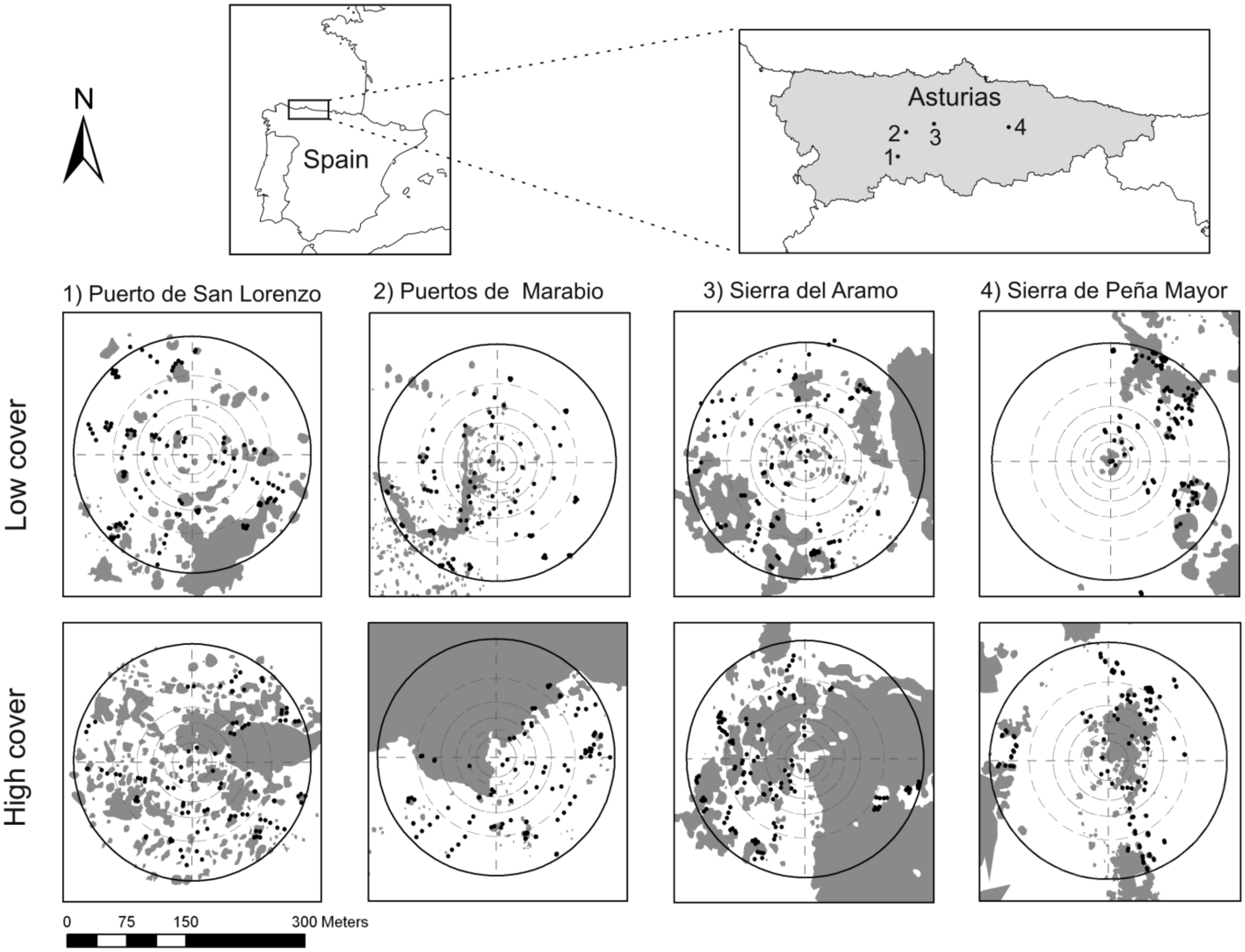
Supplementary Information

Appendix A. Spatial framework of study showing the geographical location of the four study sites (top row), and within each site the (bottom row), the two study landscapes chosen to represent low and high forest cover around the ^{15}N-marked seed source trees at the center (grey areas represent forest patches and white area open pastures). In each study landscape (circles), 108 stations (black dots) were distributed at concentric increasing distance bands to sample seed dispersal, from 0 to 150 m from ^{15}N-marked trees, and equally allocated into FF, NF and O habitat types (see Fig. 1 of main document).



