



HAL
open science

A quantitative measure of migratory connectivity

Roberto Ambrosini, Anders Pape Møller, Nicola Saino

► **To cite this version:**

Roberto Ambrosini, Anders Pape Møller, Nicola Saino. A quantitative measure of migratory connectivity. *Journal of Theoretical Biology*, 2009, 257 (2), pp.203. <10.1016/j.jtbi.2008.11.019>. <hal-00554539>

HAL Id: hal-00554539

<https://hal.science/hal-00554539v1>

Submitted on 11 Jan 2011

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.



HAL Authorization

Author's Accepted Manuscript

A quantitative measure of migratory connectivity

Roberto Ambrosini, Anders Pape Møller, Nicola Saino

PII: S0022-5193(08)00616-4
DOI: doi:10.1016/j.jtbi.2008.11.019
Reference: YJTBI5381

To appear in: *Journal of Theoretical Biology*

Received date: 12 February 2008
Revised date: 18 October 2008
Accepted date: 26 November 2008

Cite this article as: Roberto Ambrosini, Anders Pape Møller and Nicola Saino, A quantitative measure of migratory connectivity, *Journal of Theoretical Biology* (2008), doi:[10.1016/j.jtbi.2008.11.019](https://doi.org/10.1016/j.jtbi.2008.11.019)

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting galley proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.



www.elsevier.com/locate/jtbi

A quantitative measure of migratory connectivity

Roberto Ambrosini ^a, Anders Pape Møller ^b and Nicola Saino ^c

^a Dipartimento di Biotecnologie e Bioscienze, Università degli Studi di Milano Bicocca,
piazza della Scienza 2, I-20126 Milano, Italy. E-mail: roberto.ambrosini@unimib.it

^b Laboratoire de Parasitologie Evolutive, CNRS UMR 7625, Université Pierre et Marie
Curie, 7 quai St. Bernard, Case 237, F-75252 Paris Cedex 05, France. E-mail:
amoller@snv.jussieu.fr

^c Dipartimento di Biologia, Università degli Studi di Milano,
via Celoria 26, I-20133 Milano, Italy. E-mail: nicola.saino@unimi.it

Correspondence should be addressed to:

Roberto Ambrosini, Dipartimento di Biotecnologie e Bioscienze, piazza della Scienza 2,
I-20126 Milano, Italy;

Tel.: +39 0264483464, Fax: +39 0264483565.

e-mail: roberto.ambrosini@unimib.it

Total word count: 7681

22 **Abstract**

23

24 The study of the extent of the connection between areas where populations of birds breed
25 and areas where they winter has flourished in recent years mainly thanks to the
26 development of new techniques, but also due to traditional ringing and recovery schemes,
27 which allow tracking of individuals or populations linking wintering and breeding
28 distributions. Currently, studies on migratory connectivity focus on retention of breeding
29 population spatial structure on the non-breeding grounds and vice versa.

30 Here we propose a method to quantify migratory connectivity based on Mantel
31 correlation coefficients and to statistically test for deviations of the observed connectivity
32 from a random mix of individuals. In addition, we propose a procedure, based on
33 clustering algorithms, to identify whether observed connectivity depends on aggregation
34 of individuals or on rigid transference of distribution patterns between areas.

35 We applied this method to a large dataset of ringing recoveries of barn swallows
36 (*Hirundo rustica* L.) migrating from their Western Palearctic breeding areas to sub-
37 Saharan winter quarters. We show that migration of barn swallow populations connects
38 specific breeding and wintering areas, and that the 'sub-populations' quantitatively
39 identified by our method are consistent with qualitative patterns of migratory connectivity
40 identified by studies of individual geographical populations based on other methods.

41 Finally, we tested the performance of the method by running simulations under different
42 scenarios. Such simulations showed that the method is robust and able to correctly detect
43 migratory connectivity even with smaller datasets and when a strong geographical pattern
44 is not present in the population. Our method provides a quantitative measure of migratory

45 connectivity and allows for the identification of populations showing high connectivity
46 between the breeding and wintering areas. This method is suitable for a generalized
47 application to diverse animal taxa as well as to large scale analyses of connectivity for
48 conservation purposes.

49

50

51 **Keywords:** barn swallow; *Hirundo rustica*; k-means clustering; Mantel test; migration;

52

53

Accepted manuscript

53 Introduction

54

55 Migratory connectivity has been defined as the links between breeding and non-breeding
56 areas of migratory animals due to migration (Webster et al 2002). Several different
57 approaches have been proposed for the study of migratory connectivity, such as ringing
58 recoveries (Anderson et al 2000), satellite radio telemetry (Cohn 1999), stable isotope
59 analysis of feathers (e.g. Hobson and Wassenaar 1997; Marra et al 1998; Møller and
60 Hobson 2003; Hobson 2005), genetic analysis (Webster et al 2002; Lopes et al 2006), and
61 correlation analysis between point estimates of annual survival rates and proxies of
62 ecological conditions putatively affecting survival (e.g. the Normalised Difference
63 Vegetation Index; Szép et al 2006).

64 A qualitative measure of migratory connectivity has been proposed by Webster et
65 al (2002), who argued that “strong” connectivity occurs when most individuals from one
66 breeding population move to the same non-breeding location, while a “weak” or “diffuse”
67 connectivity occurs when individuals from a single breeding population spread through
68 several non-breeding grounds. Statistical approaches that have been proposed to quantify
69 migratory connectivity are based on the same procedure that is commonly used to solve
70 the problem of the “derivation of harvest” in waterfowl management (Marra et al 2006).
71 They imply the definition of transition probabilities that describe how individuals from a
72 breeding population move to each wintering location and the calculation of conditional
73 probabilities of derivation of an individual from a given site (“origin probabilities”) using
74 Bayes’ Rule (Marra et al 2006). This procedure, however, requires an *a priori*
75 identification of ‘sub-ranges’ (see Webster et al 2002; Marra et al 2006) , which may be a

76 difficult and somewhat arbitrary task, at least in species with continuous breeding and/or
77 non-breeding ranges. In addition, estimating conditional probabilities in such Bayesian
78 framework requires prior knowledge of population sizes, at least relative to each other,
79 among the various origin sub-ranges (Marra et al 2006) that may be difficult to acquire
80 unless information from intensive, large-scale population survey programs is available.
81 Finally, recovery probabilities must be assumed to be proportional to the relative
82 abundance on the origin grounds. If this assumption is unwarranted, true connectivity
83 among sub-populations may be overlooked.

84 Currently, studies on migratory connectivity focus on the retention of breeding
85 population structure on the non-breeding grounds (and vice versa) (Marra et al 2006). If
86 individuals that breed close to each other also winter close to each other (*allohiemy* sensu
87 Salomonsen 1955) then a strong connectivity occurs, while if they mix (*sinhiemy*)
88 connectivity is weak. Theoretically, if the positions of individuals, both in the breeding
89 and the wintering grounds are known, then two distance matrices can be calculated,
90 representing the distance within each pair of individuals in the breeding or the wintering
91 grounds, respectively. Besag and Diggle (1977) proposed to test for the significance of
92 the maintenance of reciprocal distribution in breeding and wintering grounds in a
93 blackbird (*Turdus merula* L) population by testing for the significance of the correlation
94 between the matrix of distances in the breeding grounds and that in the wintering
95 grounds. If individuals that breed close together also winter close together (i.e. the
96 population shows strong connectivity), then a strong positive correlation is expected
97 between the two matrices. Conversely, if individuals mix (weak connectivity) a weak
98 correlation is expected. A measure of the correlation between the two distance matrices

99 can therefore be used to quantify migratory connectivity and to statistically test for the
100 probability of observing a given pattern, under the null hypothesis of a random mix
101 among individuals at the breeding and/or the non-breeding grounds, a condition for which
102 a correlation coefficient of zero is expected. Such a test, known as the Mantel test (Sokal
103 and Rohlf 1995), does not imply an a priori identification of separated sub-ranges in the
104 breeding and wintering grounds. We emphasize that, in this approach, the length of the
105 migration journey is not taken into account as the only measures involved in the analysis
106 are distances between individuals in the breeding and in the wintering grounds.

