Government of Canada Gouvernement du Canada NOAA FISHERIES NATIONAL MARINE FISHERIES SERVICE

CERT
Comité d'évaluation des ressources transfrontaliéres

TRAC
Transboundary Resources Assessment Committee

Document de travail 2014/37

Working Paper 2014/37

Ne pas citer sans autorisation des auteurs

Not to be cited without permission of the authors

Catchability estimates using Habcam images as a measure of absolute abundance

Daniel R. Hennen

Northeast Fisheries Science Center, NMFS, NOAA, Woods Hole, MA

Ce document est disponible sur l'Internet á:

This document is available on the Internet at:

http://www.mar.dfo-mpo.gc.ca/science/TRAC/trac.html

Canada

Introduction

Both the scallop dredge survey and Habcam "capture" yellow tail flounder (YTF). They have been experimentally paired in time and space 126 times since 2007 (Figure 1). A comparison of the "catch" of YTF made by Habcam and the number of YTF captured by the dredge may provide an experimental estimate of the catchability of the dredge. The Habcam theoretically captures all YTF in the field of view of the camera. Therefore, Habcam "catch" could be considered an absolute measure of density when divided by the area of the field of view. Therefore, the density of YTF in the scallop dredge divided by the density of YTF measured using Habcam may provide an estimate of the efficiency or "catchability" of the scallop dredge.

In order for the catchability estimate $\hat{c}$ to be an unbiased statistic of the true catchability c, the following assumptions must be met: 1) All YTF in field of view of photo are observed by Habcam, 2) YTF are randomly distributed spatially 3) YTF are randomly distributed temporally. A violation of 1) would result in bias in the estimate of absolute density ($\hat{d} < d_{abs}$). A violation of 2) or 3) would result in bias as well, though with unknown direction.

Each of these assumptions was violated to some degree. The path of the Habcam did not entirely overlap the path of the dredge and they did not occur simultaneously. Furthermore, it is clear from close inspection of the Habcam images that at least a few YTF caught on film were in the process of fleeing the Habcam. Therefore we must assume that some fraction of the available fish managed to move away before they could be caught. Ideally, the distribution of YTF in time and space would be locally uniform, so that a Habcam tow that approximately overlapped a dredge tow would be sampling from an identical population. Unfortunately, we know that YTF are patchy in spatial and temporal distribution. There may be however, a distance over which they appear to be minimally patchy. That is, a sample that is broad enough in time and space will reduce the variation in density that is due to spatial and temporal patchiness.

Differences in solar zenith angle appear to be important. YTF seem more likely to swim to avoid the Habcam array during the day time. This potentially further complicates the comparison of some paired tows if they were mismatched in terms of solar zenith angle.

Methods

Data

Initial examination of (Figures 2 and 3) indicated that images captured from within 2 nm from a corresponding paired dredge tow were minimally variable in terms of density. There was no clear pattern in the variability of Habcam density estimates relative to the difference in absolute hours between dredge and Habcam tows. Therefore the definition of paired tows were bounded by 2 nm and unbounded in absolute hours difference.

Analysis

Density in the dredge ($d_{D,i}$) was

$$d_{D,i} = \frac{n_{D,i}}{a^2_{D,i}} \tag{1}$$

where $n_{D,i}$ was the number of YTF caught in the dredge in tow i and $a^2_{D,i}$ was the area swept by the dredge in tow i.

Density measured by Habcam ($d_{H,i}$) was

$$d_{H,i} = \frac{n_{H,i}}{a^2_{H,i}} \tag{2}$$

where $n_{H,i}$ was the number of YTF observed in Habcam images associated with station i and $a^2_{H,i}$ was the total field of view of all of the annotated images for station i.

Therefore catchability ($\hat{c}_i$) was estimated from the set of $\frac{d_{D,i}}{d_{H,i}}$ observations of catchability at each site.

These values were bootstrapped 100000 times using a weighted bootstrap procedure in which the weights were proportional to $a^2_{x,i}$ associated with each estimate (Figure 4). A bounded (0,1) log normal distribution was fit to the bootstrapped data set (Figure 5). The catchability estimate $\hat{c}$ was the lognormal mean of this distribution.

