# Supplementary Material 2.

**Bayesian occupancy modelling of benthic Crustacea and the recovery of the European spiny lobster, *Palinurus elephas*.**

Jackson, A.C.

## Graphical assessment of potential sources of bias

The occupancy model used reduces bias from temporal variation in detection probabilities. It does not correct for other potential confounding patterns in data collection such as spatial bias, temporal variation in spatial coverage, increasing sampling intensity over time, etc. The R package occAssess (Boyd *et al.*, 2021) was used to visualise patterns in these aspects of the data. Whilst it may not yet be possible to correct the model for such biases, they can at least be examined visually to see whether observed patterns in occupancy could be explained by patterns in bias rather than just by any actual change in occupancy. This was not done for all areas modelled, but only for a subset of areas for which there were notable or interesting trends in occupancy, specifically England, Scotland, Wales, Cornwall, South Devon and Dorset. To reduce the number of plots produced by some of tools in occAssess, data were aggregated into time periods (3 years) coarser than the unit of analysis (1 year). It is possible that this smoothed over some biases.

### Sampling intensity through time

The number of taxon records per year collected by Seasearch increased steadily between 2000 and 2010 (Fig S2.1). Since then it has remained fairly steady around 36,000 records per year (after exclusion of records not suitable for the occupancy model). The one exception was 2020 where surveys were greatly reduced due to the COVID-19 pandemic. When broken down into the three countries of Great Britain and considered in blocks of 3 years, a similar pattern was apparent for England (Fig S2.2). Records for Scotland increased in all periods whilst those from Wales dipped again in the two most recent time-periods. Of areas with interesting trends in occupancy (Cornwall, South Devon, Dorset), patterns in Cornwall and Dorset were qualitatively similar to that for England, but in South Devon, the number of records has been declining over 15 years (Fig.S2.3).

Most importantly for this study, it is difficult to explain the observed rapid upturn in occupancy by *Palinurus elephas* in England, Cornwall or South Devon over the last 5 years (Figs. 3 & 4) as being caused by the actual sampling intensity. In Dorset, declines in occupancy for *Cancer pagurus* and *Hommarus gammarus* over the last 10 years (Fig. 4) do not match patterns of sampling intensity over the same time period.

C:\MCS\OneDrive - Marine Conservation Society\R\Data\Sparta\figures\Records_per_year.tif

Fig S2.1 Total number of taxon records collected by Seasearch per year from 2000 – 2020.

C:\MCS\OneDrive - Marine Conservation Society\R\Data\occAssess\figures\GBnRec.tif

Fig S2.2 Number of taxon records collected by Seasearch per country (England, Scotland or Wales) in seven time-periods of 3 y from 2000 – 2020.

C:\MCS\OneDrive - Marine Conservation Society\R\Data\occAssess\figures\CountynRec.tif

Fig S2.3 Number of taxon records collected by Seasearch per country (England, Scotland or Wales) in seven time-periods of 3 y from 2000 – 2020.

### Taxonomic coverage

The number of different taxa recorded by Seasearch in England, Scotland or Wales has generally increased through time, likely to be driven by better resources for identification and increasingly cheap availability of underwater cameras. The number of taxa recorded in England and Scotland declined slightly in the most recent period (Fig S2.4), although this was likely caused by the greatly diminished sampling effort in 2020 (Fig. S.1). Patterns of change for the number of species recorded varied among Cornwall, South Devon and Dorset (Fig. S2.5). None of these patterns, with associated effects on inferred detectability can explain clearly the upturn in occupancy for *P. elephas* in England, Cornwall or South Devon, nor the declines for *C. pagurus* and *H. gammarus* in Dorset.

C:\MCS\OneDrive - Marine Conservation Society\R\Data\occAssess\figures\GBnSpec.tif

Fig S2.4 Number of different taxa recorded by Seasearch in England, Scotland or Wales in seven time-periods of 3 y from 2000 – 2020.

