Using Side-by-Side Haul Data to Estimate NMFS EBS Survey Catchability for Tanner Crab

William Stockhausen (NOAA/AFSC)

Scott Goodman, Madi Heller-Shipley (BSFRF)

January 15, 2025

What's "catchability" & why is it important?

- Proportionality, C, between
  - "what's caught in a haul" or "what's caught in a survey" and
  - "what's on the bottom"

- Size-specific stock assessments, like the Tanner crab assessment, need to estimate (or assume) size-specific catchability, C_z, to relate size-specific model population estimates to survey catch data

- For surveys, can think about 2 types:

haul-specific

$$N_z^{haul} = C_z^{haul} \cdot N_z^{area\ swept}$$

survey-specific (aggregated)

$$N_z^{survey} = C_z^{survey} \cdot N_z^{population}$$

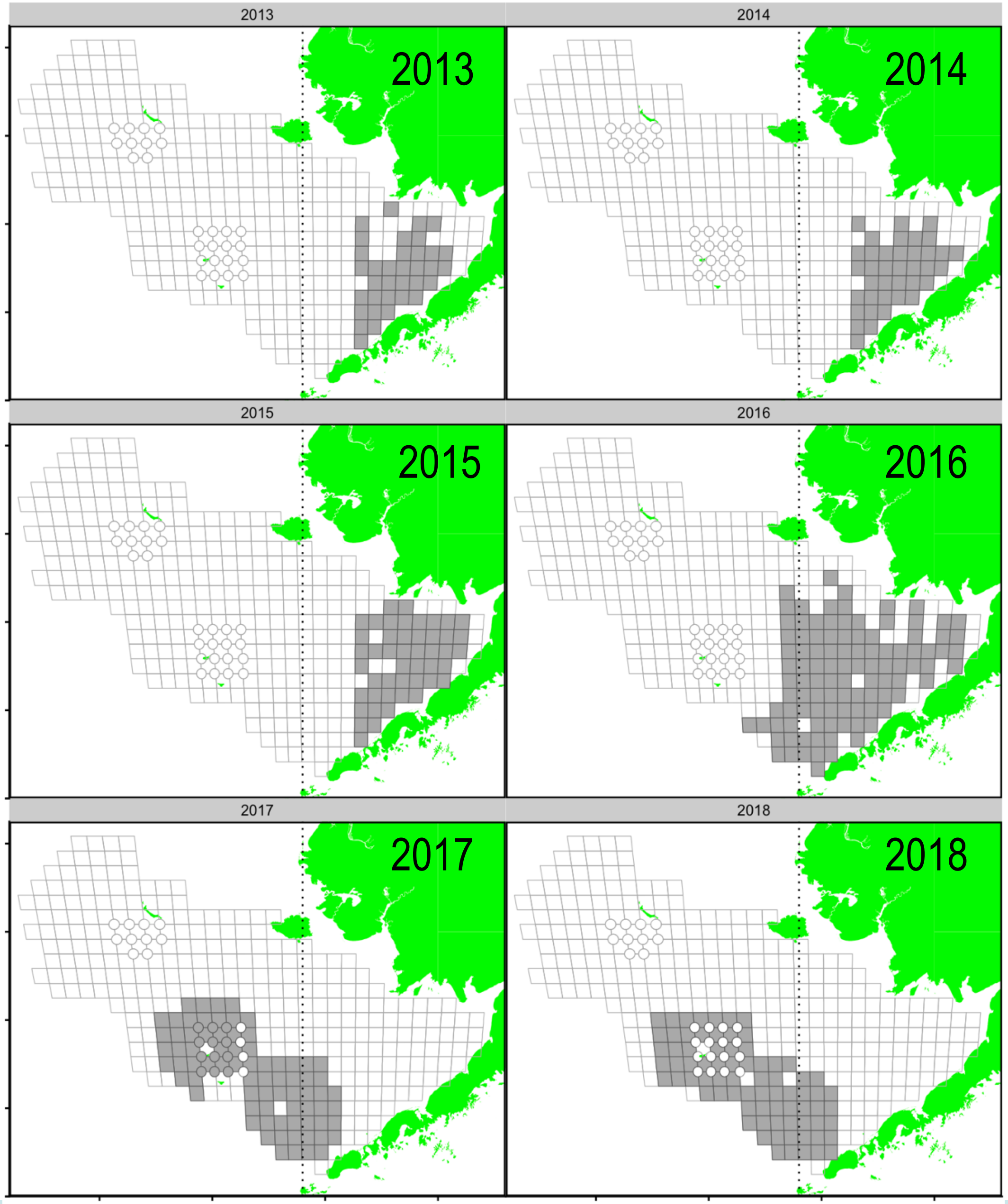
  - z: size (sex, maturity state, shell condition, etc.)

2013-2018 Catch Comparison Studies

- Compare survey catches from target gear with ***unknown*** catchability C_z^U to reference gear with ***known*** catchability C_z^K
- BSFRF-NMFS collaborative studies to estimate NMFS survey catchability for Tanner crab using paired hauls conducted "side-by-side" at several EBS survey stations each year
- Synoptic surveys across same study area allow estimates of NMFS **survey-level** catchability, *relative* to the BSFRF gear, within the study area (integrating across environmental effects)
- Side-by-side (SBS) paired hauls allow estimates of NMFS **haul-level** catchability, *relative* to the BSFRF gear (possibly estimating environmental effects)
- BSFRF gear **assumed to catch all crab in area swept**

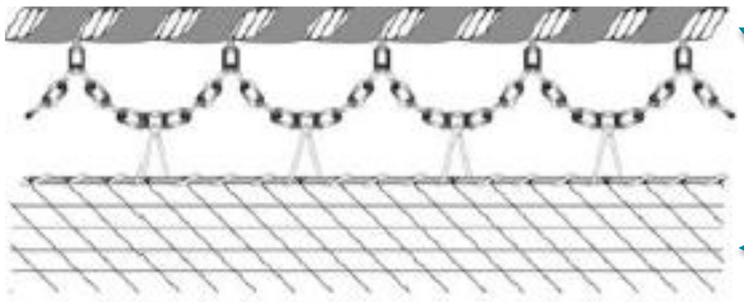
$$C_z^K = C_z^{BSFRF} = 1$$

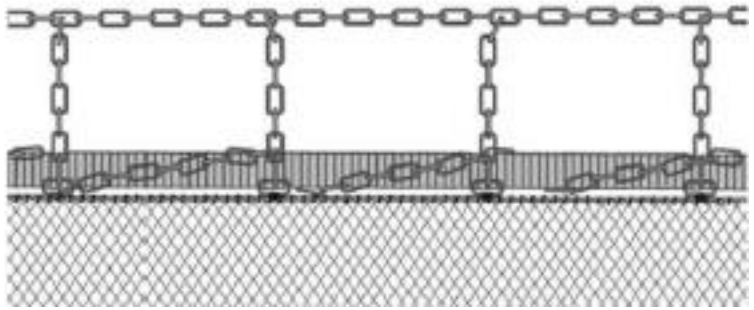
so estimated NMFS catchability is **absolute**



The Gear

NMFS	BSFRF
83-112 Eastern Trawl	*Nephrops* Trawl





tickler chain

34.1m

footrope

27 m

net

net mesh

NMFS		BSFRF
10.2 cm	body	8.0 cm
8.9 cm	cod end	6.0 cm
3.2 cm	liner	5.0 cm

tow characteristics

NMFS		BSFRF
15 – 19 m	net spread	10 – 14 m
3 kts	tow speed	2 kts
30 min	duration	5 min

Area swept ratio ~ 6*x*

tow separation

0.1 – 0.2 nmi

source: Somerton et al., 2013.

Survey/Study-level (Aggregated) Catchability

Does not utilize side-by-side nature of paired hauls

Survey/study area-level (aggregated) catch comparisons

N_z^U: estimated study area (expanded) abundance(or mean CPUE) at size using target gear with **unknown** catchability (NMFS)

N_z^K: estimated study area (expanded) abundance(or mean CPUE) at size using reference gear with **known** catchability (BSFRF)

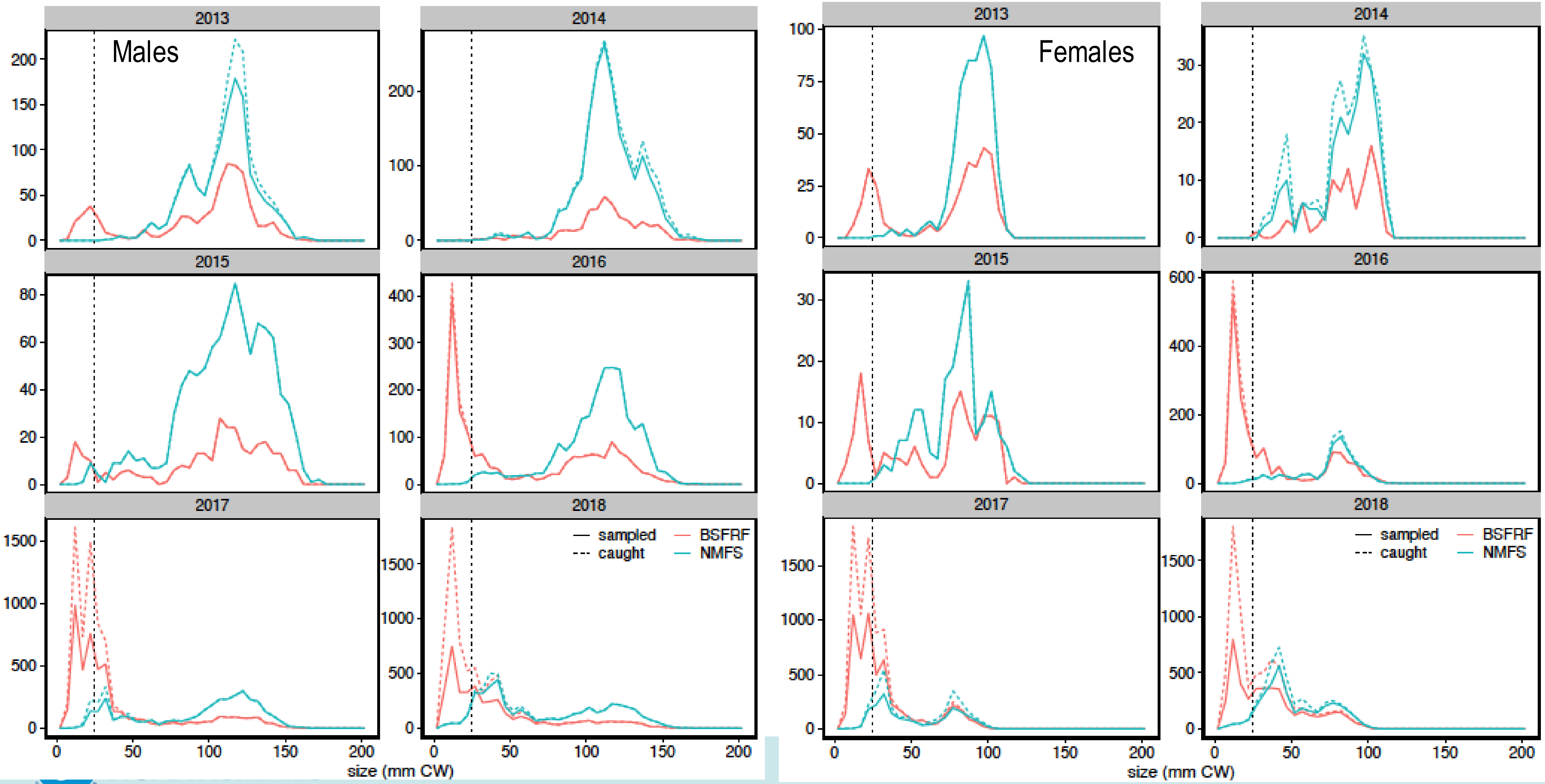
$$N_z^U \equiv C_z^U \cdot N_z^{area}$$
$$N_z^K \equiv C_z^K \cdot N_z^{area}$$

$$\frac{N_z^U}{N_z^K} = \frac{C_z^U \cdot N_z^{area}}{C_z^K \cdot N_z^{area}} = \frac{C_z^U}{C_z^K} = R_z \equiv relative\ catchability$$

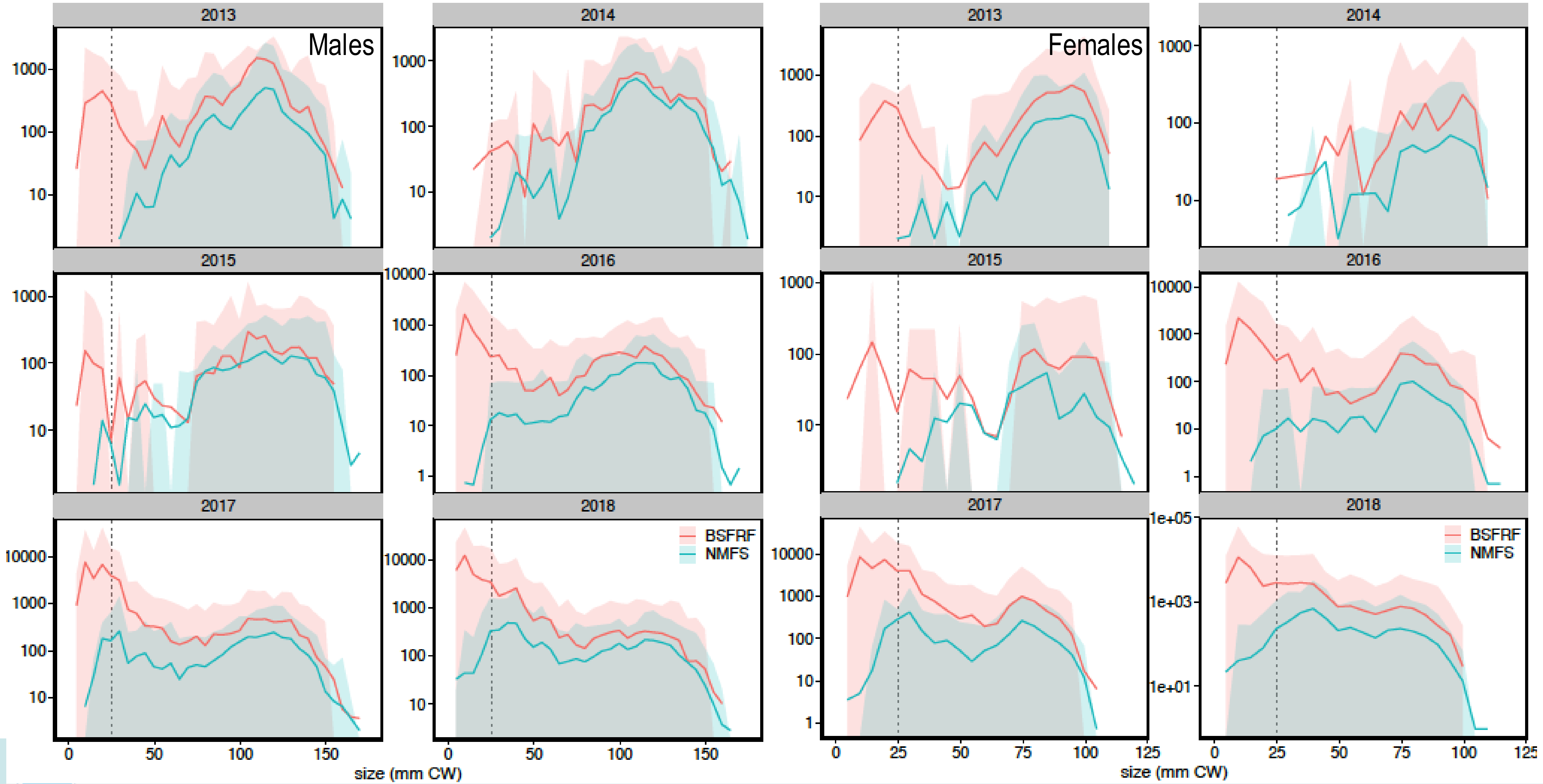
with BSFRF catchability assumed = 1: $C_z^U = R_z = \frac{N_z^U}{N_z^K}$

so want to use observations of $\frac{N_z^U}{N_z^K}$ to estimate R_z as a smooth function of size