107 A significant correlation between the distance matrices, however, can result from
108 two different processes. The first consists of individuals that maintain the same reciprocal
109 position in the breeding and wintering grounds. Following Besag and Diggle (1977) we
110 refer to this process as to “distribution pattern transfer”. We note that the correlation
111 between the distance matrices will not vary under translation, rotation and scale change
112 of the pattern of distribution of individuals. Alternatively, if the population under study
113 consists of groups of individuals that stay close together both in the breeding and in the
114 wintering grounds, then a significant correlation between the distance matrices is also
115 expected, as within-group distances will, on average, be smaller than distances between
116 individuals of different groups.

117 A procedure for the quantitative study of migratory connectivity should ideally
118 allow disentangling these two processes and, if necessary, to identify the clusters of
119 individuals that breed and winter together. These goals can be achieved by applying a
120 clustering algorithm to a distance matrix obtained from the combination of the two
121 matrices of distances between individuals at the breeding grounds and in the winter

122 quarters. A significant connectivity in a non-structured (i.e. non-clustered) population
123 indicates that migratory connectivity arises from a process of distribution pattern
124 transference. Conversely, if cluster analysis reveals a structure in the population, then
125 grouping of individuals that always stay together contributes to migratory connectivity.
126 Within (main) clusters, individuals may then show pattern transference or further
127 grouping, resulting in sub-clusters.

128 In this study, we thus propose a novel approach to quantify migratory
129 connectivity. The method is based on Mantel test to assess the strength of migratory
130 connectivity and cluster analysis to investigate the process that generates it and to unravel
131 the clustered structure of the population.

132 We first assessed the performances of the method by running simulations under
133 three different scenarios. In the first scenario, the connectivity was due to a process of
134 distribution pattern transference. In the second, connectivity arose from the clustered
135 structure of the simulated population, but no pattern transference occurred for the
136 individuals of the same cluster. In the third, we simulate a condition where both processes
137 occur simultaneously. In addition, to exemplify the use of this method, we apply it to a
138 large dataset of recoveries of barn swallows (*Hirundo rustica* L) ringed in their breeding
139 range in Europe and recovered in their sub-Saharan wintering quarters or vice versa.

140 Data from recoveries of ringed birds as a tool in the study of connectivity have
141 been criticized on several bases, such as the large biases in the probability of recovery of
142 individuals between different parts of the world (Webster et al 2002). Still, they represent
143 the largest (and largely under-utilised) datasets currently available on numerous
144 migratory species. The results obtained from the analysis of the present dataset should

145 therefore be considered with this *caveat* in mind, and they will be compared to other
146 available information about the migration of barn swallows, including stability of the
147 genetic variance-covariance matrix of different populations (Roff et al 2004) and stable
148 isotope analyses of feathers of different populations (Evans et al 2003; Møller and
149 Hobson 2003).

150 The barn swallow is a small passerine bird that migrates each year from Europe to
151 Africa south of the Sahara desert. A considerable effort has been spent in the last decades
152 to capture and ring barn swallows both in Africa and Europe mainly within the EURING
153 and SAFRING ringing schemes. Recoveries and recaptures of barn swallows are
154 available from the entire breeding and wintering range of the species, encompassing an
155 extremely wide geographical area and a very long time span. To the best of our
156 knowledge, this is one of the largest available data sets of ringing recoveries and we are
157 convinced that the amount of information it contains largely overcomes the ‘noise’ due to
158 differences in ringing and capture efforts in different locations and years.

159

160

160 **Materials and Methods**

161

162 The method we are proposing is based on the coordinates of individuals recorded in the
163 areas occupied in different phases of the annual cycle. We assume that such grounds are
164 geographically separated and individuals move annually between them during migrations.
165 Since the large majority of the data about migrants come from ornithological studies, we
166 will refer to these areas as the breeding and the wintering grounds, albeit these definitions
167 may not be strictly applicable to other migrants (e.g. whales, for which ‘breeding’ and
168 ‘feeding’ grounds may be more appropriate). In all the analysis, including the
169 simulations, we always started from datasets where the position of individuals was
170 expressed by geographical coordinates, since in real datasets position of individuals is
171 usually recorded in this way. Nevertheless, this method is applicable to any other
172 coordinate system as it basically requires only distances between individuals.

173

174 *Analysis of migratory connectivity*

175

176 From geographical coordinates of individuals both in the breeding and wintering grounds,
177 we calculated two matrices of orthodromic distances (i.e. the minimum path on the
178 surface of the Earth), between all pairs of barn swallows both in the breeding and the
179 wintering quarters (see Fig. 1 for further details).

180 Mantel correlation coefficient (r_M) was used to test for the correlation between
181 orthodromic distance matrices. It corresponds to a simple Pearson product moment
182 correlation coefficient between the two matrices, whose significance is assessed with a

183 randomization procedure because distances between individuals within each distance
184 matrix are not statistically independent (Sokal and Rohlf 1995). We randomly permuted
185 the position of individuals at the breeding grounds 999 times; for each permutation, a
186 distance matrix was calculated and its correlation coefficient with the actual distance
187 matrix of individuals at the wintering grounds was calculated. The significance of the
188 observed Mantel coefficient was determined based on its rank in the set of the
189 coefficients from the randomization procedure as $P = [(1000 - \text{rank}) / 1000]$.

190 The analysis of the spatial structure of the dataset was based on a cluster detection
191 approach. First, we combined the two distance matrices into one by computing, for each
192 pair of individuals, the square root of the sum of their squared orthodromic distances at
193 the breeding and wintering grounds. This measure can be considered an overall distance
194 index similar to the Euclidean distance between two points that also takes the round
195 shape of the Earth into account as it is based on orthodromic distances.

196 Cluster detection was then performed with the *pam* algorithm in S-Plus 4.5
197 applied to the overall distance indices matrix. The number of clusters was identified as
198 the number that maximized the *overall average silhouette width* (hereafter *oasw* for
199 brevity) (Rousseeuw 1987). The *pam* procedure (fully described in Kaufman and
200 Rousseeuw 1990) is a partitioning clustering algorithm where the number of clusters
201 must be indicated *a priori*. The *oasw* is a measure of the goodness of the overall
202 classification of points in a given number of clusters. It is a dimensionless coefficient of
203 the extent of structure of clustering that has been identified. It is defined as the average
204 value of the *silhouettes* $s(i)$ computed over all objects in the dataset where
205

$$s(i) = \frac{b(i) - a(i)}{\max\{a(i); b(i)\}}$$

207

208 $a(i)$ being the average dissimilarity of i to all other objects of the cluster to which i
209 belongs and $b(i)$ the average dissimilarity of i to all objects of the nearest cluster to which
210 i does not belong. $s(i)$ values range between -1 and 1 and the larger the value the better
211 the classification of an object in a given cluster (Rousseeuw 1987). The best number of
212 clusters corresponds to the number that maximizes the *oasw*. *Oasw* values lower than
213 0.25 indicate that no substantial structure has been found in the data, values between 0.26
214 and 0.50 indicate a “weak” structure, between 0.51 and 0.70 a “reasonable” structure and
215 values larger than 0.71 a “strong” structure (Rousseeuw 1987). Each cluster detected in
216 the first step was investigated further with the same procedure, and, if a significant
217 structure was recognized, it was divided into sub-clusters, numbered as (dataset).(main
218 cluster).(sub-cluster).

219

220 *Simulations*

221

222 In all simulations we assumed a uniform spherical surface with no limits to the movement
223 of individuals. The centre of the cluster(s) of individuals in the breeding grounds were
224 always set at 30° N, while at 30° S in the wintering grounds. Longitude of the cluster
225 centre(s) varied during simulations (see below). Positions of individuals were generated
226 as a random determination from a bivariate Gaussian distribution with unit standard
227 deviation (SD) centred on the centre(s) of the cluster(s). The distance between cluster
228 centres and the random noise added to individual positions (see below) were expressed as

229 standard deviations (SD) of the bivariate Gaussian used to generate individual position
230 within clusters. At each step of the simulation, all the analyses were run according to the
231 method described above, and the significance of the Mantel correlation coefficient, the
232 *oasw* value, the number of clusters identified and the proportion of individuals correctly
233 classified were recorded.

