

Supplementary materials

Table S1. The number of overlapping estimates of breeding success in two consecutive years for all colony pairs, which is the sample size for calculating synchrony in the form of r_{diff} . Numbers in the first column correspond to the colony numbers in the first row. Note that the values in the diagonal are the number of between-year differences in breeding success that could be estimated for each colony. When calculating synchrony as the correlation between untransformed time series of breeding success (r) the sample size is larger as it does not require consecutive estimates

	1. Buchan Ness	2. Dunbar Coast	3. Fair Isle	4. Farne Islands	5. Flamborough Head	6. Foula	7. Fowlsheugh	8. Gultak	9. Handa Island	10. Hermaness	11. Isle of May	12. Lowestoft	13. Marwick Head	14. Mull Head	15. North Hill	16. North Sutor	17. Noss	18. Row Head	19. Saltburn Cliffs	20. St Abb's Head	21. Sumburgh Head	22. Whale Wick
1	19	15	19	19	17	17	18	13	19	17	19	19	18	17	15	15	19	18	17	19	19	17
2		27	24	24	25	21	22	15	25	25	25	26	22	18	16	23	23	18	18	27	24	14
3			29	26	27	22	25	18	27	24	27	29	25	22	17	22	26	23	22	28	27	18
4				28	26	23	25	17	27	24	28	28	26	21	17	22	27	22	21	28	28	18
5					30	21	25	17	27	25	27	29	25	20	15	23	26	21	20	29	27	16
6						23	20	16	23	21	23	23	21	19	17	20	23	19	19	23	23	16
7							27	15	25	22	26	27	24	19	14	20	25	20	19	26	26	18
8								18	16	14	17	18	18	18	14	14	18	18	18	17	18	11
9									29	26	28	29	25	20	17	24	26	21	20	29	27	18
10										27	25	26	22	17	17	24	23	18	17	27	24	16
11											29	29	26	21	17	23	27	22	21	29	28	18
12												31	27	22	17	24	28	23	22	30	29	18
13													27	22	17	21	26	23	22	26	27	16
14														22	17	17	22	22	22	21	22	15
15															17	16	17	17	17	17	17	13
16																25	21	17	17	25	22	15
17																	28	23	22	27	28	18
18																		23	22	22	23	16
19																			22	21	22	15
20																				31	28	18
21																					29	18
22																						18

Table S2. Final cluster structure for each algorithm when synchrony was measured as r_{diff} . (1 = Scottish east coast, 2 = Shetland, 3 = Orkney, see Fig. 1c). Colonies that were not consistently assigned to the same cluster are indicated in bold (these are shown in grey in Fig. 1c). Colony numbers correspond to numbers in Fig. 1a.

Colony number	Colony name	<i>agnes</i>	<i>diana</i>	<i>pam</i>
1	Buchan Ness	1	1	1
2	Dunbar Coast	1	1	1
3	Fair Isle	2	2	2
4	Farne Islands	1	1	1
5	Flamborough Head	2	2	2
6	Foula	2	2	2
7	Fowlsheugh	1	1	1
8	Gultak	3	3	3
9	Handa Island	4	4	1
10	Hermaness	2	2	2
11	Isle of May	1	1	1
12	Lowestoft	5	5	4
13	Marwick Head	3	3	3
14	Mull Head	3	3	3
15	North Hill	2	2	2
16	North Sutor	5	5	4
17	Noss	2	2	2
18	Row Head	3	3	3
19	Saltburn Cliffs	2	2	2
20	St Abb's Head	1	1	1
21	Sumburgh Head	2	2	2
22	Whale Wick	2	2	2

