At-sea movements of wedge-tailed shearwaters during and outside the breeding season from four colonies in New Caledonia

Henri Weimerskirch*, Sophie de Grissac, Andreas Ravache, Aurélien Prudor, Alexandre Corbeau, Bradley C. Congdon, Fiona McDuie, Karen Bourgeois, Sylvain Dromzée, John Butscher, Christophe Menkes, Valérie Allain, Eric Vidal, Audrey Jaeger, Philippe Borsa

*Corresponding author: henri.weimerskirch@cebc.cnrs.fr

Marine Ecology Progress Series 633: 225–238 (2020)

Supplement 1

Test for pseudo-replication in the tracking dataset

Our dataset include single individuals tracked for multiple short trips (2 to 4). If these individuals show high site fidelity, pseudo-replication in the dataset can bias the results of the spatial analyses.

For both colonies Canard and Mato, only 2 individuals performed 2 successive trips. Since it is not enough to test statistically site fidelity, we simply plotted the tracks and made sure the trips were different. All 4 individuals visited clearly different areas during their 2 successive short trips so we kept all trips.

For colonies Pindai and Temrock we performed a test for pseudo-replication following the method designed by and described in Lascelle et al. 2016. This test compares the similarity of foraging locations of a single tracked bird with those of the rest of the data group. For individual that performed multiple trips the test computes the kernel utilization distribution (UD) for each trip, and then calculate the Hausdorff distance between these areas for every combination of trips, to quantify proximity (Munkres, 2018). The resulting individual distance matrix was compared to a data group reference distribution. To calculate the reference distribution, we randomly selected one trip from each tracked individual of the same colony (including individuals with only one track), and calculated the Hausdorff distance for foraging areas (UD) between individuals. The within individual distances were then compared against the population-level distances using a Mann–Whitney U-test. This examined whether the null hypothesis – that the proximity of core areas from a single individual is similar to the proximity of core areas between different individuals of the same population and life history stage – could be rejected, in which case there was some indication of site-fidelity and thus pseudo-replication. This process was repeated 100 times to account for possible bias in the random sample, and the mean p-value calculated (see Table S1 below).

For individuals for which the p-value was < 0.1 (indicative of pseudo-replication) we randomly selected one single trip for use in spatial analyses (UD computation and overlap calculation); otherwise, all data were retained.

References

Lascelles BG, Taylor PR, Miller M G, Dias MP, Oppel S, Torres L, Hedd A, Le Corre M, Phillips RA, Shaffer SA, Weimerskirch H, Small C (2016), Applying global criteria to tracking data to define important areas for marine conservation. Diversity Distrib, 22:422–431. doi:10.1111/ddi.12411

Munkres, James R. Elements of algebraic topology. CRC Press, 2018.

Table S1. Number of trips per individuals tracked multiple times in Pindai and Temrock and mean p-value of the pseudo-replication test (100 replication of Mann–Whitney U-test comparing within-individual Hausdorff distances and inter-individuals Hausdorff distances for each colony). Lines in bold highlight the individuals that show pseudo-replication and for which a single trip was randomly selected in further analyses.

Colony	ID	No trips	mean p-value
Temrock	**Temrock_38305**	**2**	**0.094**
Temrock	Temrock_38328	2	0.711
Temrock	Temrock_38309	2	0.193
Temrock	Temrock_38310	2	0.960
Temrock	Temrock_38344	2	0.413
Temrock	Temrock_38322	2	0.184
Temrock	Temrock_38362	3	0.799
Pindai	Pindai_GPS5.2	2	0.252
Pindai	Pindai_SNA08	4	0.185
Pindai	Pindai_SNA09	2	0.246
Pindai	Pindai_SNA13	2	0.148
Pindai	Pindai_SNA16	2	0.230
Pindai	Pindai_SNA22	3	0.003
Pindai	Pindai_SNA23	2	0.347
Pindai	Pindai_Tag16475	2	0.946
Pindai	**Pindai_Tag16489**	**2**	**0.093**
Pindai	Pindai_Tag16514	2	0.300
Pindai	Pindai_Tag16521	3	0.644
Pindai	**Pindai_Tag16541**	**3**	**0.021**
Pindai	Pindai_Tag16545	3	0.932

Supplement 2

Text S2. Statistical analysis of foraging areas overlaps

We use the randomization procedure described in Cecere et al. 2018 (see also Lascelles et al. 2016) to quantify the overlap between the foraging areas of individuals belonging to the four different colonies. The following description is based on that of Cecere et al. 2018.

