The prediction of random effects in hurdle models

Eva Cantoni,
Joanna Mills-Flemming
and Alan Welsh

University of Geneva,
Dalhousie University
The Australian National University

http://dsc.discovery.com/tv-shows/shark-week/photos/hammerhead-sharks-pictures.htm

Bycatch Data

For many endangered marine species, the only data on their abundance are counts of when they are caught as *bycatch*. These data

- are clustered because hauls are clustered within trips which may also be clustered within vessels and
- typically involve a larger number of zero counts (indicating that none were caught as bycatch in a particular haul) and very few positive counts (obtained if one or more are caught as bycatch in a haul).

Accurately estimating the probability of bycatch (and other cluster specific targets) and identifying bycatch *hotspots* is often the first step in trying to protect endangered marine species.

We consider hammerhead shark bycatch data (see Baum, 2007; Myer: et al., 2007) from the U.S. National Marine Fisheries Service Pelagic Observer Program (http://www.sefsc.noaa.gov/pop.jsp). There are 1825 observations on 75 vessels (1 to 86 observations per vessel) and $c = 292$ different trips (1 to 21 hauls per trip). We treat each trip as a cluster; vessels potentially add another layer.

The model

We describe the cluster structure in the data by unobserved independent random variables u_i and v_i, called random effects. Given u_i and v_i, the observed counts y_{ij} are independent with density

$$\langle y_{ij} \mid u_i, v_i \rangle = \begin{cases} 1 - p(\mathbf{x}_{ij}, u_i) & y_{ij} = 0 \\ p(\mathbf{x}_{ij}, u_i) f(y_{ij}, \lambda(\mathbf{z}_{ij}, u_i, v_i), \boldsymbol{\nu}) & y_{ij} = 1, 2, 3, \ldots \end{cases}, \qquad (1)$$

where

- $p(\mathbf{x}_{ij}, u_i)$ is the probability of observing a positive count
- $f(y_{ij}, \lambda(\mathbf{z}_{ij}, u_i, v_i), \boldsymbol{\nu})$ is the conditional density of $y_{ij}|y_{ij} \geq 1$ (the density of a discrete distribution defined on the positive integers) with parameters
  - λ which is a function of the covariates, and
  - $\boldsymbol{\nu}$ possibly additional nuisance parameters.

We parametrize p and λ as

$$\operatorname{logit}\{p(\mathbf{x}_{ij}, \mathbf{u}_i)\} = \mathbf{x}_{ij}^T \boldsymbol{\alpha} + \sigma_u u_i \tag{2}$$

and

$$\log\{\lambda(\mathbf{z}_{ij}, \mathbf{u}_i, \mathbf{v}_i)\} = \mathbf{z}_{ij}^T \boldsymbol{\beta} + \{\gamma \sigma_u u_i + \sigma_v v_i\}, \tag{3}$$

where

- $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ are regression parameters
- σ_u and σ_v are non-negative spread parameters and
- γ is a dependence parameter which controls dependence between p and λ.

The logit and log links can be replaced by other link functions if needed.

Let $u_i \sim g_i(u_i, \boldsymbol{\tau})$ and $v_i \sim h_i(v_i, \boldsymbol{\kappa})$ to allow for non-Gaussian random effect distributions with additional nuisance parameters $\boldsymbol{\tau}$ and $\boldsymbol{\kappa}$. Under standard Gaussian distributions, the nuisance parameters $\boldsymbol{\tau}$ and $\boldsymbol{\kappa}$ are redundant.

We can extend the model to more complicated random effect structures.

Prediction

There are various quantities we might want to predict at the cluster specific level or estimate marginally with respect to the clusters. For example, in addition to u_i and v_i, we may be interested in predicting the probability of presence

$$P(Y_{ij} > 0 | u_i) = p(\mathbf{x}_{ij}, u_i),$$

or the expected abundance (given presence)

$$\mathrm{E}(Y_{ij} | Y_{ij} > 0, u_i, v_i) = M\{\lambda(\mathbf{z}_{ij}, u_i, v_i), \boldsymbol{\nu}\},$$

or, the expected abundance

$$\mathrm{E}(Y_{ij} | u_i, v_i) = p(\mathbf{x}_{ij}, u_i) M\{\lambda(\mathbf{z}_{ij}, u_i, v_i), \boldsymbol{\nu}\}.$$

Here $M\{\lambda(\mathbf{z}_{ij}, u_i, v_i), \boldsymbol{\nu}\}$ is the mean of the distribution f of the positive observations. We may also be interested in the analogous marginal estimation problems, for instance, the probability of bycatch, the expected bycatch given non-zero bycatch, or the expected bycatch. These are obtained by integrating the analogous cluster specific quantities over u and v.