Total numbers sampled/caught



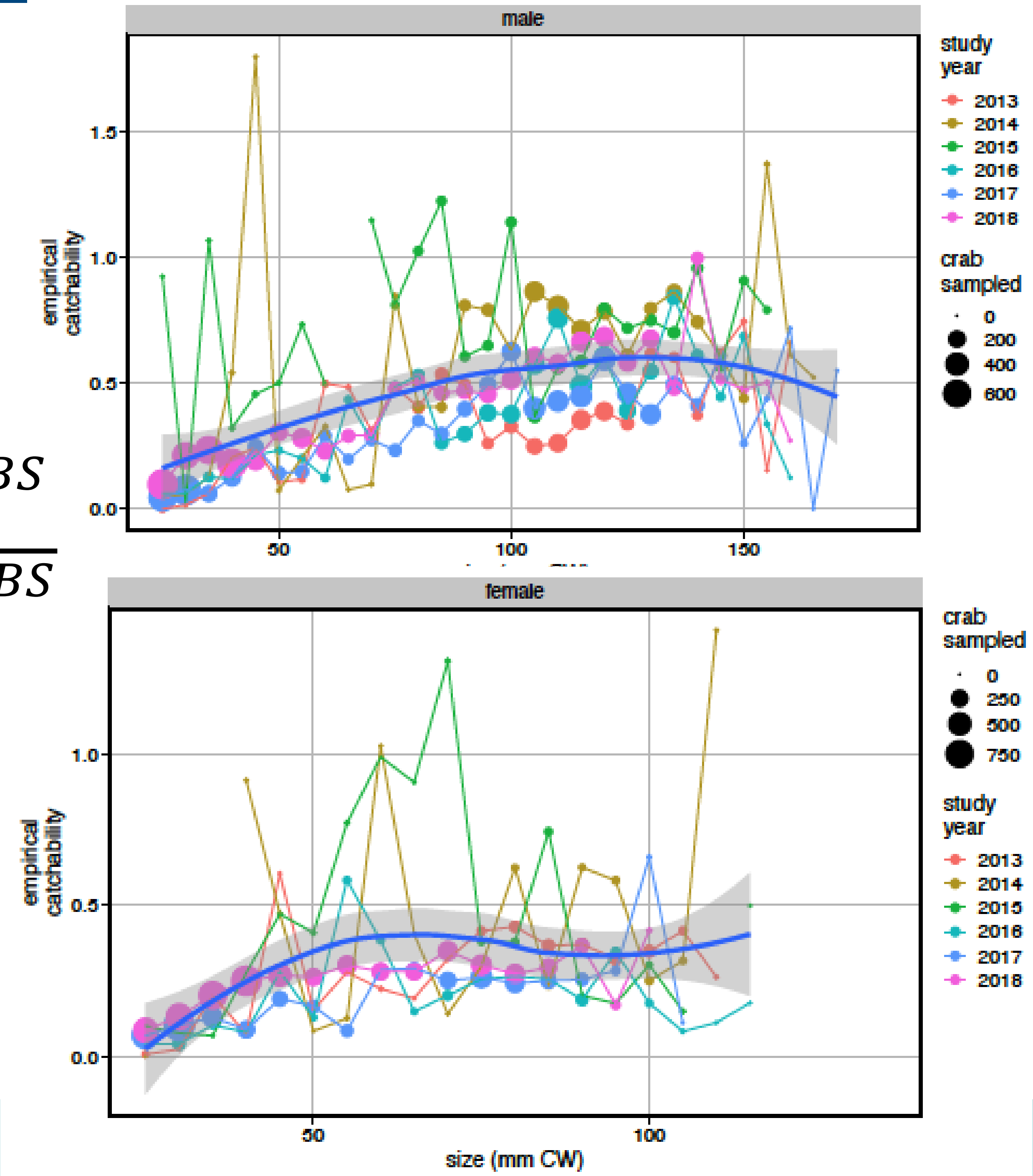
Mean CPUE (and 90% CIs)



Raw “survey-level” catchability

$$C_{z,y}^{NMFS\ SBS} = R_{z,y} = \frac{N_{z,y}^{NMFS\ SBS}}{N_{z,y}^{BSFRF\ SBS}}$$

- ignores any haul-level environmental effects



Model fitting

- survey-level catchability modeled as Tweedie-distributed function of size

$$R_{z,y} \sim Tw(\mu_{z,y}, \phi) \quad V(R_{z,y}) = \beta \cdot \mu_{z,y}^{\phi}$$

- with a log-link function

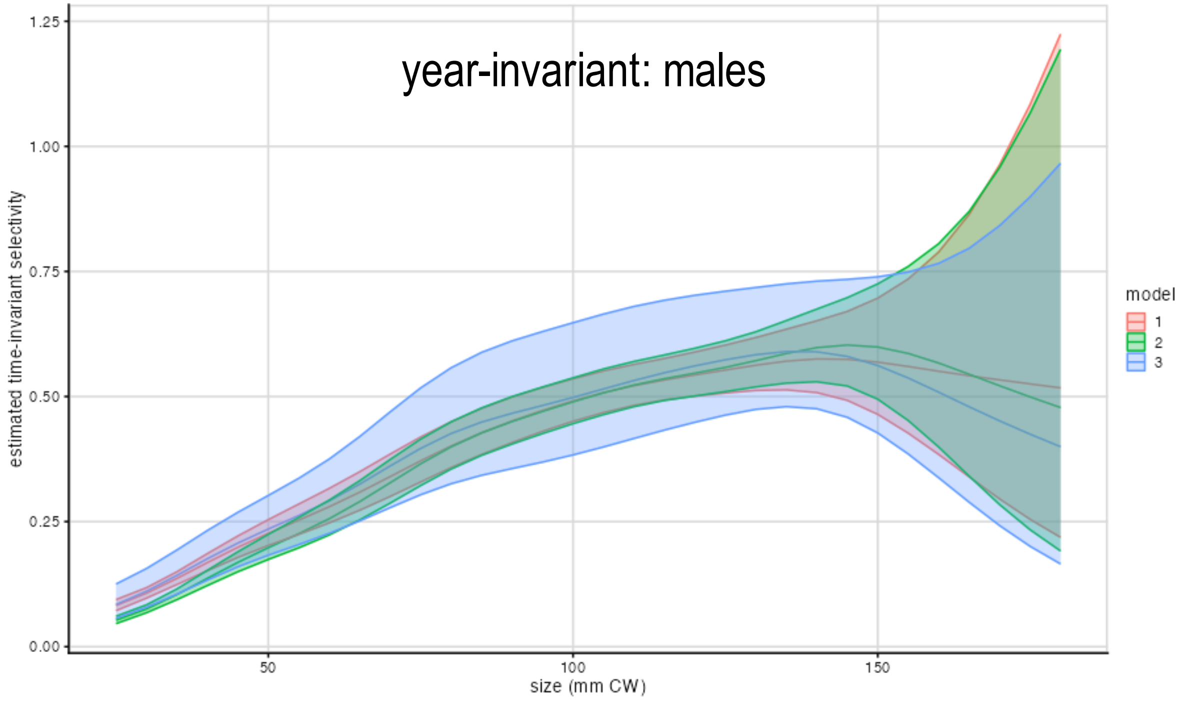
Model 1: $\ln\left(\mu_{z,y} = E\left[R_{z,y}\right]\right) = f_y(z) = \alpha + s(z)$

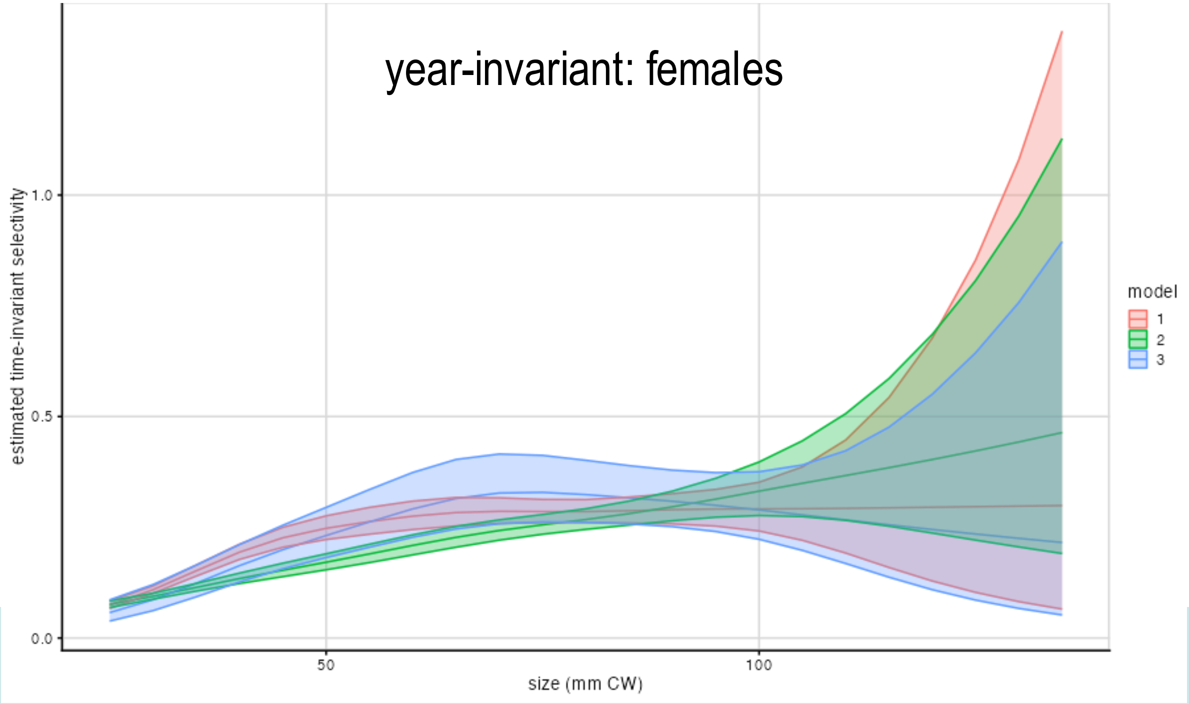
Model 2: $\ln\left(\mu_{z,y} = E\left[R_{z,y}\right]\right) = f_y(z) = \alpha + s(z) + t(z|y)$

Model 3: $\ln\left(\mu_{z,y} = E\left[R_{z,y}\right]\right) = f_y(z) = \alpha + s(z) + t_{RE}(z,y)$