C:\MCS\OneDrive - Marine Conservation Society\R\Data\occAssess\figures\CountynSpec.tif

Fig S2.5 Number of different taxa recorded by Seasearch in Cornwall, South Devon or Dorset in seven time-periods of 3 y from 2000 – 2020.Taxonomic uncertainty

Changes in taxonomic uncertainty could reflect a change in taxonomic expertise over time. For example, if taxonomic skill of recorders declines, then some species may occur earlier but not later in the dataset. The proportion of taxa identified to species level is used as a proxy for taxonomic expertise. In each country, the proportion of taxa identified to species is large, ca. 90% (Fig. S2.6, S2.7), but has declined marginally over 20 years. I suspect that this decline is less to do with decreases in taxonomic skill and more to do with increasing recognition that some taxa can not be identified to species in the field (or without microscopic or genetic examination, e.g. species of maerl, *Ulva,* *Codium*, *Spirobranchus,* etc.). Whilst they may have been identified over-optimistically to species in the past, Seasearch guidance now recommends that they are identified only as far as Genus (or sometimes even higher). Patterns here do not appear to provide any alternative explanation for the observed patterns in occupancy (Fig 3-4).

C:\MCS\OneDrive - Marine Conservation Society\R\Data\occAssess\figures\GBpropID.tif

Fig S2.6 The proportion of taxa identified to species by Seasearch recorders in England, Scotland or Wales during 3 y periods of time from 2000-2020.

C:\MCS\OneDrive - Marine Conservation Society\R\Data\occAssess\figures\CountypropID.tif

Fig S2.7 The proportion of taxa identified to species by Seasearch recorders in Cornwall, South Devon or Dorset during 3 y periods of time from 2000-2020.

### Taxonomic bias

Taxonomic bias in a dataset can be assessed as the degree of proportionality between species' range sizes (by proxy, the number of grid cells in which it was recorded) and the total number of records. Residuals from a linear regression of the number of records against range size give a measure of how over-or under-sampled a species is given its prevalence. Linear regressions are done for each time period and the *r*2 is used as an index of proportionality between range sizes and number of records. Values closer to one indicate that species are sampled in closer proportion to their range sizes, smaller *r*2 indicate that some species are over- or under-sampled.

Values for the taxonomic bias index are between 0.2 and 0.6, i.e. a moderate level of bias where some taxa are over- or under-sampled. The rise through time indicates that species are increasingly being recorded in proportion to their commonness, which may reflect the increase in number of records (Figs. S2.2 & 2.3) and the taxonomic coverage (Figs. S2.4 & S2.5).

C:\MCS\OneDrive - Marine Conservation Society\R\Data\occAssess\figures\GBtaxBias.tif

Fig S2.8 An index of taxonomic bias for Seasearch records in England, Scotland or Wales in 3 y periods of time from 2000-2020. The index is the residual from linear regressions of number of records on range size (number of grid cells where present) for each country and time-period.

C:\MCS\OneDrive - Marine Conservation Society\R\Data\occAssess\figures\CountytaxBias.tif

Fig S2.9 An index of taxonomic bias for Seasearch records in Cornwall, South Devon or Dorset in 3 y periods of time from 2000-2020. The index is the residual from linear regressions of number of records on range size (number of grid cells where present) for each country and time-period.

### Spatial coverage

Records were gridded at a spatial resolution of 0.2 degrees then mapped in geographic space (Figs S2.10--x). In England, the intensity of recording has increased through time, but has done so in a fairly even way around the coastline. Although there are gaps in coverage (e.g. parts of Kent, the Humber), the distribution of records is similar among the time periods and sampling around the coastline has increased relatively evenly (Fig. S2.10). The same patterns are broadly true in Cornwall, South Devon and Dorset (Figs. S2.11-13). There is no obvious pattern in coverage of sampling that might explain increasing occupancy by *P. elephas*, nor the recent declines by *C. pagurus* and *H. gammarus* in Dorset (Fig. 4).

C:\MCS\OneDrive - Marine Conservation Society\R\Data\occAssess\figures\EnglandspatCov.tif

Fig S2.10 Spatial coverage (grid = 0.2 degrees) of records collected by Seasearch in England through seven time periods (p1 – 2000-2; p2 - 2003-5; p3 - 2006-8, p4 – 2009-11; p5 – 2012-14; p6 – 2015-17; p7 – 2018-20).