Appendix B

Marking trees with ^{15}N.

We used a portable 10L hand spray pump to apply solutions of ^{15}N-urea $(^{15}NH_2)_2CO$ (98.9% ^{15}N) on inflorescences and leaves at a concentration of $0.5 gL^{-1}$. The spray of the pump easily reached the canopy of most trees, but with taller trees we used a ladder to reach the canopy. We added one drop of a non-ionic detergent (Uniqema Americas LLC, New Jersey, USA) per liter to make solutions stick to plant surfaces. Seeds of holly and hawthorn became enriched with ^{15}N as a result of spray applications of ^{15}N urea on flowering trees that were located at the center of the eight study landscapes. Marked holly seeds had an average $\delta^{15}N$ value of 213.13 (±19.01 SE) as compared to a value of 1.87 (± 0.25 SE) of unmarked control seeds (Table B1, also see Appendix C for more details on the distribution of values). Similarly, marked hawthorn seeds showed an average $\delta^{15}N$ value of 700.94 (±75.06 SE) as compared to 3.90 (±0.14 SE) of unmarked control seeds (see also Appendix C). In the fall, the size of the ^{15}N-labeled crops at the center patches ranged from 21,693 - 108,656 fruits (3-4 seeds per fruit) for holly, and from 1,692 - 21,032 fruits (1 seed per fruit) for hawthorn (Table 2B).

Appendix B

Table B1. ^{15}N values and number of trees and seeds sampled from ^{15}N isotope-marked (enriched) and unmarked (control) holly (*Ilex aquifoli*um) and hawthorn (*Crataegus monogyna*), at the center of eight study landscapes located at four sites (i.e., two plots per site) of the Cantabrian Range in Asturias, Spain. At the center of each landscape, 1-7 trees were marked in spring 2009 with solutions of 0.5 L^{-1} ^{15}N urea (98% atom).

Species	Group (No. Trees)	No. of seeds	Mean ^{15}N atom%	Std. Dev. (Atom%)	Mean δ^{15}N(‰)	Total ^{15}N fruits on centers
Holly	Unmarked (40)	128	0.3669880	0.0010399	1.87	--
Holly	^{15}N-marked (24)	74	0.4439919	0.0595465	213.13	380,200
Hawthorn	Unmarked (40)	86	0.3677306	0.0004724	3.90	--
Hawthorn	^{15}N-marked (26)	80	0.6208789	0.2423820	700.94	69,015