Unified treatment

To develop a unified treatment notice that the cluster specific prediction targets of interest are all of the form

$$t(u, v, \mathbf{x}, \mathbf{z}, \boldsymbol{\theta}), \tag{4}$$

and the marginal estimation targets are then of the form

$$\int\int t(u, v, \mathbf{x}, \mathbf{z}, \boldsymbol{\theta}) g(u, \boldsymbol{\tau}) h(v, \boldsymbol{\kappa}) \mathrm{d}u\mathrm{d}v. \tag{5}$$

We will focus on the cluster specific targets.

Empirical Best Predictor (EBP)

A more satisfactory approach is based on the minimum mean squared error predictor or "best predictor" of Jiang and Lahiri (2001) which is

$$\begin{aligned} T_t(\mathbf{x}, \mathbf{z}, \boldsymbol{\theta}; \mathbf{y}_i) &= \int\!\!\int t(u, v, \mathbf{x}, \mathbf{z}, \boldsymbol{\theta}) \langle u, v | \mathbf{y}_i \rangle \mathrm{d}u\mathrm{d}v \\ &= \frac{\int\int t(u, v, \mathbf{x}, \mathbf{z}, \boldsymbol{\theta}) \langle \mathbf{y}_i | u, v \rangle g(u, \boldsymbol{\tau}) h(v, \boldsymbol{\kappa}) \mathrm{d}u\mathrm{d}v}{\int\int \langle \mathbf{y}_i | u, v \rangle g(u, ta) h(v, \boldsymbol{\kappa}) \mathrm{d}u\mathrm{d}v} \end{aligned}$$

by Bayes' Theorem.

The expression for $T_t(\mathbf{x}, \mathbf{z}, \boldsymbol{\theta}; \mathbf{y}_i)$ shows that the best predictor for $t(u, v, \mathbf{x}, \mathbf{z}, \boldsymbol{\theta})$ is a ratio of integrals. Using Monte Carlo approximation we obtain the empirical best predictor (EBP)

$$\hat{T}_t(\mathbf{x}, \mathbf{z}, \hat{\boldsymbol{\theta}}; \mathbf{y}_i) = \frac{\sum_{k=1}^{K} t(u_k^*, v_k^*, \mathbf{x}, \mathbf{z}, \hat{\boldsymbol{\theta}}) \langle \mathbf{y}_i | u_k^*, v_k^* \rangle}{\sum_{k=1}^{K} \langle \mathbf{y}_i | u_k^*, v_k^* \rangle},$$

where u_k^* and v_k^* are randomly sampled from $g_i(u_i, \hat{\boldsymbol{\tau}})$ and $h_i(v_i, \hat{\boldsymbol{\kappa}})$ respectively.

Mean Squared Error of Prediction

We used the parametric bootstrap to estimate the mean squared error of prediction in the following way:

- Compute the estimate $\hat{\boldsymbol{\theta}}$ from the data.
- For $b = 1, \ldots, B$
  - use the bootstrap to generate $\hat{\boldsymbol{\theta}}_b^*$
  - generate u_{b1}^*, u_{b0}^* from $g(u, \hat{\boldsymbol{\tau}})$, v_{b1}^*, v_{b0}^* from $h(v, \hat{\boldsymbol{\kappa}})$ and y_{bi}^* from the density $\langle y_i | u_{b1}^*, v_{b1}^* \rangle$.
- Compute the bootstrap estimate of $\text{msep}_i(\hat{T}_t, t)$ as

$$\frac{1}{B} \sum_{b=1}^{B} \left[\hat{T}_t(\mathbf{x}_i, \mathbf{z}_i, \hat{\boldsymbol{\theta}}_b^*, y_{bi}^*) - t(u_{b0}^*, v_{b0}^*, \mathbf{x}_i, \mathbf{z}_i, \hat{\boldsymbol{\theta}}) \right]^2 .$$

Note that we randomly generate u, v and y_i to take into account all the sources of variability (they are all random variables) in the mean squared error of prediction.