The density estimates from each gear did not appear to be correlated by site (Figure 6). Therefore, the probability of obtaining useful estimates of catchability through the methods already described was low. Additional methods for providing inference on catchability were considered.

The density estimates from each gear were uncorrelated by site. They might therefore be considered independent estimates of a broader scale density. Given that $d_{H,i}$ and $d_{D,i}$ are independent, an alternative estimate of catchability for the region sampled is

$$\hat{c} = \frac{E[d_D]}{E[d_H]} \tag{3}$$

the variance of $\hat{c}$ is

$$\sigma^2_{\hat{c}} = \frac{E^2[d_D]}{E^2[d_H]}(\frac{\sigma^2_{d_D}}{E^2[d_D]} - 2\frac{cov(d_D, d_H)}{E[d_D]E[d_H]} + \frac{\sigma^2_{d_H}}{E^2[d_H]}) \tag{4}$$

A weighted mean, weighted median and weighted variance where

$$\sigma^2_{\hat{c}^w} = \frac{E^2[d^w_D]}{E^2[d^w_H]}(\frac{\sigma^2_{d^w_D}}{E^2[d^w_D]} - 2\frac{cov(d^w_D, d^w_H)}{E[d^w_D]E[d^w_H]} + \frac{\sigma^2_{d^w_H}}{E^2[d^w_H]}) \tag{5}$$

were calculated as well. Because the density estimated by Habcam was usually less than the density estimated by the dredge, $\frac{d_{D,i}}{d_{H,i}} > 1$ for most sites. Two additional sets of $\hat{c}$ statistics were produced using data truncated at $\frac{d_{D,i}}{d_{H,i}} = 1$.

Diel effects

Examination of plots illustrating diel effects (Figure 7) showed that correlation between the density measured by Habcam and the density measured in the dredge were not strongly affected by diel effects. That is, none of the plots in Figure (7) show much indication of a correlation. Therefore all the data was included in analysis.

Diagnostics

The lack of correlation between densities estimated using Habcam and those estimated using the scallop dredge was investigated with a diagnostic exercise. Residuals composed of the minimum distance between each $(d_{H,i}, d_{D,i})$ and the $d_H = d_D$ line were calculated. These residuals were standardized and then plotted against the mean distance and time separating the dredge and Habcam tows, against $a^2_{H,i}$, and finally against the mean difference in zenith angle during the two tows at each site.

Results

It was not possible to estimate catchability using the densities measured by Habcam and the scallop survey dredge with the initial approach described here. The density estimated by Habcam was usually less than the density estimated by the dredge and $\frac{d_{D,i}}{d_{H,i}} > 1$ for 70% of sites. The densities estimated by each gear were not correlated by site (Figure 6). The lognormal fit to the bootstrapped set of catchability estimates did not converge (Figure 5).

The additional statistical estimates of catchability ranged from 0.46 to 0.83 though the variance was high (Table 1).

Diagnostic plots revealed no obvious cause for the lack of correlation between (d_H and d_D) (Figure 8).

Discussion

The fact that $\frac{d_{D,i}}{d_{H,i}} > 1$ for 70% of sites probably indicates that current $d_{H,i}$ are not sufficient estimates of absolute abundance at each site. This could result from several potential causes, including: spatial or temporal patchiness in YTF abundance, detection problems in Habcam images, fish fleeing Habcam at a higher than expected rate, a positive bias in $d_{D,i}$, or other factors. Furthermore, the densities estimated by Habcam and the scallop survey dredge were not correlated, despite being taken from roughly the same area at approximately the same time. The lack of correlation was not explained by distance or time between the tows made by the two types of gear, the area "swept" by the Habcam or mean difference in zenith angle when the tows using the two gears were made (Figure 8). Thus there was not clear path to removing covariates that might be disrupting the expected signal (a positive correlation) in the data. This indicates that treating the ratio of densities at each site as an individual estimate of catchability is not likely to produce a useful statistical distribution for catchability of YTF in the scallop survey dredge.