234 The first simulation aimed at testing the ability of the method we propose to
235 detect migratory connectivity arising from a process of distribution pattern transference
236 (Besag and Diggle 1977). One cluster with 200 individuals was generated in the breeding
237 ground. In a process of pure pattern transference, the reciprocal position of individuals in
238 the wintering grounds should strictly reflect that in the breeding grounds except for
239 random variation. This process was simulated by generating positions in the wintering
240 grounds corresponding to those in the breeding grounds plus a bivariate normal random
241 noise added to the position of each individual. At each step of the simulation, the standard
242 deviation of the noise increased by 0.1 SD from 0 to 4 SD. Hence, this simulation was
243 run 41 times.

244 The second simulation aimed at testing the ability of the method to detect
245 migratory connectivity arising from a clustered structure in the population. Two clusters
246 of 100 individuals were generated. The position of individuals within each cluster both in
247 the breeding and in the wintering grounds was randomly determined. Hence, no pattern
248 transference existed within each cluster. At each run of the simulation, the distance
249 between cluster centres in the wintering and/or in the breeding grounds increased by 1 SD
250 from 0 (i.e. clusters superimposed both in the breeding and in the wintering grounds) to
251 10 SD, so that this simulation was run 121 times.

252 The third simulation investigated a situation when both clustering and pattern
253 transference occurred. This third simulation followed the general scheme of the second,
254 involving two clusters of individuals with increasing distance between cluster centres,
255 but, as in the first simulation, the reciprocal position of individuals within clusters in the
256 wintering grounds reflected that in the breeding grounds plus a random Gaussian noise
257 with $SD = 0.2$. This value was arbitrarily chosen as the first simulation demonstrated that
258 the amount of noise does not influence the performance of the method (see below).

259

260 *The barn swallow dataset*

261

262 The whole dataset consisted of recoveries of individually marked barn swallows between
263 1911 and 1998. Only first recoveries were selected in order to avoid pseudo-replication.
264 Individuals breeding east of longitude 60° E (approximately corresponding to the Ural
265 Mountains) were excluded. A first dataset (hereafter called “All” for brevity) of 1103
266 barn swallows included all individuals that were within the breeding range of the species
267 between April and September and within the wintering range between October and
268 March. We then quantitatively assessed the potential noise arising from recoveries of
269 individuals during migration by selecting a second dataset (“FP”) of 225 barn swallows
270 that were within the breeding range in May-June and within the winter quarters in
271 December-February, i.e. in the focal periods of reproduction and wintering, respectively.
272 Natal dispersal in the barn swallow is larger than breeding dispersal (Turner 1994). This
273 may bias the results if the proportion of ringed adults and young varies geographically. In
274 addition, in Africa a large number of barn swallows was ringed in a rather small number

275 of sites so that cluster detection could be biased by the small distance between ringing
276 sites of many individuals. Hence, a third dataset (“AE”) was selected, only including barn
277 swallows ringed in Europe as adults. Finally, the two previous selection criteria were
278 combined to obtain a fourth restricted dataset (“R”) of 71 barn swallows ringed as adults
279 in Europe in May-June and recovered in Africa in December-February. We then
280 compared the results of the analyses based on the four datasets by investigating
281 consistency of classification of individuals into clusters.

282

283

Accepted manuscript

283 **Results**284 *Simulations*

285 In all the 41 runs of the first simulations, a highly significant Mantel correlation
286 coefficient was found ($r_M < 0.001$ in all cases). No cluster structure was present in the
287 data and therefore classification of individuals into clusters was not relevant.

288 Fig. 2 summarizes the results of the second simulation. In 14 cases, the method
289 failed to detect connectivity, i.e. both the Mantel correlation coefficient was not
290 significant and the *oasw* value was lower than 0.5. In all these cases, the distance between
291 cluster centres in one area was 0 or 1 SD. Clusters generated as random determination
292 from a bivariate Gaussian are roughly circular and about 90% of individuals lie within 2
293 SD from the centre. Hence, a distance of 1 SD between cluster centres implies that at
294 least 60% of individuals lie in the common area. We note that no within-cluster pattern
295 transference occurred in this simulation, so that non-significant Mantel coefficients were
296 expected when clusters were close together. In another 14 cases the *oasw* value did not
297 indicate a reasonable structure in the data, but the Mantel coefficient was significant. In
298 all these cases the two clusters partly overlapped at least in one area. However, also in
299 these cases more than 90% of individuals were correctly classified in two clusters.

300 Mantel correlation coefficient was significant in all the 121 runs of the third
301 simulation. In 24 cases, always with partly overlapping clusters, *oasw* values did not
302 reveal a reasonable structure in the data albeit only in 10 cases cluster analysis correctly
303 classified less than 90% of individuals.

304

305 *Barn swallow dataset*

306 The Mantel correlation coefficient for the “All” dataset was 0.0247 (Fig. 3a). In the
307 randomization procedure we obtained 54 times a correlation coefficient larger than the
308 observed one. Migratory connectivity should therefore be considered marginally non-
309 significant ($P = 0.054$). However, the *oasw* value suggested a “reasonable” structure with
310 two clusters that also emerged from the analysis of the data subsets (see below). For this
311 reason, we identified two clusters that are shown in Fig. 4a and pictorially in Fig. 5a. The
312 corresponding silhouette plot is shown in Fig. 6. Mean orthodromic distance between
313 individuals in Europe was 1358.16 (941.673 SD) km (range 0-5400.34 km) while in
314 Africa it was 1493.67 (1553.880 SD) km (range 0-6192.57 km). Hence, distances in both
315 ranges were similar, and the pattern of distribution in the two ranges had similar effects
316 on cluster detection and composition. Cluster All.1 mainly included barn swallows
317 breeding in South-Western Europe and wintering from Liberia to Uganda, while cluster
318 All.2 was mainly constituted of barn swallows from Northern Europe that winter south of
319 the Equator, from Zaire to South Africa. The clusters partly overlap at the breeding
320 grounds while they are well separated in the wintering grounds.

321 Mantel tests showed a significant migratory connectivity for cluster All.2 and a
322 marginally non-significant connectivity for cluster All.1 that, however, showed an *oasw*
323 of 0.433 (i.e. reasonably high) (Fig. 3). We therefore analyzed the structure of both
324 clusters (see also Discussion) that could be divided into two sub-clusters each (All.1.1,
325 All.1.2, All.2.1 and All.2.2, pictorially shown in Fig. 5b). In the next step of the analysis,
326 only clusters All.1.2 and All.2.1 could be further divided in two and six third-level sub-
327 clusters, respectively. However, the results at this fine level were inconsistent between
328 data subsets (see below), and they are therefore not presented in detail.

329 The “FP” subset showed a significant connectivity ($n = 255$, $r_M = 0.0701$, $P =$
330 0.005). The *oasw* reached its maximum value ($= 0.582$) for two clusters, thus indicating a
331 reasonable structure in the data. At the second step of analysis, cluster FP.1 could be
332 divided into two second-level sub-clusters based on *oasw* value ($n = 32$, $r_M = 0.0768$, $P =$
333 0.140 , $oasw = 0.484$), while cluster FP.2 was significantly structured with two sub-
334 clusters ($n = 223$, $r_M = 0.0670$, $P = 0.002$, $oasw = 0.510$), none of which could be further
335 subdivided (details not shown). These results are consistent with the results of the
336 analysis of the “All” set of data. Indeed, all individuals classified in cluster FP.1 had been
337 classified in cluster All.1 and all individuals in FP.2 had been classified in All.2. At the
338 second level of analysis, only 3 individuals were misclassified.