New Fast Bootstrap

Our estimator $\widehat{\boldsymbol{\theta}}$ satisfies the estimating equation $\mathbf{0} = \boldsymbol{\psi}(\widehat{\boldsymbol{\theta}})$.

- The estimating equation can be turned into a fixed-point equation with fixed-point function $\mathbf{g}(\boldsymbol{\theta}) = \boldsymbol{\theta} + \mathbf{A}(\boldsymbol{\theta})\boldsymbol{\psi}(\boldsymbol{\theta})$.
- We can simplify differentiating $\mathbf{g}(\boldsymbol{\theta})$ by setting $\mathbf{A}(\boldsymbol{\theta}) = \mathbf{I}$, so $\partial_{\boldsymbol{\theta}}\mathbf{g}(\boldsymbol{\theta}) = \mathbf{I} + \partial_{\boldsymbol{\theta}}\boldsymbol{\psi}(\boldsymbol{\theta})$, and hence $\mathbf{I} - \partial_{\boldsymbol{\theta}}\mathbf{g}(\boldsymbol{\theta}) = -\partial_{\boldsymbol{\theta}}\boldsymbol{\psi}(\boldsymbol{\theta})$, which is already available from the initial maximum likelihood estimation.

Then we can apply the fast bootstrap using

$$\widehat{\boldsymbol{\theta}}^*_A = \widehat{\boldsymbol{\theta}} - \{\partial_{\boldsymbol{\theta}}\boldsymbol{\psi}(\widehat{\boldsymbol{\theta}})\}^{-1}\{\widehat{\boldsymbol{\theta}}^{(1)*} - \widehat{\boldsymbol{\theta}}\},$$

with $\widehat{\boldsymbol{\theta}}^{(1)*} = \widehat{\boldsymbol{\theta}} + \boldsymbol{\psi}^*(\widehat{\boldsymbol{\theta}})$, or

$$\widehat{\boldsymbol{\theta}}^*_A = \widehat{\boldsymbol{\theta}} - \{\partial_{\boldsymbol{\theta}}\boldsymbol{\psi}(\widehat{\boldsymbol{\theta}})\}^{-1}\boldsymbol{\psi}^*(\widehat{\boldsymbol{\theta}}),$$

We can compute $\boldsymbol{\psi}^*$ from analytic derivatives or using a finite difference approximation based on evaluating the bootstrap sample log-likelihood at two points near $\widehat{\boldsymbol{\theta}}$. This is simple when we can compute the log-likelihood quickly.

