

Appendix 1

Assessment methods (technical supporting information)

The area swept by hauls clearly has an influence on the amount of litter collected. We want to be able to control for this variable in our assessments to make sure that any apparent differences in litter amounts are real effects and are not caused by differences in area swept.

We can calculate the area swept for a haul by the wingspan of the haul multiplied by the distance towed. For beam trawls, the wingspan is simply the width of the gear.

For GOV hauls, there are many situations in which wingspan is not recorded. Thus, for GOV hauls we used the formula provided by O'Donoghue and Van Hal (2018) to impute wingspan from doorspan for Dutch hauls:

$$\text{Wingspan} = 5.8728 + 0.1887 * \text{Doorspan} \quad (1)$$

This formula is being used as an approximation for all GOV hauls in this assessment. Future assessments need to consider the ICES Swept Area Calculation Algorithms reports which describe mathematical functions for estimation of the missing values of doorspan, wingspan and distance for each country and for each survey. These mathematical functions are based on observed values over the years of the survey and are provided by the national institutes.

According to the Manual of the IBTS North-Eastern Atlantic Surveys Version 4.0 (ICES, 2017) the mean wingspan for NCT hauls (PT-IBTS) is 15.1. We thus used this figure for our data.

For our original data set, 48% of the 17 100 hauls had missing wingspan values. After the imputation procedures described above, this reduced to 5 856, or 34%, of the hauls. These hauls were not used in the analyses.

In terms of the gear characteristics of the UK NS-IBTS survey, we received the following advice from the operatives:

- The net used is a GOV (Grand Ouverture Verticale) 36/47 Otter Trawl, where the 36/47 is the length of the top rope and the fishing rope, respectively.
- The net is rigged to IBTS (International Bottom Trawl Survey) standard to allow for comparative catches across all IBTS surveys.
- Focusing the dataset on 2015-2020 is advised, as significant changes were made to the gear just prior to this, relating to the net construction, as well as introducing a change to the floatation of the net.

With regard to the sample locations for the UK NS-IBTS survey:

- Surveyors always try to fish the same tow line as the previous year, if it was successful. However, this is not always possible due to issues such as tide. Where this happens, they always try to bisect the previous year's tow line.
- The survey requires only that a tow is in the correct ICES rectangle. Therefore, if the tow needs to be moved (due to previous gear damage, cables laid since the previous year, or an active seismic vessel etc.) it will be. This probably happened on fewer than 5 out of the 77 tows each year.

Statistical Methods

As explained, the challenge with the different gear types is that they all have different abilities to catch litter. Whilst we can control for the area swept of the gear, there will be other characteristics of the gears that will induce different litter catchability rates. The consequence of our data containing different gears is that we do not know if levels of litter amounts are a function of the haul location, or the gear type used. Having said that, we want to use all the available data, so that we can maximise our spatial coverage and intensity of sampling. We return to this issue below.

Another issue is that of unequal sampling effort in space. If one part of a region is sampled more heavily than another part, when we take a simple mean of all the results our mean will be biased towards the area with the greater sampling effort. Thus, for example, if we sample more heavily in an area with low litter levels than in parts of the region with higher litter levels, a simple mean will underestimate the mean litter levels per km² over the whole region.

For both the assessment of the probabilities of hauls containing litter items and the UK NS-IBTS demonstration study of litter counts, we adopted a statistical modelling approach to accommodate the three potential problems (area swept, gear type, and spatial sampling effort) described above.

Probabilities of hauls containing a litter item

The first step was to place a (virtual) square grid over the region. The reason for doing this was so that we could generate a grid of points evenly located over the region. For each year, we then established whether a grid point was less than 20 km from a sampling point. If this was the case, then that grid point was used as a point at which modelled values were estimated.

We fitted a Generalised Additive Model (Wood, 2017) to each year's data using the function *gam* in the *mgcv* R package. This model's link function was of the form:

$$f(p) = \alpha + \beta \text{areaswept} + \text{gear}_j + s(\text{lat}, \text{long}) \quad (2)$$

where p is the probability of one or more litter items being present in a haul, $f(p)$ is the standard logit function for binary data such that $f(p) = \log\left(\frac{p}{1-p}\right)$, α is an intercept term, $\beta \text{areaswept}$ is a linear function of haul area swept, gear_j is an estimate of the j th gear effect, and $s(\text{lat}, \text{long})$ is a smooth two-dimensional function of latitude and longitude that includes both the two main effects and the interaction. For the Bay of Biscay and Iberian Coast assessment, only one gear was used – thus, for that assessment, the gear_j term in (2) was not needed.