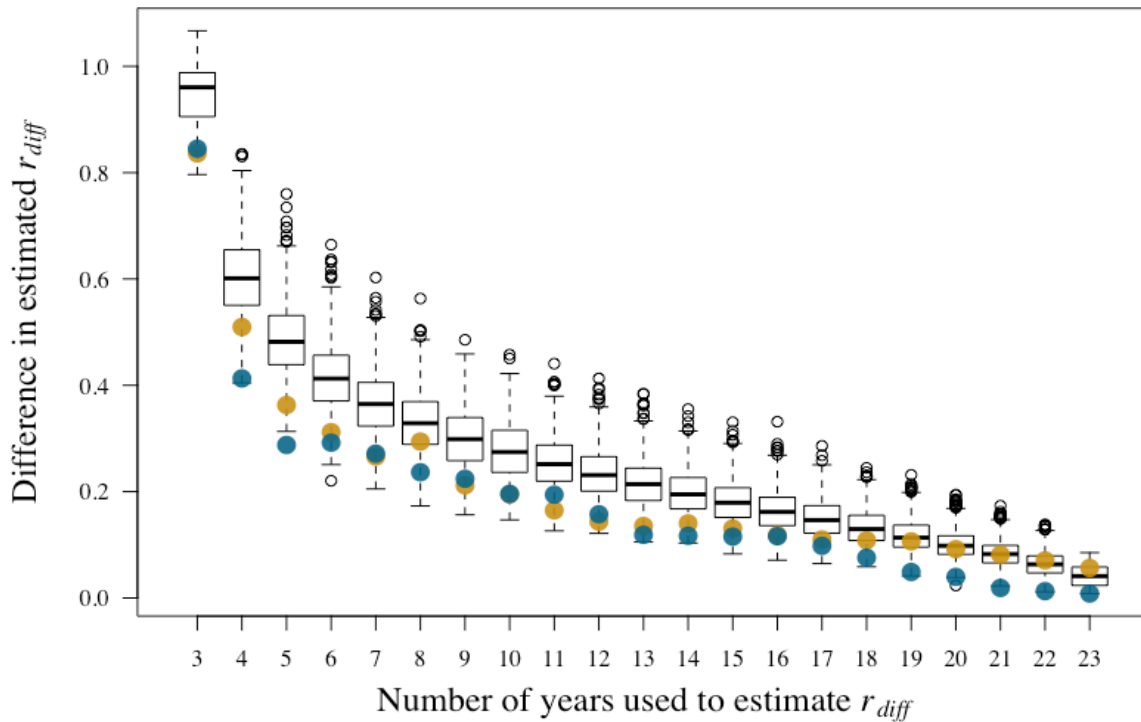


Fig. S1. In order to investigate the role of sample size we picked the 11 UK colonies with the most complete time series, which had a total of 24 years of estimated breeding success in common, and determined how the estimated r_{diff} changed as years were randomly removed from the data. This process was repeated 1000 times, where the order in which the years were removed was randomised each time. The plot shows the number of years used to estimate r_{diff} against the distribution of the 1000 estimated average absolute differences between the estimated r_{diff} and the r_{diff} estimated from the complete 24 year time series for the different colonies. It is clear that as the number of years increase, the difference decreases, being 0.180 (95% CI: 0.177; 0.183) on average at the cut-off of 15 (see Materials and Methods) (but it should be noted that the estimate based on 24 years is not a “true” estimate of correlation strength). When synchrony was measured as r rather than r_{diff} , differences were generally smaller than shown here. If a longer sequence of years is missing at the beginning or the end of the time series, which is the case in some of the colonies (see Materials and Methods), this could potentially have a different impact on the estimated synchrony if synchrony patterns has changed over time. For this reason, we also looked at how the estimated synchrony changed when removing years sequentially from the beginning or the end of the time series (yellow and blue dots, respectively). There was no evidence that this resulted in larger differences in the estimated correlation strength when synchrony was measured as r_{diff} . However, when synchrony was measured as r , the differences when years were removed sequentially were generally slightly larger than the average for when years were removed randomly, which is to be expected as this measure also reflects synchrony in long-term trends.

