Designing a survey around fishery-dependent data

FISH 507B project proposal

John Best

Introduction

Model-based indices of abundance are increasing in popularity, particularly those that incorporate spatiotemporal dynamics (Thorson, Maunder, and Punt 2020). When available, model-based indices are usually based on fishery-independent observations. However, one advantage of model-based indices is the ability to incorporate multiple sources of data. Many fisheries have extensive observer programs that record detailed fishery-dependent catch information. Naturally, there are many more fishery-dependent observations than could feasibly be collected during a survey. Incorporating these fishery-dependent observations into indices of abundance has the potential to increase the precision of these indices.

Moving from a fishery-independent index of abundance to one that also includes fishery-dependent data is not without pitfalls. Fishery-dependent data are not collected using a defined sampling procedure. This may result in higher variance of catches (e.g. among vessels). Fishing locations are also chosen for their potential profit rather according to a specified design. This targeting behavior may result in a preferential sample, where only areas with higher biomass of the targeted species are sampled by the fishery. This may bias resulting estimates of abundance (Gelfand, Sahu, and Holland 2012; Conn, Thorson, and Johnson 2017). It is possible to account for the effects of preferential sampling post hoc, but this comes at the cost of considerable model complexity (Diggle, Menezes, and Su 2010; Dinsdale and Salibian-Barrera 2019). A simpler option (from a modeling perspective though admittedly not a logistical perspective) would be to design a survey that accounts for the weaknesses of the fishery-dependent observations.

At one extreme, survey locations could be chosen so that overall sampling effort is roughly uniform throughout the domain, eliminating the observation location preference. This would generally require an excessive amount of survey effort. A second option would be to overlay a standard survey grid over the domain of interest, as is currently done. This would ensure samples outside the area where fishery-dependent data is abundant, but apportions survey effort in a way that does not account for the fishery-dependent samples. In between these options, a survey design with greater sampling intensity in under-fished areas can fill in information where it is most needed. Any successful survey that intends to integrate fishery-dependent and -independent observations will need to allow enough spatial overlap between the two so that differences in catchability may be estimated. Sampling away from a species' area of highest abundance can also aid in establishing the limits of a species' range, and provide additional information about species that are not the target of commercial fisheries. Survey design under preferential sampling has been studied in the case where the preference is meant to be preserved (da Silva Ferreira and Gamerman 2015), designing a survey specifically to *counteract* the effects of preferential sampling has not been addressed.

Methods

A simulation study will be undertaken to compare indices of abundance that include preferentially sampled fishery dependent data as well as fishery-independent data sampled according to two different design strategies. The first, reference strategy, overlays a uniform survey grid over the domain, similar to current practice. The second chooses survey locations at random, where locations with higher fishery-dependent fishing effort are less likely to be chosen for the survey. These will be compared over multiple levels of total survey effort (e.g. 50, 100, 150, and 200 total trawls).

Catches will be simulated using the `FisherySim.jl` software package. This package generates fishery observations on a grid. A spatially correlated habitat covariate determines fish movement. Overall population dynamics follow a region-wide biomass dynamics model, in this case a Beverton-Holt. Multiple fleets remove biomass from each cell. Each fleet may have different catchability and/or targeting behavior. This allows the different survey strategies to be implemented. To simplify the simulations all of the fishery-dependent fishing locations will be chosen prior to fishing, rather than applying a dynamic preference that accounts for within-cell depletion. On a 100×100 grid this should have minimal impact on the resulting catches. Pre-specifying the fishery-dependent fishing locations allows the survey locations to be pre-specified as well.

Indices of abundance will be estimated using the `spatq` R package, which includes a spatiotemporal index standardization model written in Template Model Builder (Kristensen et al. 2016). Indices will be evaluated for bias and total error, particularly relative to total survey effort.

Extensions

To ensure that this project can feasibly be completed in the required timeframe, a number of interesting extensions will not be addressed. The first issue is that the focus here is on estimating an index of abundance; the sampling strategies and evaluation criteria do not account for other data types that may not be available from fishery-dependent data. Second, practical implementation of these survey designs would require strategies for specifying survey locations *before* fishery-dependent effort distribution is known. This may be trivial in fisheries that don't change much year-to-year, but it may be necessary to include uncertainty in fishery-dependent location preference in order to make the strategy robust to spatial shifts in the fishery. Third, this simulation study focuses on a single species, which exclusively drives the fishery-dependent fishing location preference. Survey designs in this case may or may not be optimal when the location preference is driven by other factors such as a different species. It would also be interesting to see how these strategies fare when the goal is to produce indices of abundance for multiple species.

References

Conn, Paul B., James T. Thorson, and Devin S. Johnson. 2017. "Confronting Preferential Sampling When Analysing Population Distributions: Diagnosis and Model-Based Triage." *Methods Ecol. Evol.* https://doi.org/10.1111/2041-210X.12803.

da Silva Ferreira, Gustavo, and Dani Gamerman. 2015. "Optimal Design in Geostatistics Under Preferential Sampling." *Bayesian Anal.* 10 (3): 711–35. https://doi.org/10.1214/15-BA944.

Diggle, Peter J., Raquel Menezes, and Ting-li Su. 2010. "Geostatistical Inference Under Preferential Sampling." *J. Royal Stat. Soc. B* 59 (2): 191–232. https://doi.org/10.1111/j.1467-9876.2009.00701.x.

Dinsdale, Daniel, and Matias Salibian-Barrera. 2019. "Methods for Preferential Sampling in Geostatistics." *J. Royal Stat. Soc. B* 68 (1): 181–98. https://doi.org/10.1111/rssc.12286.

Gelfand, Alan E., Sujit K. Sahu, and David M. Holland. 2012. "On the Effect of Preferential Sampling in Spatial Prediction." *Environmetrics* 23 (7): 565–78. https://doi.org/10.1002/env.2169.

Kristensen, Kasper, Anders Nielsen, Casper W. Berg, Hans Skaug, and Brad Bell. 2016. "TMB: Automatic Differentiation and Laplace Approximation." *J. Stat. Softw.* 70 (5). https://doi.org/gf4mhz.

Thorson, James T., Mark N. Maunder, and André E Punt. 2020. "The Development of Spatio-Temporal Models of Fishery Catch-Per-Unit-Effort Data to Derive Indices of Relative Abundance." *Fish. Res.* 230 (October): 105611. https://doi.org/gg2bx4.