Simulation Study

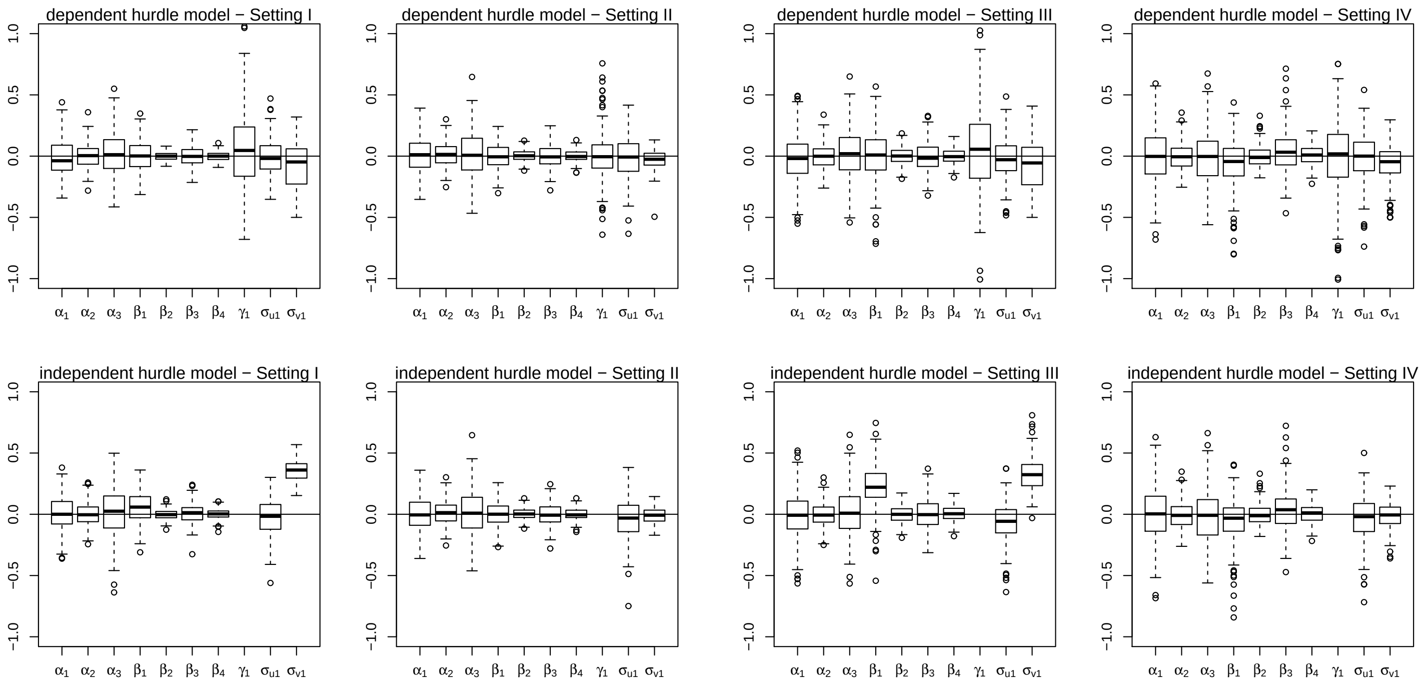
We simulated data from our model (1)–(3) using a truncated Poisson distribution for the positive counts and the random intercepts u_i and v_i having $N(0,1)$ distributions. Each simulated data set contained $c = 100$ clusters, half with 5 and half with 10 measurements per cluster, for a total of 750 observations.

We included in $\mathbf{x}_{ij}$ an intercept, a $N(0,1)$ covariate and a Bernoulli(1/2) covariate. The covariates $\mathbf{z}_{ij}$ included the same variables plus another $N(0,1)$ variable. For the parameters, we considered four settings:

	Proportion of zeros	
	Low (0.3)	High (0.7)
Dependent ($\gamma \neq 0$)	I	III
Independent ($\gamma = 0$)	II	IV

For each of 200 simulations, we used $K = 1000$ for the Monte Carlo approximation to the likelihood, 10 different starting points for its numerical optimisation and took $B = 1000$ in the fast bootstrap.

Results for parameter estimation



Settings I–IV: boxplots of the parameter estimates with their true values subtracted off.

Incorrectly assuming independence between the random parts of the model produces biases in the estimates, in particular the intercept for the positive part.

When $\gamma_1 = 0$ (Settings II and IV), our dependent hurdle model performs as well as the independent hurdle model.

	bias	se	$\sqrt{msep}$	$\sqrt{msep_t^*}$
u_{i1}	-0.007	0.770	1.237	1.281
v_{i1}	-0.029	0.415	1.005	1.119
$P(Y_{i1} > 0)$	0.000	0.129	0.269	0.211
$P(Y_{i2} > 0)$	-0.001	0.125	0.237	0.204
$P(Y_{i3} > 0)$	-0.004	0.110	0.199	0.181
$P(Y_{i4} > 0)$	-0.006	0.090	0.194	0.150
$E(Y_{i1}\|Y_{i1} > 0)$	0.002	0.109	2.595	0.329
$E(Y_{i2}\|Y_{i2} > 0)$	0.002	0.143	2.569	0.439
$E(Y_{i3}\|Y_{i3} > 0)$	0.021	1.643	3.683	4.492
$E(Y_{i4}\|Y_{i4} > 0)$	0.039	4.507	9.674	11.589
$E(Y_{i1})$	-0.004	0.209	2.178	0.434
$E(Y_{i2})$	-0.006	0.236	2.119	0.521
$E(Y_{i3})$	-0.081	1.542	2.842	4.124
$E(Y_{i4})$	-0.223	4.266	7.320	10.808

We made predictions in two clusters with $n_i = 5$ and two with $n_i = 10$.