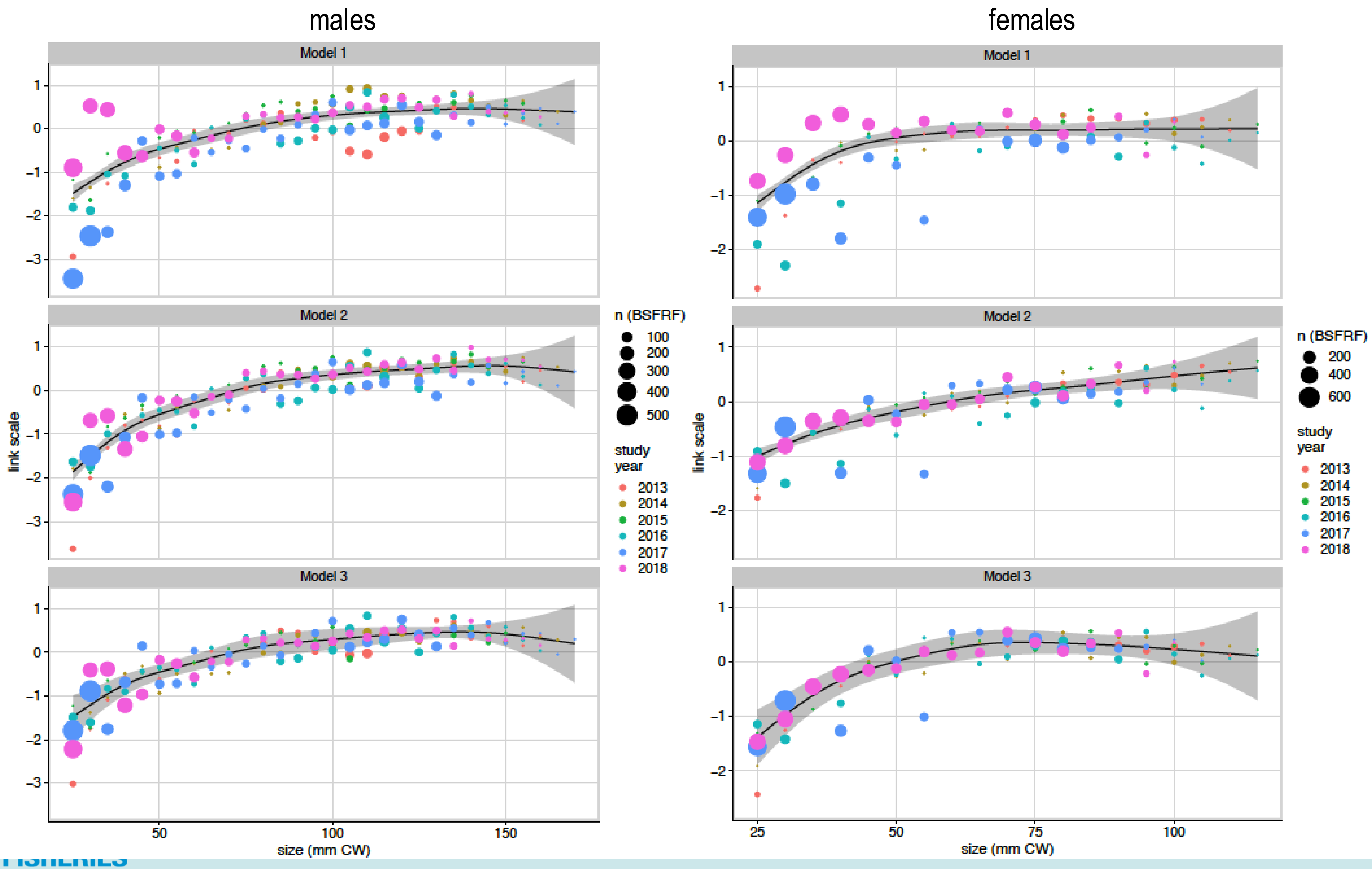
$$t_{RE}(z,y) \sim N(0, \sigma_z^2)$$

- so $C_{z,y}^{NMFS\ SBS} = E\left[R_{z,y}\right] = \exp\left(f_y(z)\right)$

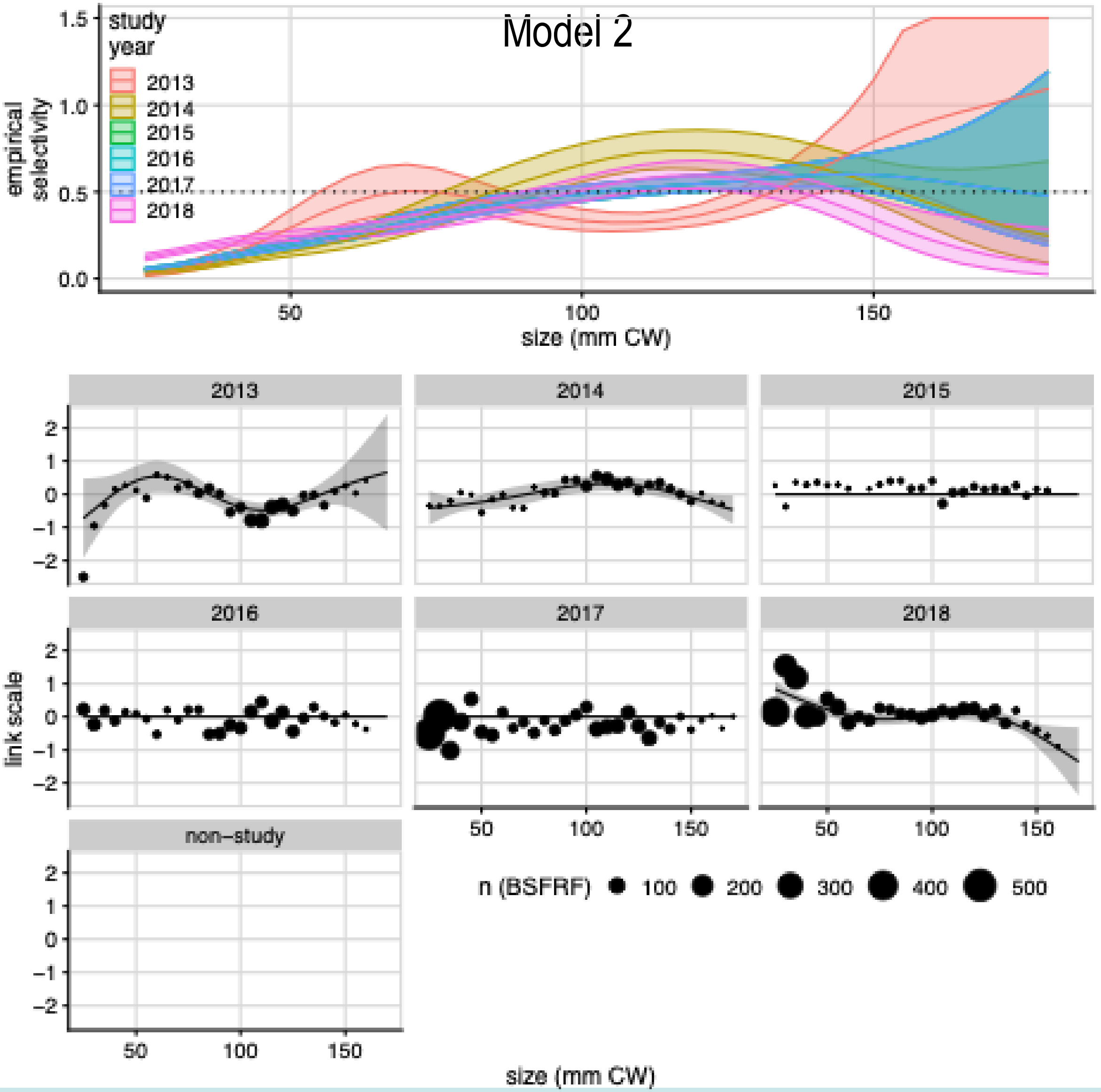


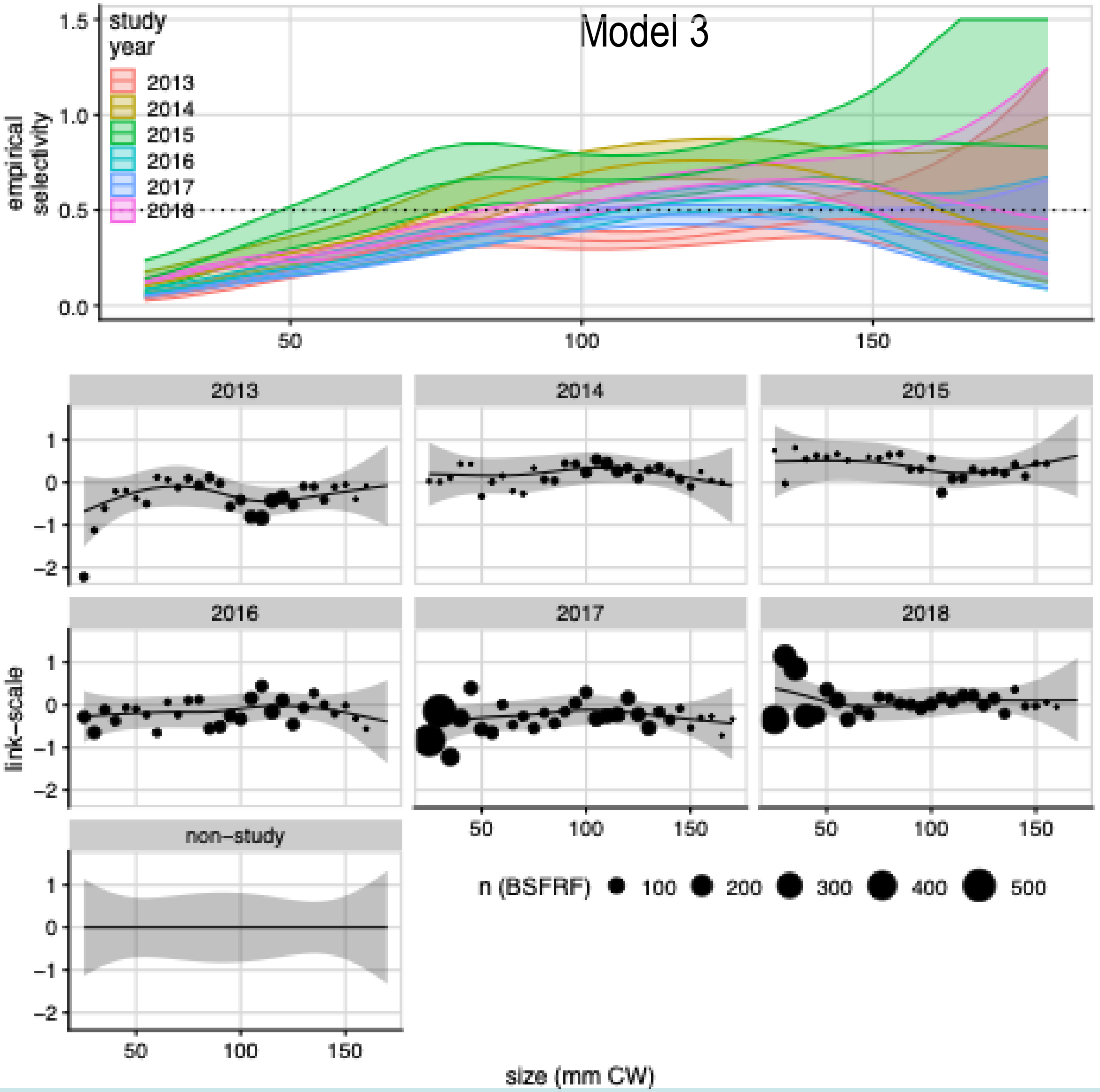


Partial residuals: s(z)