339 The details of the results of the analysis of the “AE” set of barn swallows are
340 shown in Fig. 3b. The individuals showed significant connectivity and could be grouped
341 into two clusters that, in turn, were both structured in two sub-clusters. Sub-clusters
342 AE.1.1 and AE.1.2 showed non-significant connectivity, while AE.2.1 and AE.2.2 were
343 structured and could be divided into two and four third-level sub-clusters, respectively
344 (details not shown). However, as we did for the “All” dataset, we cautiously considered
345 the results only to the second step of the analysis. The results obtained from this selection
346 criterion were highly consistent with the results of the analysis of the “All” dataset at the
347 first level of analysis, and at the second level for cluster AE.1. Indeed only 3 out of the 86
348 individuals classified in AE.1 were classified in All.2, and 3 individuals of AE.1.2 were
349 classified in All.2.1. The partitioning of cluster AE.2 was not consistent with that
350 obtained from the “All” dataset. This was due to the small number of barn swallows (32)
351 in cluster All.2.1 that were ringed in Europe as adults and thus included in the “AE”

352 subset.

353 The most restricted subset of data (“R”) only included 71 barn swallows. The
354 Mantel test indicated a significant connectivity ($r_M = 0.2243$, $P < 0.001$), and the *oasw*
355 value (= 0.646) showed that it could be divided in two sub-clusters. The first cluster
356 could be divided in two sub-clusters according to *oasw* value (R.1: $n = 32$, $r_M = 0.0768$, P
357 = 0.155, $oasw = 0.484$) while the second was not structured (R.2: $n = 39$, $r_M = -0.0371$, P
358 = 0.555, $oasw \leq 0.367$). All individuals in R.1 had previously been classified in All.1 and
359 all individuals in R.2 in All.2. Consistent results emerged also at the second level with no
360 individual being misclassified.

361 Thus, the four levels of selection of the dataset, despite differing in composition
362 and number of individuals, led to highly consistent clustering of individual barn
363 swallows, whereby the large majority of individuals were assigned to corresponding
364 clusters in the different analyses.

365 These largely consistent results could arise because of the geographical structure
366 that the barn swallow population showed in the winter quarters, with two latitudinally
367 well-separated main clusters. To further investigate the robustness and the generality of
368 our method we then re-ran the first level analysis on a partly simulated dataset. We
369 assigned to barn swallows classified in cluster All.2 a new position in the wintering
370 ground by adding to their actual latitude the difference in latitude between the centres of
371 clusters All.1 and All.2, while leaving unchanged longitude in the wintering grounds as
372 well as position in the breeding area (see Fig. 5). This partly de-structured our dataset.
373 Mantel correlation coefficient for this partly simulated dataset increased to 0.0519, $P =$
374 0.002. The cluster algorithm identified three rather than two first level clusters ($oasw =$

375 0.5067), which however corresponded to cluster All.1, All.2.1 and All.2.2, respectively,
376 with only 31 out of 1103 individuals being misclassified. Hence, despite the weaker
377 geographical structure, our method was able to detect the cluster structure present in the
378 data.
379

Accepted manuscript

379 **Discussion**

380

381 In this study, we propose a novel approach to quantitatively and objectively investigate
382 migratory connectivity, which can be applied to any organism where clearly distinct
383 geographical ranges, corresponding to different phases of the life cycle, can be identified.
384 We also propose a method to quantify migratory connectivity based on the Mantel
385 correlation coefficient and a method to distinguish between two processes that can
386 generate the observed connectivity, *i.e.* distribution pattern transfer *versus* clustering of
387 individuals. In the event of clustering, we also propose a method to identify ‘sub-
388 populations’ of individuals that tend to associate during the two phases of their life cycle
389 (e.g. breeding and wintering).

390 The number of studies focusing on ‘migratory connectivity’ has been increasing
391 rapidly in recent years, mainly due to an expanding set of techniques, but also due to new
392 analyses of data from ‘traditional’ mark-recapture methods, allowing identification of
393 breeding and wintering areas of several bird species. To date, however, quantitative
394 methods to estimate migratory connectivity have been based on the calculation of origin
395 probability of individuals from *a priori* identified geographical populations or areas.
396 These methods do not test statistically for the deviation of the observed patterns of
397 reciprocal distribution of individuals from a random mix. The method we propose, which
398 is based on relatively simple calculations and statistics, may thus contribute to fill this
399 gap.

400 A prominent feature of our approach is that the investigation of migratory
401 connectivity is not based on an *a priori* identification of breeding and wintering ‘sub-

402 ranges', which may be difficult because several species show continuous distributions
403 both in the breeding and the wintering quarters. We propose that the identification of
404 highly connected ranges should be based on the actual distribution pattern of individuals.
405 This has the double advantage of reducing the subjectivity of the analysis and increasing
406 its biological realism by relying on inherent patterns present in the data.

407 In addition, we emphasize that our approach subtly differs from that based on the
408 *a priori* identification of separate sub-ranges. In fact, following that approach, the level of
409 connectivity would appear to differ depending on which of the different breeding and
410 wintering populations is currently under focus. Imagine, for example, a single group
411 (population) of individuals that breeds in an area (B) and winter in two well separated
412 areas (W1 and W2). B would be defined as moderately connected to either W1 or W2,
413 whereas both W1 and W2 would be defined as strongly connected to B. Thus, an
414 asymmetry is implicit in that approach, with a larger relevance usually given to the
415 breeding areas. Our approach solves this problem as equal relevance is given to each
416 distance matrix, and number and composition of groups of individuals is assessed *a*
417 *posteriori* by means of cluster analysis. In our approach, identification of sub-ranges and
418 highly connected areas follows from the identification of clusters of individuals that
419 connect regions by means of their migration.

420 Some of the features of our method deserve close consideration. First, the method
421 we used to combine the two distance matrices into the overall Euclidean distance matrix
422 implies that, for a given distance between two data-points in one range, their chances to
423 be classified in the same cluster declines as their distance in the other range increases.
424 This is desirable because, according to the symmetric approach that we are proposing

425 (see above), only individuals that tend to both breed and winter together should be
426 classified into the same group. Second, the clustering procedure we chose involves the
427 comparison of the clustering efficiency, based on the *oasw*, when data are forced into 2,
428 3, ..., n clusters and the univocal identification of the best number of clusters. As a
429 drawback of the method, however, if the pattern of distribution of sampling locations
430 varies markedly between the breeding and the wintering areas, some clusters may be
431 obscured and forced by the clustering algorithm to join with other clusters. Finally, high
432 connectivity does not imply, *per se*, a highly clustered population. If, for example,
433 uniformly dispersed individuals in the breeding areas tend to maintain the same position,
434 relative to the other individuals, in the wintering areas (i.e. in the case of distribution
435 pattern transfer), the connectivity will be high even in the absence of a clear grouping of
436 individuals. Thus, our approach allows inferring connectivity arising from distribution
437 pattern transfer when the Mantel correlation coefficient is significant, but the whole
438 population cannot be partitioned into well-defined clusters.

439 The simulations we run under three different scenarios give further insights into
440 the interpretation of the results and the robustness of this method. First of all, the Mantel
441 correlation coefficient is able to detect connectivity due to distribution pattern
442 transference even when a large random noise is added to the data. This clearly emerges
443 from the results of the first and the third simulations. In the second and third simulation,
444 *oasw* values lower than the suggested threshold of 0.5 were recorded in several runs
445 where a cluster structure could then be identified, as indicated by the fact that the same
446 clustering algorithm was able to detect the correct number of clusters and to correctly
447 classify more than 95% of individuals in all but 18 simulations of the second and third

448 type. In all cases where the method failed to detect connectivity, the distance between
449 cluster centres was not larger than 3 SD in one area and not larger than 2 SD in the other,
450 implying that the grouping structure was obscured by the fact that about 13% of
451 individuals mixed in one ground and about 60% in the other (details of this calculations
452 are not shown for brevity). On the other hand, Mantel correlation coefficients seemed
453 rather sensitive in detecting a structure in a population also when it arose from grouping
454 of individuals. Based on these results, we suggest to always investigate the cluster
455 structure of a population when the Mantel test is significant, and to use the value of the
456 Mantel correlation coefficient as a measure of migratory connectivity. In addition, we
457 tentatively suggest to investigate the structure of a group of individuals when the *oasw* is
458 larger than 0.4, since, based on our simulations, the suggested threshold of 0.5 may be too
459 high. We admit, however, that further investigation is needed to identify an optimal
460 threshold for the *oasw* value.

