the scope for evaluating population-level dispersal kernels (Murray & Fuller 2000; Wikelski *et al.* 2007). There is therefore a pressing need to improve our ability to derive unbiased dispersal kernels from 'classically marked' individuals and beyond the scale of a single study area.

The rise of citizen-based science offers a unique opportunity to take up this challenge by
95 providing observations over a large spatial scale, typically nation-wide scale. In birds, large spatial scale monitoring programs dedicated to the study of dispersal have started to emerge, such as in seabirds, waders or raptors (Etheridge, Summers & Green 1997; Gill, Norris & Sutherland 2001; Barlow *et al.* 2013). Yet, increasing the spatial scale of the study inevitably comes at the expense of high spatial heterogeneity in marking effort and in the probability of
100 detecting returning individuals, representing the second major issue when studying dispersal. Indeed, over large spatial scales, the probability of detecting a marked animal is typically low

and cannot be considered constant across space. Data on distribution of the observation effort, that could help estimating this bias, are however usually missing in a citizen-science context.

Different statistical methods have been proposed for estimating the bias in the observation
105 process when investigating dispersal. Multi-states models within the Capture-Mark-Recapture
framework (CMR) provide a way to estimate dispersal probability among a set of discrete and
finite number of study areas (Spendelov *et al.* 1995; Schwarz & Arnason 1996). This
approach is appropriate for species with highly patchy and restricted distributions such as rare
colonial seabirds. However the number of parameters exponentially increases with the
110 number of sites included, and moreover, this framework provides movement probabilities, not
dispersal kernel (e.g. Lagrange *et al.* 2014). For more continuously distributed species, the
method classically used consists in increasing the scale of the sampling area and weighting
the number of observed dispersal events (i.e. recruitments of individuals previously marked)
by the total number of individuals caught or recruited (i.e. individuals not previously marked)
115 at different distance classes. Obtaining such information on unmarked individuals entails
substantial additional costs and thus reduces the scope of this method otherwise efficient in
providing unbiased dispersal kernels (Baker, Nur & Geupel 1995; van Noordwijk 1995;
Winkler *et al.* 2005; Van Houtan *et al.* 2007). Van Noordwijk (2011) proposed the use of
distance-dependent recruitment rates to test whether the dispersal rule exhibited by the species
120 differs from a uniform law, when the observation process is incomplete across space.
However, this method does not provide an estimation of dispersal kernels (see Appendix A
where we demonstrate that the form of the distance-dependent recruitment rates depends on
the spatial distribution of the observation effort). A promising method recently proposed is the
development of a spatial extension of the Cormack-Jolly-Seber CMR model that allows one
125 to jointly estimate true instead of apparent survival, dispersal and observation processes
(Schaub & Royle 2014). However, a critical assumption of this model (primarily dedicated to

the estimation of true instead of apparent survival) is that recapture probability is homogeneous across the study area, a clearly unrealistic assumption in nation-wide citizen-based programs. Therefore to date, there is no satisfactory methods producing unbiased dispersal kernels for species widely distributed over continuous space.

Here we developed such a method to estimate unbiased dispersal kernels at any spatial scale (including large nation-wide scale) and for any species distribution type (patchy or continuous), relying on three successive steps: 1) we minimised the bias induced by the border effect by considering the set of resighted individuals whose departure point lies within an eroded spatial window; 2) we estimated dispersal kernel parameters using a contrast function (i.e. an estimation function whose expectation is minimum at the true parameter set; Dacunha-Castelle & Duflo 1983) which converges to a weighted likelihood, up to a multiplicative constant proportional to the average recapture rate. We accounted for spatial heterogeneity in observation effort by weighting each dispersal event by the inverse of the density of departure points (i.e. marking effort). By doing this, we avoided the direct modelling of observation effort, typically intractable when dealing with citizen-science data; 3) finally, we provided an unbiased dispersal kernel by estimating the residual bias in dispersal parameters using a simulation and extrapolation method (Cook & Stefanski 1994). As a first approach, we assumed that habitat quality was homogeneous across space and that all individuals had a similar average dispersal behaviour (but see Discussion for the possibility to relax these hypotheses).

The objectives of this paper are twofold. After elaborating on the method, its assumptions and mathematical developments, we checked for its robustness (validity, properties) with a simulation study based on a heavy-tailed distribution of dispersal events (log-sech) and spatially heterogeneous marking and observation efforts. Thereafter, we applied this method to a dataset gathered by a large-scale citizen-based science program conducted in France and

designed to estimate the dispersal kernel of the Montagu's harrier *Circus pygargus*, a widespread species breeding in cereal intensive farming landscapes of Western Europe and expected to disperse over several hundreds of kilometres.

155

METHODS

DISPERSAL KERNEL ESTIMATION

Modelling dispersal data accounting for observation effort

Let consider I individuals located at their departure points (e.g. birth nests) $X = (x_1, \dots, x_I)$ in an observation area $W \in R^2$. We suppose that X follows a spatial Poisson process described by its intensity $\lambda(x)$. The individual i located at x_i undergoes a single dispersal event to the arrival point y_i with respect to a dispersal kernel g_{θ_0} such that $p(y|x) = g_{\theta_0}(y - x)$, where θ_0 is the set of parameters of the dispersal model g . The dispersal kernel is thus supposed to be stationary over W . The individual i is observed at its arrival point when $y_i \in W$ and when an observer detects it (either by direct capture or observation at distance) with a probability $\omega(y_i)$, or the detection probability linked to observation effort. We define $o(i) = 1$ if the individual i is 'captured' by an observer in W , $o(i) = 0$ if not. Let $J = \sum_1^I o(i)$ the number of individuals observed at their arrival points. The probability to observe an individual at y is therefore:

170 with the set of arrival points in W following a spatial Poisson process of intensity:

$$p(y) = \omega(y) \frac{\int_{x \in W} \lambda(x) g_{\theta_0}(y - x) dx}{\int_{x \in W} \lambda(x) dx}$$

$$\mu(y) = \omega(y) \int_{x \in W} \lambda(x) g_{\theta_0}(y - x) dx$$

Maximum Likelihood method (ML) leads to a biased estimation of dispersal kernel due to spatial heterogeneity in observation effort

A classical way of estimating the dispersal kernel is through the maximisation of the log-

175 likelihood $LL = \sum_{i/o(i)=1} \log \left[\frac{g_{\theta}(y_i - x_i)}{\int_{y \in W} g_{\theta}(y - x_i) dy} \right]$ (Clark, Lewis & Horvath 2001; Gerber *et al.*

2014). The law of large numbers implies that, when $J \rightarrow \infty$, $\frac{LL}{J}$ tends towards:

$$\frac{\int_{x \in W} \int_{y \in W} \lambda(x) \omega(y) \frac{g_{\theta_0}(y - x)}{\int_{z \in W} \omega(z) g_{\theta_0}(z - x) dz} \log \left(\frac{g_{\theta}(y - x)}{\int_{z \in W} g_{\theta}(z - x) dz} \right) dy dx}{\int_{x \in W} \int_{y \in W} \lambda(x) \omega(y) g_{\theta_0}(y - x) dy dx}$$

180 However, because of the weightings by $\lambda(x)\omega(y)$, the ratio $\frac{LL}{J}$ does not converge to a Kullback-Leibler divergence (up to an additive constant) and asymptotic unbiasedness is therefore not ensured (Dacunha-Castelle & Duflo 1983). Note that if ω is constant, that is to say when the observation process is spatially homogeneous (*i.e.* all individuals are detected with equal probability in any location of W), then $\frac{LL}{J}$ tends towards a weighted Kullback-Leibler divergence (up to an additive constant) of the form:

$$\frac{LL}{J} \rightarrow \frac{\int_{x \in W} \lambda(x) \int_{y \in W} \frac{g_{\theta_0}(y - x)}{\int_{z \in W} g_{\theta_0}(z - x) dz} \log \left(\frac{g_{\theta}(y - x)}{\int_{z \in W} g_{\theta}(z - x) dz} \right) dy dx}{c \int_{x \in W} \int_{y \in W} \lambda(x) g_{\theta_0}(y - x) dy dx}$$

Thus, maximizing $\frac{LL}{J}$ leads to asymptotically biased estimators of dispersal parameters unless observation process or departure point process is homogeneous.

185 *Estimation of dispersal kernel using weighted Maximum Likelihood method (wML)*

Below are the three successive steps we propose to account for the sources of bias identified above: the border effect and imperfect detection due to spatial heterogeneity in observation and marking efforts.

Step 1- Eroding the spatial window to minimise the bias induced by the border effect.

190 Let assume that the majority of dispersal distances are small relatively to the size of the spatial area W considered, i.e. in a context of large-scale citizen science program. Let $W_0 \in W$ be the result of an erosion of W by a distance d so that the probability of an individual i coming from W_0 and dispersing outside W_0 is small. We considered the subset of birds whose departure point lies in W_0 and arrival point in W (we show in step 3 below how
195 to correct for this potential bias).

Step 2- Estimating dispersal kernel parameters using weighted likelihood and accounting for local density of departure points.

The observation process ω cannot be considered as constant across space. To circumvent this issue, we weighted each dispersal event by the inverse of the density of departure points $\lambda(x)$.

200 By doing this, all points x in W_0 have now a density probability proportional to $g_{\theta_o}(y-x)$ to be the departure point of the bird captured at y . Let $Z_1(\theta) = \sum_{i/o(i)=1} \frac{\log(g_{\theta}(y_i-x_i))}{\lambda(x_i)}$ be the log-likelihood of observed dispersal distances weighted by the densities of departure points. Then $Z(\theta) = \frac{1}{I} Z_1(\theta)$ tends towards:

$$\frac{\int_{x \in W_0} \int_{y \in W} \omega(y) g_{\theta_o}(y-x) \log(g_{\theta}(y-x)) dy dx}{\int_{x \in W} \lambda(x) dx}$$

$$\approx \frac{\int_{y \in W} \omega(y) dy}{\int_{x \in W} \lambda(x) dx} \int_{u \in \mathbb{R}^2} g_{\theta_o}(u) \log(g_{\theta}(u)) du$$

which is a Kullback-Leibler divergence (up to a multiplicative constant $\frac{\int_{y \in W} \omega(y) dy}{\int_{x \in W} \lambda(x) dx}$ and an
 205 additive one $\int_{u \in R^2} g_{\theta_0}(u) \log(g_{\theta_0}(u)) du$). The minimum of this divergence is obtained for
 $\theta = \theta_0$ which ensures asymptotic unbiasedness, with u a generic point anywhere across R^2 .
 Here we want to estimate the parameter set θ by maximising $Z(\theta)$.

Random variables $U_i = o(i) \frac{\log(p_\theta(y_i - x_i))}{\lambda(x_i)}$ are independent and identically distributed. Let
 $\hat{\theta}$ be the value maximizing $Z(\theta)/I$. Using a development similar to the ML method, one
 210 obtains $\theta_0(W) = \text{argmin} \int_{y \in W} \omega(y) \int_{x \in W_0} p_\theta(y - x) \log(p_\theta(y - x)) dx dy$ and $\sqrt{I}(\hat{\theta} - \theta_0(W))$
 asymptotically converging towards a Gaussian distribution $N(0, D^{-1} V J D'^{-1})$ where D is the
 second derivative of $E(U_i)$ such as:

$$D_{k,l} = \int_{x \in W} \int_{y \in W_0} \omega(y) \frac{\frac{\partial p_\theta(y-x)}{\partial \theta_k} \frac{\partial p_\theta(y-x)}{\partial l}}{p_\theta(y-x)} dy dx$$

with V , the variance matrix of the U_i gradient with respect to θ , such as:

$$V_{k,l} = \int_{x \in W} \int_{y \in W_0} \frac{\omega(y)}{\lambda(x)} \frac{\frac{\partial p_\theta(y-x)}{\partial \theta_k} \frac{\partial p_\theta(y-x)}{\partial l}}{p_\theta(y-x)} dy dx$$

Step 3- Correcting dispersal kernel parameters for the bias due to border effect

215 Finally, $\theta_0(W) \rightarrow \theta_0$ when $W \rightarrow R^2$. In practice, the confidence interval around $\theta_0(W)$ is
 computed by bootstrapping the dispersal data set, whereas the bias $\overline{E(\hat{\theta})} - \theta_0(W)$ is
 estimated as the difference between $\theta_0(W)$ and the mean bootstrap parameter value (Cook &
 Stefanski 1994).