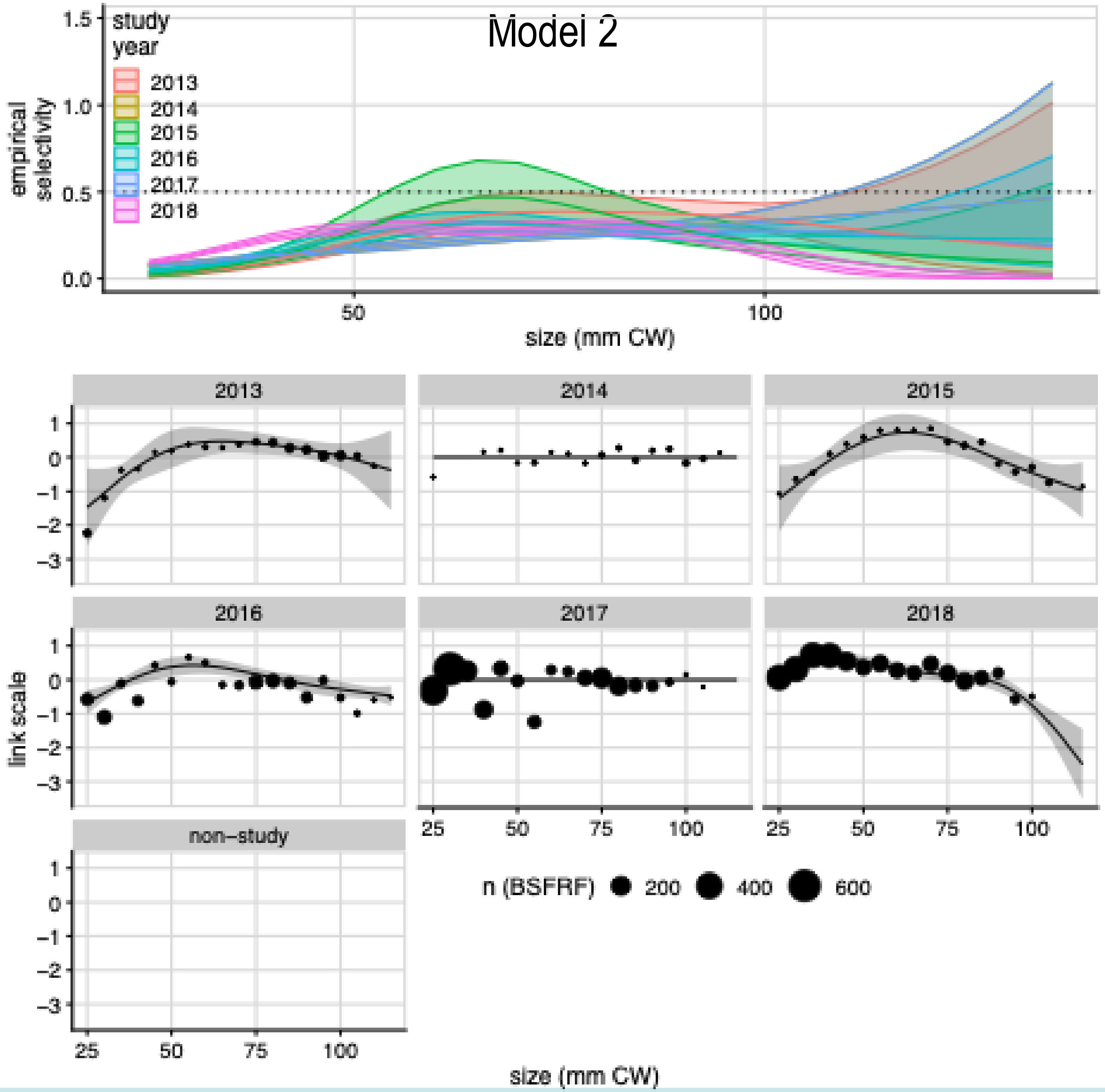


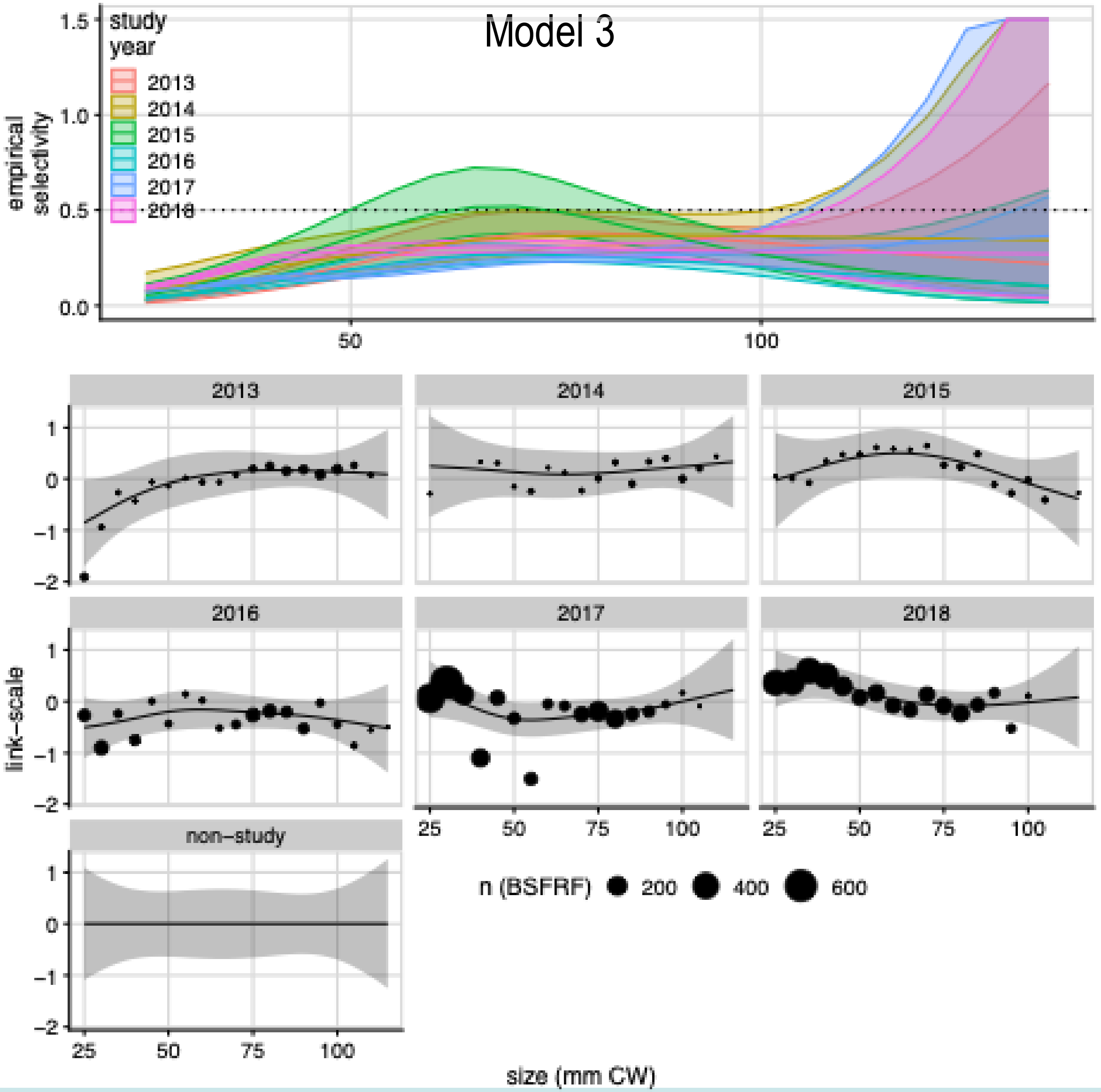
Year effects: males



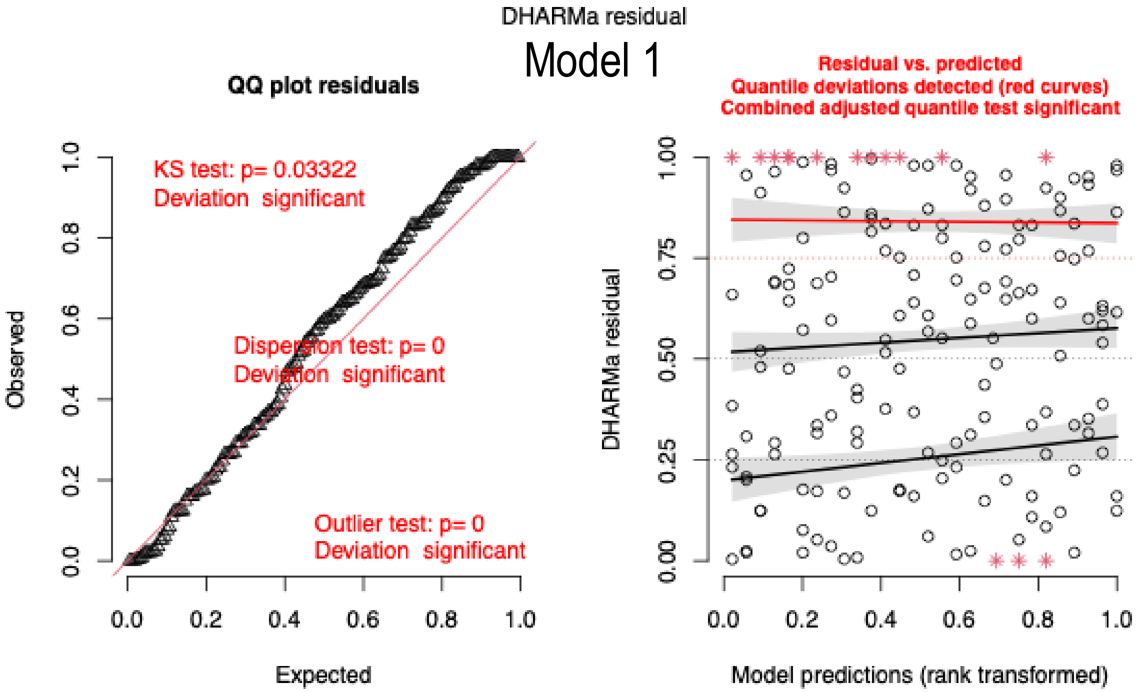


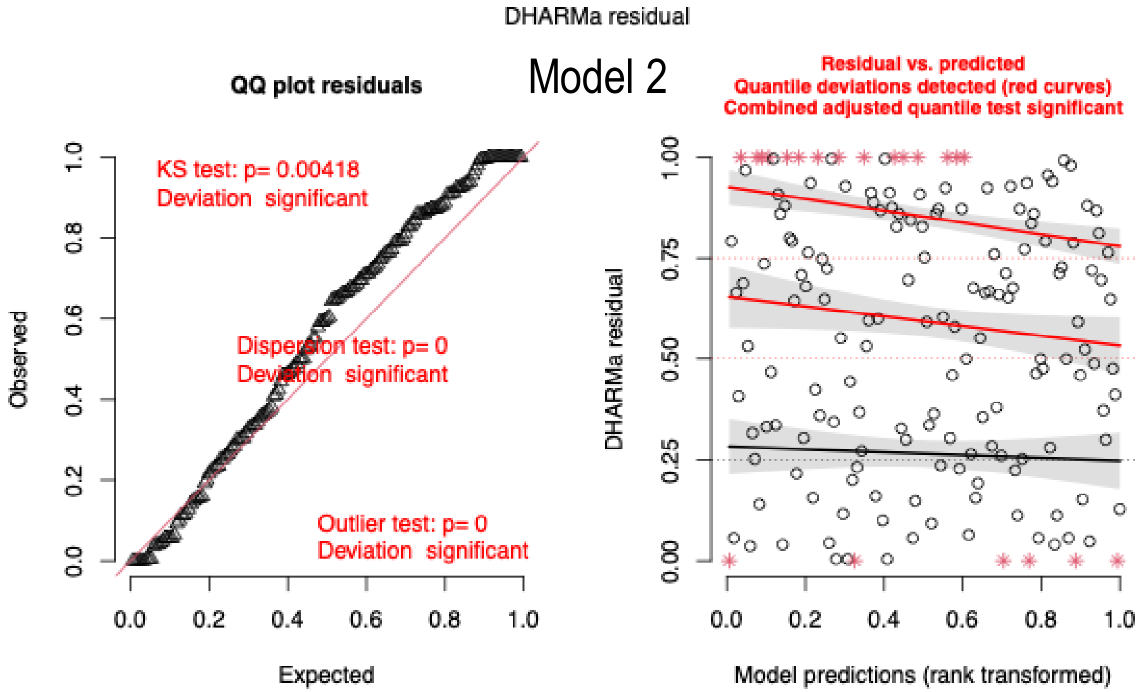
Year effects: females

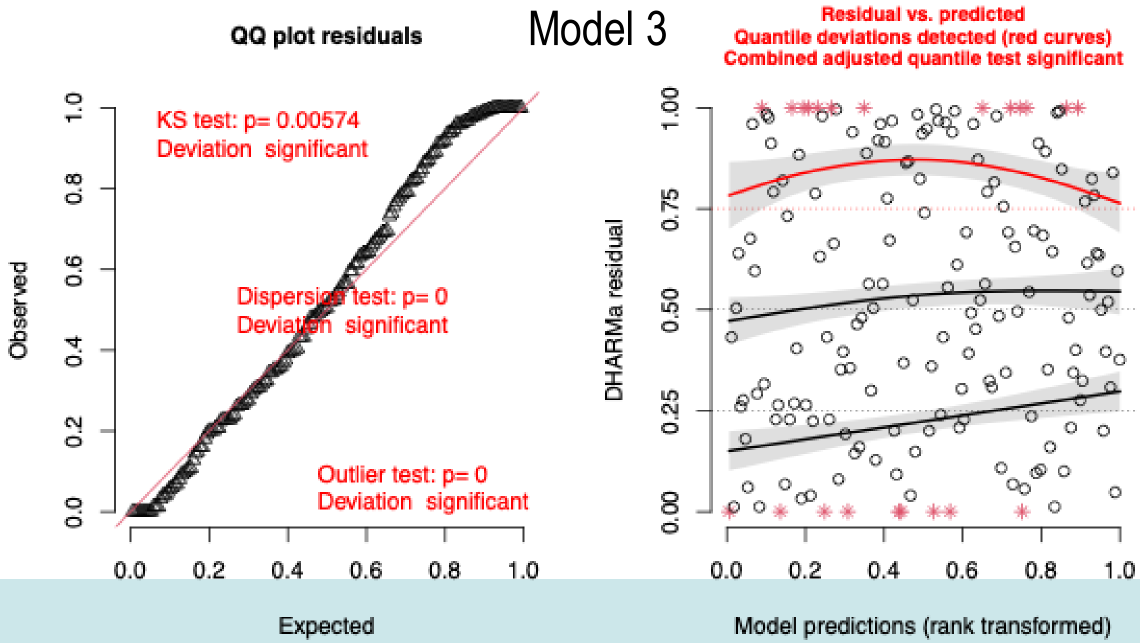




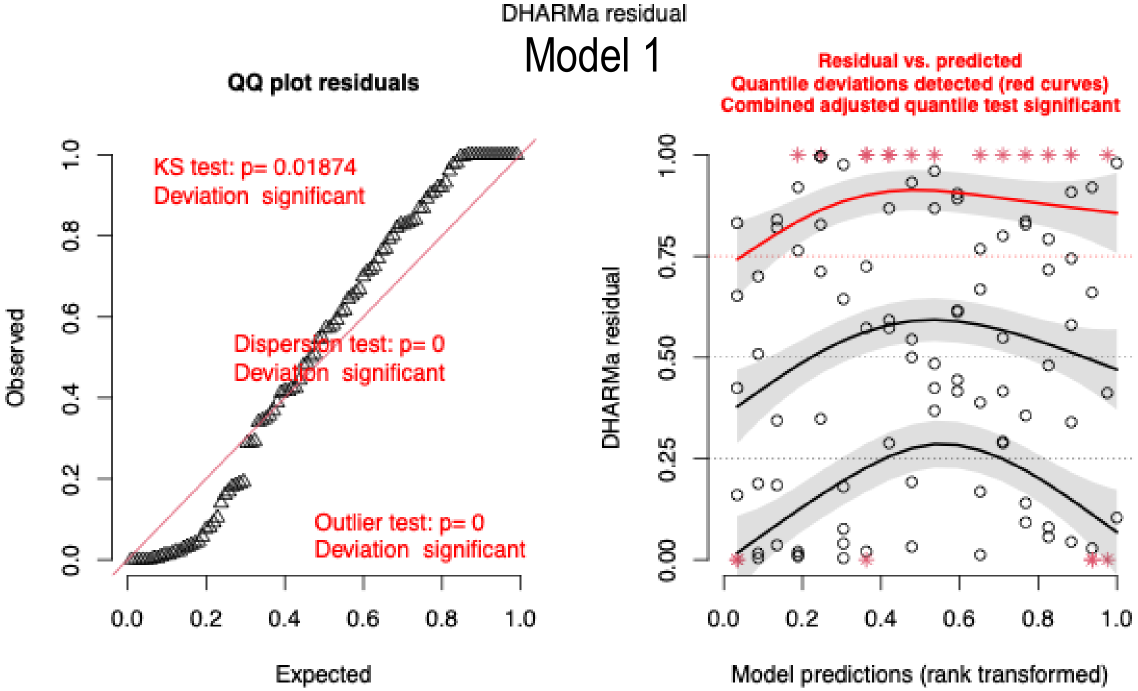
Residuals analysis: males

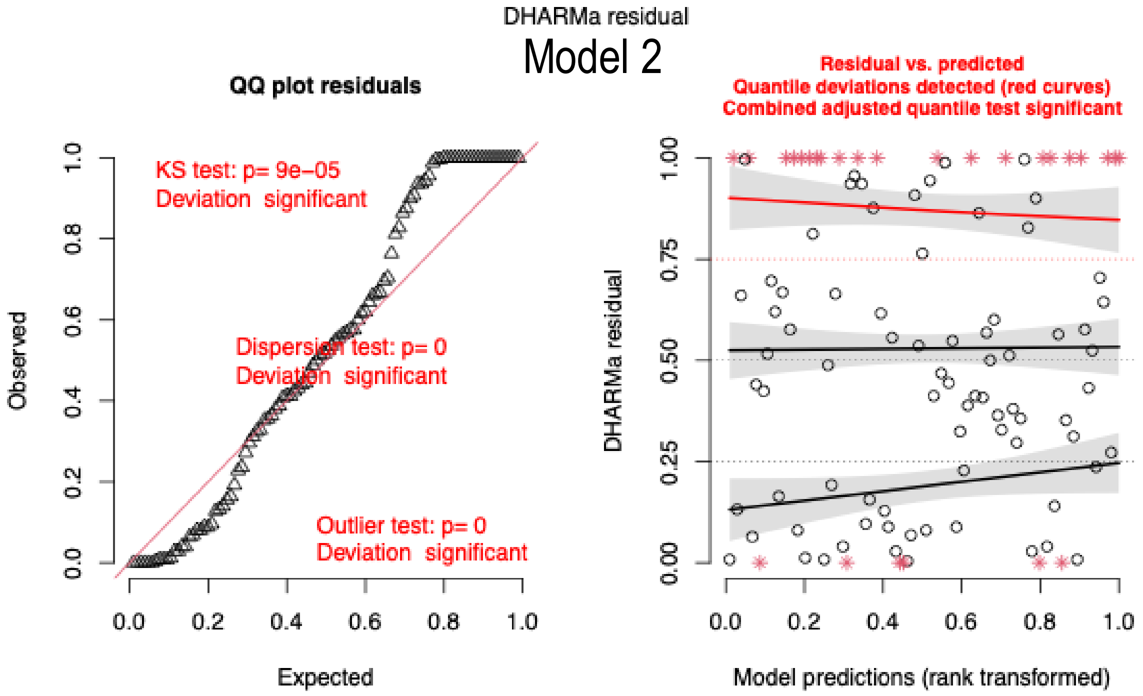


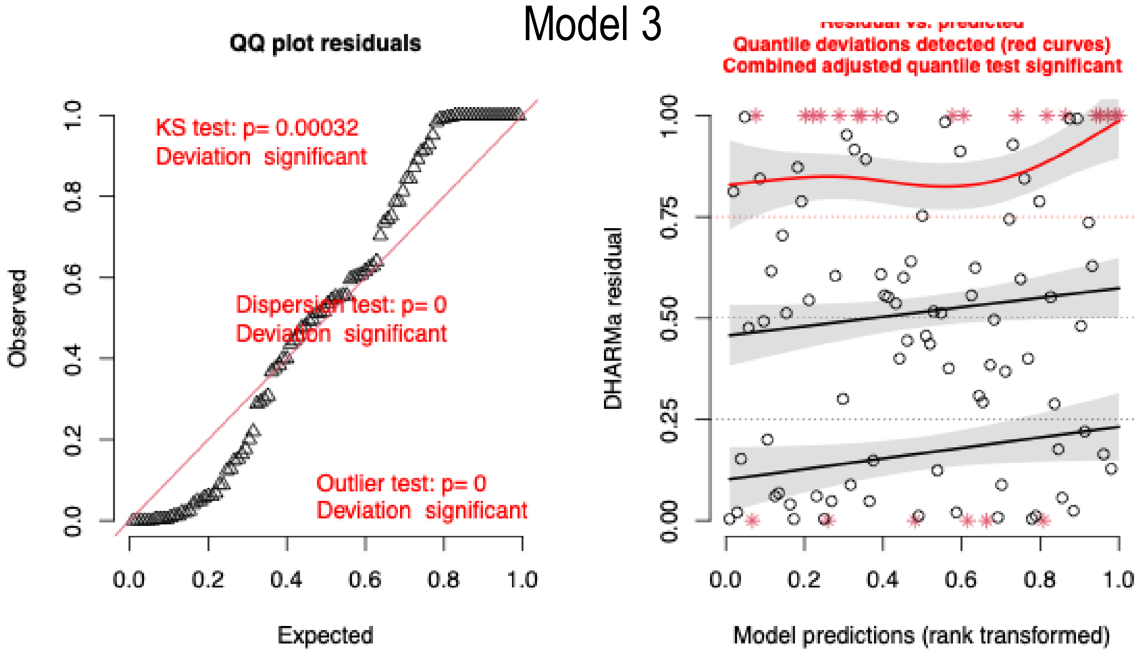




Residuals analysis: females







Statistical analysis

males

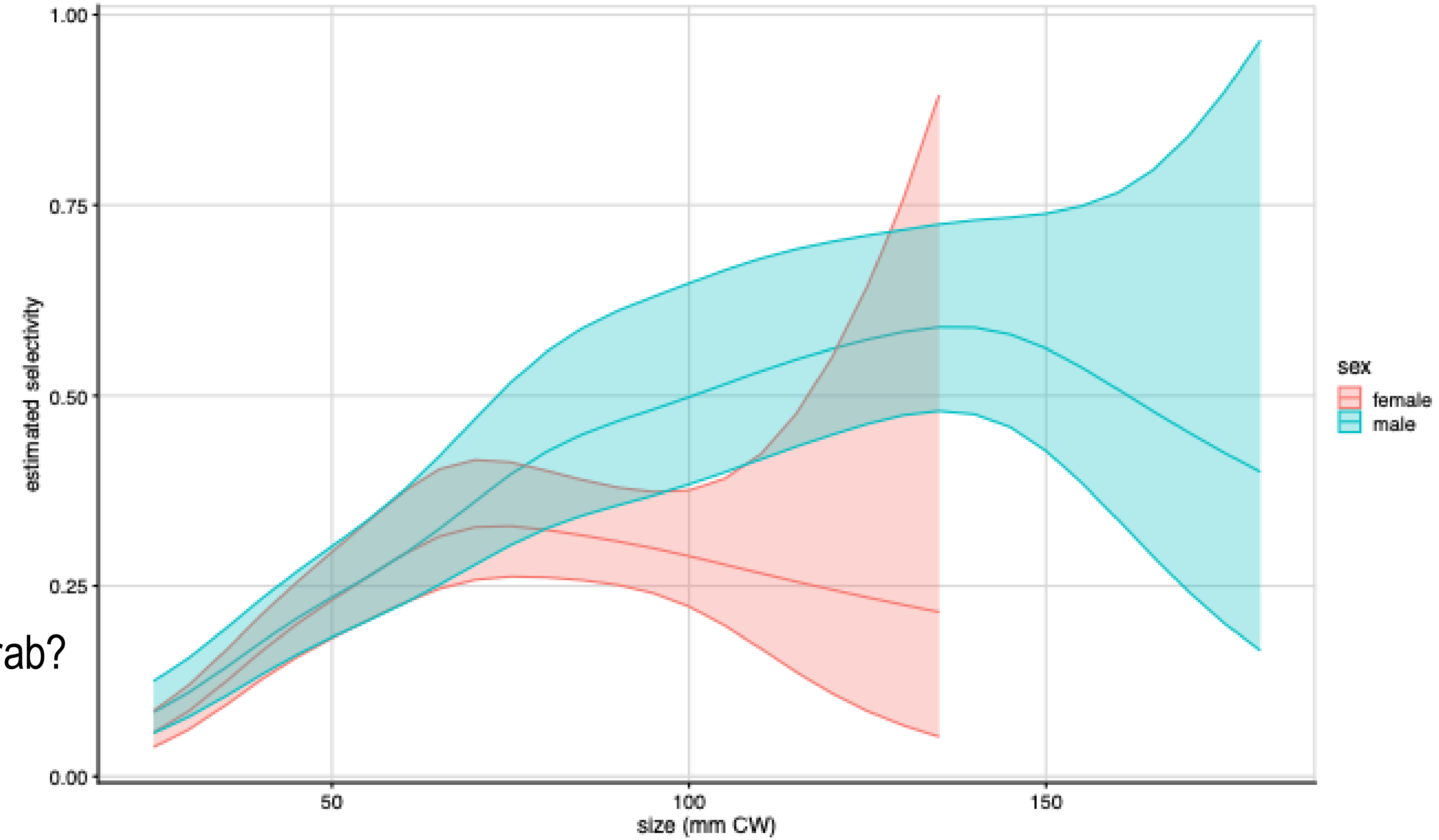
quantity	Model 1	Model 2	Model 3
family	Tweedie p=1.436	Tweedie p=1.537	Tweedie p=1.467
AIC	-330.11	-412.35	-449.83
BIC	-302.83	-352.09	-356.90
deviance	13.42	9.13	6.10
adj. R ^2	0.66	0.75	0.82
df	5.75	14.86	22.29
N(params)	10	34	80
N(obs)	169	169	169

females

quantity	Model 1	Model 2	Model 3
family	Tweedie p=1.642	Tweedie p=1.753	Tweedie p=1.723
AIC	-397.91	-469.71	-484.92
BIC	-376.57	-418.41	-417.40
deviance	7.36	4.55	3.51
adj. R ^2	0.55	0.69	0.71
df	5.70	14.81	20.79
N(params)	10	34	80
N(obs)	104	104	104

Best “survey-level” catchability: Model 3 for both sexes

- Best AIC, adjusted R^2
  - all models had poor residual patterns
- Allows extension to non-study years
  - not valid for Model 2
- Fully-selected catchability (“*q*”)
  - 0.59: males
  - 0.33: females



Big (unanswered) questions:

- Why is catchability low even for large crab?
- Are the dome shapes real?

Haul-level Catchability

Utilizes side-by-side nature of paired haul
Allows incorporation of environmental covariates

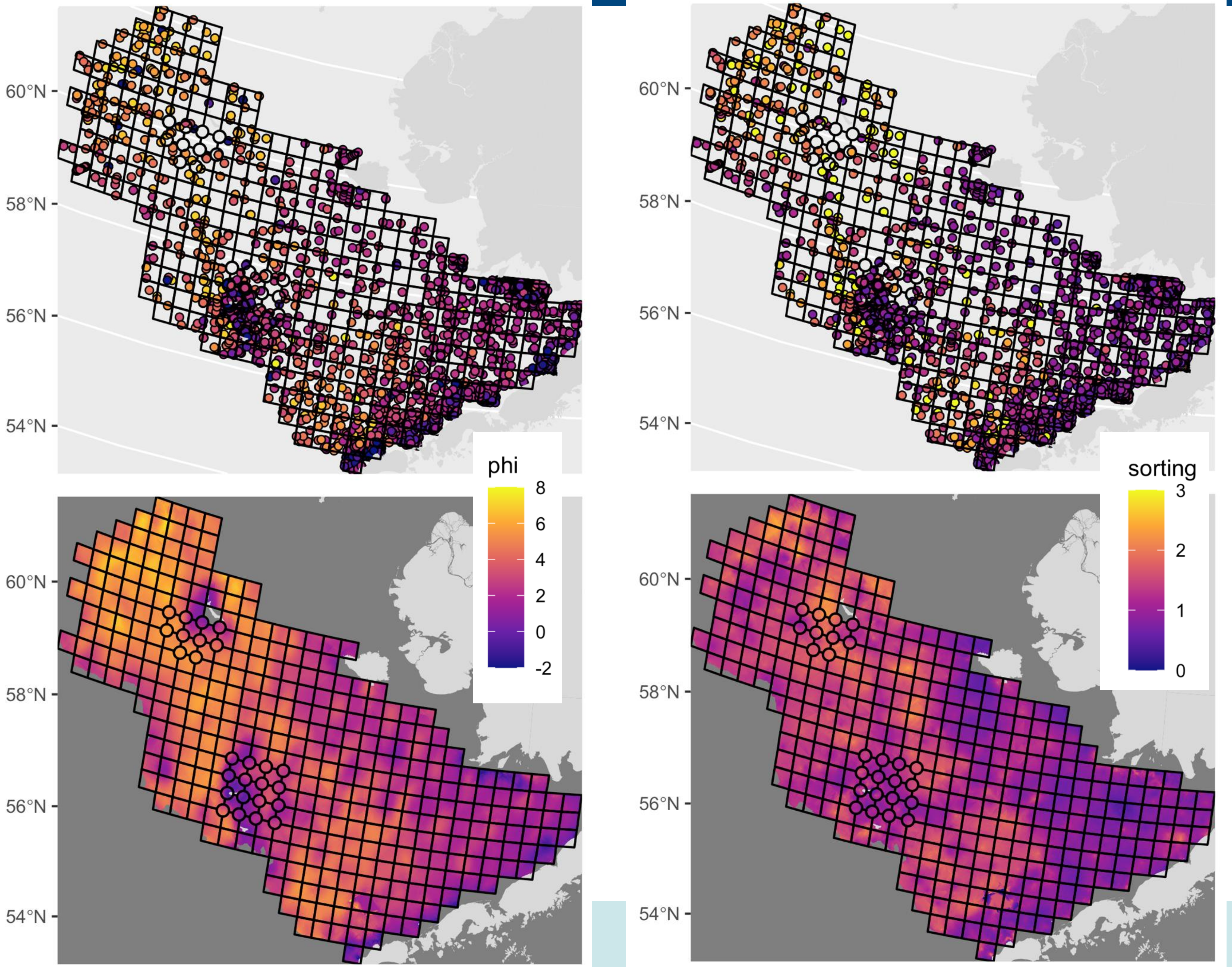
Potential environmental covariates

phi (f): measure of
ln-scale mean grain size

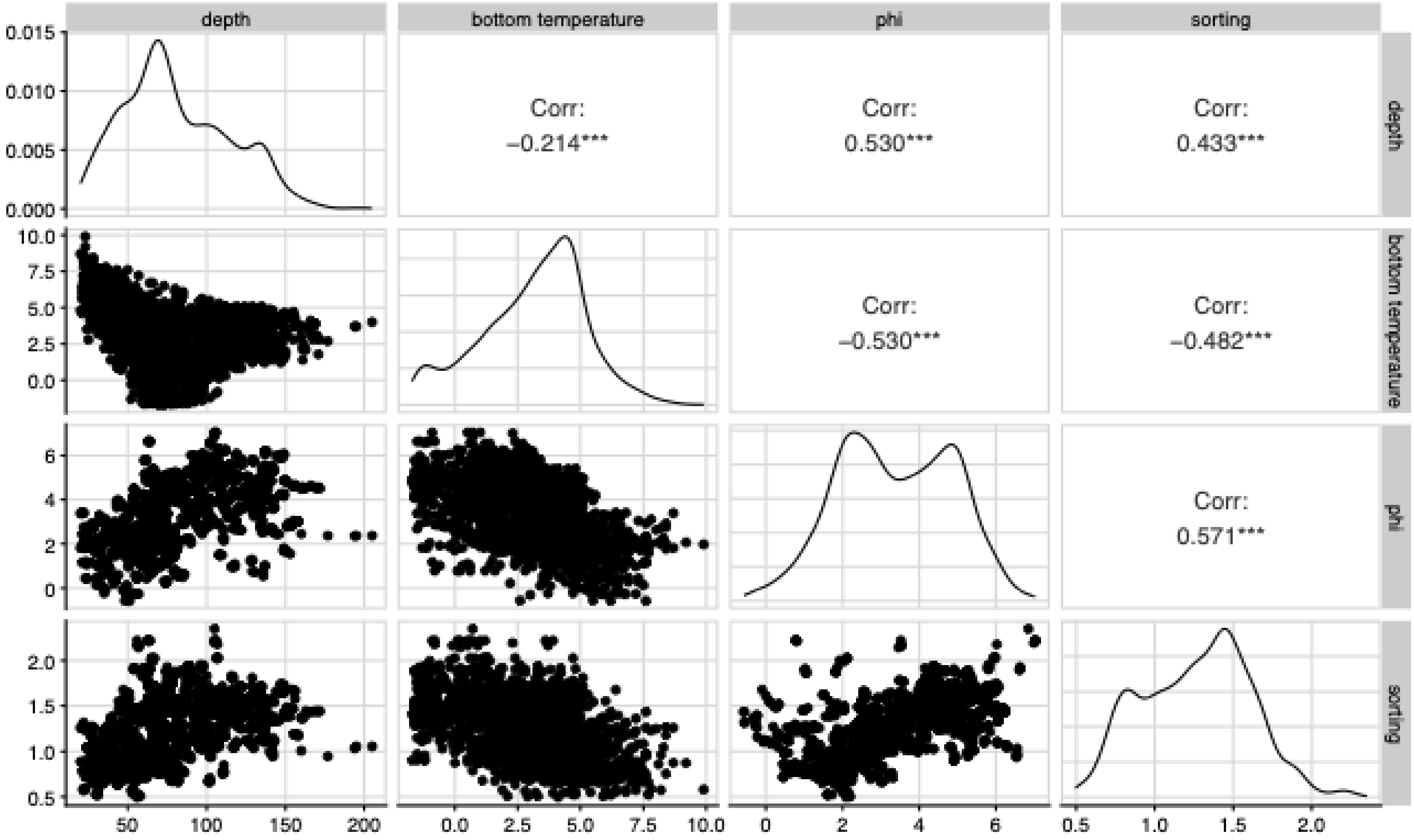
sorting (s): measure of
ln-scale grain size variance

depth (d)

bottom temperature (t)