Some additional methods were considered, including: shifting the density metric to a proportion (*i.e.* $\hat{c_i} = \frac{d_{H,i}}{(d_{H,i}+d_{D,i})}$, and using length based methods to estimate $\hat{c_{i,L}}$ where L is length. These methods were ultimately not pursued because they appeared unlikely to produce materially different results.

In the absence alternatives, treating the set of all paired tows as observations from a single broader distribution of catachability over all of Georges Bank might be the preferred approach. The statistics in Table (1) provide some limited information on the efficiency of the scallop survey dredge for YTF. It is however, important to incorporate the cv's of these estimates as they are highly imprecise.

Incorporating the Habcam data in its current extent probably does not materially improve estimates of YTF catchability in the scallop survey.

Tables

	Mean	CV	Median
Standard	0.83	2.06	0.50
Weighted	0.77	1.55	0.41
Truncated	0.49	3.46	0.50
Wt & Trunc.	0.46	2.60	0.41

Table 1: Statistics for catchability. Standard mean, cv and median use all the data with equal weight. Weighted mean, cv and median use weights that are proportional to the area swept by each gear. Truncated mean, cv and median use unweighted data that is truncated so that all values are between 0 and 1. Weighted and truncated use both of the latter two options.

Figures

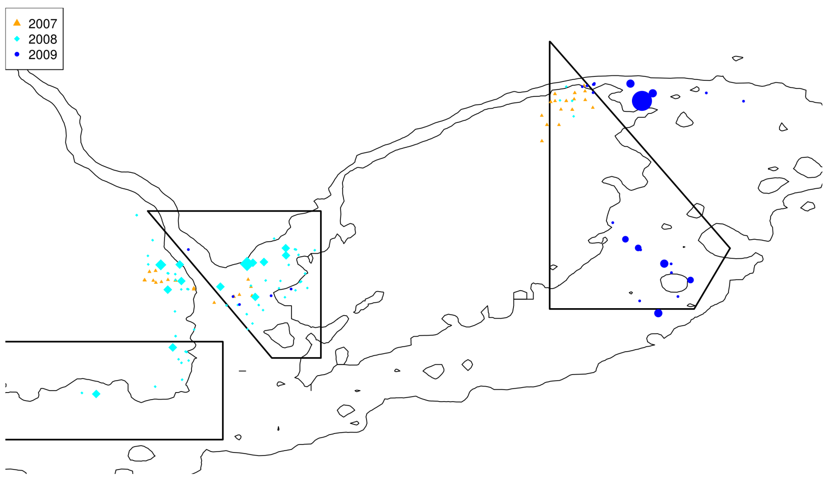

Figure 1: Locations of Habcam/dredge paired tow experiments.

Image credit: Burton Shank.

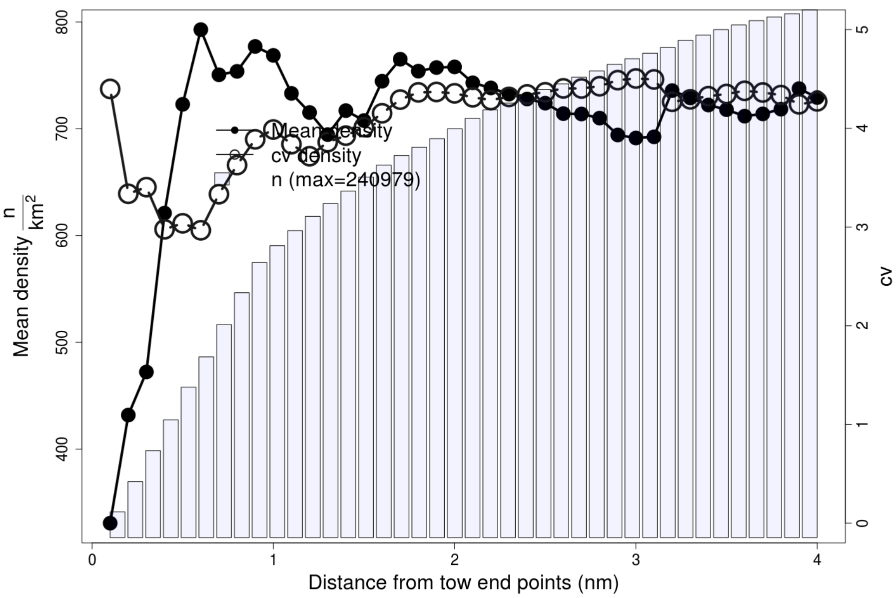

Figure 2: Density estimated by Habcam by distance between paired dredge tows and Habcam tows. The solid circles are the density estimates, the open circles are the cv of density and the solid bars are the total number of annotated images that make up the sample.