Parameters in (2) for each year, 2012-2019, were estimated using the default method of cross-validation (Wood, 2017). Essentially, cross-validation works by leaving out each point in turn and calculating the mean ability of the remaining points to predict the left-out datum. Parameter estimates are chosen that give the best mean prediction. We did not restrict the degrees of freedom of the smooth terms (the degrees of freedom determine how flexible – or wiggly – the smooth terms are).

Once models of the form (2) had been fitted, we used the parameter estimates from the models to predict p , the probability that the grid point would contain a litter item. The latitude and longitude in the model were simply the location of the grid point, the area swept was set to a constant value of 57 000 m² for all grid points and the gear was set to the GOV gear if there was more than one gear in the region. The exception was for the small analysis conducted for the Iberian Peninsula. The NCT gear was used here, with a smaller area swept. Thus, for this analysis, the area swept was standardised to 44 000 m².

There is a caveat in terms of the area swept variable. For some analyses, area swept was not a statistically important variable in explaining litter proportions. For example, sometimes the parameter β was estimated as negative. We are not sure why this is the case; it could be that the area swept for hauls for those analyses were similar and it was difficult for the model to pick up any relationship between increasing area swept and increasing proportions of litter

found. It is important to be wary of using the area swept variable for predictions in these situations. For example, if a negative β is used, this will impact on our predicted values of p . Thus, we adopted the approach below.

Once the model (2) had been fitted, we assessed the parameter β . If β was not statistically significant ($p < 0,05$) or if the estimate of $\beta < 0$ then the model was re-fitted, but without the area swept term. Prediction then used only latitude, longitude and, if relevant, gear type.

A similar issue arose if the spatial element of the model was not statistically significant. In such situations, showing differences in predicted litter levels over space is misleading given that we do not have evidence that such spatial differences exist. In these circumstances, we did not use the spatial term in our predictions. In situations where the area swept effect is not statistically significant and there is only one gear, this results in our predictions being the mean values of the litter (0, 1) variable for the region – i.e., the variable has value 0 if a litter item is not found and value 1 if at least one litter item is found.

For mapping purposes, we simply used the predicted values at the grid points. We also calculated standard errors, that reflect the precision of these predictions, but these have not been used for the current assessments.

To get an estimate of p for a particular region and year, the mean of all predicted points on the grid was used. However, so that fair comparisons could be made between years, only grid points that were within 20 km of a sampled point in *all* eight years of the study were used (whereas, for the maps, all grid points for a particular year that were within 20 km of a sampled point were used). To get a 95 % confidence interval for this estimate we used the procedure described on pages 342-343 of Wood (2017). Essentially, this involves simulating from the posterior distribution of the parameters - assuming that they have a Multivariate Normal distribution, calculating the linear predictor (LP, the right-hand side of equation (2)) for each simulation and then transforming back to the original scale using the standard $p = \frac{\exp(LP)}{1 + \exp(LP)}$ back transformation. For repeat simulations (we used 1 000), the 95% confidence interval is the 2,5th and 97,5th percentiles of these simulated p s. Essentially, we might think of this process as a form of parametric bootstrapping of the parameters.

Litter counts

We have not undertaken a full modelling assessment on counts following guidance from the working group. However, we have done a demonstration study for NS-IBTS surveys conducted in the Greater North Sea by the UK. We have also done some preliminary, exploratory analyses of catchability of different litter types by gears. Clearly, these results should be interpreted with caution for the same reasons that we have not done a full counts assessment. However, we believe that they provide some interesting insights and so we report our initial findings in this assessment.

For the demonstration study of counts in the GNS, we modelled the total litter counts by a GAM model, with link function:

$$\log(E[C]) = \alpha + \beta \text{areaswept} + s(\text{lat}, \text{long}) \quad (3),$$

where $E[C]$ is the mean count. We used the raw data to compare Negative Binomial and Poisson error distributions for the counts. Theoretically, we would expect the Negative Binomial distribution to be appropriate if litter items were clustered on the sea floor; the Poisson distribution would be appropriate if litter items were spatially random.

For the study of catchability of different litter types, we transformed the original counts to count per unit effort by dividing the count by the area swept (in m^2) and then multiplying the answer by 10^6 . We thus have counts per 1 km square.

To calculate the conversion factor between Beam Trawl (BT) and GOV hauls for the GNS, eighty spatial squares covering the GNS Region were chosen. The mean of the ratio (MOR) of BT to GOV counts per unit effort was calculated over these 80 squares for each of the litter types defined (this was done for all data points and was not done separately

for each year). Thus, to standardise counts per unit effort to GOV hauls, all BT counts per unit effort values were divided by the appropriate MOR. For the Celtic Seas, 90 squares were used, and ratios calculated between BT and GOV, and between PORB and GOV. For the Iberian Coast and Bay of Biscay, 80 squares were used, and ratios calculated between BAK and GOV gears.