220 *SIMULATING DISPERSAL DATA*

We simulated a set of 5000 departure points by drawing coordinates from a non-stationary Poisson process including a gradient with densities declining eastward, within a 100×100 unit area. The detection probability was also spatially heterogeneous and set proportionally to the y -coordinates, therefore also producing a gradient with declining effort from north to south.

225 The combination of these two gradients (density of departure points, detection probability) created four distinct conditions spread over four quarters (starting from the north-west quarter, clockwise): 1. high nest density/high detection probability, 2. low nest-density/high detection probability, 3. low nest density/low detection probability, 4. high nest density/low detection probability (Fig. 1). These four conditions encompass the full range of spatial
230 heterogeneity in data collection typically observed in large-scale monitoring schemes of animals. Dispersal distances were generated using a log-sech distribution ($a = 0.75$, $b = 1$; generating a median of 8.3 units). Dispersal events occurring outside the observation square were not considered. This simulation process was repeated 200 times to analyse the stability of parameter estimates.

235

A REAL CASE STUDY: DISPERSAL OF MONTAGU'S HARRIERS ACROSS FRANCE USING CITIZEN-SCIENCE DATA

The Montagu's harrier and its monitoring in France

The Montagu's harrier *Circus pygargus* is a medium-sized (ca. 300 g) ground-nesting raptor
240 inhabiting open farming landscapes, within the Western Palearctic, and wintering in the Sahel region in Africa. About 3900-5100 breeding pairs are found in France (Millon, Bretagnolle & Leroux 2004), essentially nesting in winter-sown cereal fields (wheat and barley). A substantial proportion of the broods (typically 2-5 chicks) does not fledge before crop harvesting takes place (20-50%; Millon et al., 2002). In response to this conservation issue, an

245 extensive network of amateur ornithologists emerged from 1975, which undertakes the protection of the nests in close collaboration with farmers (Santangeli *et al.* 2015). Between 2008 and 2015, an average of 1083 ± 191 nests were monitored annually (range: 681-1287).

Data from several small-scale ringing schemes suggested that if most birds tended to return in the vicinity of their natal nests, both natal and breeding dispersal can incur distances of several hundreds of kilometres. In 2007 was launched a wing-tagging program across the whole country, with the objective of unravelling the demography of the species at large spatial scale, including the estimation of dispersal kernels. Building upon the existing network of volunteers, a total of 5129 chicks have been wing-tagged from 1778 nests, between 2007 and 2010, using coloured plastic tags attached to the patagium of each wing (see www.busards.com for a complete description of the scheme & Appendix B).

Dispersal data

A dedicated website, including a Google© map interface with automatic recording of the geographical coordinates, has been launched so as to ensure an efficient collection of resightings across France and more widely across Europe and Africa (www.busards.com). We filtered the >6000 resightings recorded in the database between 2008 and 2013 so as to retain a single resighting per year (excluding the year of birth) for each bird observed during the breeding season (15 April-15 July, i.e. excluding movements during the migration period). In the case of multiple resightings of a same bird during the same breeding season, we selected the first occurrence, so that each bird can only produce one dispersal distance per year, but up to six dispersal distances over the study period. Birds from a same nest were considered as independent in regard to dispersal event. The actual reproductive status and sex of the birds when resighted were not considered. Our observation area was defined as France, thus

observations made outside France were excluded from the dataset (5%, $N = 68$), and we ended up with a dataset gathering 1215 dispersal events.

270 We applied the aforementioned method to the harrier dispersal data. W was defined as the minimum convex polygon containing all harrier nests. $W_0 \in W$ was an eroded area including nests located at ≥ 150 km from the border of W . The threshold of 150 km was chosen from the dispersal distribution, so as to include ca. 80% of the observed dispersal events. We chose a heavy-tailed distribution, namely the log-hyperbolic secant, or log-sech distribution, to
 275 describe harrier dispersal distribution (Van Houtan *et al.* 2007). We assumed that dispersal distances $d \geq 0$ follow a log-sech distribution such that:

$$g(d) = \frac{2/(\pi br)}{(r/a)^{1/b} + (r/a)^{-1/b}}$$

where $a > 0$ is the scale parameter (expressed in distance units), and $b > 0$ sets the shape of the kernel (no dimension). We excluded the locations at sea from the area W . Local nest density was calculated by a two-dimensional kernel density estimator (Venables & Ripley
 280 2002). Simulations and analyses were performed in R 3.1.3 (R Development Core Team 2015). Descriptive statistics (mean or median) are given together with their 95% confidence intervals. The code for conducting such analyses is fully available upon request.

RESULTS

285 *ASSESSING THE PERFORMANCE OF THE METHOD USING SIMULATED DATA*

As expected, spatial heterogeneity in marking and observation efforts (Fig. 1) led to a biased dispersal kernel, with relatively long dispersal events being under-represented (Fig. 2). The two parameters of the estimated log-sech distribution were biased (scale: $a = 0.85$, shape: $b = 0.88$) compared to the parameters used to simulate the data ($a = 0.75$, $b = 1$). Application of

290 our method accounting for spatially heterogeneous marking and observation efforts did
largely correct for this bias (mean value with 95% CI: $a = 0.69$ [0.49-0.93], $b = 0.99$ [0.94-
1.05]). If the scale parameter a appeared as slightly underestimated, this had only little
consequences in terms of departure from the simulated data (Fig. 2).