Appendix B

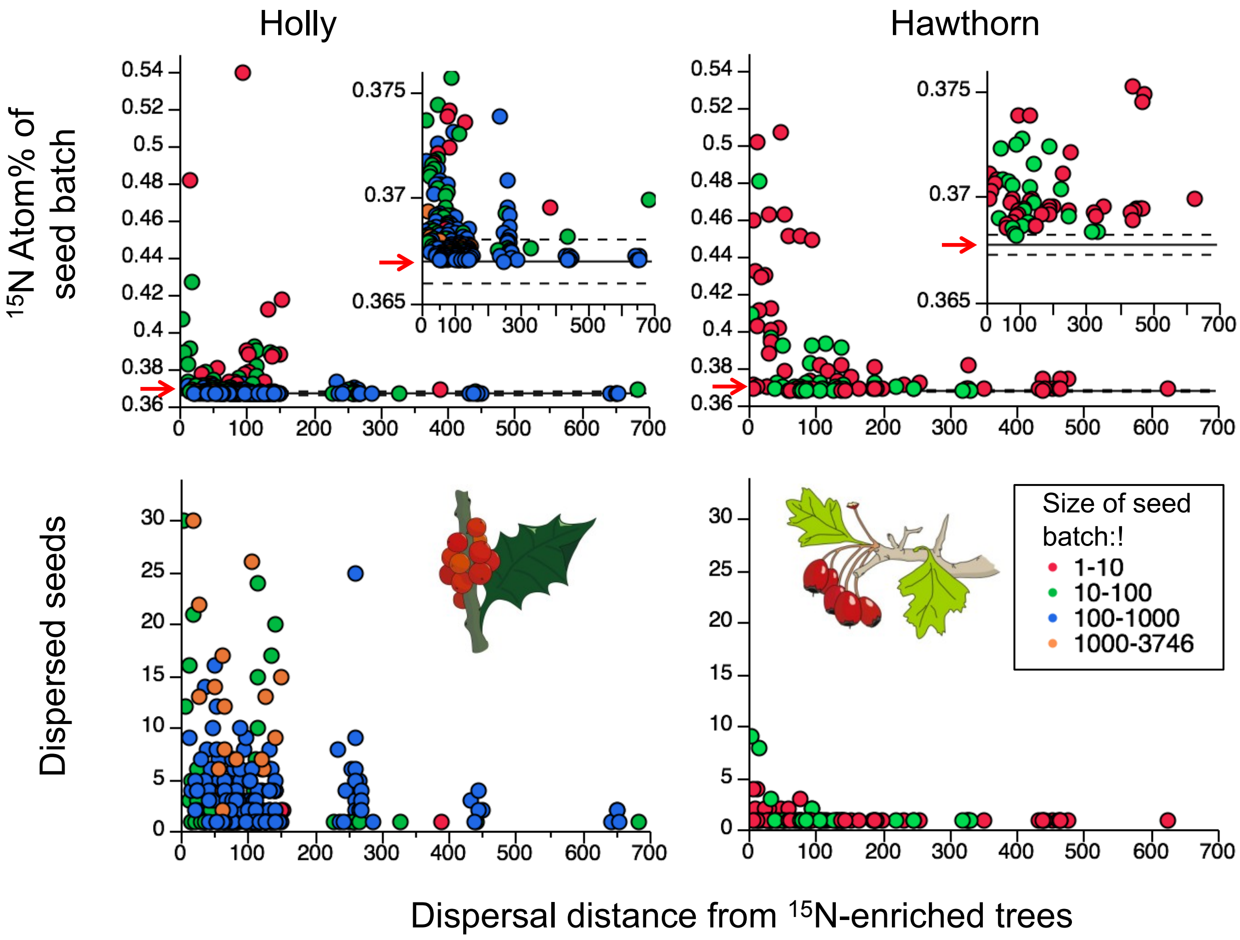
Table B2. Description of study sites and marked fruit crop sizes (i.e. the sum of the crop sizes of all marked trees). In each tree, we counted the number of fruits in 15 randomly selected branches and the number of branches per tree. Individual crop size was estimated by multiplying the average number of fruits per branch by the number of branches per tree.

Site	UTM		% Forest Cover	No. Forest Patches	Forest cover class	^{15}N-marked fruit crop size (no. marked trees)	
	X	Y				Holly *(Ilex aquifolium)*	Hawthorn *(Crataegus monogyna)*
Sierra del Aramo	261466	4791571	58.9	44	High	21693 (2)	1692 (3)
Sierra del Aramo	262244	4791436	27.0	106	Low	25646 (2)	4036 (3)
Puertos de Marabio	733456	4789197	42.05	58	High	29964 (5)	11907 (5)
Puertos de Marabio	734534	4788686	7.55	133	Low	77333 (2)	13048 (7)
Sierra de Peña Mayor	297132	4796681	13.25	82	High	54147 (3)	7690 (3)
Sierra de Peña Mayor	296455	4798087	6.69	48	Low	26995 (3)	6130 (1)
Puerto de San Lorenzo	727896	4780821	36.30	158	High	35776 (4)	15180 (2)
Puerto de San Lorenzo	728976	4779727	17.72	81	Low	108656 (3)	21032 (2)

Appendix B

Figure B1. Scatter plots showing the ^{15}N atom % of seed batches with ^{15}N-marked seeds (top panels), including the predicted quantity of marked seeds (bottom panels), for holly and hawthorn relative to the distance to the marked source trees. The color of dots indicates a range for the number of seeds ground and homogenized in order to form one sample for mass spectrometry analysis. Solid black lines (and red arrow mark) indicate the mean value of unmarked seeds and the dashed lines indicate 1 SD from the mean (top inset panels zoom the values close to mean value ± SD). Overall, hawthorn seed batches display higher ^{15}N values than holly batches (top panels). This is because hawthorn seeds were more enriched with ^{15}N than holly seeds (Table B2), and also because of the quantities of seeds in the batches differed greatly between species: larger seed batches require smaller increases in ^{15}N to reveal the presence of marked seeds. For example, 10 marked seeds in a batch of 50 seeds produce a batch signature that is higher than in a batch of 500 seeds (Carlo et al. 2009). Thus, almost all of the hawthorn batches having marked seeds were much smaller than the holly batches, and also displayed higher ^{15}N values.