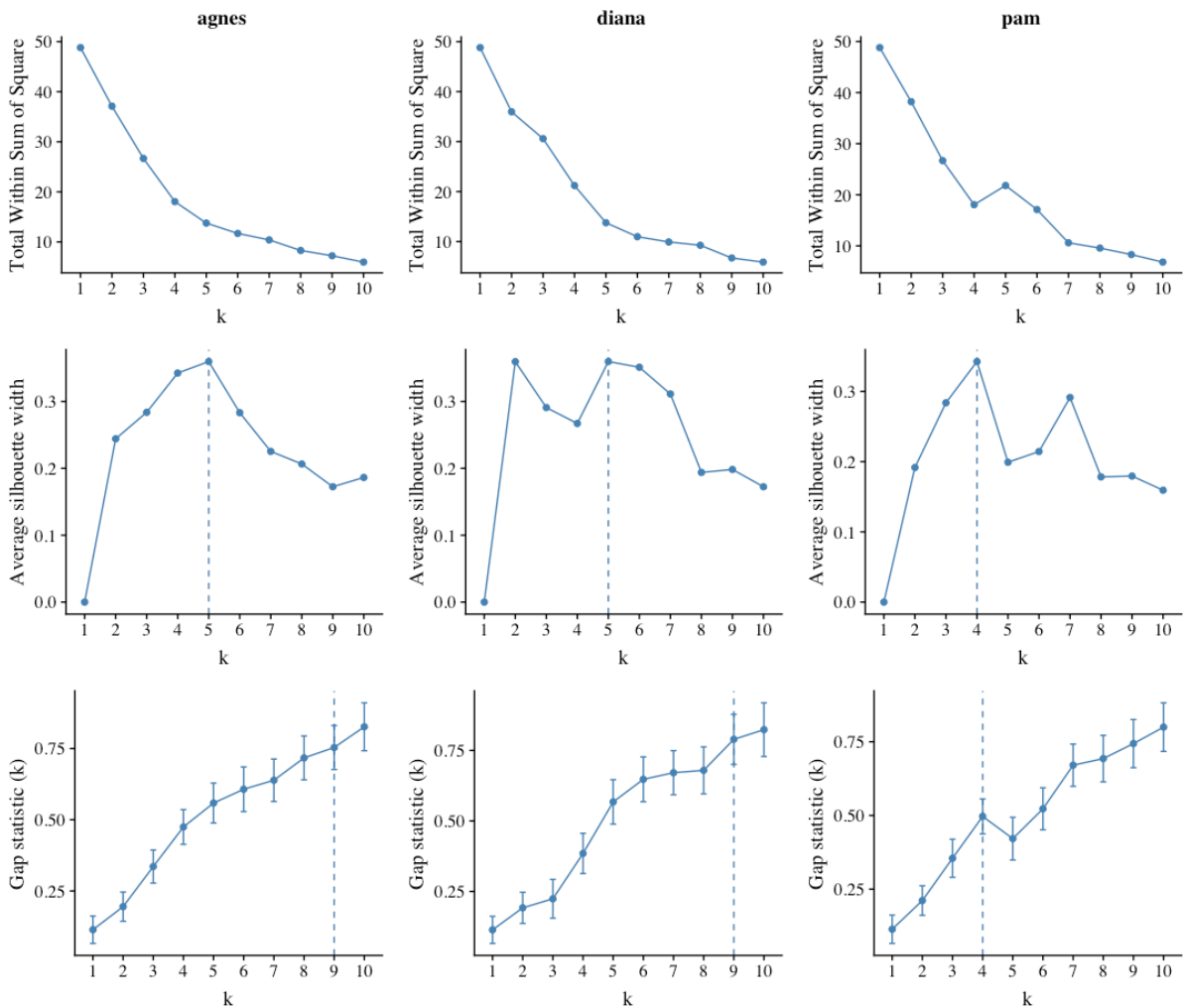


Fig. S2. Results for the three metrics used to determine the appropriate number of clusters k to use when synchrony was measured as r_{diff} . Each row represents a different metric and each column a different clustering algorithm (*agnes*, *diana* and *pam* – note that *fanny* was used primarily to look at affinities and for that reason not included). The first metric is the total within sum of squares, which shows how within-cluster variation decreases as the number of clusters increase, where the appropriate number of clusters is often considered to be point where the decrease in the total within sum of squares shows a marked change in the rate of decline, known as the “elbow criterion”. Here, this seems to suggest a k of 4 for *pam*, while it is less clear for *agnes* and *diana*, but around 4-5. The next metric is the average silhouette width, which measures the strength of within-cluster similarity as compared to the similarity with the next-closest cluster, which here suggests a k of 5 for *agnes* and *diana* and a k of 4 for *pam*. Finally, we looked at the gap statistic, which compares the intra-cluster variation for each value of k with their expected values under a null reference distribution. The optimal k is taken to be the smallest k for which the gap-statistic is not more than one standard error away from the first local maximum (standard errors are indicated by error bars). This method suggested the optimal values 9 for *agnes* and *diana* and 4 for *pam*. All methods thus suggest a k of 4 for *pam*, which is what we use for this algorithm. The values to use for *agnes* and *diana* were less obvious. Since we were more interested in large-scale patterns than fine-scale structure, we focused on the lower values suggested by the total within sum of squares and the average silhouette width rather than the higher values suggested by the gap-statistic. We chose a k of 5 for both as this was suggested as optimal based on the total within sum of

squares and also seemed reasonable for the average silhouette width. We want to note that we are not interested in k in itself but rather the spatial configuration of the clusters. Based on a similar reasoning to the one outlined here, we used a k of 3 for all algorithms when synchrony was measured as r .

