Advancing marine species density surface modelling – Public Webinar, 1st June 2022.

Q&A written responses

- Can we please see an e.g. of how the data is input in distance for DSM? *Answer: The distance sampling examples website has a number of examples using DSM, each of which has details of the data setup, see https://examples.distancesampling.org/ and specifically https://examples.distancesampling.org/dsm-data-formatting/dsm-data-formatting.html for information on data processing.*

- Does the process of calculating these uncertainties (with multiple values for each covariate for each day, for example) require substantially higher computational power? *Answer: It doesn't require more processing power, as we make a single prediction for each posterior sample and time slice, but we need to store all of those predictions, so that can take up a lot of disk space, so it's worth making sure you have space before you start. Because we make lots of predictions and those calculations are independent, we can use multiple processors on one computer and run this in parallel.*

- If I have seasonal line transect data but a pooled detection function (because the sample size per season is not sufficient), is it still possible to fit a dsm model? Can I get a season specific abundance estimate with pooled detection function? *Answer: While there is no requirement that the detection function has seasonally/temporally-varying covariates to fit a DSM, if the sample size is not large enough to allow for covariates then we might worry about whether we have enough data to fit the DSM. Also, if detectability does vary over time and this is not accounted for in the detection function model, then this will bias the season-specific estimates; similarly, if detectability varies over space and this is not allowed for in the detection function model (by including appropriate spatially-referenced covariates) then this will bias the density surface.*

- It is great seeing the uncertainty along with the density surfaces. However, will folks like regulators understand them or still focus only on the densities? *Answer: Firstly, we can't speak for the regulator. How the regulator might use uncertainty depends on the question being addressed, e.g., for environmental impact they could use an upper quantile, or for directing future monitoring effort an uncertainty quantification could be used to direct where future effort is best spent. The question being addressed also determines the best way to display/summarize the uncertainty: CV, SE, CIs, movies showing realizations from the uncertainty surface. It also determines the components of uncertainty to include: e.g., do we want average predictions averaging over future years (so environmental uncertainty averaged out) or for some particular year (in which case we want to include uncertainty in what the conditions will be that year).*

- You talk about data coming from line transects, distance sampling, etc. What about data coming from ad-libitum survey (e.g., no transects, no fixed routes)? Are detection function models still valid and calculable? Related question: I was wondering how sensitive DSMs are to non-systematic data collection provided that a comprehensive range of environmental

conditions are surveyed? *Answer: The first question appears to concern* <u>*detection functions*</u> *specifically. Here we need to go back to the basic assumptions of distance sampling. In order for data to be usable under this framework, perpendicular distances from the transect line to sightings need to be collected in order to fit a detection function. If this is not possible, some suitable proxy from another study could be applied – but only if observers search in the same way and all other factors affecting detectability is similar (which is often far from true). Observer effort should be tracked as the survey proceeds, so it is known when observers were actually watching for animals. Ideally, observers should not be advised of the presence or absence of animals ahead of time, so that there is less risk that their attention level will be biased based on this information. In addition to this, a basic assumption of distance sampling is that the distribution of animals (sighted and unsighted) is uniform with respect to distance from the transect line – that way we can attribute the drop-off in number of sightings with increasing perpendicular distance as being a detectability effect and not an animal density effect. We typically ensure this assumption is met via random line placement. In most opportunistic surveys it is also met, but it would not be in, for example, surveys from a boat going along the shoreline with one-sided effort looking out to sea where animal density could easily increase (or decrease) with increasing distance from shore.*
The second question concerns the <u>*spatial models*</u>*. Non-systematic designs will yield heterogeneous sampling in space and time. The purpose of the model is, in some sense, to handle this by fitting smoothed relationships to covariates. To some degree it can handle "oversampling", in which regions of covariate space that are sampled heavily will not necessarily be overweighted in influencing the model. But if ranges of covariates are poorly sampled, DSM methodology will not allow you to evade the problem, and model relationships may be highly uncertain or incorrect. Ideally, surveys should be conducted and transects laid out independent of known presence or absence of animals. If, for example, surveys are directed only towards known aggregations of animals (for example where survey boats are directed towards known aggregations of animals reported by members of the public), the results will likely be biased. In addition, it would be particularly important to include an analysis of potential extrapolation when using highly non-systematic data.*

- Pertaining to data integration, is there any work on incorporating acoustic information from gliders where you only have an idea of the detection range of the instrument, not the actual location of a calling critter, into SDMs? *Answer: Potentially yes, but there a number of issues that one would need to consider when trying to combine acoustic data that doesn't have distance information into any model (spatial or otherwise). If one assumes that detection is certain out to some detection distance, one could treat the glider data as a strip transect and incorporate as one would any data source of that type – but one would have to be sure that this assumption was reasonable. A close alternative would be if one had an estimate (from another study) of the effective strip width of the glider (see distance sampling literature for the definition of effective strip width). (In other work, members of CREEM have done trials using a glider in combination with buoyed sensors to determine the glider effective strip width using a spatial capture recapture framework – however this work is preliminary and not yet published.) Secondly, the issue of call rates would need to be tackled. Marques et al (2013; https://onlinelibrary.wiley.com/doi/10.1111/brv.12001) highlights these and more issues when using passive acoustic data.*

- With reference to the talk that discussed sperm whales and availability bias correction for when whales are diving. Does it matter if the species dive coordinated (e.g. Md) vice as singular animals (e.g. solitary large male Pm)? *Answer: In essence, single animals' vs groups of animals will not matter, since we are detecting and localizing single animals, not groups from acoustics. We are applying the availability correction on a per animal bias. What will matter, however, is the vocal type/social behaviour. For example, slow clicks vs. regular clicks or codas. Individuals or animals producing slow clicks or codas would have different detectability, availability, and perhaps different relationships to covariates given they are using their environment in different ways based on what they are doing, e.g., foraging vs. socializing. Ideally, data from tagged animals producing the vocalization of interest would be used to obtain availability bias corrections.*

- In general when we do spatial models outside of distance sampling a key issue that folks worry about is spatial autocorrelation. Here the models seem to assume that (corrected) counts in segments are independent, but these segments are actually usually clustered, say, into transects. Have you looked into that and can you comment about what if any are the downsides of assuming independence in the observations if they are not independent, say? *Answer: Autocorrelation is a complicated issue. We assume that the observations are correlated in a DSM (which is why we fit spatial models), and then we test for residual autocorrelation after fitting the model to check to see if the model did a good job of handling the correlation that was in the data. If it doesn't then we can think about including more complex correlative structures in the data.*

- Are you collaborating with the Navy folks who work on NAEMO to explore ways the modeled uncertainty could be incorporated into the NAEMO model?
  Answer: Yes – we had a representative from the NAEMO team at our second working group meeting and discussed strategies then; we have had some continuing contact since then. One potential way to include uncertainty from the density surface model is to use the model to generate multiple realizations of the density surface (as mentioned in Dave's talk on uncertainty); in NAEMO multiple runs are made of the same scenario so for each run a different density surface realization could be picked at random. However, there are many other factors the NAEMO team need to consider when deciding what extensions to add and how to do this, and of course such decisions are up to the Navy.