C:\MCS\OneDrive - Marine Conservation Society\R\Data\occAssess\figures\CornwallspatCov.tif

Fig S2.11 Spatial coverage (grid = 0.2 degrees) of records collected by Seasearch in Cornwall through seven time periods (p1 – 2000-2; p2 - 2003-5; p3 - 2006-8, p4 – 2009-11; p5 – 2012-14; p6 – 2015-17; p7 – 2018-20).

C:\MCS\OneDrive - Marine Conservation Society\R\Data\occAssess\figures\S_DevonspatCov.tif

Fig S2.12 Spatial coverage of records collected by Seasearch in South Devon through seven time periods (p1 – 2000-2; p2 - 2003-5; p3 - 2006-8, p4 – 2009-11; p5 – 2012-14; p6 – 2015-17; p7 – 2018-20).

C:\MCS\OneDrive - Marine Conservation Society\R\Data\occAssess\figures\DorsetspatCov.tif

Fig S2.13 Spatial coverage of records collected by Seasearch in Dorset through seven time periods (p1 – 2000-2; p2 - 2003-5; p3 - 2006-8, p4 – 2009-11; p5 – 2012-14; p6 – 2015-17; p7 – 2018-20).

### Spatial bias

There may be spatial bias in the distribution of sampling if the data depart from a random distribution in the area of interest or if there is a change through time in any tendency towards randomness. To assess spatial bias, the average nearest neighbour distance (Clark & Evans, 1954) of actual records is compared with the average nearest neighbour distances of simulated random distributions at the same density. This requires a mask layer that defines the area in which randomly placed points can be generated. For Seasearch records, this area clearly must be, at least, below mean high water of spring tides (MHWS). Here, bias was assessed only at the scale of all waters around the British Isles. The full extent of the marine environment is, however, not accessible to divers and snorkellers. Thus the mask consisted of polygons bounded by MHWS and a contour representing 40m below chart datum. Random points for simulated nearest neighbour distances were placed within this. A nearest neighbour index (NNI) is provided by the ratio of average observed nearest neighbour distance to the average simulated nearest-neighbour distance. This process was repeated for each of 50 simulations allowing calculation of confidence intervals around a mean for each time period. The ‘degrade’ option was set to TRUE, where duplicated records in the same time period (i.e. those with identical with positions) were disregarded. Multiple Seasearch divers often visit the same place on the same day with identical or near identical positions. Inclusion of all points (degrade = FALSE) greatly inflates the degree of clustering (NNI closer to zero) preventing any assessment of change. The NNI increased through time from 0.12 to nearly 0.3, indicating that records are more clustered than a random distribution, but become less clustered. It is likely that similar levels and trends in clustering also occur at the smaller spatial scales considered in this study. A clustered distribution for locations of records from divers is not in the slightest bit surprising given variation in accessibility, exposure to unfavourable conditions, biodiverse habitats and distribution of features of interest such as pinnacles or wrecks. Thus, for Seasearch records which are always going to be clustered, it is not so much the absolute level of clustering that is important, but whether it changes through time. The decreasing degree of clustering observed is an improbable mechanism to drive the upturn in occupancy by *P. elephas* since 2014. Whilst moving from more-clustered to less-clustered could explain a decrease in occupancy, the temporal patterns of change observed do not correspond well with decreases in *C. pagurus* and *H. gammarus* in Dorset. Future studies may benefit from:

1. using masks at smaller spatial scales, such that NNI can be assessed at those scales;
2. considering clustering at slightly coarser scales than individual Lat-Long positions and/or taking into account the spatial uncertainty of the records.

C:\MCS\OneDrive - Marine Conservation Society\R\Data\occAssess\figures\spatBias.tif

Fig S2.14 A time-series of mean (± 90% CI, n = 50 random simulations) nearest-neighbour index calculated for Seasearch records from all waters of the British Isles in seven 3y periods of time from 2000-2020.

## References

**Boyd, R.J., Powney, G., Carvell, C., and Pescott, O.L.** (2021). occAssess: An R package for assessing potential biases in species occurrence data. *Ecology and Evolution* 11, 16177–16187. DOI: 10.1002/ece3.8299

**Clark, P.J., and Evans, F.C.** (1954). Distance to nearest neighbor as a measure of spatial relationships in populations. *Ecology* 35, 445–53.