295 *DISPERSAL OF MONTAGU'S HARRIERS ACROSS FRANCE*

Considering the whole dataset of dispersal events occurring within France ($N = 1215$),
Montagu's harrier dispersed over relatively long distance with median dispersal distance of 32
km and maximal dispersal distance reaching 627 km (Fig. 3). When spatially restricting the
spatial scale of the dataset so as to limit the border effect (W_0 , $N = 557$), the distribution of
300 dispersal events remained largely unchanged with a median distance slightly increasing up to
37 km (max. distance = 602 km; see Appendix C comparing histograms and quantile
distributions between the two datasets). Parameters of the log-sech distribution fitted to the
observed dispersal data (W_0) were $a = 11.5 \times 10^2$ km and $b = 0.66$.

Implementing our correction method however led to a strong increase of the shape
305 parameter $b = 0.95$ [0.90-1.03], whereas the scale parameter $a = 11.5 \times 10^2$ km [10.0×10^2 -
 12.7×10^2] did not change. Corrected median dispersal distance was considerably increased
and reached 123 km [105-160] (Fig. 4). The interquartile range lay between 34 and 473 km.
Note that the cumulative frequency only reached 0.8 at 700 km, i.e. at roughly the maximal
dispersal distance observed within France, suggesting that another 20% of dispersal events are
310 likely to be recorded further away (maximum fixed at 2000 km). In order to more directly
compare the corrected kernel with the dispersal events observed (Fig. 3), we restricted the
analysis up to a distance of 700 km. At this reduced spatial scale, the corrected median
dispersal distance was 78 km [70-101] (interquartile range: 27-227 km, Fig. 4).

315 **DISCUSSION**

Here we proposed a method for deriving unbiased dispersal kernels over continuous space from data collected at large spatial scale. Such data typically suffer from high spatial heterogeneity in both marking and observation efforts, in addition to a border effect. Our method successfully corrected dispersal kernel parameters by accounting for heterogeneity in the distribution of departure points (i.e. local nest density) within an eroded spatial window. Using simulations, we checked that this method was indeed able to provide unbiased and precise estimates of a probability density function fitted to dispersal distances. We further applied it to data collected on Montagu's harriers from a nation-wide citizen-science program and showed that corrected median dispersal distances were 2.4-3.8 larger than the median of observed dispersal distances.

ACCOUNTING FOR SPATIAL HETEROGENEITY IN MARKING AND OBSERVATION EFFORTS, AND BORDER EFFECT WHEN FITTING DISPERSAL KERNELS

From a technical viewpoint, the method we used is akin to a weighted likelihood method but present major improvements for dealing with the two classical issues encountered when modelling dispersal data: i) the finite nature of a study area (border effect) and ii) the spatial heterogeneity in detection probability of marked animals linked to marking and observation efforts. Indeed, all dispersal events within the study area cannot be observed. Therefore, in the contrast estimation function we used, the likelihood of each dispersal event was weighted by the inverse of the local nest density around the departure points, a critical advance since departure points (i.e. marking locations) are known while observation effort, the parameter necessary in alternative methods, is not.

The logic here was to account for the local nest density (i.e. potential departure points) when modelling the probability of birds to arrive at a given point. Indeed, by weighting the number of observed dispersal events by the number of departure points, and not the number of recruitments (see e.g. Baker, Nur & Geupel 1995), we avoid the need of collecting data on unmarked individuals and the direct modelling of observation effort, typically intractable using citizen-science data. Data on recruitment are particularly costly to collect, even for species nesting in a defined set of nestboxes (e.g. van Noordwijk 1995; Winkler *et al.* 2005). Our method does not require information on recruitment other than for marked individuals. Alternative methods such as the distance-dependent recruitment rate allow to test for departure of the dispersal rule from a uniform law (i.e. birds chose their breeding location at random), and is robust to spatial heterogeneity in observation effort (van Noordwijk 2011). However, this method cannot be used to derive dispersal kernels (Appendix A).

Second, the bias due to the finite nature of the observed area was corrected by a bootstrap procedure so as to get the unobserved tail of the distribution, assuming the form of the distribution was scale-independent. The advantage of this approach is that inference can be made on the frequency of dispersal occurring outside the study area. This however comes at the expense of relying on a particular statistical distribution. We used the log-sech distribution for its ability to fit heavy-tailed dispersal kernels (Van Houtan *et al.* 2007). Note however that our method can be fitted with alternative distributions (e.g. negative exponential, gamma, log-normal), and that a procedure can be performed to select the most appropriate distribution.

LIMITATIONS AND POSSIBLE EXTENSIONS OF THE METHOD

Our method relies on a non-stationary spatial Poisson distribution of departure points but can be extended to incorporate any point process distribution. However, the bootstrap step will be

more tedious as it will need the simulation of the chosen distribution of departure points. We also assumed that the rules governing dispersal processes are stationary, i.e. are constant across space or across ecological conditions encountered by individuals, and that all
 365 individuals have a similar average dispersal behaviour. If dispersal behaviour varies among individuals, for example according to age or density of conspecifics in the natal area, the model can then be extended to investigate the effect of a covariate on the kernel parameters. Let $v(x)$ be this driving variable (continuous or categorical), the new kernel then becomes $g'_{\phi}(x, y) = g_{\theta(\phi, v(x))}(x, y)$ where ϕ is the parameter set linking θ to $v(x)$. A comparison of
 370 dispersal kernels between birds born in low-density areas vs. birds born in high-density areas could then be made. Our method is also flexible enough to explore whether habitat selection, the final stage of the dispersal process (Clobert *et al.* 2001), depends on local environmental suitability. In this case, let $h(\phi, u(y))$ be the probability to settle in site y with environmental suitability $u(y)$. The model can then be extended by using:

$$g'_{\phi}(x, y) = \frac{g_{\theta(x, y)h(\phi, y)}}{\int_W g_{\theta(x, z)h(\phi, z)} dz}$$