Potential haul-specific environmental covariates



based on NMFS haul

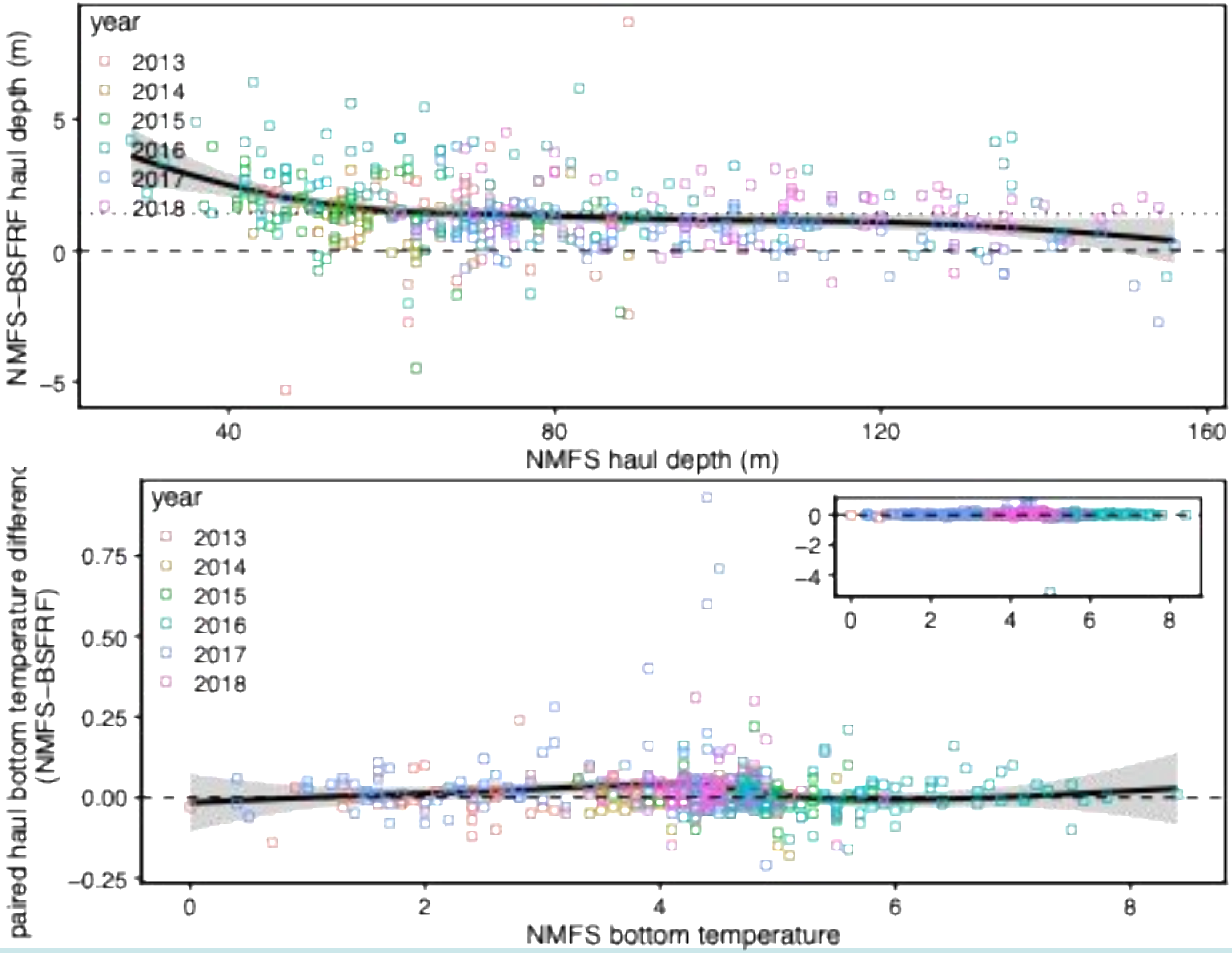
bottom depth (d)

bottom temp. (t)

phi (f): mean ln-scale grain size

sorting coefficient (s): measure of ln-scale grain size variance

Potential environmental covariates



Haul-level catchability: approach 1(a)

$$n_z^U \sim \Pr(c_z^U \cdot A_s^U \cdot \lambda_z^{loc} = \bar{n}_z^U, variance)$$

$$n_z^K \sim \Pr(c_z^K \cdot A_s^K \cdot \lambda_z^{loc} = \bar{n}_z^K, variance)$$

$$n_z^T \sim \Pr((c_z^K \cdot A_s^K + c_z^U \cdot A_s^U) \cdot \lambda_z^{loc} = \bar{n}_z^T, variance)$$

A_s^g : area swept by gear *g*

λ_z^{loc}: mean density at *loc*

$\bar{n}_z^g$: expected num caught by gear *g* at *loc*

$n_z^T = n_z^U + n_z^K$

if crab are distributed completely randomly (i.e., Poisson-distributed), the n_z^U, conditional on n_z^T, is binomially-distributed as $n_z^U | n_z^T \sim Bin(n_z^T, p_z)$ where p_z is the expected proportion of the catch in the gear with unknown catchability and

$$p_z = \frac{\bar{n}_z^U}{\bar{n}_z^K} = \frac{c_z^U \cdot A_s^U \cdot \lambda_z^{loc}}{c_z^U \cdot A_s^U \cdot \lambda_z^{loc} + c_z^K \cdot A_s^K \cdot \lambda_z^{loc}} = \frac{c_z^U}{c_z^U + c_z^K \cdot \frac{A_s^K}{A_s^U}} = \frac{r_z}{1 + r_z \cdot \frac{A_s^K}{A_s^U}}$$

where $r_z = \frac{c_z^U}{c_z^K} \equiv$ the selectivity ratio ($= c_z^U$ under the assumption $c_{z,h}^{BSFRF} = 1$)

or, rearranging a bit: $\ln\left(\frac{p_z}{1 - p_z}\right) = ln\left(r_z \cdot \frac{A_s^U}{A_s^K}\right) = ln(r_z) + \ln\left(\frac{A_s^U}{A_s^K}\right) = logit(p_z)$

Haul-level catchability: approach 1(b)

rearranging that last equation a bit

$$logit(p_z) = \ln\left(\frac{p_z}{1-p_z}\right) = ln\left(r_z \cdot \frac{A_s^U}{A_s^K}\right) = ln(r_z) + \ln\left(\frac{A_s^U}{A_s^K}\right)$$

which suggests using a **binomial model with a logistic link** to estimate a smooth function of size and potential environmental covariates for $ln(r_z)$ as

$$\tilde{p}_{z,h} = \frac{n_{z,h}^U}{n_{z,h}^U + n_{z,h}^K} \quad \text{and} \quad logit\left(E\left[\tilde{p}_{z,h}\right]\right) \sim f(z, d_h, t_h, f_h, s_h) + \ln\left(\frac{A_h^U}{A_h^K}\right)$$

such that (under the assumption $c_{z,h}^{BSFRF} = 1$)

$$c_{z,h}^{NMFS} = r_{z,h} = \exp\left(f(z, d_h, t_h, f_h, s_h)\right)$$

- Similar to Somerton's (2013) approach for snow crab

Somerton et al. 2013.
https://doi.org/10.1139/cjfas-2013-0100

Haul-level catchability: approach 2

$$n_z^U \sim \Pr\left(c_z^U \cdot A_s^U \cdot \lambda_z^{loc} = \bar{n}_z^U, variance\right) \Rightarrow \overline{CPUE_z^U} = \bar{n}_z^U / A_s^U = c_z^U \cdot \lambda_z^{loc}$$

$$n_z^K \sim \Pr(c_z^K \cdot A_s^K \cdot \lambda_z^{loc} = \bar{n}_z^K, variance) \Rightarrow \overline{CPUE_z^K} = \bar{n}_z^K / A_s^K = c_z^K \cdot \lambda_z^{loc}$$

$$\tilde{r}_{z,h} = \frac{c_{z,h}^U}{c_{z,h}^K} = \frac{c_{z,h}^U \cdot \lambda_{z,h}^{loc}}{c_{z,h}^K \cdot \lambda_{z,h}^{loc}} = \frac{\frac{\bar{n}_z^U}{A_h^U}}{\frac{\bar{n}_z^K}{A_h^K}} = \frac{\overline{CPUE_{z,h}^U}}{\overline{CPUE_{z,h}^K}}$$

so use $r_{z,h} = \frac{CPUE_{z,h}^U}{CPUE_{z,h}^K}$ as observations

- model haul-level $r_{z,h}$ as Tweedie-distributed smooth function of size and local environmental covariates

$$r_{z,h} \sim Tw(\mu_{z,h}, \phi) \quad V(r_{z,h}) = \beta \cdot \mu_{z,h}^{\phi}$$

- with a log-link function (and BSFRF assumption)

$$\ln\left(\mu_{z,h} = E\left[r_{z,h}\right]\right) = f(z,t,d,f,s) \Rightarrow \tilde{c}_{z,h}^{NMFS} = E\left[r_{z,h}\right] = \exp\left(f(z, d_h, t_h, f_h, s_h)\right)$$

- Similar to Kotwicki et al. (2017)

Kotwicki et al. 2017.
https://doi.org/10.1016/j.fishres.2017.02.012

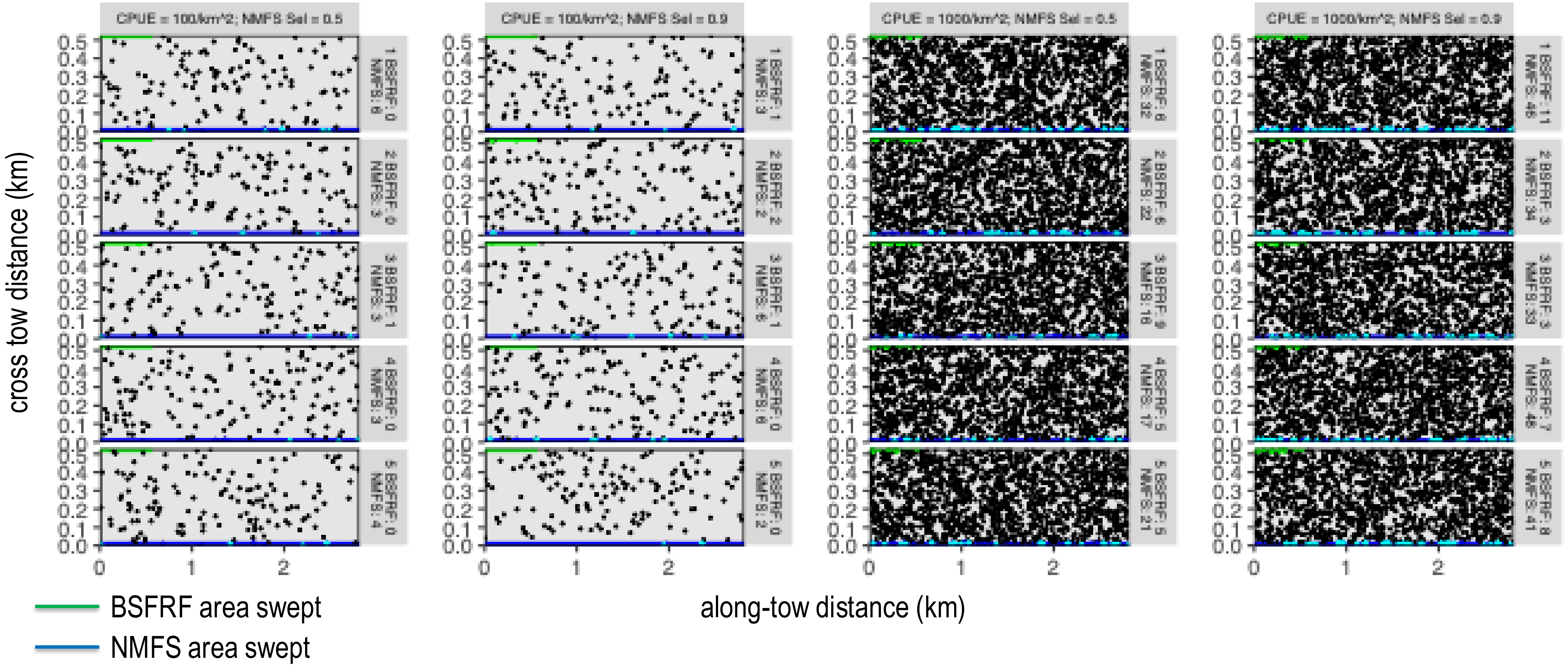
The Devils in the Details

- hauls are random samples of local abundance
  - best case: truly random spatial distributions (i.e., Poisson-distributed)
  - different areas swept by gears imply
    - expected numbers caught are different
    - associated variances are different

- "side-by-side" hauls offset by 0.1-0.2 nmi
  - small-scale patchiness? <- affects sampling distribution

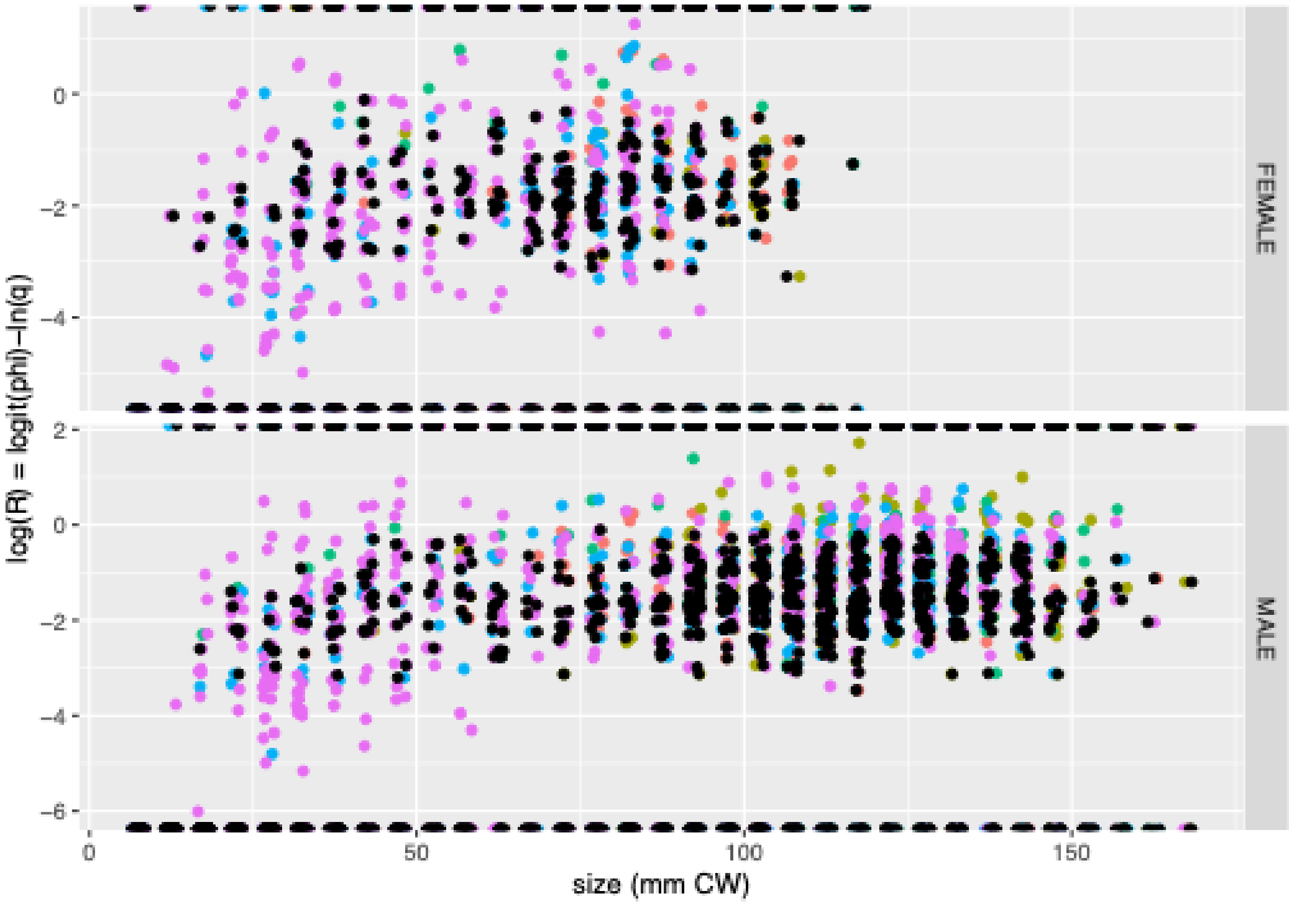
- sampling distribution of catch ratios??