The bias is generally quite small, but more often negative. This is due to the underestimation of the spread parameters and the built-in shrinkage effect in optimal prediction.

Coverages of nominal 95% prediction intervals for u_i and v_i.

	$\text{CI}(\hat{u}_i, u_i)$				$\text{CI}(\hat{v}_i, v_i)$			
Setting	I	II	III	IV	I	II	III	IV
Cluster 1	0.950	0.970	0.985	0.970	0.885	0.935	0.905	0.910
Cluster 2	0.925	0.950	0.970	0.960	0.910	0.955	0.910	0.950
Cluster 51	0.950	0.960	0.950	0.940	0.885	0.970	0.925	0.945
Cluster 52	0.950	0.950	0.960	0.940	0.915	0.935	0.910	0.915

These coverages are very good for u_i, but not so good for v_i, when $\gamma \neq 0$ (Settings I and III). Recall that v_i is estimated from a smaller sample.

Hammerhead Shark Bycatch Data

The observations are counts of bycatch y_{ij} for $j = 1, \ldots, n_i$ hauls on $i = 1, \ldots, c$ trips. We have 85% zeros and the positive counts range from 1 to 46. The covariates are

- YEAR, from 1 to 14, representing the period 1992-2005
- AVGHKDEP, from 6.40 to 182.88, average or median hook depth
- AREA, 4=South Atlantic Bight (reference) and 5=Mid Atlantic Bight
- SEASON, with 464 observations in autumn (reference), 543 in spring, 525 in summer and 293 in winter
- TOTHOOK, from 25 to 1548, the log number of hooks to measure catch effort.

The covariates or subsets of the covariates are represented by $\mathbf{x}_{ij}$, $\mathbf{z}_{ij}$.

Presence-absence fit

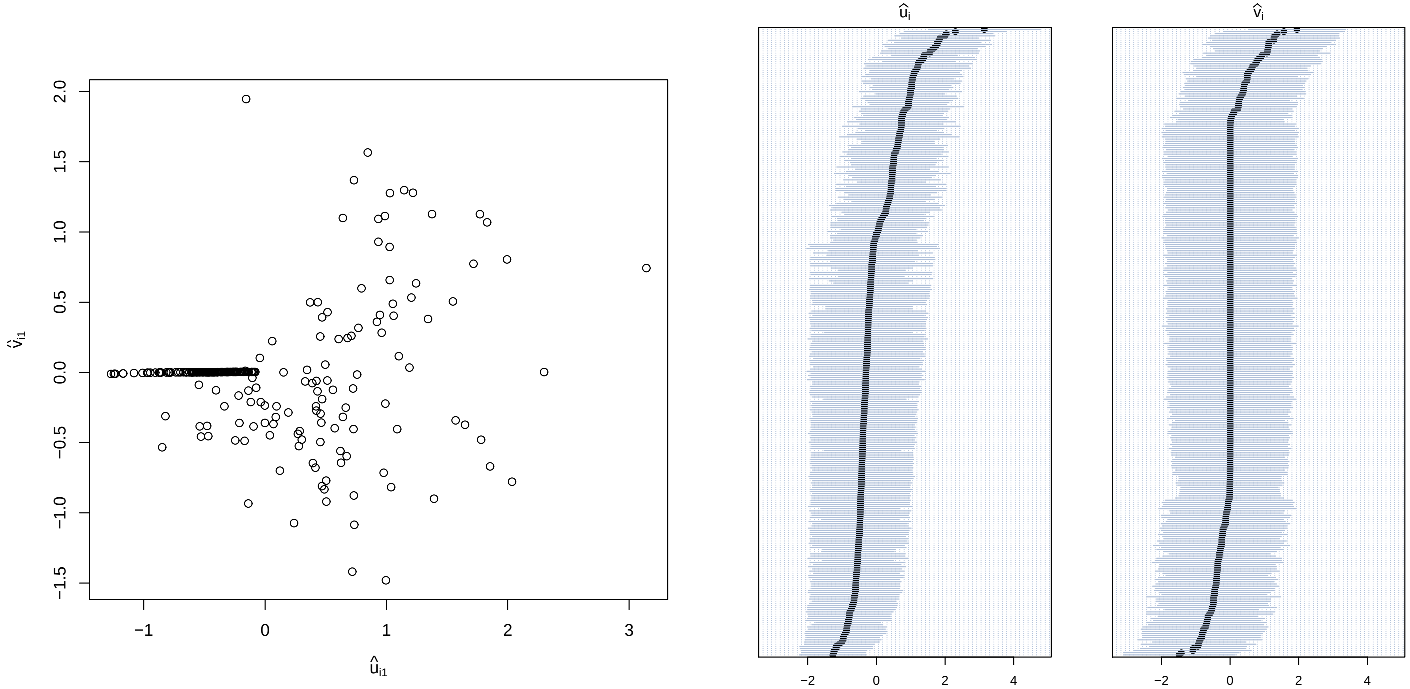
	Dependent model	Independent model
Variable	Coeff. (SE-H, SE-b)	Coeff. (SE)
Intercept	-2.123 (1.453,1.636)	1.951 (1.508)
YEAR	-0.059 (0.028,0.027)	-0.044 (0.030)
AVGHKDEP	-0.011 (0.011,0.016)	0.007 (0.010)
AREA5	-0.241 (0.242,0.317)	-0.053 (0.254)
SEASONspring	1.609 (0.341,0.351)	1.630 (0.358)
SEASONsummer	0.074 (0.362,0.372)	0.096 (0.366)
SEASONwinter	1.068 (0.369,0.342)	0.950 (0.393)
log(TOTHOOK)	-0.008 (0.212, 0.206)	-0.170 (0.222)