Appendix B

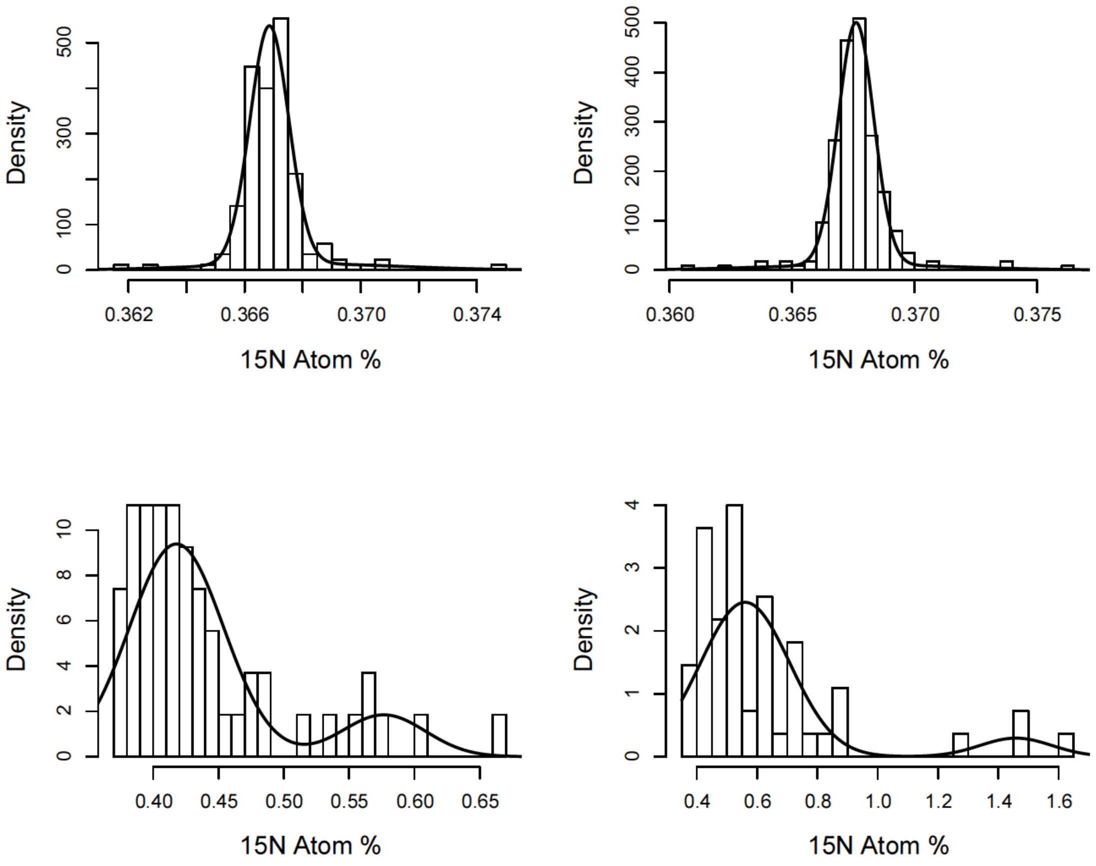

Holly
Hawthorn
15N Atom% of seed batch
0.54
0.52
0.5
0.48
0.46
0.44
0.42
0.4
0.38
0.36
0.375
0.37
0.365
0
100
200
300
400
500
600
700
Dispersed seeds
30
25
20
15
10
5
0
Size of seed batch:!
● 1-10
● 10-100
● 100-1000
● 1000-3746
Dispersal distance from 15N-enriched trees