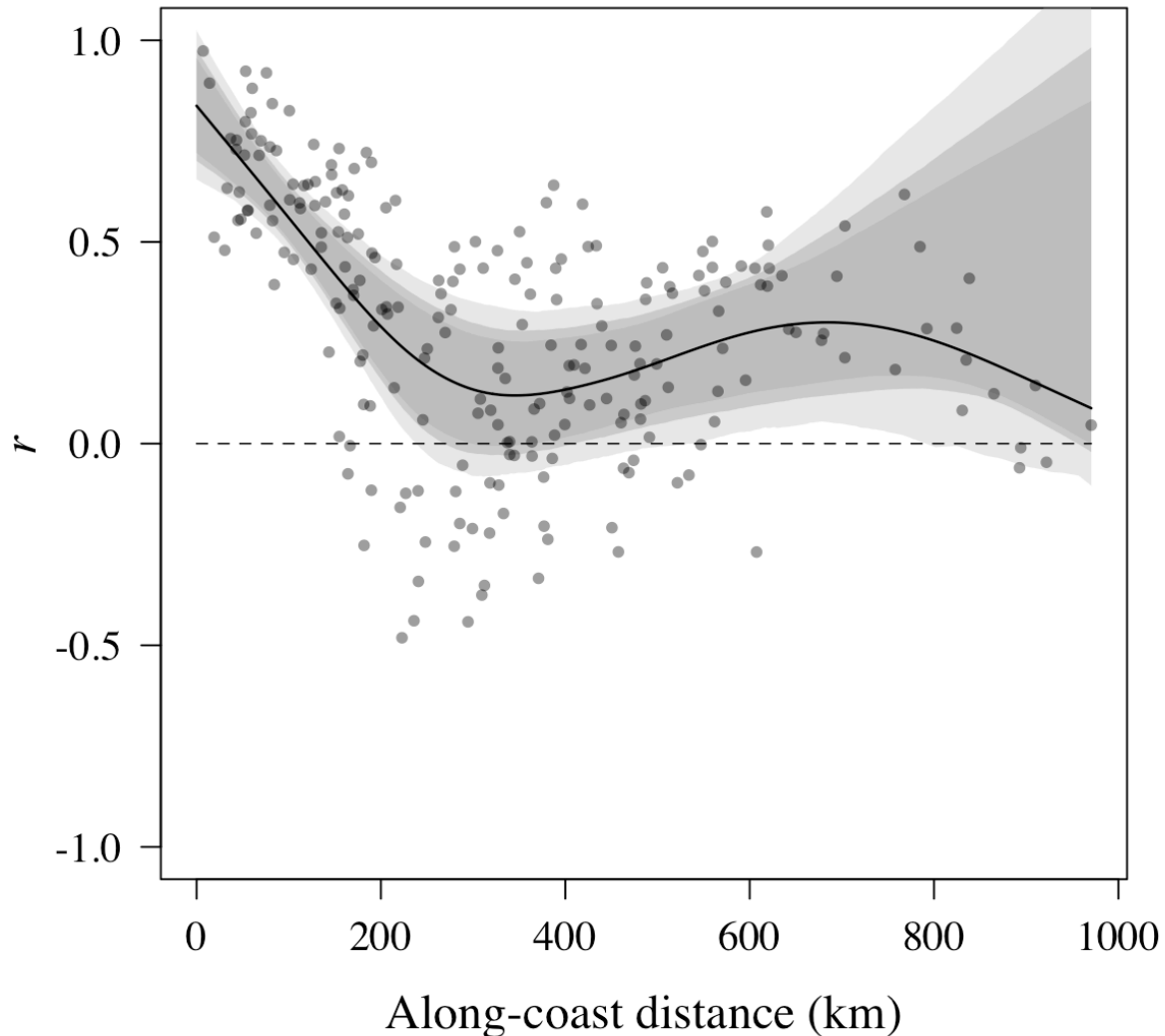


Fig S3. This plot is equivalent to Fig. 2 but measuring synchrony as the strength of the correlation between the untransformed time series of breeding success (r), rather than the correlation of the differences between consecutive years (r_{diff}). The black line shows the cubic smoothing spline fitted between the along-coast distance and synchrony for each colony pair. The dashed line indicates $r = 0$. Grey shading shows 99%, 95% and 90% confidence intervals as estimated from boot-strapping. Each point shows the synchrony of a colony pair against the along-coast distance between the two colonies.