SBS sampling randomly-dispersed crab

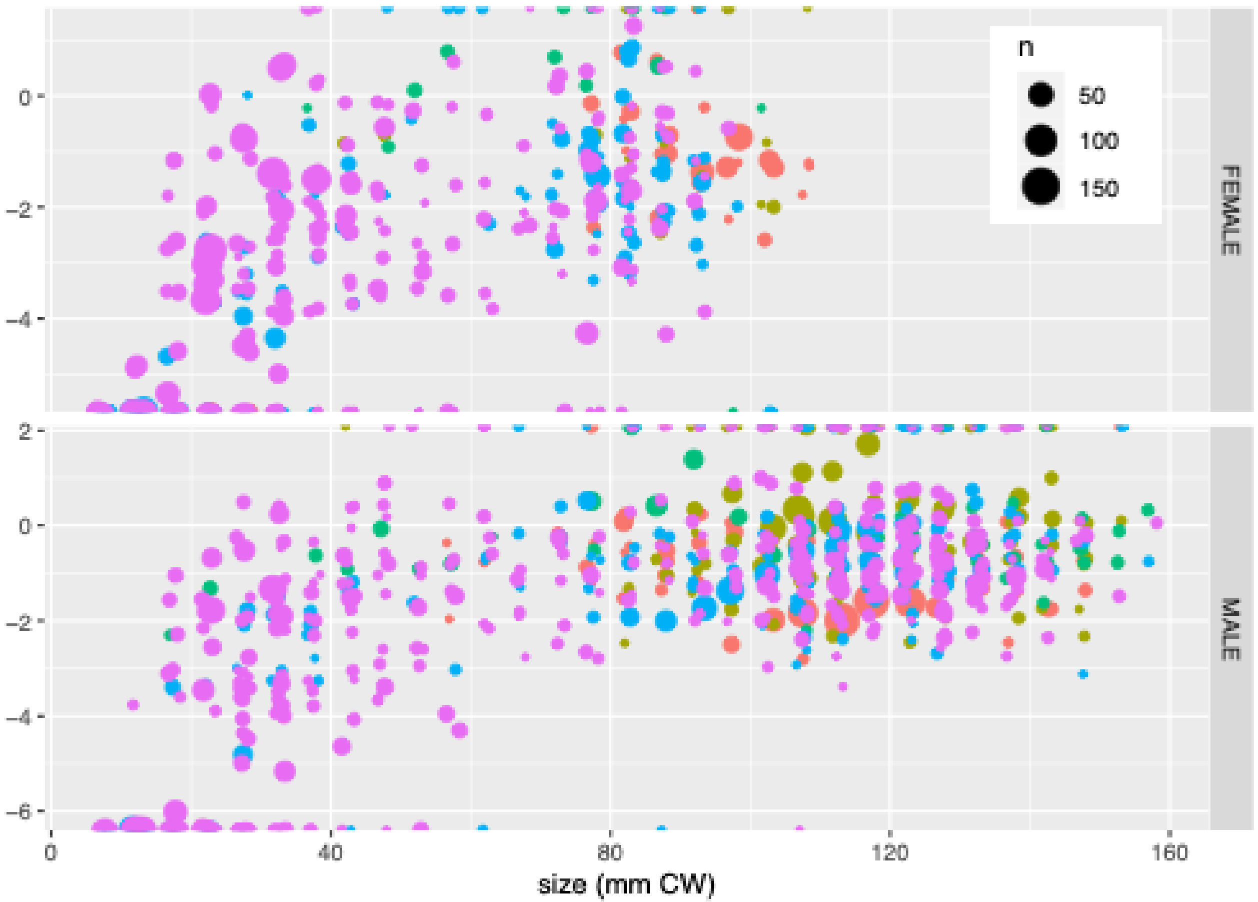


Paired haul data

“raw”

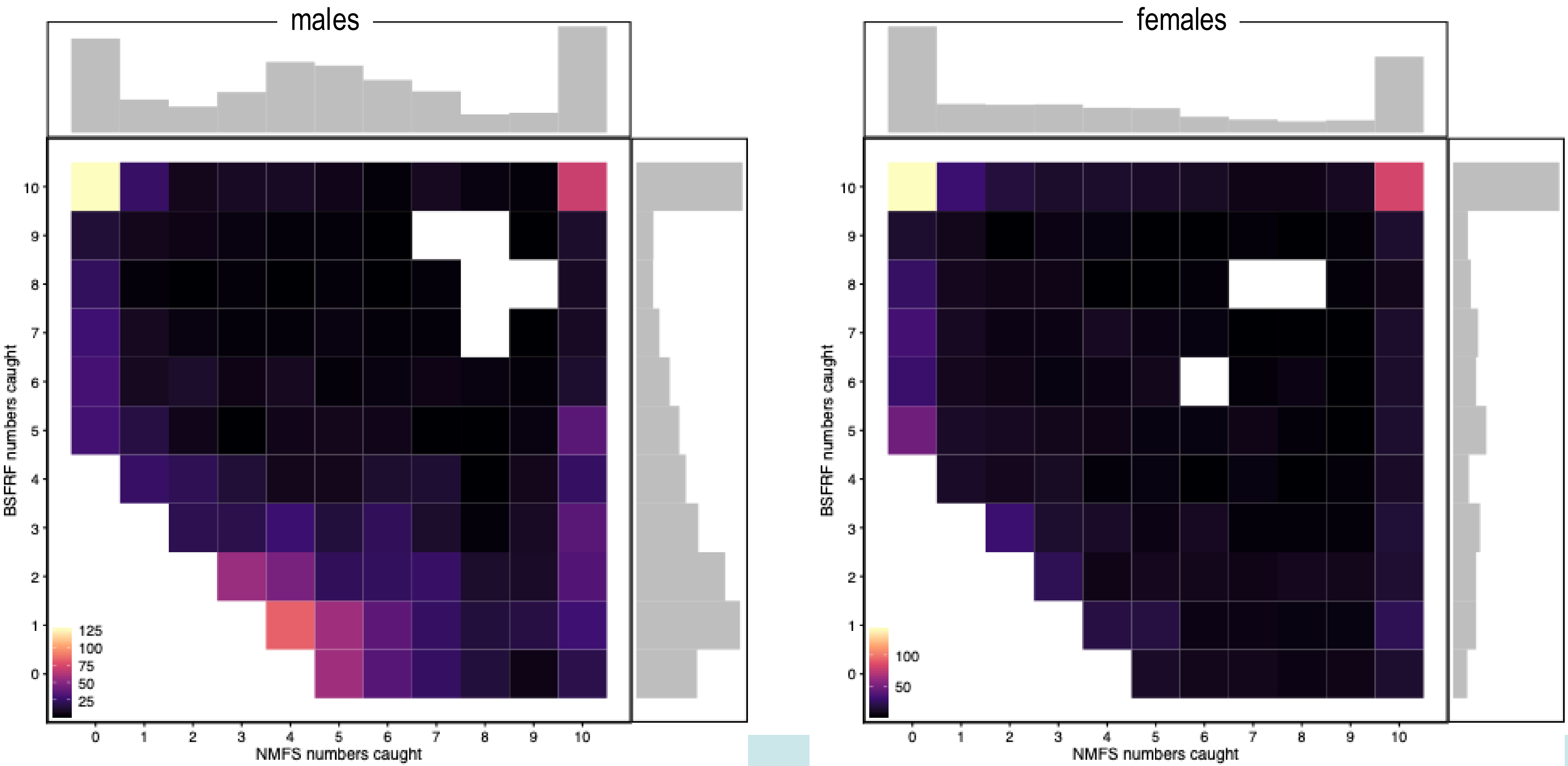


“trimmed”



- removed observations with n_T<5, missing environmental data

Joint histograms of numbers caught in paired tows



GAM models with environmental covariates

MGCV "gam" model formulae

link(response) ~ f(z,d,t,f,s) = intercept + ti(z,bs="ts") +
ti(d,bs="ts") + ti(t,bs="ts") + ti(f,bs="ts") + ti(s,bs="ts") +
ti(z,d,bs="ts") + ti(z,t,bs="ts") + ti(z,f,bs="ts") + ti(z,s,bs="ts");

Binomial distributions with logit link function

$response_{haul,z}$ = proportion of crab caught in 5-mm size bin by NMFS gear in paired haul
$weight_{haul,z}$ = total number in 5-mm size bin caught in paired haul
$offset_{haul,z}$ = ln-scale ratio of areas swept in paired haul

Tweedie distributions with log link function

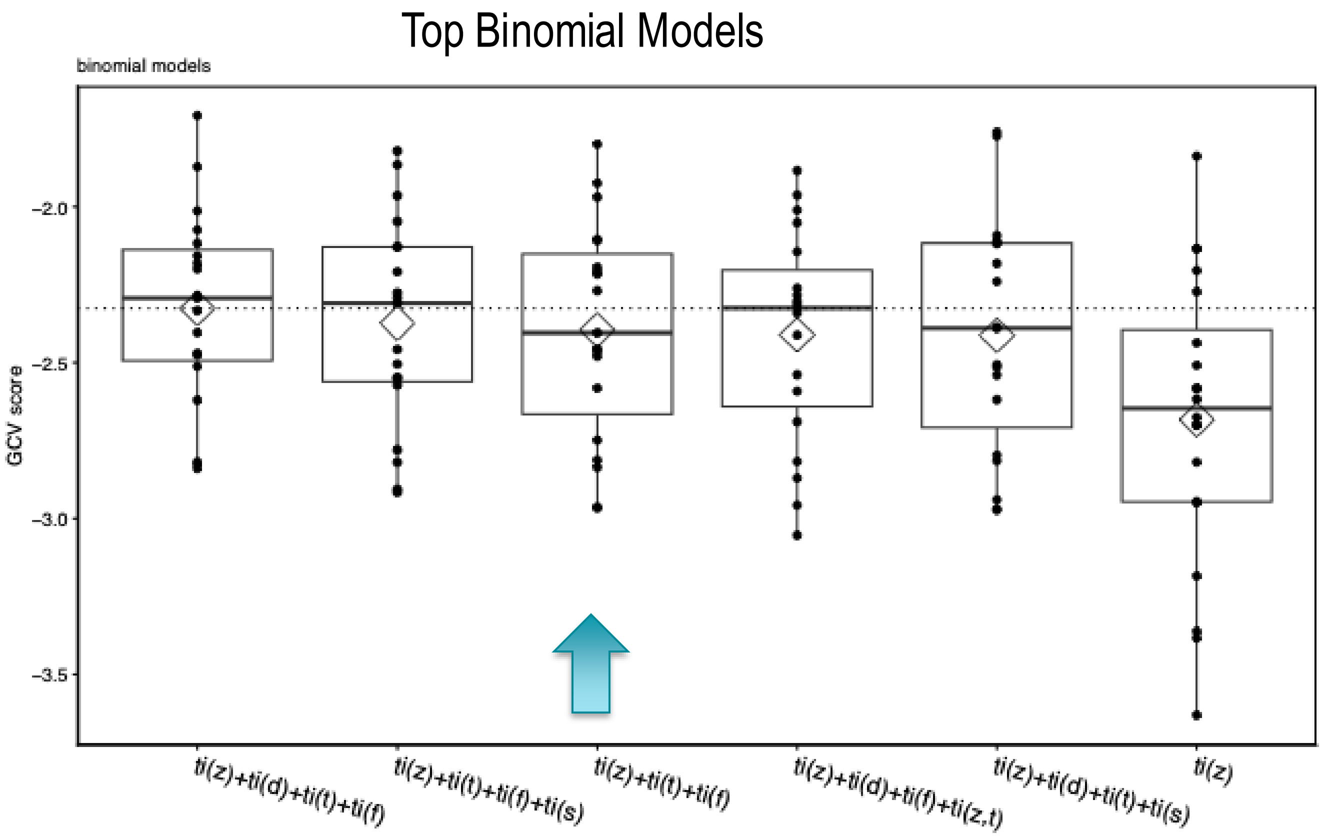
$response_{haul,z}$ = ratio of CPUEs for caught in 5-mm size bin in paired haul
$weight_{haul,z}$ = none
$offset_{haul,z}$ = none

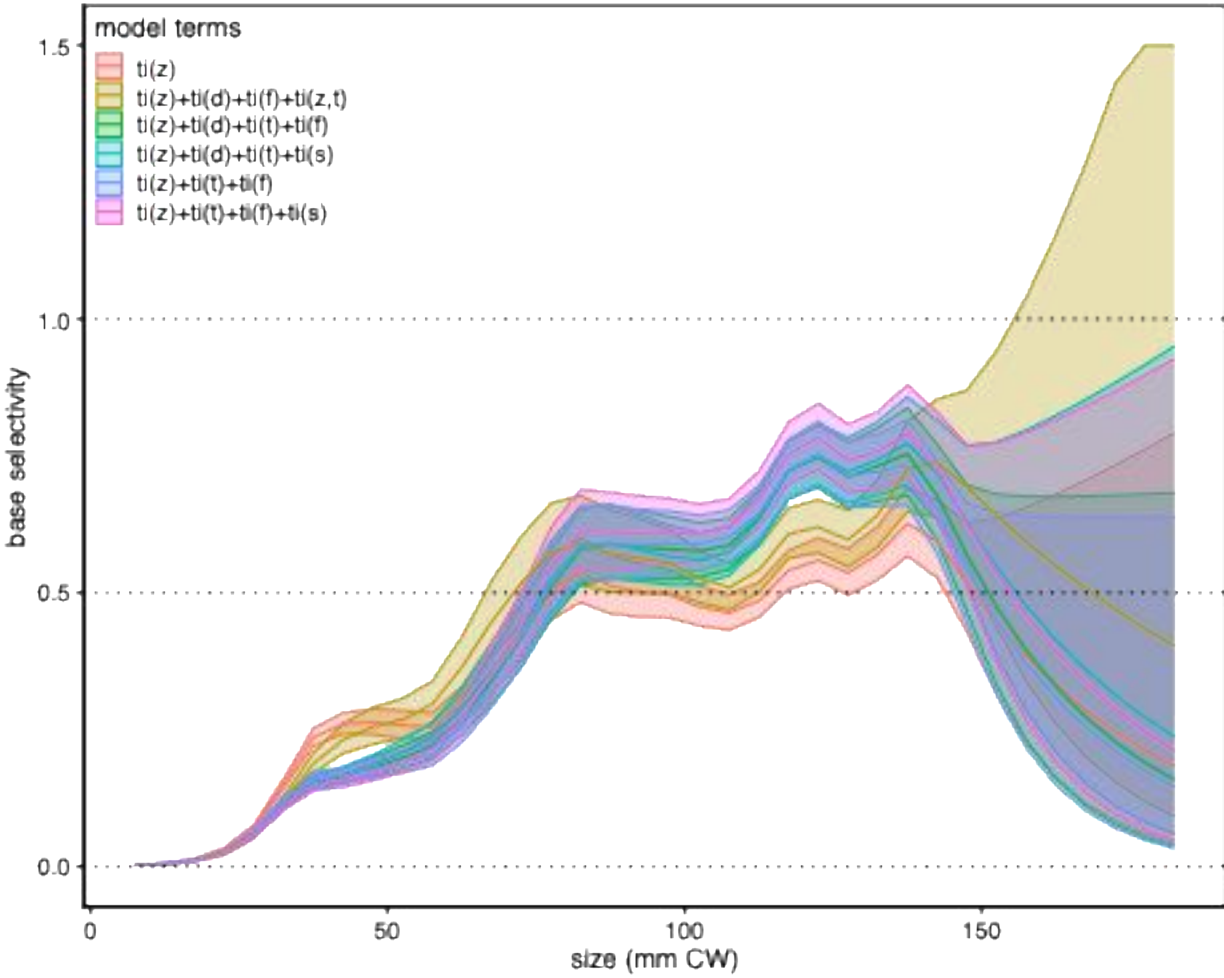
Model evaluation and selection

- used mgcv "gam" function to evaluate models
- for each distribution
  - evaluated every combination (256) of "intercept + ti(z)" with the environmental covariate terms
  - performed k-fold cross validation using 20 folds. for each fold:
    - randomly select 95% of observations as "training set"
    - fit (each) model
    - use fitted model to predict observations in "testing set" (i.e., remaining 5% of observations)
    - calculate predictive ability score by evaluating mean likelihood of predicted responses
  - selected "best" model based on (Yates et al. 2022)
    - mean prediction score,
    - absence of significant concurvity across model terms
    - simplicity of model