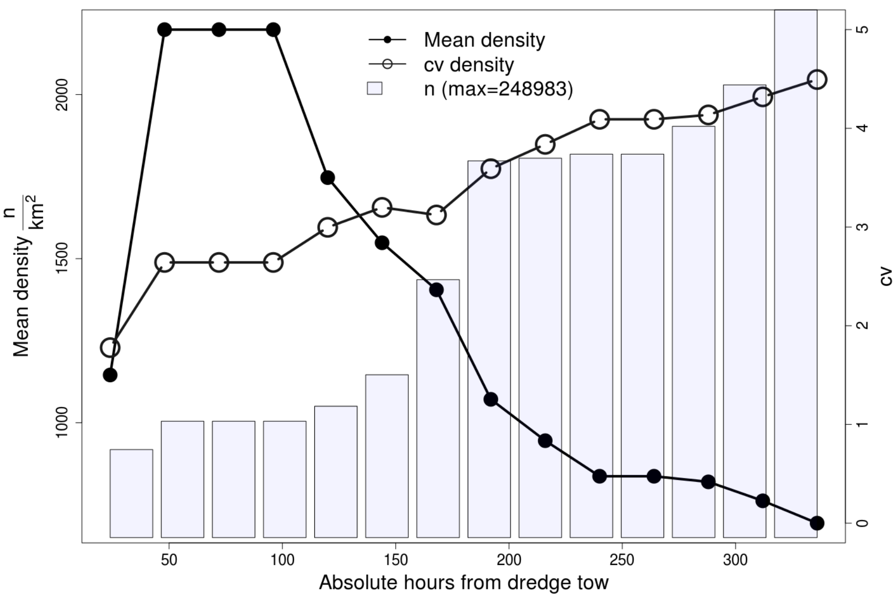

Figure 3: Density estimated by Habcam by absolute hours between paired dredge tows and Habcam tows. The solid circles are the density estimates, the open circles are the cv of density and the solid bars are the total number of annotated images that make up the sample.

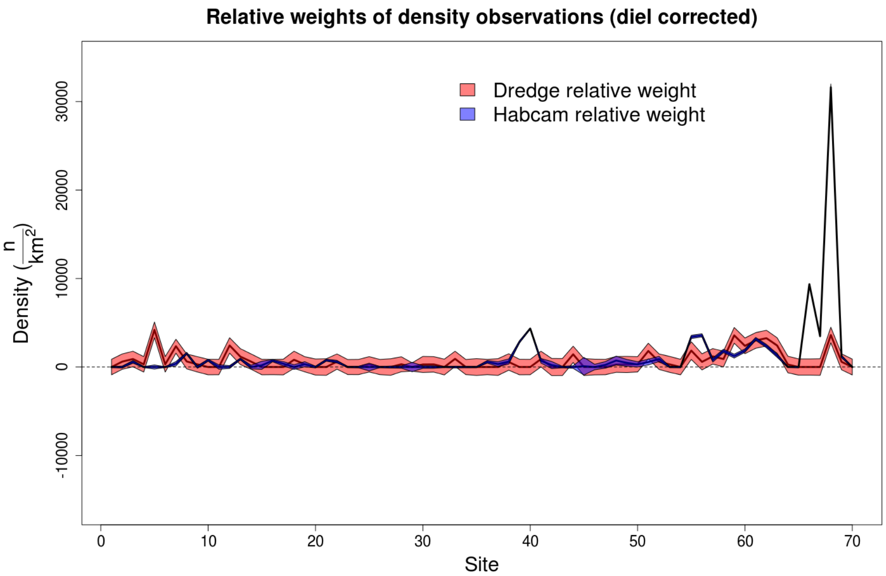

Figure 4: Comparison of density estimates from dredge (pink) and Habcam (purple). The polygons are weights based on area swept, not confidence intervals. That is, the wider the polygon around a each point, the more confidence we have in that estimate of density.