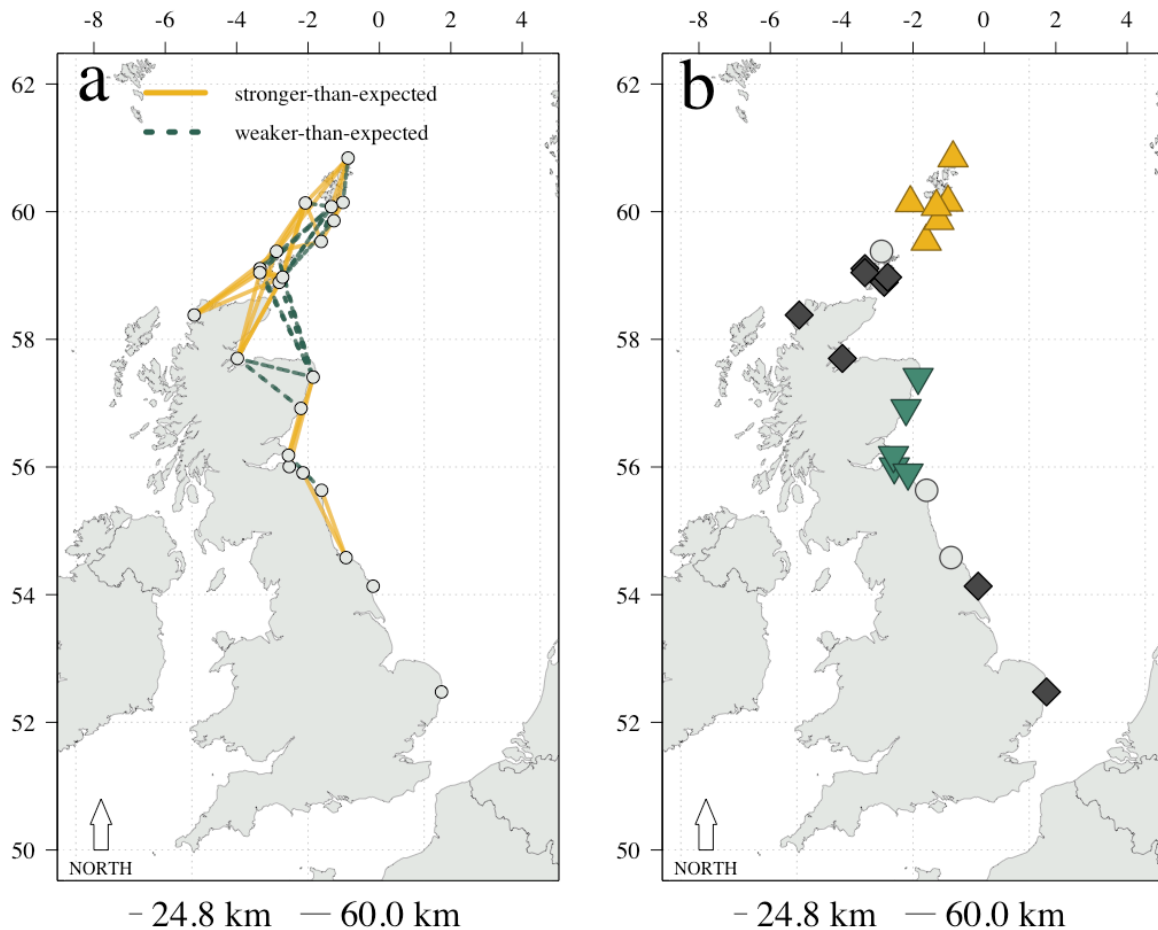


Fig S4. These plots are equivalent to Fig. 1b and Fig. 1c but measuring synchrony as the strength of the correlation between the untransformed time series of breeding success (r), rather than the correlation of the differences between consecutive years (r_{diff}). (a) Green dashed lines indicate weaker synchrony than expected based on distance (negative residuals from the relationship between r and between-colony distance, Fig. S3) whereas yellow lines indicate stronger correlations than expected based on distance (positive residuals from the relationship between r and between-colony distance, Fig. S3). (b) Combined results from the cluster analysis. Round grey symbols indicate colonies that were assigned to different clusters by the different algorithms, whereas the other coloured symbols indicate clusters that were consistently identified by all algorithms. The average silhouette widths for all algorithms were around 0.36–0.37, indicating a slightly weaker cluster structure compared to when clustering was based on synchrony measured as r_{diff} . The average proportion of non-overlap ranged from 2 to 18%, indicating a more stable cluster structure compared to when clustering was based on synchrony measured as r_{diff} .