375 *A LARGE-SCALE CASE STUDY USING CITIZEN-SCIENCE DATA: MONTAGU'S HARRIER DISPERSAL ACROSS FRANCE*

Accounting for spatial heterogeneity in both marking and observation efforts increased the median dispersal distance from 32-37 km (using observed data) to 78-123 km, depending on the chosen spatial scale. Whether considering France or Western Europe, the interquartile
 380 range of dispersal approximately doubled, illustrating the necessity to clearly specify the population of interest when summarising dispersal propensity. Actually, the cumulated frequency of dispersal distances derived from the corrected log-sech distribution did not reach one at the maximal observed dispersal distance and the corrected log-sech distribution predicted that roughly 10-20% of dispersal events may occur beyond this distance (Fig. 4).

385 This may appear as substantial but it is important to note here that a number of dispersal
events have indeed been recorded in neighbouring countries during the breeding season
(Netherlands, Belgium, Germany, Switzerland, Scotland, Ireland, Czech Republic, Hungary,
Spain) and were excluded from our analysis. Forty-eight dispersal events from 34 different
individuals have been actually recorded in neighbouring countries (4% of the data; Appendix
390 D). Twelve individuals dispersed at >700 km from their birth nest (median = 561) with a
maximal distance recorded at 1366 km (a bird resighted in Scotland).

ON THE USE OF PASSIVE MARKS TO STUDY DISPERSAL

Estimating animal dispersal has been identified as a major methodological challenge in
ecology for decades (Doligez & Pärt 2008; Schaub & Royle 2014). Capture-recapture
395 estimates based on marked animals have been criticised because of spatial sampling bias.
Alternative approaches have thus been proposed, but all have their limits. For instance,
despite dispersal can be inferred from genetic data and provide insightful information about
the spatial scale of gene flow, such methods cannot help to infer the shape of dispersal kernels
and to assess the demographic contribution of dispersal, partly because they integrate
400 dispersal events over several generations. Satellite tracking (Argos and GPS systems) has
been presented as the panacea for the study of dispersal (Koenig et al. 1996). However,
economic cost limits the ability to investigate juvenile dispersal (since young individuals
suffer higher mortality rates compared to adults), and device weights further restrict the
number of species that can afford to carry such material (Wikelski *et al.* 2007). Therefore we
405 argue here that passive marks remain a major tool for studying dispersal, especially
combining citizen-science program allowing studies to be conducted at very large spatial

together with the statistical method we propose here which correct for sample bias including heterogeneity in marking and observation efforts as well as for border effect.

410

ACKNOWLEDGMENTS

We are immensely thankful to the network of volunteers working on Montagu's harriers in France and sorry not to name the hundreds of them that contribute to the conservation of the species. This work has greatly benefited from the critical review of E. Gabriel and R.

415 Senoussi. We also thank A. van Noordwijk for his constructive comments.

DATA ACCESSIBILITY

Harrier data and R code are available from the Dryad Digital Repository:

<http://dx.doi.org/10.5061/dryad.g4n5q>.

420

AUTHORS CONTRIBUTIONS STATEMENT

JC, AM, VL & VB designed the research. JC did the mathematical development and ran analyses and simulations. AM wrote the paper. JC & VB contributed to revisions of the manuscript. JLB, TP, BVH, AM & VB ensured the coordination of the volunteers' network and provided the data.

425

REFERENCES

Baker, M., Nur, N. & Geupel, G.R. (1995) Correcting biased estimates of dispersal and survival due to limited study area : theory and an application using wrentits. *The Condor*,

430 **97**, 663–674.

Barlow, E.J., Daunt, F., Wanless, S. & Reid, J.M. (2013) Estimating dispersal distributions at multiple scales: within-colony and among-colony dispersal rates, distances and directions in European Shags *Phalacrocorax aristotelis*. *Ibis*, **155**, 762–778.

Barrowclough, G.F. (1978) Sampling bias in dispersal studies based on finite area. *Bird*
435 *Banding*, **49**, 333–341.

Bellard, C., Bertelsmeier, C., Leadley, P., Thuiller, W. & Courchamp, F. (2012) Impacts of climate change on the future of biodiversity. *Ecology letters*.

Cain, M.L., Milligan, B.G. & Strand, A.E. (2000) Long-distance seed dispersal in plant populations. *Am. J. Botany*, **87**, 1217–1227.

440 Chadœuf, J., Millon, A., Bourrioux, J.-L., Printemps, T., van Hecke, B., Lecoustre, V. & Bretagnolle, V. (2017) Data from: Modelling unbiased dispersal kernels over continuous space by accounting for spatial heterogeneity in marking and observation efforts. *Dryad Digital Repository*. doi:10.5061/dryad.g4n5q

Clark, J.S., Lewis, M. & Horvath, L. (2001) Invasion by extremes: population spread with
445 variation in dispersal and reproduction. *The American Naturalist*, **157**, 537–554.

Clobert, J., Danchin, E., Dhont, A.A. & Nichols, J. (2001) Dispersal – Causes, Consequences and Mechanisms of Dispersal at the Individual, Population and Community Level. , 452.

Cook, J.R. & Stefanski, L.A. (1994) Simulation-extrapolation estimation in parametric measurement error models. *Journal of the American Statistical Association*, **89**, 1314–
450 1328.

Dacunha-Castelle, D. & Duflo, M. (1983) *Probabilités et Statistiques: 2. Problèmes À Temps Mobile*. Masson, Paris.

Dhondt, A.A. (1979) Summer dispersal and survival of juveniles Great Tits in southern

Sweden. *Oecologia*, **42**, 139–157.