Appendix C

Two-Source Mixing Model. The concentration of ^{15}N in a mixture of marked and unmarked seeds depends clearly on the proportion of marked seeds in the mix but also on the possible levels of ^{15}N present in each seed. Technically, the distribution of ^{15}N values of a mixture of n_l marked and n_u unmarked seeds is the convolution of a total of $n_l + n_u$ distributions, with the distribution of marked values represented n_l times and that of unmarked values represented n_u times. These convolutions are not problematic if the distributions involved are “nice” such as Gaussian. Our control data however, showed that both marked and unmarked seeds had no simple distributions and hence we fitted a combination of two Gaussian distributions to each (Fig. S2 and Table S1). Furthermore, and in order to be conservative, we decided to truncate the fitted distributions for positive controls in order to avoid any overlap with the observed values in the unmarked seeds. The range of ^{15}N atom % found in unmarked seeds of the two species combined varied from 0.361 to 0.376, so we truncated the distributions for marked seeds to have values larger than 0.377.

Our goal was to infer the number of marked seeds in a mixture of n seeds based on the observed ^{15}N atom %. For this we looked for the combination of marked and unmarked seeds that maximized the likelihood of observing a value similar to the one in the sample at hand. The distributions of ^{15}N from a particular combination of marked and unmarked seeds were estimated by first simulating 10 E06 (a million) such seed mixtures by sampling from the combinations of Gaussian distribution fitted to the positive and negative controls (with truncation for the marked ones). An empirical cumulative distribution function was then computed from these simulated values (*ecdf* function of package "stats" in R). From this empirical cumulative distribution we computed the likelihood of the observed ^{15}N atom %. The estimated number of marked seeds in a batch corresponds to that which maximizes the likelihood. Confidence intervals were found by looking at those combinations of marked and un-marked seeds that where 2 likelihood units from the maximum.

Figure C-1 Unmarked (top panels) and marked (bottom panels) ^{15}N values (atom %) for seeds of *Ilex aquifolium* (left column) and *Crataegus monogyna* (right column). A mix of two Gaussian distributions (black lines) was fitted to each dataset. For *Ilex* there were 170 seeds analyzed from umarked trees, and 54 for marked trees. For *Crataegus* there were 228 seeds from unmarked trees and 54 from marked ones.



Parameters and their Std. Dev. for mixtures of two Gaussians fitted to negative and positive controls of *Ilex aquifolium* and *Crataegus monogyna* seeds (see Figure S2).

	Unmarked *Ilex*	Unmarked *Crataegus*	Marked *Ilex*	Marked *Crataegus*
P[1]	0.882 (0.041)	0.893 (0.036)	0.269 (0.091)	0.088 (0.037)
P[2]	0.118 (0.042	0.107 (0.036)	0.731 (0.091)	0.912 (0.037)
mean[1]	0.367 (0.00007	0.368 (0.00006)	0.516 (0.030)	1.458 (0.060)
mean[2]	0.368 (0.0009	0.368 (0.00096)	0.406 (0.005)	0.560 (0.021)
sigma[1]	0.0007 (0.00005	0.0007 (0.00005)	0.077 (0.019)	0.119 (0.021)
sigma[2]	0.0034 (0.0009	0.0042 (0.0009)	0.026 (0.005)	0.148 (0.015)

Dispersal Kernel fitting. In order to explore how the probability of seed arrival changed with habitat type we fitted dispersal kernels to the seed dispersal data (^{15}N marked seeds) where the effective distance from source to seed sampling station was a function of habitat type. Seed dispersal was modeled with a 2Dt kernel with fixed scale (*u*) and shape (*p*) parameters. The expected number of seeds at the *i*th sampling station was a function of distance to the source (r_i), the number of seeds at the source (*Q*) and the area of the seed trap or sampling station (A_i):

$$\hat{s}_i = \frac{\quad}{\pi \quad + \quad \times} \times \quad \times \tag{1.1}$$

We model the effect of habitat type on the probability of seed arrival as a change in "movement space" (Schurr *et al.* 2008):

$$w = \quad + \quad \tag{1.2}$$

Where "open" (O) habitat serves as the reference habitat and I_f and I_{nf} are indicator variables for "fleshy" (FF) or "non-fleshy" (NF) habitats with b_f and b_{nf} being the corresponding coefficients.