461 Our method can be widely applied to different kinds of data about migratory birds
462 (and other migratory animals). Indeed, this method of analysis is applicable to any pair of
463 matrices of distance indices between individuals and not to geographical distances only.
464 For example, a large number of studies about migratory connectivity is based on the
465 analysis of the isotopic composition of feathers (see e.g. Hobson 2005). Differences in
466 the isotopic composition are considered to be related to the distance in the locations
467 where individuals moult their feathers at least in areas where specific geographical
468 gradients in isotopic abundance exist (Hobson 2005). Hence, a measure of the migratory
469 connectivity for a population can be obtained by correlating the matrix of geographic
470 distances between places where individuals were captured and the matrix of differences

471 in the isotopic composition of their feathers. During moult, birds simultaneously wear
472 both recently moulted and old feathers when in the wintering or in the breeding areas.
473 The two matrices derived from isotopic profiles of newly moulted and old feathers should
474 be readily suitable for the analyses proposed here. Likewise, our method could be applied
475 to a matrix of isotopic profiles produced in winter and a second matrix of genetic markers
476 for the same individuals reflecting the population genetic structure of individuals at the
477 breeding grounds. We emphasize, however, that our method could be applied to distances
478 in the isotopic profiles and/or genetic distances provided that they are univocally related
479 to geographic distances between individuals. For example, deuterium values in North
480 America vary along a latitudinal gradient and were therefore used to study migration of
481 birds moving in a north-south direction (Hobson 2005), but this may not be the case for
482 other areas or other isotopes, like ^{13}C and ^{15}N in Africa, which provide habitat-specific,
483 rather than geographic area-specific markers (Møller and Hobson 2003), so that distance
484 in isotopic composition may not reflect true geographic distance between individuals. As
485 concerns genetic distances, these usually fulfil the criterion because isolation by distance
486 is a common feature of such distances.

487 We applied our method to a large ringing and recovery dataset of a Palearctic
488 migratory bird species that has been subject to intensive ringing programmes. We found
489 that the Western Palearctic breeding population of barn swallows is structured into two
490 main 'sub-populations'. The first population breeds in South-Western Europe and winters
491 in an elongated belt from Liberia to Uganda, while the second population breeds in
492 Northern Europe and winters south of the Equator. A second level of analysis showed
493 that both main groups could be split into two sub-groups. Barn swallows breeding in

494 South-Western Europe could be divided according to segregation in the wintering
495 quarters with a first sub-group including barn swallows wintering from Liberia to
496 Cameroon and a second including barn swallows that winter from Cameroon to Uganda.
497 Conversely, the second main group can be divided according to segregation in the
498 breeding grounds, with a first sub-cluster that includes barn swallows mainly breeding in
499 North-Eastern Europe and a second group including barn swallows that breed in North-
500 Western Europe (Fig. 5b).

501 This pattern was identified based on the entire dataset (1103 individuals),
502 although the analyses were possibly confounded by the fact that either breeding and
503 wintering ringing or recovery could partly refer to migrating individuals. In addition,
504 differences in ringing effort at each ringing location may skew the results as the number
505 of ringed birds recovered at each location can be considered proportional to capture
506 effort. This problem could be exacerbated by the fact that larger ringing effort may be
507 devoted in areas with higher bird densities. Unfortunately, no quantitative measure of
508 ringing effort was available so it was impossible to correct for this potential bias.
509 However, the analyses based on different subsets of data identified according to
510 increasingly conservative criteria gave largely consistent results to those based on the
511 entire dataset, suggesting that the analysis was robust to the interference of different
512 potential sources of bias and even sample size. However, for several bird species less
513 than 200 recoveries may be available and small sample size may reduce the power of the
514 tests. Further theoretical as well as simulation studies are therefore needed to assess the
515 effect of sample size on the detection of migratory connectivity. However, some
516 preliminary results indicate that this method is robust when at least 35 recaptures are

517 available (R. Ambrosini, unpublished results). In addition, we emphasize that our main
518 goal was to propose a method to quantify migratory connectivity, and that this large set of
519 real data was only used to exemplify its application.

520 Despite these potential drawbacks, the results of the present analysis of
521 connectivity can be qualitatively compared to the results obtained using different
522 approaches. The subdivision of the Western Palearctic barn swallow population in two
523 main sub-populations is roughly consistent with the observation of differences in the
524 quantitative genetic variance-covariance matrix (G-matrix) between barn swallows from
525 Northern and Southern Europe (Roff et al 2004). In addition, in an analysis of the isotopic
526 composition of feathers, Evans et al (2003) identified segregation in the wintering
527 quarters of barn swallows breeding in England and Switzerland. This is consistent with
528 our results, as British barn swallows are mainly classified in a group that winters in
529 Southern Africa, while Swiss barn swallows were included in the cluster wintering north
530 of the Equator. Moreover, stable isotope composition of feathers of barn swallows
531 indicates that birds that breed in Denmark winter in at least two different areas with
532 different levels of $\delta^{15}\text{N}$ (Møller and Hobson 2003). This is consistent with our results,
533 since barn swallows breeding in Denmark were classified partly in the first and partly in
534 the second first-level clusters, whose African wintering grounds are separated. High $\delta^{15}\text{N}$
535 values in willow warbler (*Phylloscopus trochilus* L) feathers sampled in South Africa
536 was reported by Bensch et al. (2006). If we assume that isotopic profiles in barn swallow
537 feathers are not dissimilar to those in the willow warbler, we can go further and
538 tentatively suggest that $\delta^{15}\text{N}$ enriched cluster of Møller and Hobson should correspond to
539 cluster All.2 and the $\delta^{15}\text{N}$ depleted one to cluster All.1.

540 These results suggest that analyses of stable isotopes or other markers of group
541 identity of barn swallows should provide evidence of clusters similar to those reported in
542 Fig. 4, both when based on captures in the winter quarters in Africa or at the breeding
543 grounds in Europe. Hence, the connectivity pattern we found based on ringing recoveries
544 is largely consistent with the picture provided by studies of individual breeding
545 populations based on different approaches. In addition, although the identification of a
546 structure of the barn swallow population under focus is not novel, our analysis provides a
547 first quantitative measure of migratory connectivity.

548 In conclusion, we have proposed a quantitative measure of migratory connectivity
549 and a method for identification of ‘sub-populations’ that are amenable to inferential
550 statistical analysis and open the possibility of quantitatively investigating the
551 relationships between different areas due to the movement of migrants between them.
552 This measure can be quickly calculated for several species as large sets of data from
553 recoveries are already available at least for the most common species, and datasets on
554 genetic and isotopic distances between populations are rapidly growing. In addition, a
555 quantitative measure of migratory connectivity allows the comparison of migration
556 strategies of different species and/or populations and thus allows further investigation of
557 the ecological bases of the evolution of migratory systems. Finally, this method allows
558 estimates of migratory connectivity between geographic regions based on assemblages of
559 species that migrate between them. Such a measure may provide a substantial
560 contribution in the planning of effective conservation strategies for migratory species.

561 **Acknowledgments**

562

563 EURING, SAFRING and numerous ringing centres in Europe kindly provided access to
564 their recoveries of ringed barn swallows. We are grateful to Dr. Riccardo Borgoni and to
565 two anonymous referees whose comments greatly improved the quality of the
566 manuscript.

567

Accepted manuscript

567 **References**

568

569 Anderson, Å., Follestad, A., Nilsson, L., Persson, H., 2001. Migration patterns of Nordic
570 Greylag Geese *Anser anser*. *Ornis Sve.* 11, 19-58.