455 Doligez, B. & Pärt, T. (2008) Estimating fitness consequences of dispersal: a road to ‘know-where’? Non-random dispersal and the underestimation of dispersers’ fitness. *Journal of Animal Ecology*, **77**, 1199–1211.

Doncaster, C.P., Clobert, J., Doligez, B., Gustafsson, L. & Danchin, E. (1997) Balanced dispersal between spatially varying local populations: An alternative to the source-sink
460 model. *American Naturalist*, **150**, 425–445.

Etheridge, B., Summers, W. & Green, R.. (1997) The Effects of Illegal Killing and Destruction of Nests by Humans on the Population Dynamics of the Hen Harrier *Circus cyaneus* in Scotland. *Journal of Applied Ecology*, **34**, 1081–1105.

Gerber, S., Chadœuf, J., Gugerli, F., Lascoux, M., Buiteveld, J., Cottrell, J., Dounavi, A.,
465 Fineschi, S., Forrest, L.L., Fogelqvist, J., Goicoechea, P.G., Svejgaard Jensen, J., Salvini, D., Vendramin, G.G. & Kremer, A. (2014) High rates of gene flow by pollen and seed in oak populations across Europe. *PLoS ONE*, **9**, e85130.

Gill, J., Norris, K. & Sutherland, W.J. (2001) The effects of disturbance on habitat use by black-tailed godwits *Limosa limosa*. *Journal of Applied Ecology*, **38**, 846–856.

470 Guisan, A. & Thuiller, W. (2005) Predicting species distribution: offering more than simple habitat models. *Ecology Letters*, **8**, 993–1009.

Hanski, I. (1999) *Metapopulation Ecology*. Oxford University Press.

Hénaux, V., Powell, L.A., Hobson, K.A., Nielsen, C.K. & Larue, M.A. (2011) Tracking large carnivore dispersal using isotopic clues in claws : an application to cougars across the
475 Great Plains. *Methods in Ecology and Evolution*, **2**, 489–499.

Van Houtan, K.S., Pimm, S.L., Halley, J.M., Bierregaard, R.O. & Lovejoy, T.E. (2007) Dispersal of Amazonian birds in continuous and fragmented forest. *Ecology letters*, **10**,

219–29.

- 480 Koenig, W.D., Van Vuren, D.H. & Hooge, P.N. (1996) Detectability, philopatry, and the
distribution of dispersal distances in vertebrates. *Trends in Ecology & Evolution*, **11**,
514–517.
- Kokko, H. & López-Sepulcre, A. (2006) From individual dispersal to species ranges:
perspectives for a changing world. *Science (New York, N.Y.)*, **313**, 789–91.
- Lagrange, P., Pradel, R., Bélisle, M. & Gimenez, O. (2014) Estimating dispersal among
485 numerous sites using capture-recapture data. *Ecology*, **95**, 2316–2323.
- Lieury, N., Besnard, A., Ponchon, C., Ravayrol, A. & Millon, A. (2016) Geographically
isolated but demographically connected: Immigration supports efficient conservation
actions in the recovery of a range-margin population of the Bonelli's eagle in France.
Biological Conservation, **195**, 272–278.
- 490 Lieury, N., Ruetten, S., Devillard, S., Albaret, M., Drouyer, F., Baudoux, B. & Millon, A.
(2015) Compensatory immigration challenges predator control: An experimental
evidence-based approach improves management. *Journal of Wildlife Management*, **79**,
425–434.
- Millon, A., Bourrioux, J.-L., Riols, C. & Bretagnolle, V. (2002) Comparative breeding
495 biology of Hen Harrier and Montagu's Harrier: an 8-year study in north-eastern France.
Ibis, **144**, 94–105.
- Millon, A., Bretagnolle, V. & Leroux, A.B.A. (2004) Busard cendré. *Rapaces Nicheurs de
France. Effectifs, Distribution & Conservation*. (eds J.-M. Thiollay), & V. Bretagnolle),
pp. 70–74. Delachaux & Niestlé, Paris.
- 500 Murray, D.L. & Fuller, M.R. (2000) A critical review of the effects of marking on the biology
of vertebrates. *Research techniques in animal ecology : controversies and consequences*

(eds L. Boitani), & T.K. Fuller), pp. 15–64. Columbia University Press, New York.

van Noordwijk, A.J. (1995) On bias due to observer distribution in the analysis of data on natal dispersal in birds. *Journal of Applied Statistics*, **22**, 683–694.

505 van Noordwijk, A.J. (2011) Measuring dispersal as distance-dependent recruitment rates: Testing the performance of DDRR on simulated data. *Journal of Ornithology*, **152**, S239–S249.

Paradis, E., Baillie, S.R. & Sutherland, W.J. (2002) Modeling large-scale dispersal distances. *Ecological Modelling*, **151**, 279–292.

510 Postma, E. & van Noordwijk, A.J. (2005) Gene flow maintains a large genetic difference in clutch size at a small spatial scale. *Nature*, **433**, 65–68.

R Development Core Team. (2015) R: A language and environment for statistical computing.

Ronce, O. (2007) How does it feel to be like a rolling stone? Ten questions about dispersal evolution. *Annual Review of Ecology, Evolution, and Systematics*, **38**, 231–253.

515 Santangeli, A., Arroyo, B., Millon, A. & Bretagnolle, V. (2015) Identifying effective actions to guide volunteer-based and nationwide conservation efforts for a ground-nesting farmland bird. *Journal of Applied Ecology*, **52**, 1082–1091.

Schaub, M., Aebischer, A., Gimenez, O., Berger, S. & Arlettaz, R. (2010) Massive immigration balances high anthropogenic mortality in a stable eagle owl population:
520 Lessons for conservation. *Biological Conservation*, **143**, 1911–1918.

Schaub, M., Jakober, H. & Stauber, W. (2013) Strong contribution of immigration to local population regulation: Evidence from a migratory passerine. *Ecology*, **94**, 1828–1838.