The first step is to build an overlap matrix of all the individual UDs from all colonies, based on GPS locations classified as "foraging". The UD overlap between pairs of individuals i,j is calculated using the Utilization Distribution Overlap Index (UDOI, Fieberg & Kochanny 2005). UDOI values range from zero (no overlap) to 1 (uniformly distributed and have 100% overlap; but can be >1 when UDs are non-uniformly distributed). A second matrix in then built where each pair of individuals i, j is coded as 0 if both individuals belonged to the same colony, and 1 if they belonged to different colonies. We calculated the correlation coefficient r_{obs} (Pearson's r) between the two matrices. Because of the coding of colony membership, highly negative values of robs indicate that 1) foraging areas (UD) of individuals belonging to the same colony are highly overlapping, and that 2) those of individuals belonging to different colonies are segregated. We then randomly and independently rotate each individual track 5000 times and calculate each time a UD overlap matrix, which is correlated with the colony membership matrix. The rotation is anchored to colony coordinates and for short trips the land barrier (mainland) is considered by restricting the angle variation according to the colony (e.g. Pindai and Temrock tracks can only be rotated on a 180° semi-circle). By this randomization bootstrap procedure, we obtain a distribution of r values representing the null hypothesis of random spatial distribution of UDs around the colonies and we can calculate the probability (p_{rand}) that r_{rand} is more negative than r_{obs} (i.e. a very small p_{rand} indicates that the UDs of individuals from different colonies were more segregated than random) as shown in figures S1. We use the same procedure to quantify segregation between (Fig. S2) and within pairs (Fig. S3 and S4) of neighbouring colonies Pindai/Temrock and Canard/Mato.

All computations were performed in R 3.3.153.

References

Cecere Jacopo G et al. "Spatial segregation of home ranges between neighbouring colonies in a diurnal raptor." *Scientific reports* 8.1 (2018): 11762.

Fieberg J, Kochanny CO (2005) Quantifying home-range overlap: the importance of the utilization distribution. *J. Wildl. Manag.* 69:1346–1359

Figure S1. UD overlap of individuals belonging to the 4 different colonies. Frequency distribution of randomized *r* values (r_{rand}) obtained from random rotations of foraging areas (UDs) for short and long trips of individuals belonging to the four colonies Canard, Mato, Pindai and Temrock. More negative *r* values denote greater spatial segregation of foraging areas between birds from different colonies. The observed *r* value (r_{obs}), resulting from the spatial distribution of foraging areas shown in Fig. 7 of the article, is highlighted with a dashed red line within each panel.