571

572 Besag, J., Diggle, P. J., 1977. Simple Monte Carlo tests for spatial pattern. *Appl. Stat.* 26,
573 327-333.

574

575 Bensch, S., Bengtsson, G., Åkesson, S., 2006. Patterns of stable isotope signatures in
576 willow warbler feathers collected in Africa. *J Avian Biol*, 37, 323-330., doi:
577 [10.1111/j.2006.0908-8857.03628.x](https://doi.org/10.1111/j.2006.0908-8857.03628.x)

578

579 Cohn, J. P., 1999. Tracking wildlife: high-tech devices help biologists trace the
580 movement of animals though sky and sea. *BioScience* 49, 12-17.

581

582 Evans, K. L., Waldron, S., Bradbury, R. B., 2003. Segregation in African wintering
583 ranges of England and Swiss Swallow *Hirundo rustica* populations: a stable isotope
584 study. *Bird Study* 50, 294-299.

585

586 Hobson, K. A., Wassenaar, L. I., 1997. Linking breeding and wintering grounds of
587 Neotropical migrant songbirds using stable hydrogen isotopic analysis of feathers.

588 *Oecologia* 109, 142-148., doi: 10.1007/s004420050068

589

- 590 Hobson, K. A., 2005. Stable isotopes and the determination of avian migratory
591 connectivity and seasonal interactions. *Auk* 122, 1037-1048.
592
- 593 Hooge, P. N., Eichenlaub, B., 2000. Animal movement extension to Arcview. Ver. 2.0. -
594 Alaska Science Centre - Biological Science Office, U.S. Geological Survey, Anchorage,
595 AK, USA. Available from: <http://www.absc.usgs.gov/glba/gistools/index.htm> (accessed
596 July 2007)
597
- 598 Jennrich, R. I., Turner, F. B., 1969. Measurement of non-circular home range. *J. Theor.*
599 *Biol.* 22, 227-237.
600
- 601 Kaufman, L., Rousseeuw, P. J., 1990. *Finding Groups in Data: An Introduction to Cluster*
602 *Analysis*. Wiley, New York.
603
- 604 Lopes, R. J., Marques, J. C., Wennerberg, L., 2006. Migratory connectivity and temporal
605 segregation of dunlin (*Calidris alpina*) in Portugal: evidence from morphology, ringing
606 recoveries and mtDNA. *J. Ornithol.* 147, 85-94. 10.007/s10336-005-0048-y.
607
- 608 Marra, P. P., Hobson, K.A., Holmes, R.T., 1998. Linking winter and summer events in a
609 migratory bird by using a stable-carbon isotopes. *Science* 288, 2040-2042., doi:
610 10.1126/science.282.5395.1884
611
- 612 Marra, P.P., Norris, D.R., Haig, S.M., Webster, M., Royle, J.A., 2006. Migratory

- 613 connectivity. In: Crooks, K. R., Sanjan, M. (Eds) Connectivity Conservation. Cambridge
614 University Press, Cambridge, pp 157-183.
- 615
- 616 Marshak, S., 2001. Earth: Portrait of a Planet. W. W. Norton & Company, New York.
- 617
- 618 Møller, A. P., Hobson, K.A., 2003. Heterogeneity in stable isotopes profiles predicts
619 coexistence of populations of barn swallows *Hirundo rustica* differing in morphology and
620 reproductive performance. Proc. R. Soc. B. 71, 1355-1362., doi: 10.1098/rspb.2003.2565.
- 621
- 622 Roff, D. A., Mousseau, T., Møller, A. P., de Lope, F., Saino, N., 2004. Geographic
623 variation in the G matrices of wild populations of barn swallow. Heredity 93, 8-14., doi:
624 10.1038/sj.hdy.6800404
- 625
- 626 Rousseeuw, P. J., 1987. Silhouettes: A graphical aid to the interpretation and validation
627 of cluster analysis. J. Comput. App. Math. 20, 53-65.
- 628
- 629 Salomonsen, F., 1955. The evolutionary significance of bird migration. Biologiske
630 Meddelelser 22, 1-62.
- 631
- 632 Sokal, R. R., Rohlf, F. J., 1995. Biometry: The Principles and Practice of Statistics in
633 Biological Research, 3rd edn. Freeman, New York
- 634
- 635 Szép, T., Møller, A. P., Piper. S., Nuttall, R., Szabó, Z. D., Pap, P. L., 2006. Searching for

636 potential wintering and migration areas of a Danish Barn Swallow population in South

637 Africa by correlating NDVI with survival estimates. *J. Ornithol.* 147, 245-253., doi:

638 10.1007/s10336-006-0060-x

639

640 Turner, A. K., 1994. *The swallow*. Hamlyn, London.

641

642 Webster, M. S., Marra, P. P., Haig, S. M., Bensch, S., Holmes, R. T., 2002. Links

643 between worlds: unravelling migratory connectivity. *Trends Ecol. Evol.* 17, 76-83., doi:

644 10.106/S0169-5347(01)02380-1.

645

Accepted manuscript

645 **Figure legends**

646

647 Figure 1. Calculation of orthodromic distance between points A and B along a spherical
648 surface. Geographical coordinates (latitude = φ and longitude = λ) are first converted
649 into three-dimensional Cartesian coordinates (x, y, z) with Eq. 1-3 in figure. r is the
650 average radius of the Earth (6371 km; Marshak 2001). Euclidean distance (d) between the
651 points is easily calculated from Cartesian coordinates by means of Pythagoras' theorem
652 and then converted into orthodromic distance (l) with Eq. 4.

653

654 Figure 2. Results from the second simulation. *Oasw*: overall average silhouette width.
655 Gray surface colour indicates a simulation with a significant ($P \leq 0.05$) Mantel
656 correlation coefficient. Black surface colour represents non-significant simulations. Plans
657 represent *oasw* values of 0.5 and 0.7 respectively. Distances between cluster centres are
658 expressed as standard deviations (SD) of the bivariate Gaussian used to generate
659 individual position within clusters.

660

661 Figure 3. Dendrogram of the analysis (a) of the whole set of recoveries and recaptures of
662 barn swallows (All) and (b) of the subset of data only including barn swallows ringed in
663 Europe as adults (AE). r_M : Mantel correlation coefficient; P : significance of the Mantel
664 test as assessed with the randomization procedure; *oasw*: overall average silhouette
665 width.

666

667 Figure 4. First level clusters for different selection criteria. a) All barn swallows . b)

668 Barn swallows recorded in the breeding and wintering grounds in the focal periods of
669 reproduction and wintering respectively. c) Barn swallows ringed in Europe as adults. d)
670 Barn swallows ringed in Europe as adults in the focal period of reproduction and
671 recovered in Africa in the focal period of wintering.