Maximum likelihood estimates for all parameters were obtained using the function “mle2” in Bolker (2010). We calculated profile likelihoods for all parameters and estimated 95% confidence intervals from them Bolker (2010).

Parameter estimates and 95% confidence intervals (CI) for 2Dt kernels.

	MLE		CI		*p*-value	
	Ilex	*Crataegus*	*Ilex*	*Crataegus*	*Ilex*	*Crataegus*
p	12.903	1.851	7.736 ; 20.635	0.458 ; 6.676	<0.00001	0.132
u	0.112	0.037	0.074 ; 0.170	0.011 ; 0.093	<0.00001	0.052
b_{f}	-3.601	-3.201	-3.819 ; -3.415	-3.776 ; -2.786	<0.00001	<0.00001
b_{nf}	-2.785	-2.967	-2.993 ; -2.608	-3.561 ; -2.542	<0.00001	<0.00001

References

Schurr, F. M., O. Steinitz, and R. Nathan. 2008. Plant fecundity and seed dispersal in spatially heterogeneous environments: models, mechanisms and estimation. *J. Ecol.*, 96, 628–641.

Bolker, B. 2010. bbmle: Tools for general maximum likelihood estimation.
http://cran.fyxm.net/web/packages/bbmle/

R code for the isotopic analysis of seeds batches.

```r
library(msm)
load("posCM.RData") # winbugs output for normal mixture of positive control (pcm.sim)
load("posIA.RData") # same for Ilex (pia.sim)
load("negCM.RData") # negative controls (ccm.sim)
load("negIA.RData")     #  cia.sim
load("all.RData")       # data from all sites


# probability density by way of simulation
mixfs <- function(nreps, nu, nl, cpars, mixpars, low){
 au <- matrix(0,nreps,1)
 al <- matrix(0,nreps,1)
 if(nu>0){
  for(i in 1:nu){
    au1 <- rnorm(nreps, mean=cpars$lambda[1], sd = cpars$sigma[1])
    au2 <- rnorm(nreps, mean=cpars$lambda[2], sd = cpars$sigma[2])
    rc <- runif(nreps) < cpars$P[2]
    au1[rc] <- au2[rc]
    au <- rowSums(cbind(au, au1))
  }}
  if(nl>0){
   for(i in 1:nl){
    alt <- rtnorm(nreps, mixpars$lambda[1],sd=mixpars$sigma[1], lower=low)
    al1 <- rtnorm(nreps, mixpars$lambda[2],sd=mixpars$sigma[2], lower=low)
    rc <- runif(nreps)  < mixpars$P[1]
    al1[rc] <- alt[rc]
    al <- rowSums(cbind(al,al1))
  }}
 satom <- rowSums(cbind(au,al))/(nu+nl)
 return(satom)
}

#-----------------------------------------------------------------------------#
#                          the analysis                                  #
#-----------------------------------------------------------------------------#

nlab <- numeric(length(N))*NA # will hold the estimated number of labelled seeds
delta <- 0.00001              # resolution to estimate likelihoods
nreps = 1000000               # number of simulated mixtures to estimate likelihoods
LKS <- vector('list', length(atom)) # will hold vectors of likelihoods to estimate CI
low = 0.377                   # truncation value for marked seeds

# run IA
cpars <- cia.sim$mean        # for the distribution of un-marked seeds
mixpars <- mia.sim$mean      # for the distribution of marked seeds
enes <- sort(unique(N[which(spp==2 & N>0)]))  # find and sort the number of seeds in each batch

for(i in 1:length(enes)){
 idx <- which(spp==2 & N==enes[i])  # find samples with a given number of seeds
 oa <- atom[idx]  # observed atom values in batch
 us <- 0:enes[i]  # possible numbers of un-marked seeds
 la <- enes[i]:0  # same for marked ones
 if(enes[i]==1){  # if there is only one seed then the comparison is among the distributions fitted to possitive and negative controls
  tmpl <- mixpars$P[1]*dtnorm(oa, mixpars$lambda[1],sd=mixpars$sigma[1],lower=low) + mixpars$P[2]*dtnorm(oa,
mixpars$lambda[2],sd=mixpars$sigma[2],lower=low)
  tmpu <- cpars$P[1]*dnorm(oa,mean=cpars$lambda[1],sd=cpars$sigma[1])+ cpars$P[2]*dnorm(oa,mean=cpars$lambda[2],sd=cpars$sigma[2])