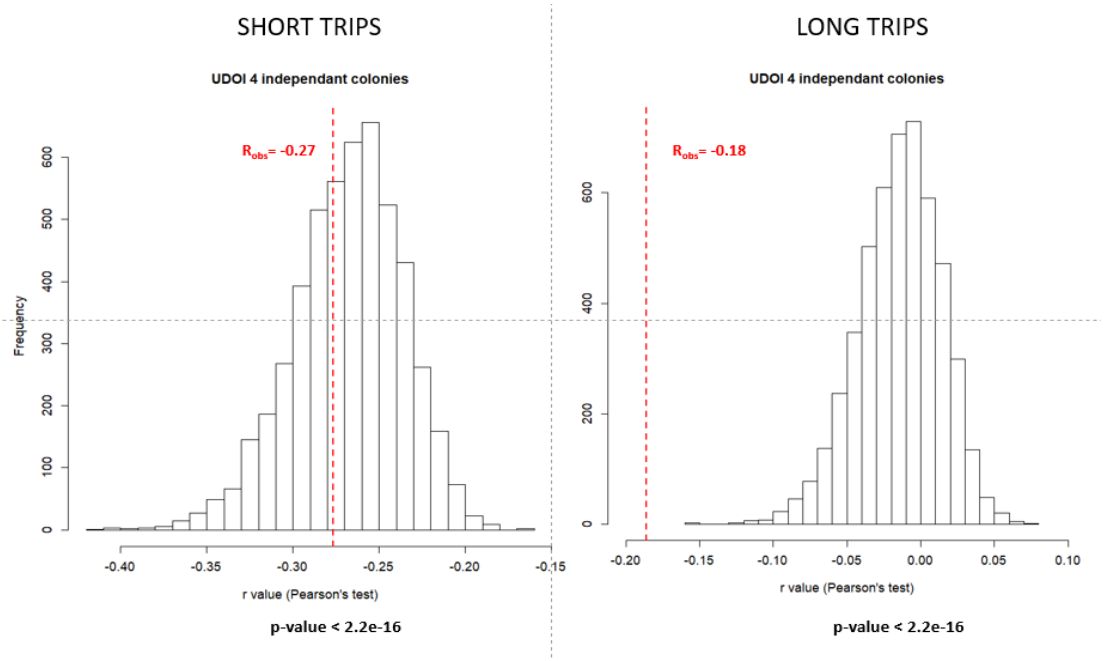


Figure S2. UD overlap of individuals belonging to pairs of neighbouring colonies (east / south). Frequency distribution of randomized *r* values (r_{rand}) obtained from random rotations of foraging areas (UDs) for short (A) and long (B) trips of individuals belonging to pairs of neighbouring colonies (i.e. southern colonies pair Canard-Mato and eastern colonies pair Pindai-Temrock). More negative *r* values denote greater spatial segregation of foraging areas between birds from neighbouring colonies. The observed *r* value (r_{obs}), resulting from the spatial distribution of foraging areas shown in Fig. 7 of the article, is highlighted with a dashed red line within each panel.

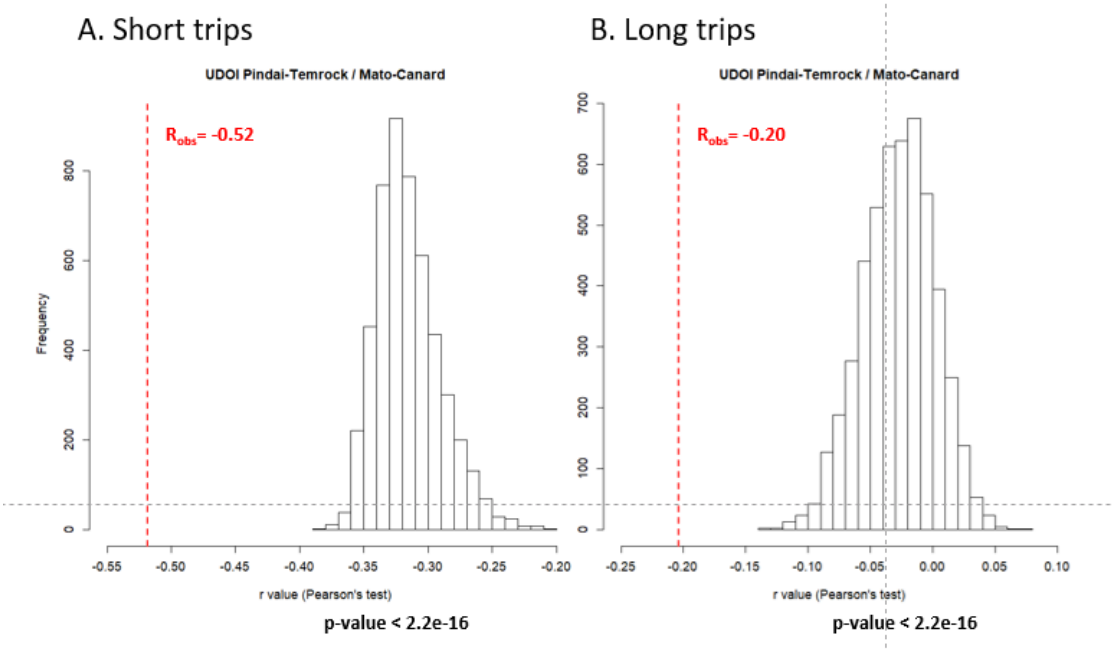


Figure S3. UD overlap of short trips of individuals belonging to neighbouring colonies. Frequency distribution of randomized *r* values (r_{rand}) obtained from random rotations of foraging areas (UDs) for **short trips** of individuals belonging to **neighbouring colonies** A) southern colonies (Canard and Mato) and B) eastern colonies (Pindai and Temrock). More negative *r* values denote greater spatial segregation of foraging areas