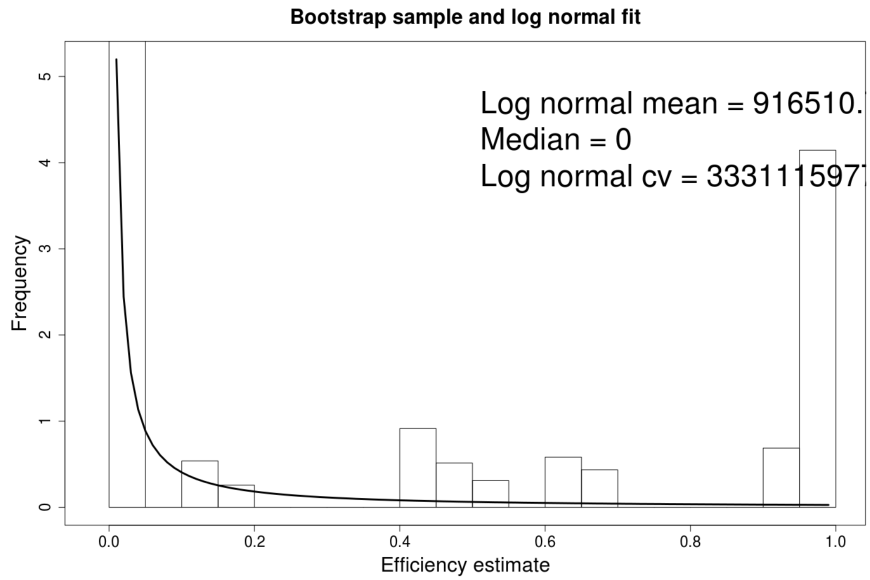

Figure 5: Lognormal fit to the sample of density ratios. The fit did not converge.

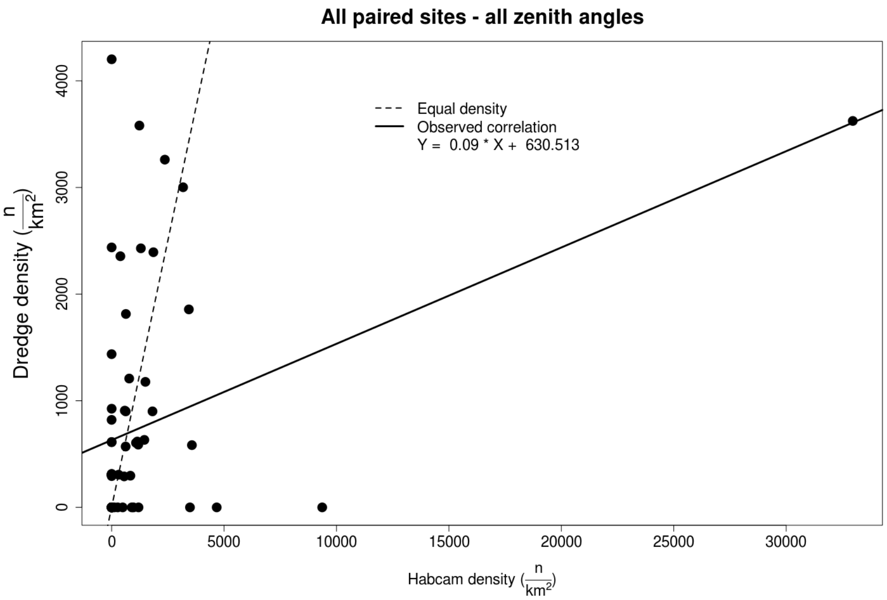

Figure 6: Correlation between density estimates at each site by gear (all data where $a^2_{H,i} > 400m^2$). The dashed line is $d_D = d_H$ and the solid line is a simple linear regression fit to the data. The data are jittered to improve visibility of individual points.