Yates et al. 2022. Cross validation for model selection: A review with examples from ecology. Ecological Monographs. https://doi.org/10.1002/ecm.1557

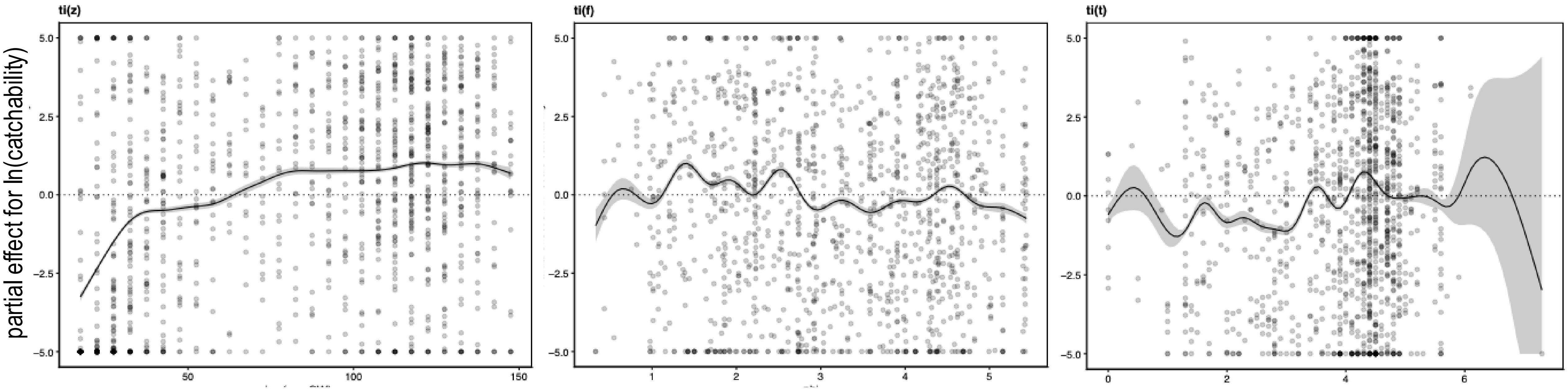
Binomial model results for males

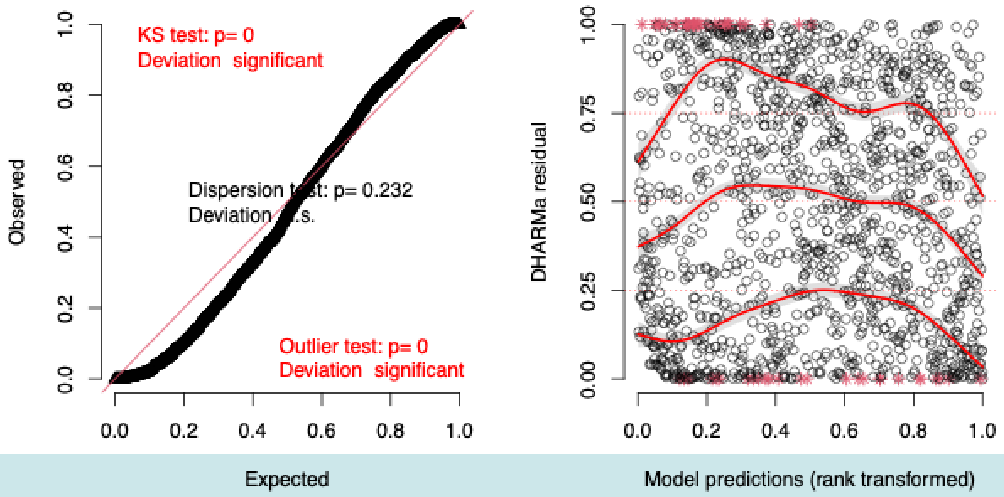




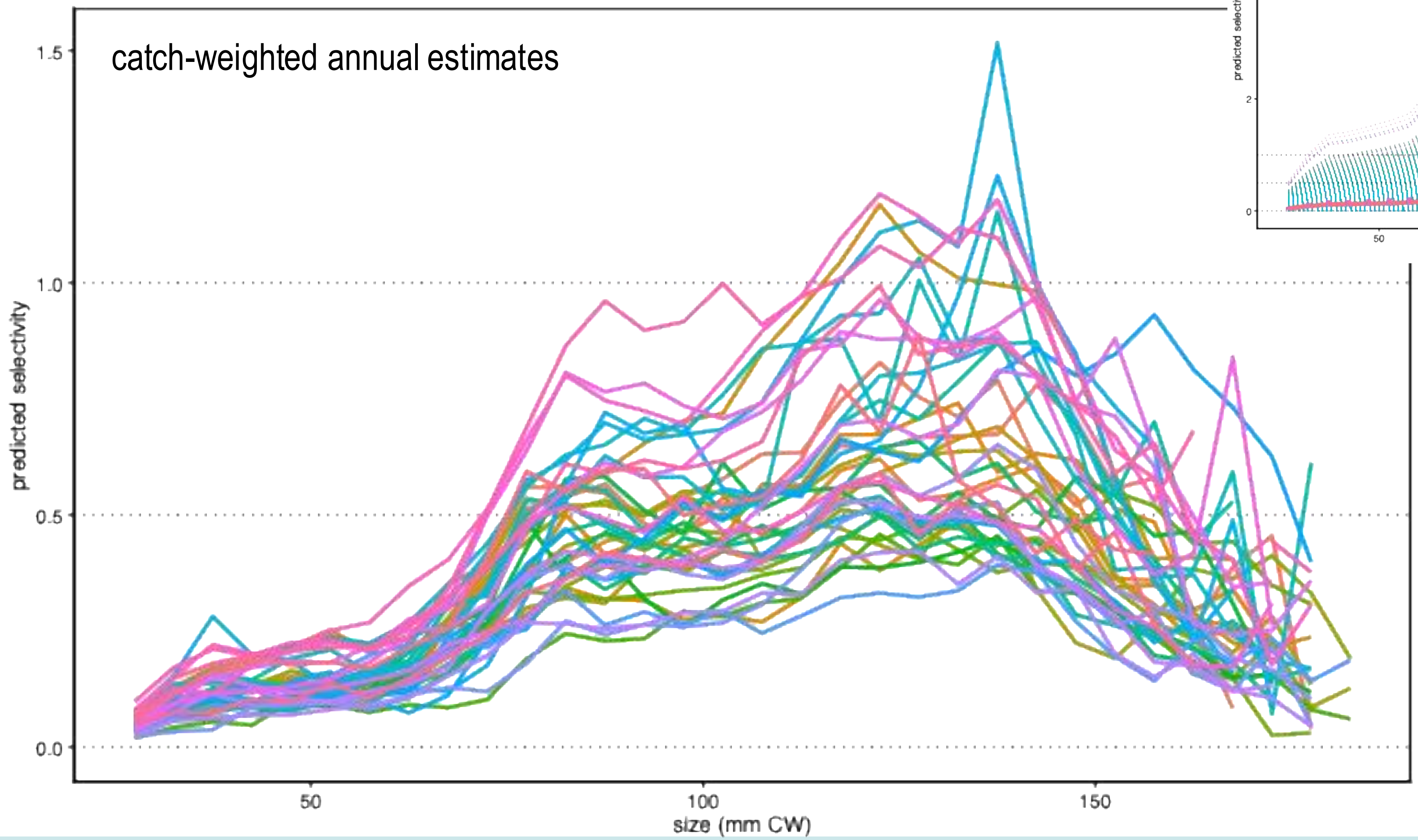
Best binomial model has covariate terms for temperature (*t*) and grain size (*f*)