between birds from neighbouring colonies. The observed *r* value (r_{obs}), resulting from the spatial distribution of foraging areas shown in Fig. 7 of the article, is highlighted with a dashed red line within each panel.

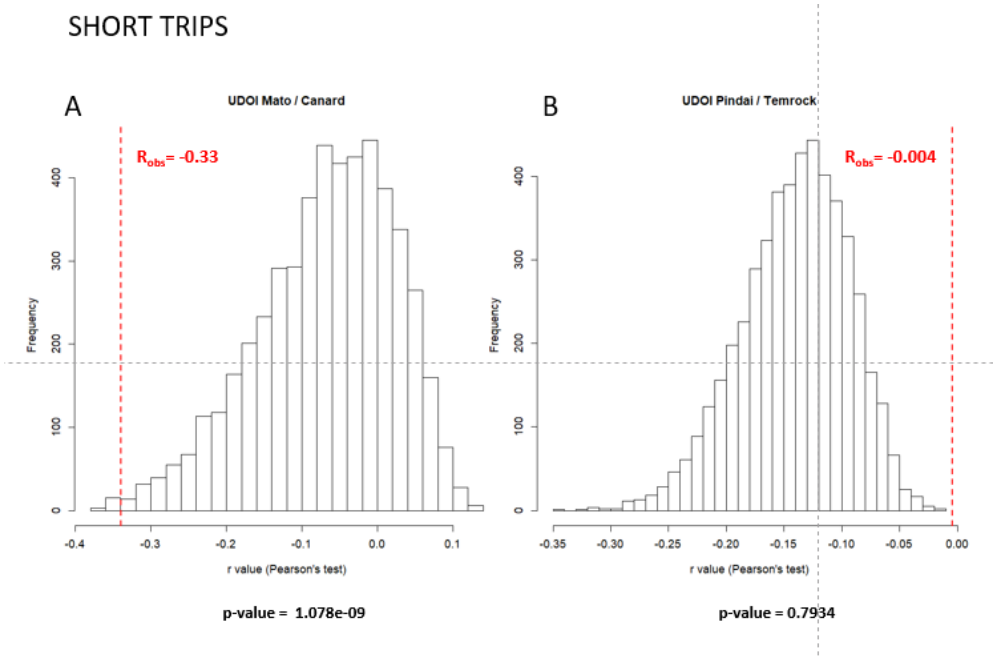
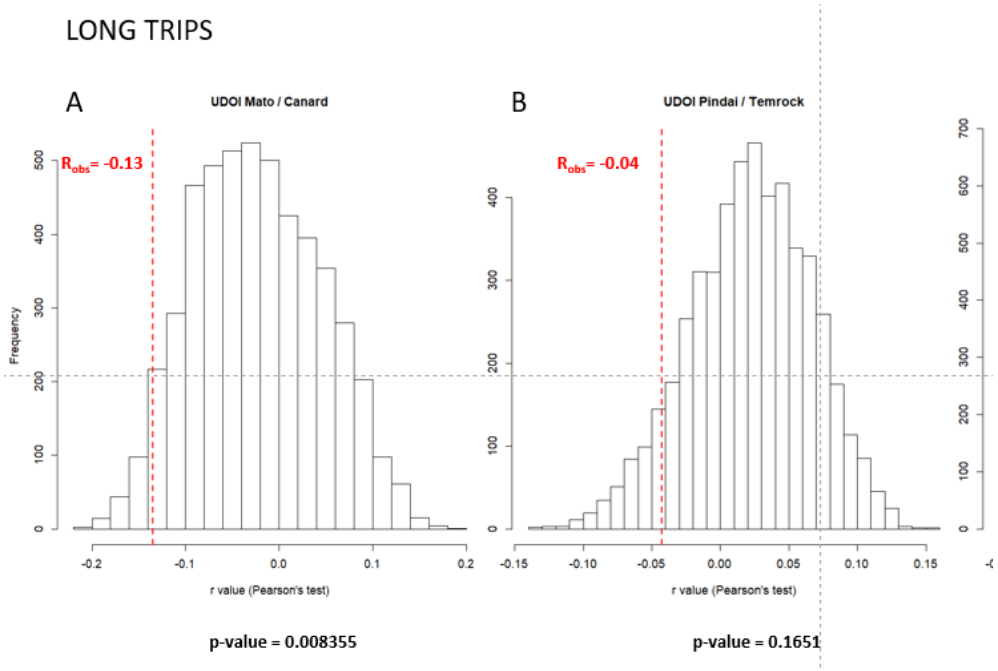