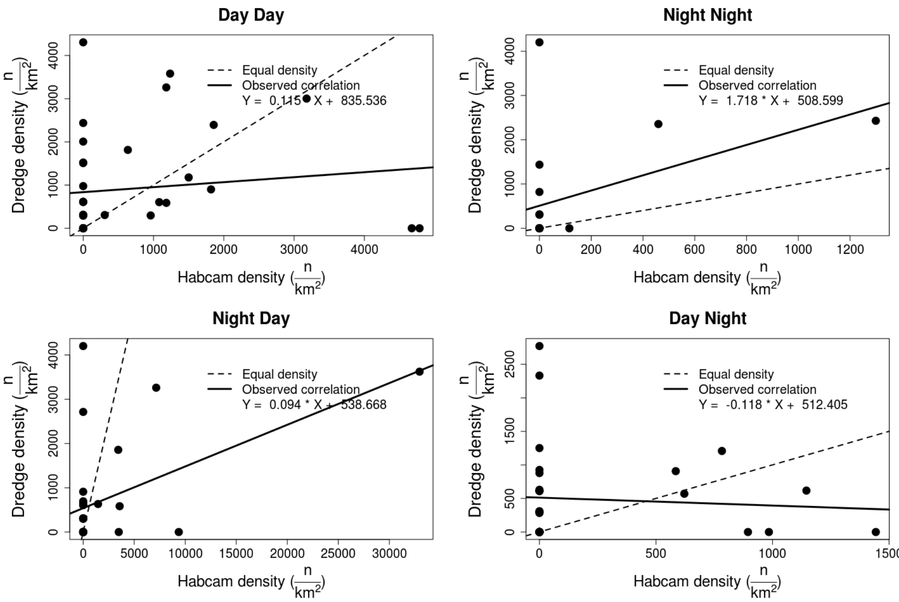

Figure 7: Correlations between the density estimated using Habcam and the density estimated using the dredge catch. The plots indicate the time of day each observation/catch was made, for example in the lower left plot the Habcam observations occurred at night, while the dredge was towed during the day. The dotted line represents 1 to 1 correlation and the solid line is a simple linear regression fit to the data.

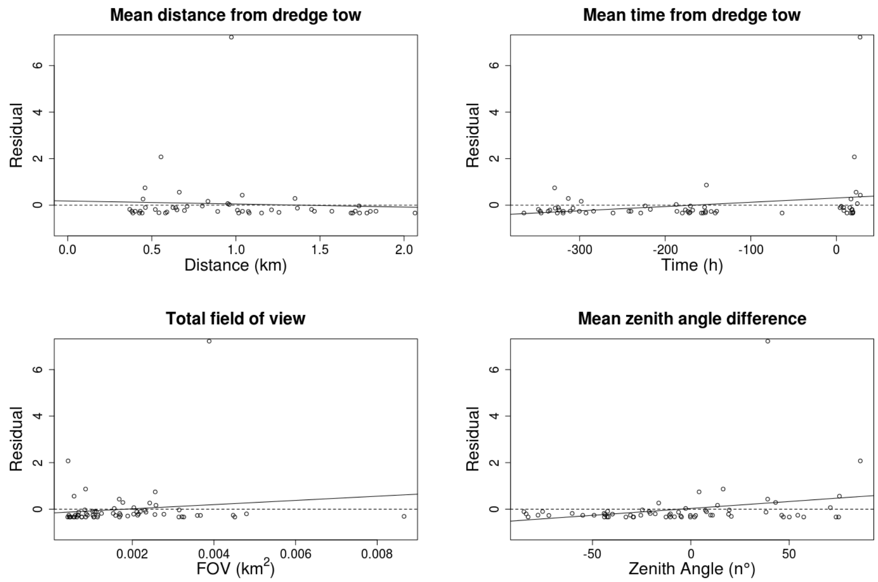

Figure 8: Plots of the standardized residuals (relative to perfect correlation; see methods) against potential drivers for the lack of correlation. The dotted line is residual=0, and the solid line is a simple linear regression fit to the data.