Schaub, M. & Royle, J.A. (2014) Estimating true instead of apparent survival using spatial Cormack–Jolly–Seber models. *Methods in Ecology and Evolution*, **5**, 1316–1326.

- 525 Schwarz, C.J. & Arnason, A.N. (1996) A general methodology for the analysis of capture-recapture experiments in open populations. *Biometrics*, **52**, 860–873.
- Spendelov, J.A., Nichols, J.D., Nisbet, I.C.T., Hays, H., Cormons, G.D. & Dec, N. (1995) Estimating annual survival and movement rates of adults within a metapopulation of roseate terns. *Ecology*, **76**, 2415–2428.
- 530 Sutherland, C., Elston, D.A. & Lambin, X. (2012) Multi-scale processes in metapopulations: contributions of stage structure, rescue effect, and correlated extinctions. *Ecology*, **93**, 2465–2473.
- Tinbergen, J.M. (2005) Biased estimates of fitness consequences of brood size manipulation through correlated effects on natal dispersal. *Journal of Animal Ecology*, **74**, 1112–1120.
- 535 Venables, W.N. & Ripley, B.D. (2002) *Modern Applied Statistics with S*, Fourth edi. Springer.
- Van Vuren, D.H. & Armitage, K.B. (1994) Survival of dispersing and philopatric yellow-bellied marmots: what is the cost of dispersal? *Oikos*, **69**, 179–181.
- Wikelski, M., Kays, R.W., Kasdin, N.J., Thorup, K., Smith, J.A. & Swenson, G.W. (2007)
- 540 Going wild : what a global small-animal tracking system could do for experimental biologists. *Journal of Experimental Biology*, **210**, 181–186.
- Winkler, D.W., Wrege, P.H., Allen, P.E., Kast, T.L., Senesac, P., Wasson, M.F. & Sullivan, P.J. (2005) The natal dispersal of tree swallows in a continuous mainland environment. *Journal of Animal Ecology*, **74**, 1080–1090.

545

Figure 1. Map of the simulated dispersal events (dark blue dots: departure points, light blue dots: arrival points). The simulations were generated so as to create spatial heterogeneity in the combination of nest density and detection probability, as typically observed in field data
550 (see Methods). Note that many short-distance dispersal events are actually hidden by the dots.

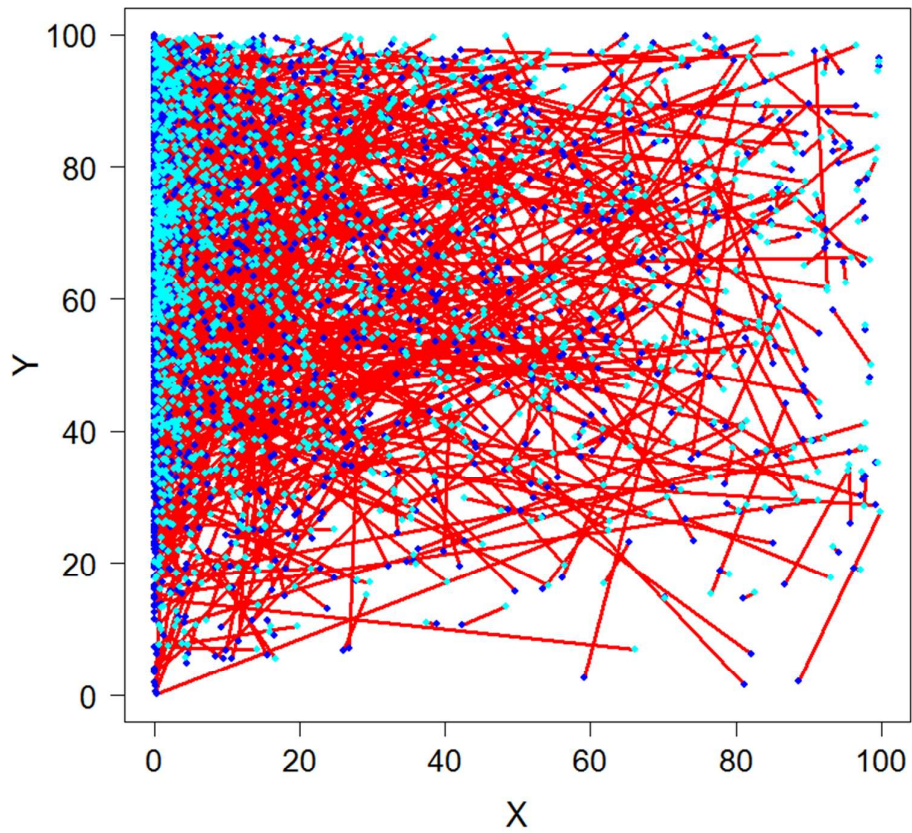


Figure 2. Cumulated frequency of dispersal distances (in arbitrary unit) i) predicted from the log-sech distribution used for the simulation ($a = 0.75$, $b = 1$; dashed green line), ii) estimated from the observed data (i.e. through the filter of the observation process; black line) and iii) predicted from the log-sech distribution fitted to the observed data but accounting for spatially heterogeneous marking and observation efforts ($a = 0.69$, $b = 0.99$; red line). Grey envelopes indicate 95% CI for black and red curves.