SE-H, standard errors from the numerical Hessian; SE-b, standard errors from the bootstrap.

Abundance fit

	Dependent model	Independent model
Variable	Coeff. (SE-H, SE-b)	Coeff. (SE)
Intercept	-4.871 (0.661,0.678)	-3.322 (1.148)
YEAR	-0.132 (0.032,0.029)	-0.105 (0.042)
AVGHKDEP	-0.067 (0.008,0.013)	-0.052 (0.013)
AREA5	-0.151 (0.230,0.257)	-0.182 (0.249)
SEASONspring	0.850 (0.362, 0.236)	-0.121 (0.519)
SEASONsummer	-0.435 (0.488,0.401)	-0.843 (0.590)
SEASONwinter	1.278 (0.342,0.226)	0.288 (0.574)
log(TOTHOOK)	1.020 (0.096,0.122)	0.944 (0.171)
γ_1	1.413 (0.159,0.157)	–
σ_{u1}	1.248 (0.145,0.134)	1.387 (n.a.)
σ_{v1}	1.116 (0.144,0.179)	1.544 (n.a.)

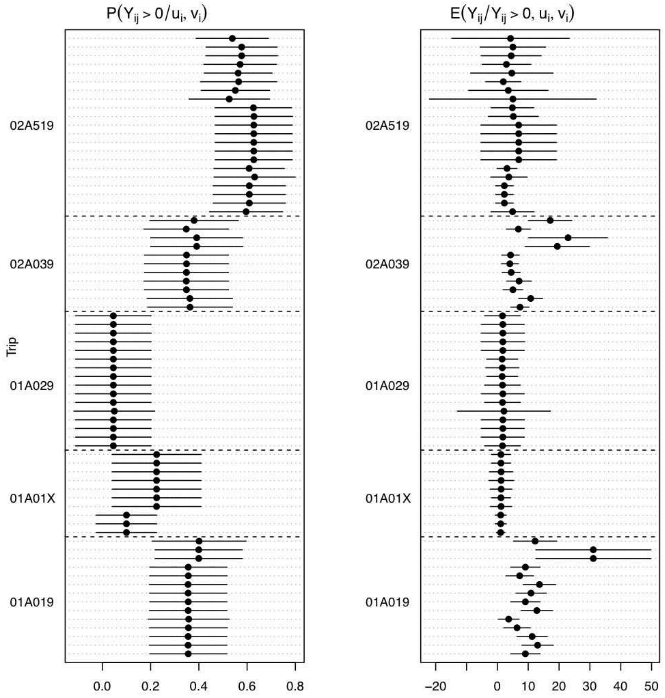
Predictions of u_i and v_i for the 292 trips



The $\hat{v}_i$ are generally smaller in magnitude than the $\hat{u}_i$. The very small $\hat{v}_i$ correspond to negative $\hat{u}_i$. The happens in 172 clusters whose responses are all zero. The $\hat{u}_i$ for these clusters are negative, to reduce the estimated probability.