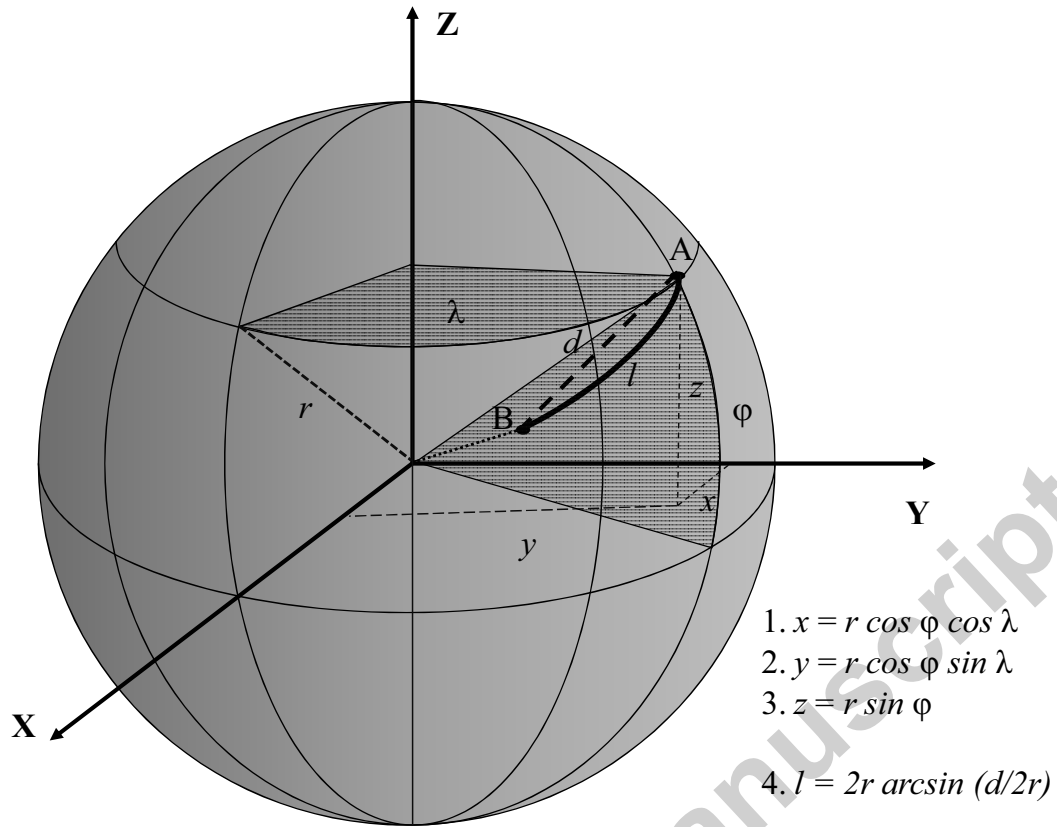
672

673 Figure 5. Pictorial representation of first (a) and second (b) level clusters. Ellipses
674 represent the Jennrich and Turner (1969) Bivariate Normal Home Range calculated with
675 the Animal Movement extension to ArcView Gis (Hooge and Eichenlaub 2000). Dashed
676 ellipses in (a) represents the simulated position of cluster All.2 used to assess the
677 robustness of the method with a less geographically structured dataset.

678

679 Figure 6. Silhouette plot showing the classification of the All dataset in two first-level
680 clusters. Each bar represents the silhouette values $s(i)$ for a single barn swallow (see also
681 Statistical methods). Within each cluster, bars are drawn in decreasing length order.
682 Large values indicate good classification.

683



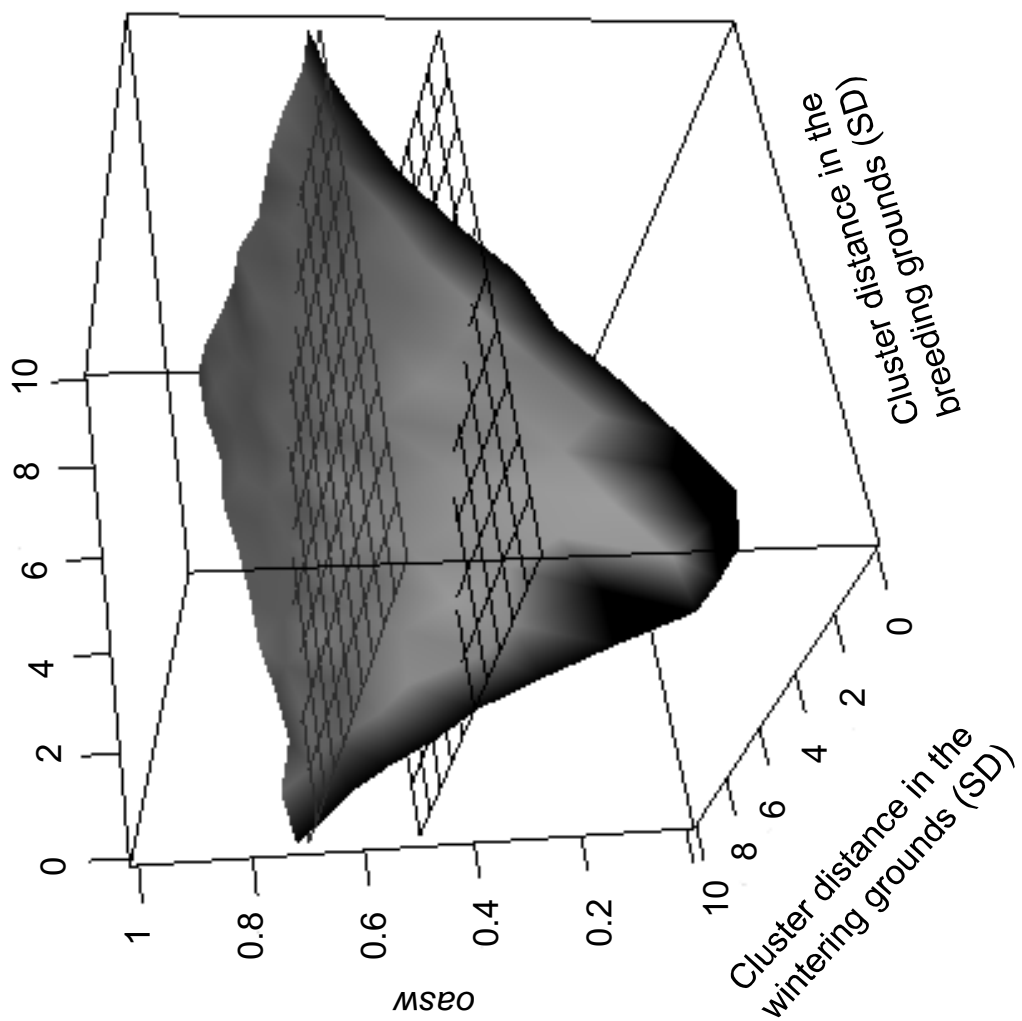
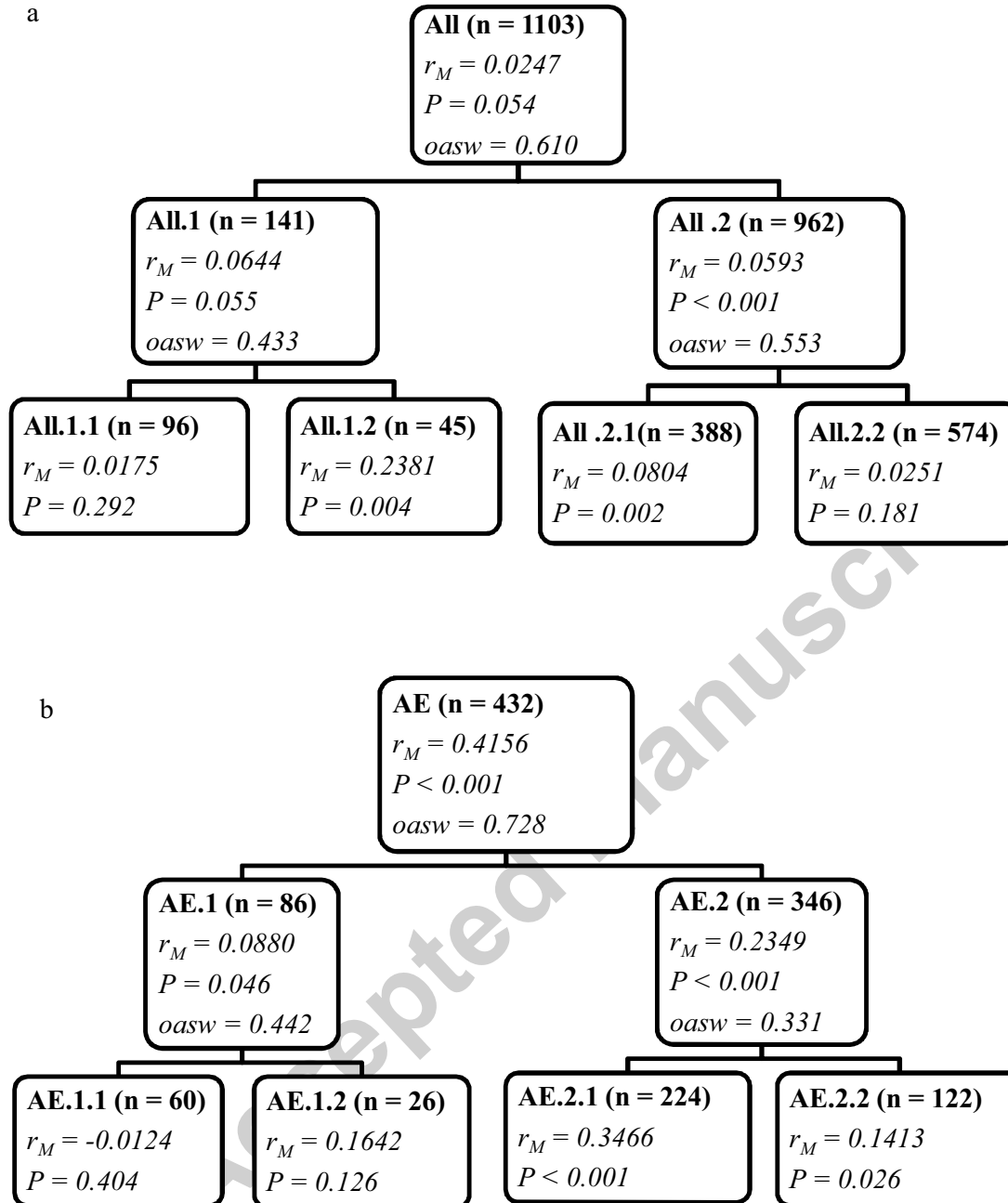
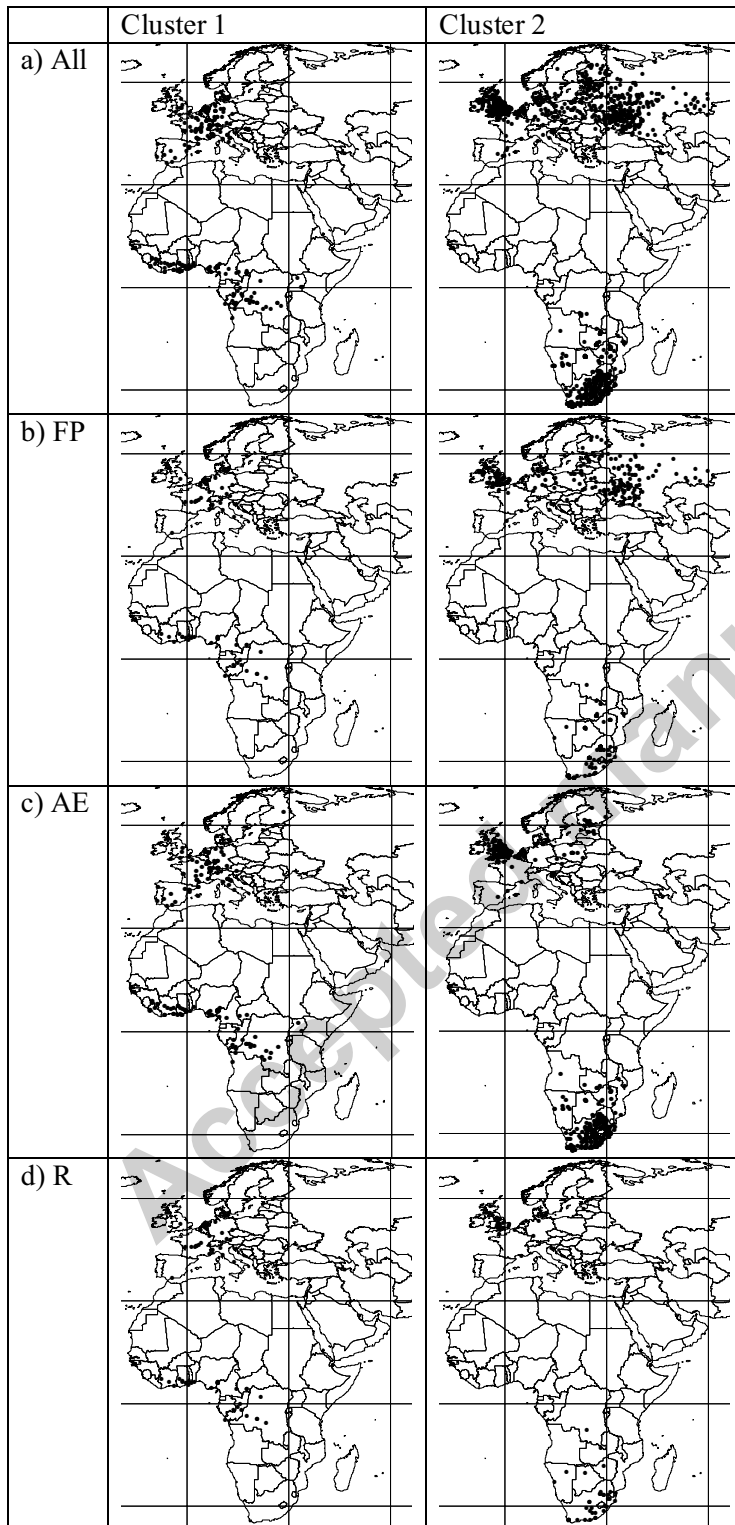