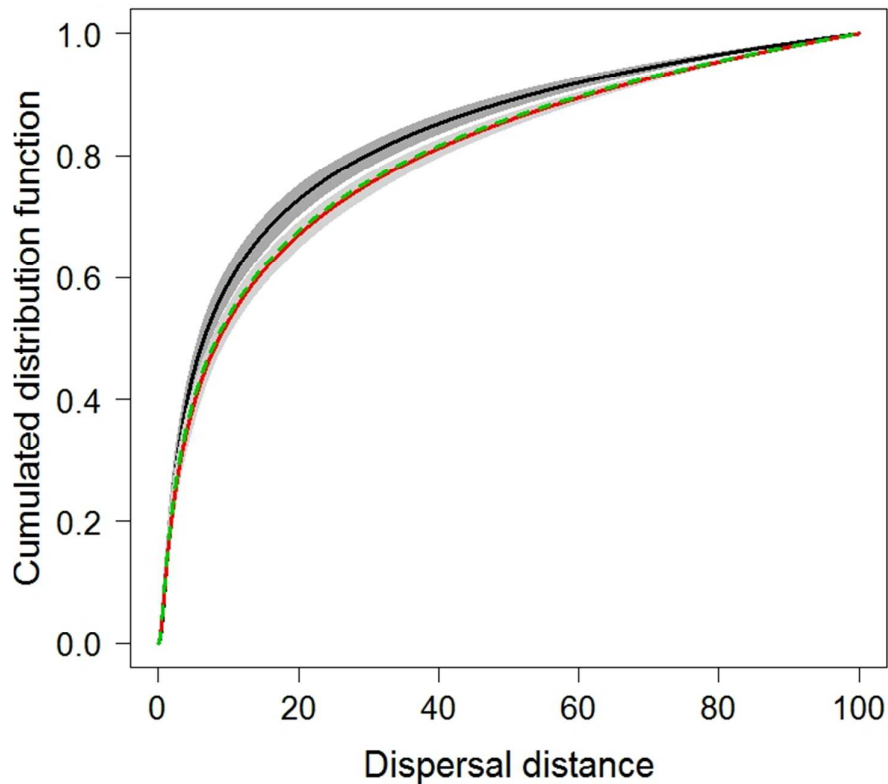


Figure 3. Observed dispersal events of Montagu's harrier between natal nests (all nest locations; dark blue dots) and subsequent re-observations in France (≥ 1 yr after birth, during the breeding season only, note that many short-distance dispersal events were actually hidden by the size of the dots), and the corresponding distribution histogram (all birth nests considered in W ; $N = 1215$ individual \times year). Median dispersal distance was 32 km.

565

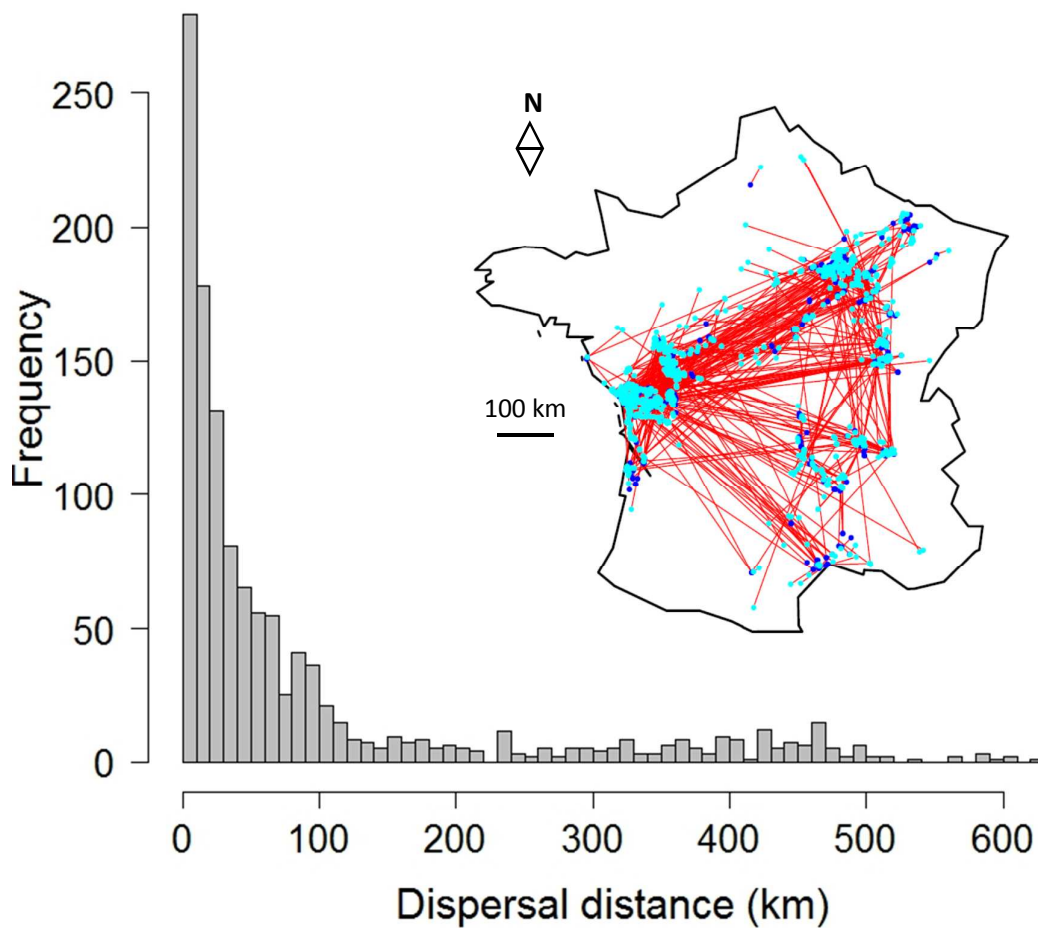


Figure 4. Observed (black line) and modelled (red line, grey envelope for 95% CI) cumulated frequency of dispersal events of the Montagu's harrier in France. Cumulated frequency of observed data reached one at 650 km, as dispersal events occurring <150 km from the French border were excluded. The median of observed dispersal distance was 32 km (black arrow). After correcting for spatial heterogeneity in marking and observation efforts, and the border effect, median dispersal distance (red arrow with dark grey rectangle showing the 95% CI around the median, the light grey rectangle indicates the interquartile range) reached 123 and 78 km respectively according to whether maximal dispersal distance was set at a) 2000 km and b) 700 km (i.e. corresponding to the maximal dispersal distance possible within the French borders).