  lk <- cbind(tmpl, tmpu)
  for(j in 1:length(oa)){
     nlab[idx[j]] <- la[which(lk[j,] == max(lk[j,]))] # find the MLE
     LKS[[idx[j]]] <- lk[j,]                          # save likelihoods for CI estimation
  }}

 if(enes[i]>1 ){
  lk <- matrix(0, length(oa), length(la))
  for(k in 1:length(us)){
   tmp <- mixfs(nreps=nreps, nu=us[k], nl=la[k], cpars=cpars, mixpars=mixpars,low=low) # simulate mixtures
   cmf <- ecdf(tmp)                                                          # compute CDF
```

```
   for(j in 1:length(oa)){
      lk[j,k] <- ( cmf(oa[j]+(delta/2)) - cmf(oa[j]-(delta/2)) ) * 1/delta   # estimate likelihood
   }
  }
  for(jj in 1:length(oa)){
   nlab[idx[jj]] <- 0
      if(max(lk[jj,])>0){
        nlab[idx[jj]]<- la[which(lk[jj,]==max(lk[jj,]))]  # find the MLE
        LKS[[idx[jj]]] <- lk[jj,]                          # save likelihoods for CI estimation
      }
  }}

  save(LKS, i, nlab, file="resMix.RData")
}

# ------------------------------------
# now for Crataegus
mixpars <- pcm.sim$mean  # parameters for mix of two Gaussians for positive controls
cpars   <- ccm.sim$mean  # parameters for mix of two Gaussians for negative controls
enes    <- sort(unique(N[which(spp==1 & N>0)]))   # find and sort number of seeds per sample

for(i in 1:length(enes)){
 idx <- which(spp==2 & N==enes[i])  # find samples with a given number of seeds
 oa <- atom[idx]  # observed atom values in batch
 us <- 0:enes[i]  # possible numbers of un-marked seeds
 la <- enes[i]:0  # same for marked ones
 if(enes[i]==1){  # if there is only one seed then the comparison is among the distributions fitted to possitive and negative controls
  tmpl <- mixpars$P[1]*dtnorm(oa, mixpars$lambda[1],sd=mixpars$sigma[1],lower=low) + mixpars$P[2]*dtnorm(oa,
mixpars$lambda[2],sd=mixpars$sigma[2],lower=low)
  tmpu <- cpars$P[1]*dnorm(oa,mean=cpars$lambda[1],sd=cpars$sigma[1])+ cpars$P[2]*dnorm(oa,mean=cpars$lambda[2],sd=cpars$sigma[2])
  lk <- cbind(tmpl, tmpu)
  for(j in 1:length(oa)){
     nlab[idx[j]] <- la[which(lk[j,] == max(lk[j,]))] # find the MLE
     LKS[[idx[j]]] <- lk[j,]                          # save likelihoods for CI estimation
  }}

 if(enes[i]>1 ){
  lk <- matrix(0, length(oa), length(la))
  for(k in 1:length(us)){
   tmp <- mixfs(nreps=nreps, nu=us[k], nl=la[k], cpars=cpars, mixpars=mixpars,low=low) # simulate mixtures
   cmf <- ecdf(tmp)                                                                   # compute CDF
   for(j in 1:length(oa)){
      lk[j,k] <- ( cmf(oa[j]+(delta/2)) - cmf(oa[j]-(delta/2)) ) * 1/delta   # estimate likelihood
   }
  }
  for(jj in 1:length(oa)){
   nlab[idx[jj]] <- 0
      if(max(lk[jj,])>0){
        nlab[idx[jj]]<- la[which(lk[jj,]==max(lk[jj,]))]  # find the MLE
        LKS[[idx[jj]]] <- lk[jj,]                          # save likelihoods for CI estimation
      }
  }}

  save(LKS, i, nlab, file="resMix.RData")
}


res <- cbind(Location, id, spp, N, X, Y, hab, Rodal, distance, atom, nlab)
write.csv(res, file="labeledmix.csv")
```

Appendix D. Results of linear models (ANCOVA) examining the effects of habitat type, distance from source tree, and landscape variables (site, forest cover) on the dispersal rates (number of ^{15}N-marked seeds dispersed relative to the size of the ^{15}N-marked crop on source trees at center patches) of holly (*Ilex aquifolium*) and hawthorn (*Crataegus monogyna*) seeds in the Cantabrian Range in Asturias, Spain. In bold are $P < 0.05$.