We looked at the predicted u_i and v_i in the lower and upper tails (ordered separately) to determine whether there was structure with respect to vessel (and other covariates). We found that the pattern was mainly seasonal.

Predictions of $P(Y_{ij} > 0|u_i, v_i)$ and $E(Y_{ij}|Y_{ij} > 0, u_i, v_i)$ for 5 trips

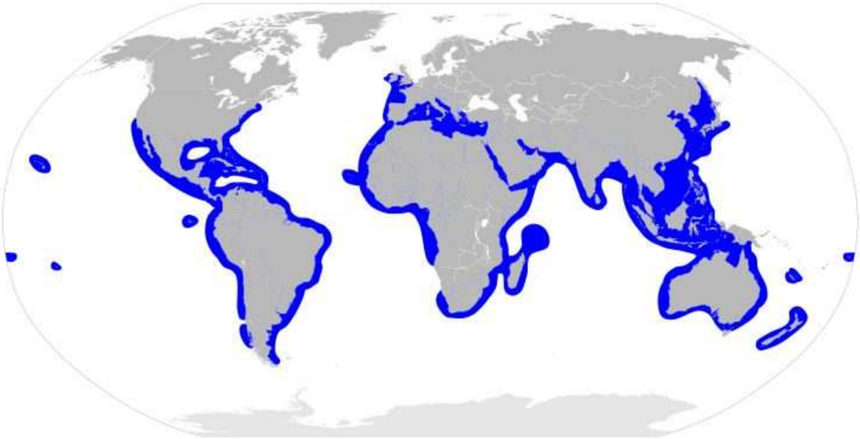


For trip 01A019 (left panel) the two groups of predictions correspond to two different values of AVGHKDEP. For trip 01A01X, the two groups of predictions correspond to two different values of YEAR.

In the right panel, the length of the confidence intervals is quite variable. For trip 01A019, the two much larger confidence intervals correspond to a different combination of AVGHKDEP and log(TOTHOOK). The longer confidence interval for trip 01A029 corresponds to one observation with a different value of AVGHKDEP. The smaller variations in length correspond to differences in log(TOTHOOK).

Hammerhead Sharks

- 8 or 9 species (one species may be subdivided into two)
- range from 0.9 – 6 m long and weigh from 3 – 580 kg
- are usually light grey and have a greenish tint
- heads are flattened and laterally extended into a ”hammer” shape called a ”cephalofoil”
- are found worldwide in warmer waters along coastlines and continental shelves
- form schools during the day and hunt alone at night



http://onebigphoto.com/a-shoal-of-hammerhead-sharks/
wikipedia

The cephalofoil

- The eye position allows 360-degree vision in the vertical plane (they can see above and below them at all times).
- The spread of electroreceptory sensory pores (ampullae of Lorenzini) over a wide area allows effective sweeping for prey.
- The head is used to pin down stingrays (eaten when they are weak and in shock).

http://animalsearths.blogspot.com.au/2011/05/great-hammerhead-shark.html

The head was previously thought to aid manoeuverability, allowing stable, sharp turning movement, but the unusual structure of its vertebrae may be more important than the head, though it also shifts and provide lift.

Diet

Hammerheads have small mouths and do a lot of bottom-hunting. Diet includes fish, squid, octopus, crustaceans, stingrays and other sharks. The Great Hammerhead, a larger and more aggressive species, is known to eat other hammerhead sharks, including their own young.

Of the nine known species of hammerhead, three can be dangerous to humans: the Scalloped, Great, and Smooth hammerheads. Up to 2010 there had been 33 reported attacks, but no fatalities.

http://animalsearths.blogspot.com.au/2011/05/great-hammerhead-shark.html
Wikipedia