Figure 2





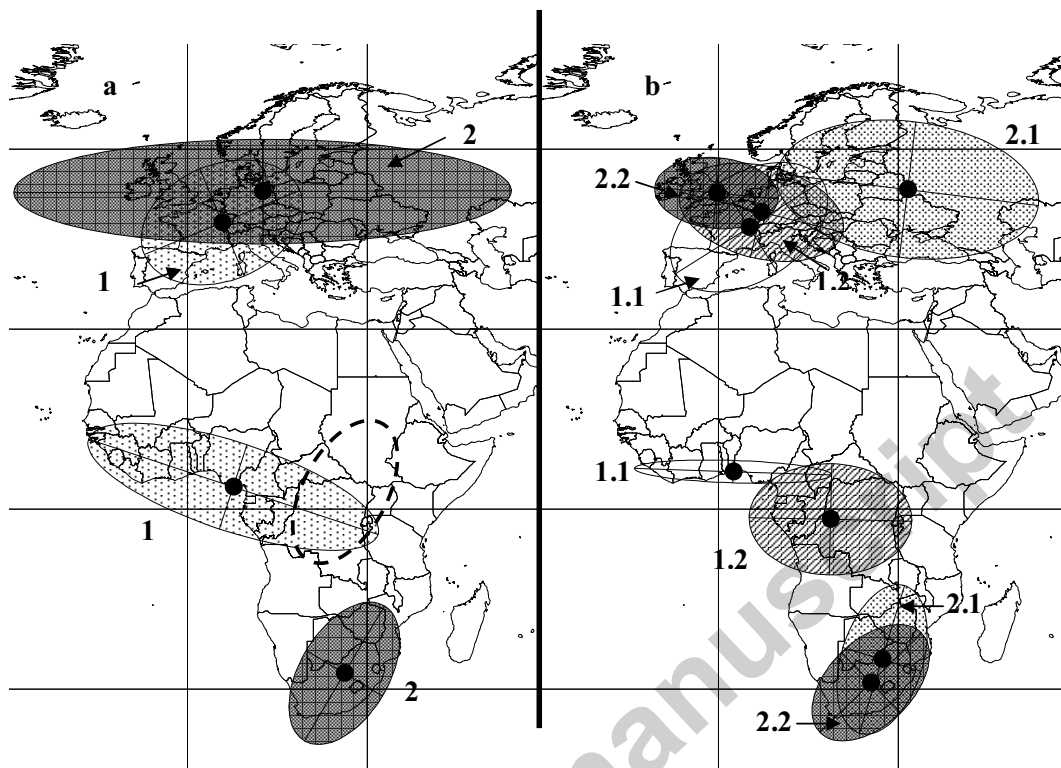


Figure 6