Effects: Holly ($r^2 = 0.33$, $F_{12, 137} = 5.58$, $P < 0.0001$)	DF	SS	*F* Ratio	$P > F$
Site	3	0.00005483	0.8391	0.4747
Habitat (O, NF, FF)	2	0.00079334	18.2103	<0.0001
Distance	1	0.00019418	8.9144	0.0034
Forest Cover (Low, High)	1	0.00008267	3.795	0.0534
Site * Distance	3	0.00024747	3.7869	0.0120
Habitat * Distance	2	0.00005014	1.151	0.3194

Effects: Hawthorn ($r^2 = 0.29$, $F_{12, 137} = 4.75$, $P < 0.0001$)	DF	SS	*F* Ratio	$P > F$
Site	3	0.00070391	3.9174	0.0101
Habitat (O, NF, FF)	2	0.00148240	12.3747	<0.0001
Distance	1	0.00074333	12.4102	0.0006
Forest Cover (Low, High)	1	0.00066092	11.0344	0.0011
Site * Distance	3	0.00044530	2.4782	0.0639
Habitat * Distance	2	0.00025341	2.1154	0.1245

Parameter estimates for linear models (ANCOVA) examining the effects of habitat type, distance from source tree, and landscape variables (site, forest cover) on the dispersal rates of holly (*Ilex* aquifolium) and hawthorn (*Crataegus monogyna*) seeds the Cantabrian Range in Asturias, Spain. In bold are $P < 0.05$.

Linear Model (ANCOVA): Holly

Parameters	Estimate	Error	Approx. *t*	P > | *t* |
Intercept	0.0100474	0.001302	7.71	<.0001
Site[Aramo]	0.0012818	0.000899	1.43	0.1564
Site[Marabio]	3.61E-05	0.000675	0.05	0.9574
Site[Peña Mayor]	-0.001098	0.000833	-1.32	0.1896
Habitat [NF-F]	-0.001914	0.000943	-2.03	0.0443
Habitat [O-NF]	-0.003597	0.00093	-3.87	0.0002
Distance	-0.000028	9.35E-06	-2.99	0.0034
Forest Cover	-0.000058	2.98E-05	-1.95	0.0534
Site[Aramo]*Distance	-4.84E-05	1.46E-05	-3.31	0.0012
Site[Marabio]*Distance	1.44E-05	8.47E-06	1.69	0.0925
Site[Peña Mayor]*Distance	1.46E-05	8.47E-06	1.73	0.0859
Habitat[NF-F]*Distance	6.50E-07	0.000011	0.06	0.9525
Habitat[O-NF]*Distance	0.0000139	1.09E-05	1.28	0.2021

Linear Model (ANCOVA): Hawthorn

Parameters	Estimate	Error	Approx. *t*	P > | *t* |
Intercept	0.0168962	0.002160	7.82	<.0001
Site[Aramo]	0.0024217	0.001491	1.62	0.1067

Site[Marabio]	0.0025094	0.001119	2.24	0.0265
Site[Peña Mayor]	-0.003589	0.001382	-2.6	0.0104
Habitat [NF-F]	-0.003888	0.001564	-2.49	0.0141
Habitat[O-NF]	-0.003783	0.001542	-2.45	0.0154
Distance	-5.46E-05	1.55E-05	-3.52	0.0006
Forest Cover	-0.000164	4.94E-05	-3.32	0.0011
Site[Aramo]*Distance	-5.43E-05	2.42E-05	-2.24	0.0266
Site[Marabio]*Distance	2.48E-05	0.000014	1.77	0.0791
Site[Peña Mayor]*Distance	3.00E-05	0.000014	2.14	0.0345
Habitat [NF-F]*Distance	0.000026	0.000018	1.44	0.1525
Habitat [O-NF]*Distance	9.89E-06	0.000018	0.55	0.5831