Fig. S4. UD overlap of long trips of individuals belonging to neighbouring colonies. Frequency distribution of randomized *r* values (r_{rand}) obtained from random rotations of foraging areas (UDs) for **long trips** of individuals belonging to **neighbouring colonies** A) southern colonies (Canard and Mato) and B) eastern colonies (Pindai and Temrock). More negative *r* values denote greater spatial segregation of foraging areas between birds from neighbouring colonies. The observed *r* value (r_{obs}), resulting from the spatial distribution of foraging areas shown in Fig. 7 of the article, is highlighted with a dashed red line within each panel.



Supplement 3

Text S3. Representativeness assessment

Only a fraction of each colony has been tracked and small sample sizes may be insufficient to capture the variability among individuals in space use (Lindberg & Walker, 2007). Therefore, we assessed the representativeness of the track samples following the method described in Lascelles et al. 2016 (see also Cecere et al. 2018) to check if we can infer population distribution from our dataset. The following explanation of the method is based on that of Lascelles et al. 2016.

To assess representativeness, we examine how foraging distribution of individuals from each colony change with increasing sample size. We randomly select individual trips iteratively and compare the randomly selected ('sampled') with the unselected ('unsampled') data. For each sample size, the 90% UD is calculated from the sampled data. We then assess what proportion of the unsampled data is located within this 90% UD (inclusion value). The procedure is repeated 100 times (100 random draws) for every sample size. Then, we fitted a non-linear regression to inclusion values and the representativeness of the tracked individuals was computed as the percentage of the estimated asymptote value reached by the highest predicted inclusion value. This test was not performed for Canard due to the small sample size. The obtained average 'inclusion value' is a metric indicating how well the sampled data explain the space use of individuals in the unsampled data for short foraging trips (Fig. S5A) and long foraging trips (Fig. S5B) for each of the four colonies. These assessments allow to determine if a tracked sample is representative of the colony.

References

Lascelles BG, Taylor PR, Miller MG, Dias MP, Oppel S, Torres L, Hedd A, Le Corre M, Phillips RA, Shaffer SA, Weimerskirch H, Small C (2016), Applying global criteria to tracking data to define important areas for marine conservation. Diversity Distrib., 22: 422-431. doi:10.1111/ddi.12411

Cecere Jacopo G, et al. "Spatial segregation of home ranges between neighbouring colonies in a diurnal raptor." *Scientific reports* 8.1 (2018): 11762.

Lindberg MS, Walker J. (2007) Satellite telemetry in avian research and management: sample size considerations. Journal of Wildlife Management, 71, 1002–1009.

Fig. S5. Results of the representativeness analysis for short (A) long trips (B) showing that the sample of tracked individuals reliably represents the variability in space use of birds for 3 colonies (Pindai, Mato, Temrock) but not for the colony Canard (n=7). Circles indicate the average proportion of out-of-sample GPS positions located within the 90% KDE areas estimated from sampled positions (Inclusion) for 100 random draws of sample sizes, from 1 to n-1 individuals. Shaded areas indicate variability of inclusion value for 100 random draws of tracked individuals, and the solid lines represents the fitted nonlinear regression lines for each colony. Inclusion rate (and thus representativeness of the tracking dataset) is based on the estimated asymptote of the nonlinear regression.