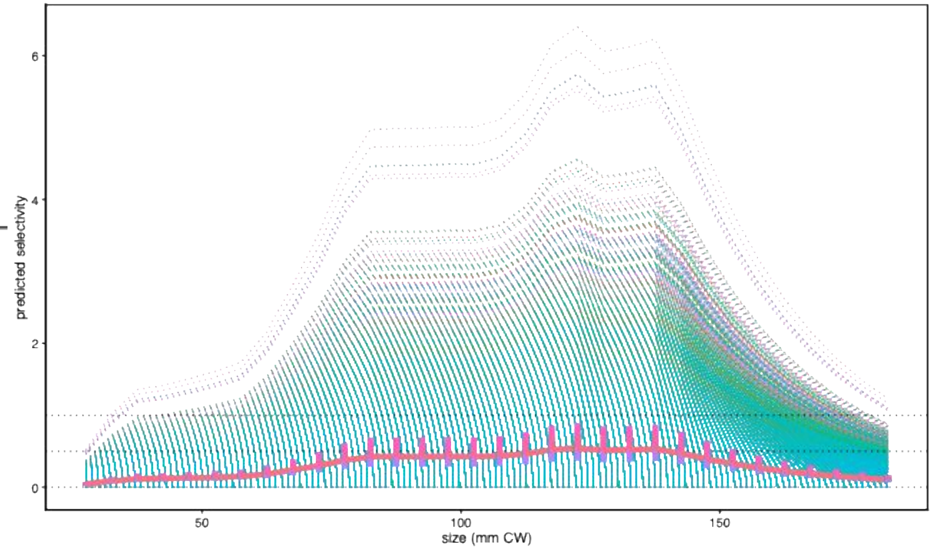
“Best” binomial model for males results





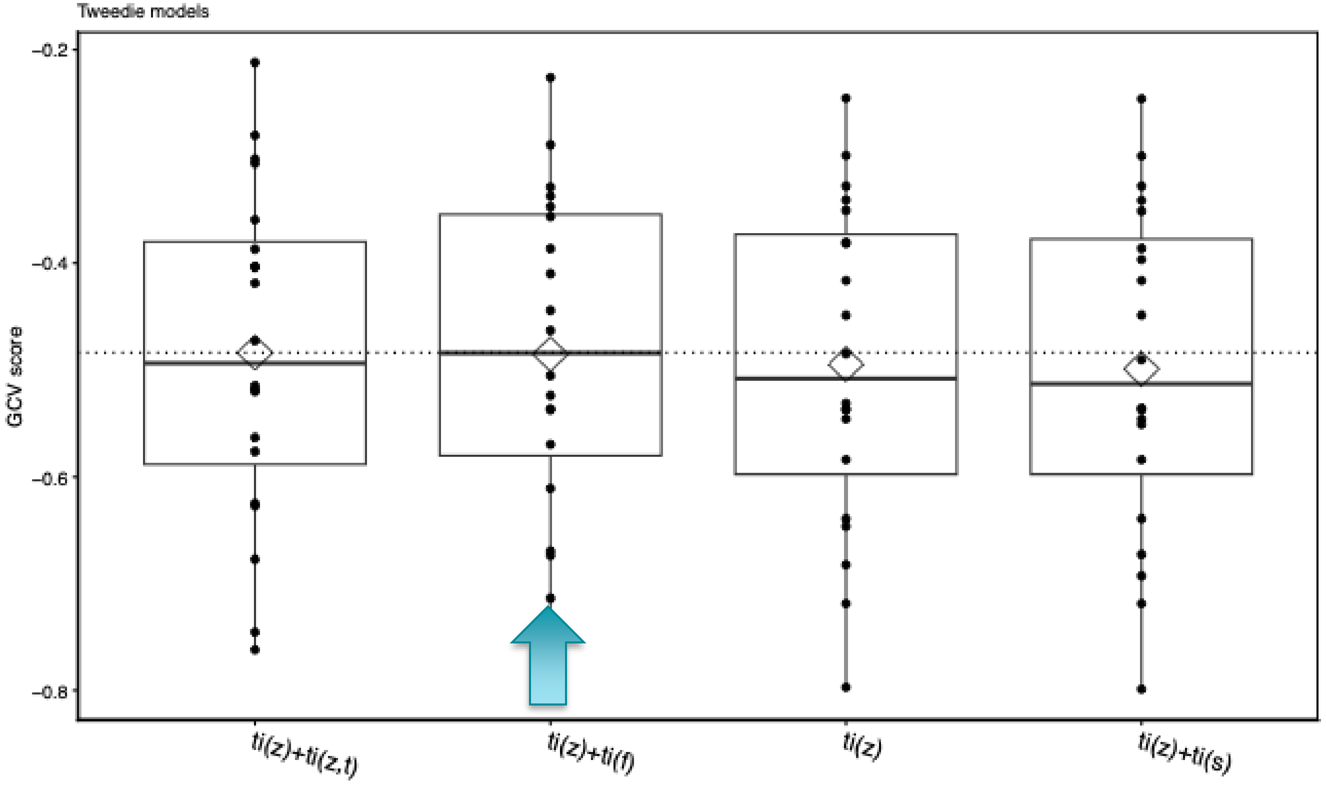
“Best” binomial model for males results



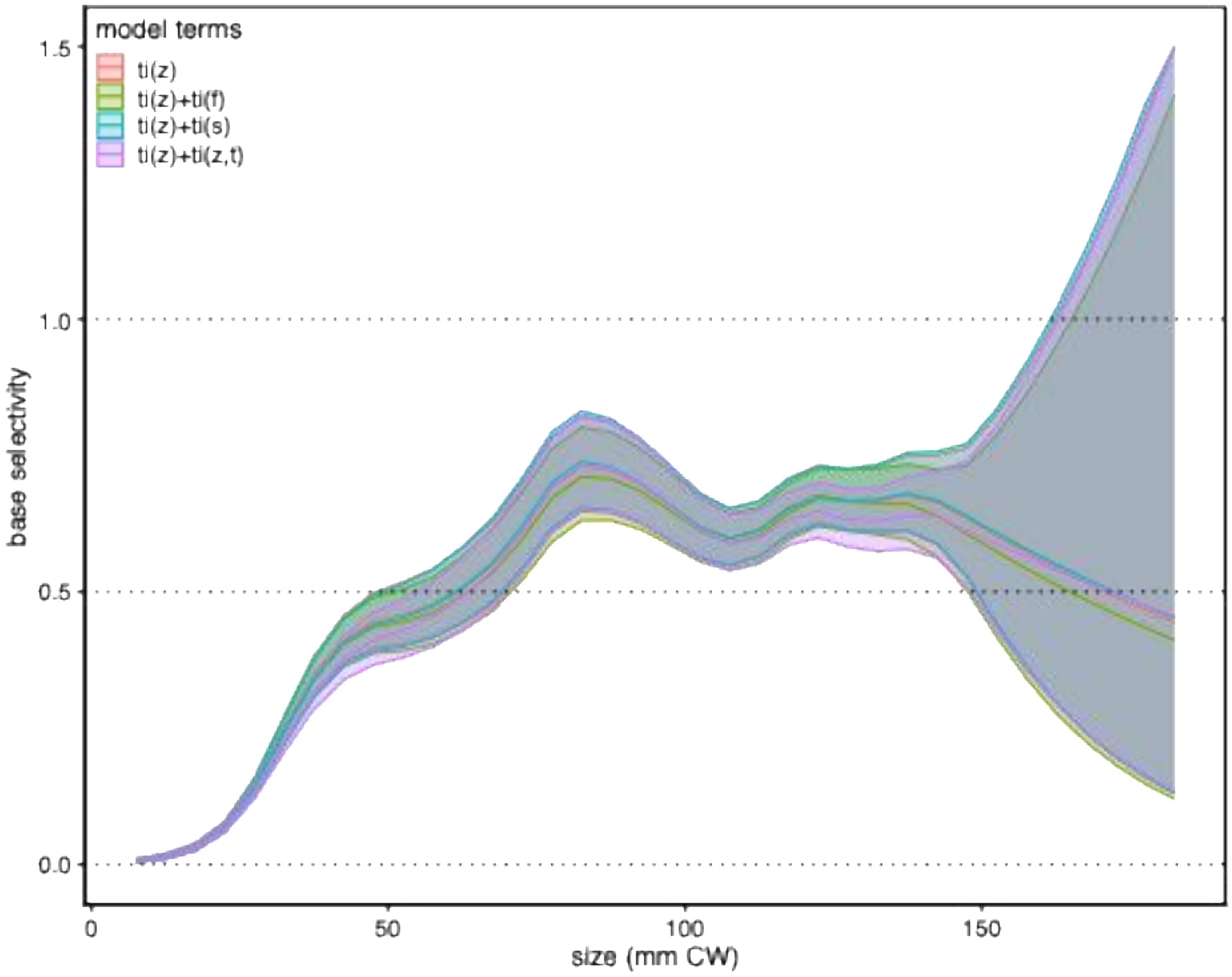


Tweedie model results for males

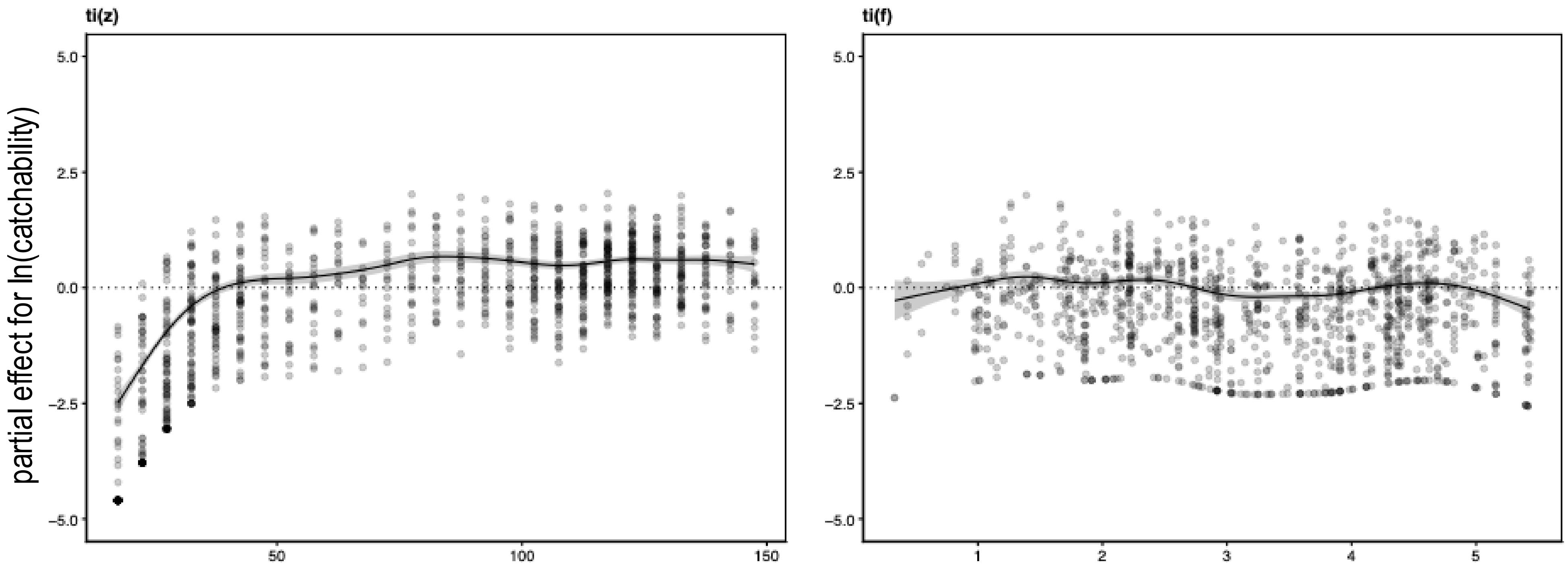
Top Tweedie Models

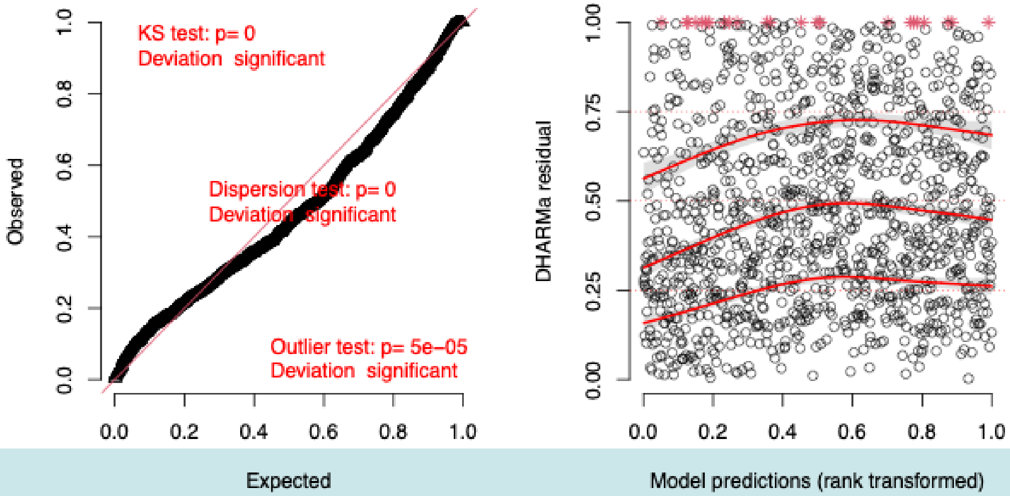


Best Tweedie model has covariate term for grain size (*f*)

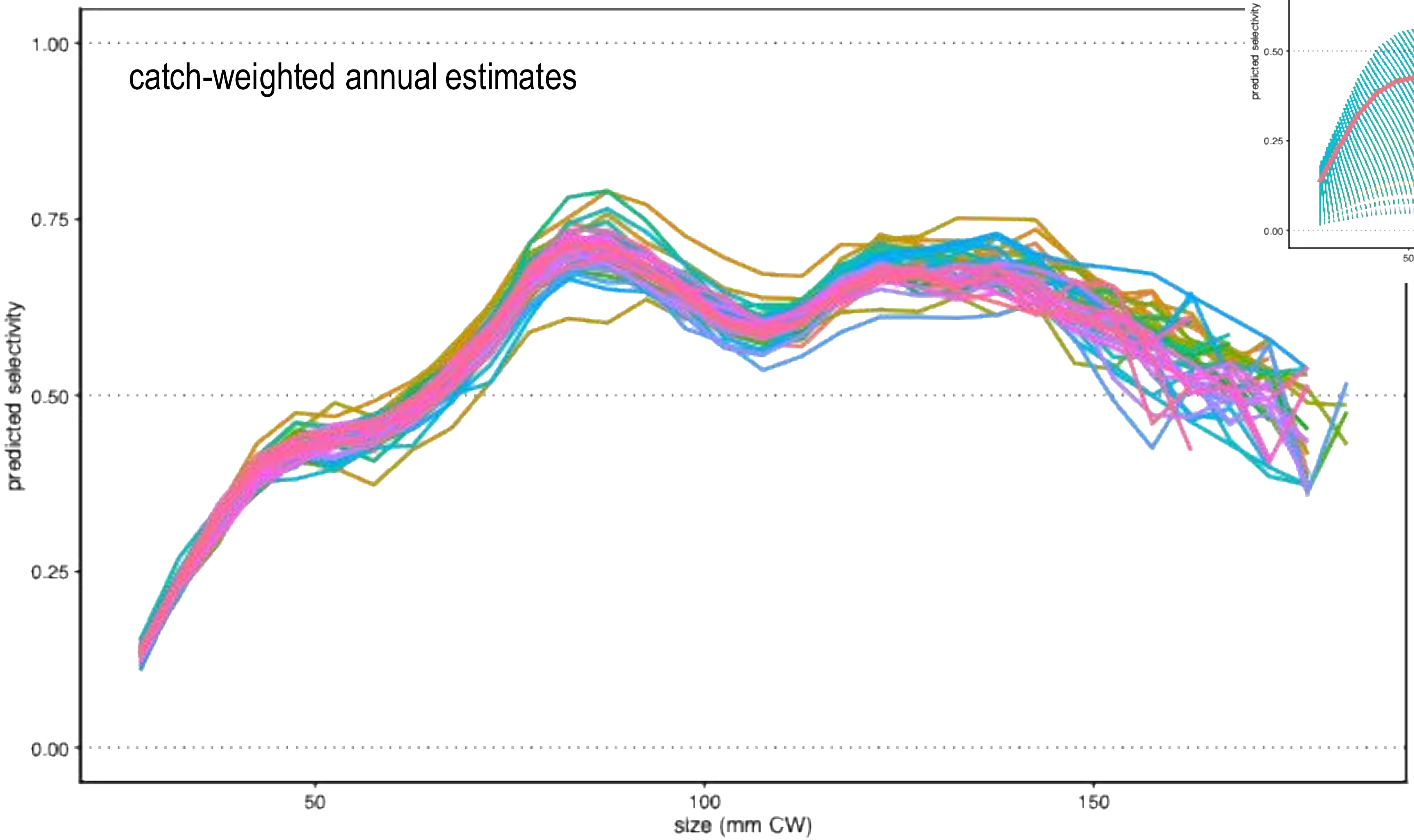


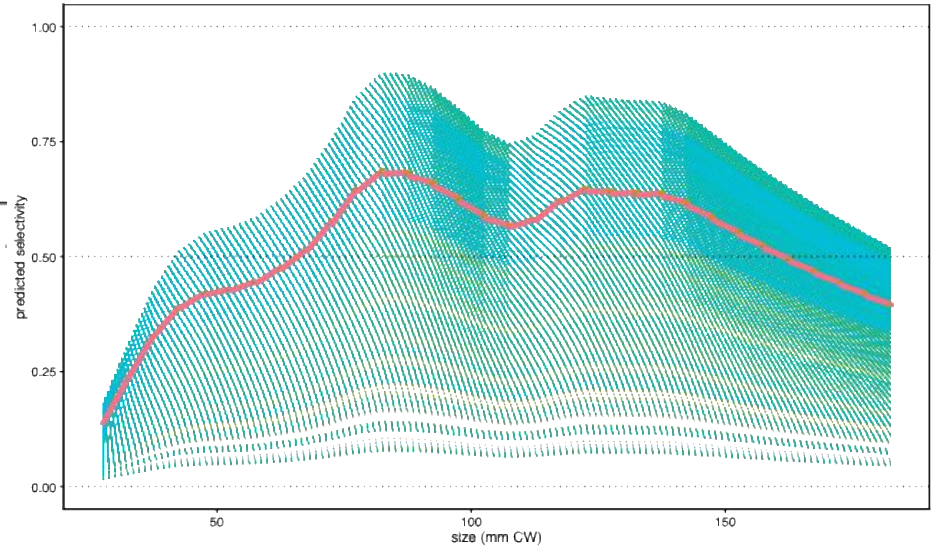
“Best” Tweedie model for males results



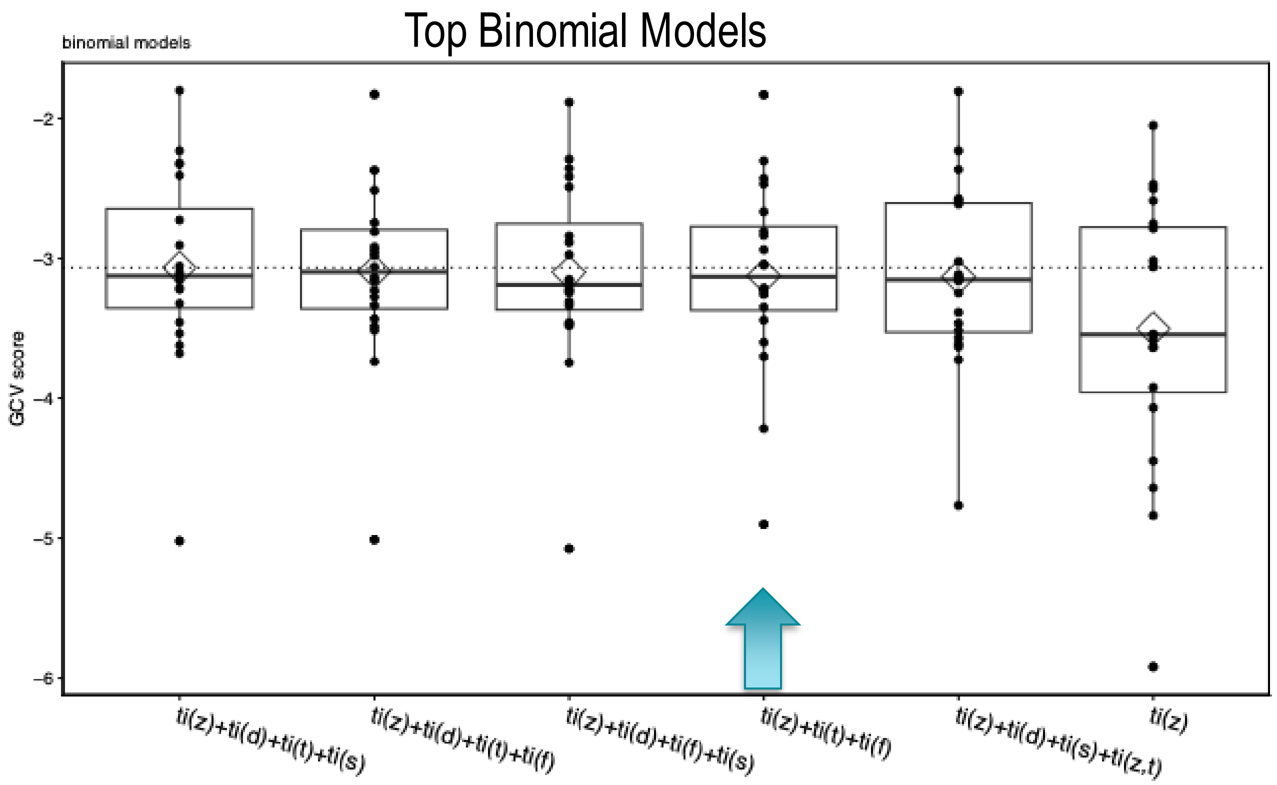


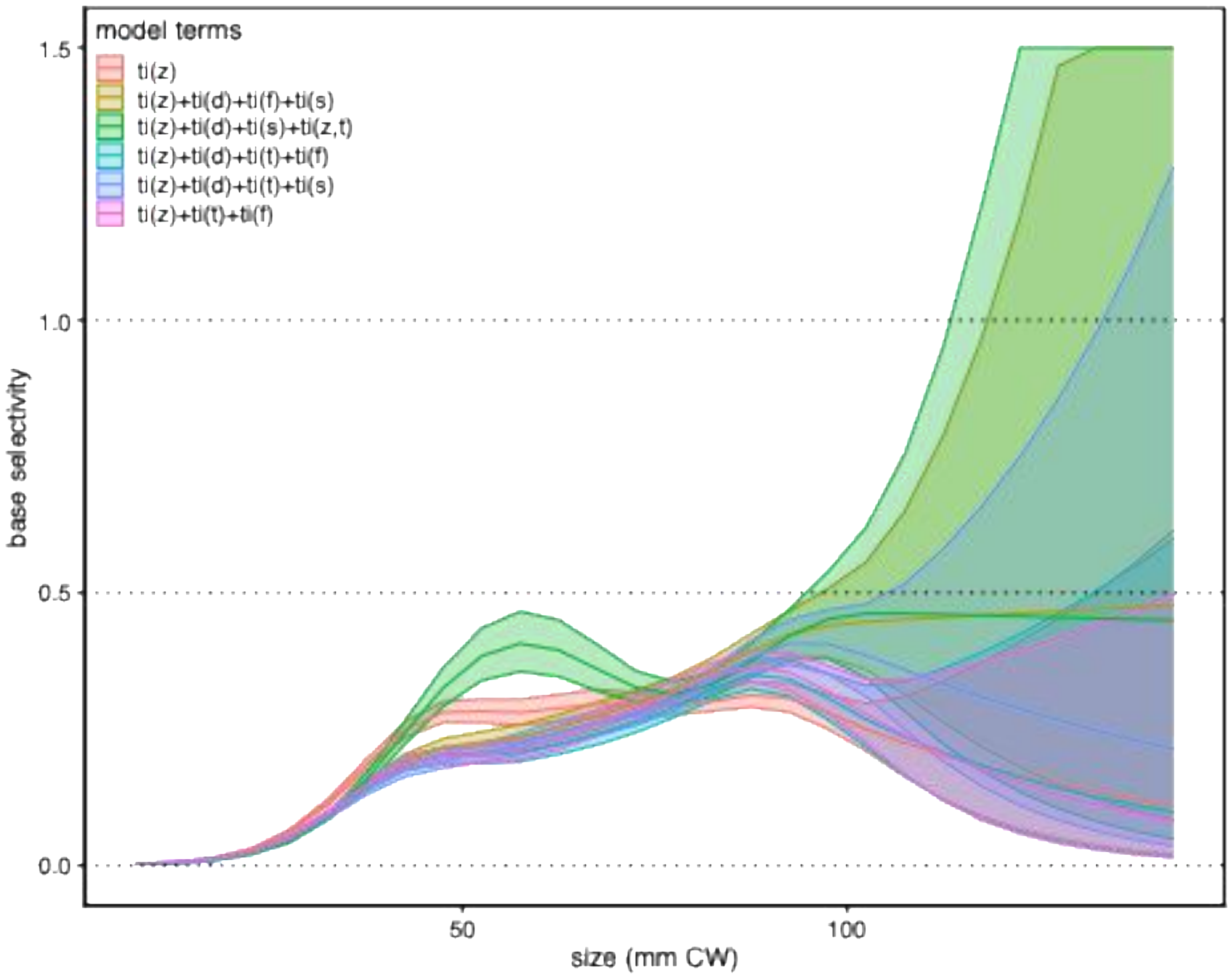
“Best” Tweedie model for males results





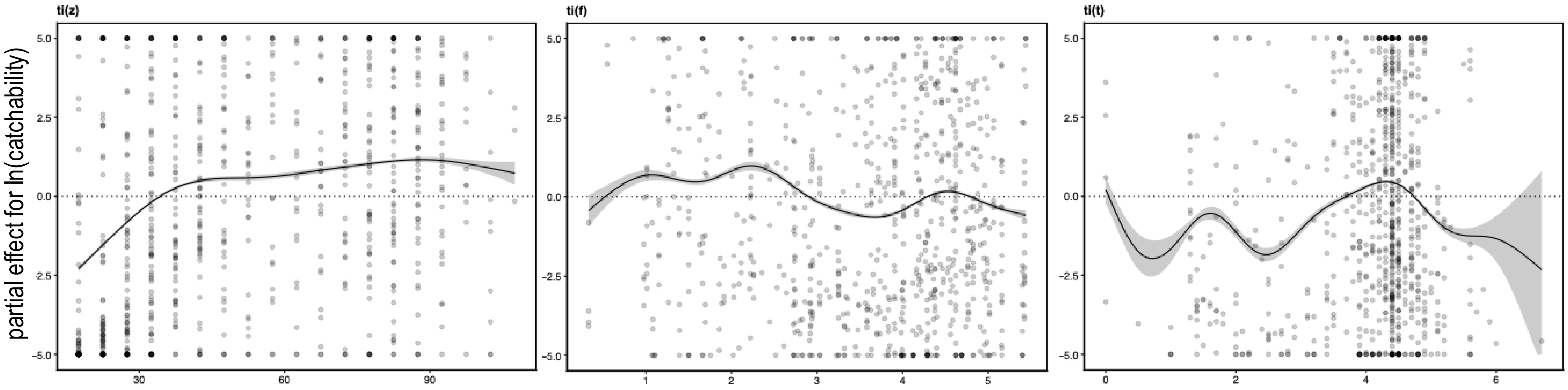
Binomial model results for females

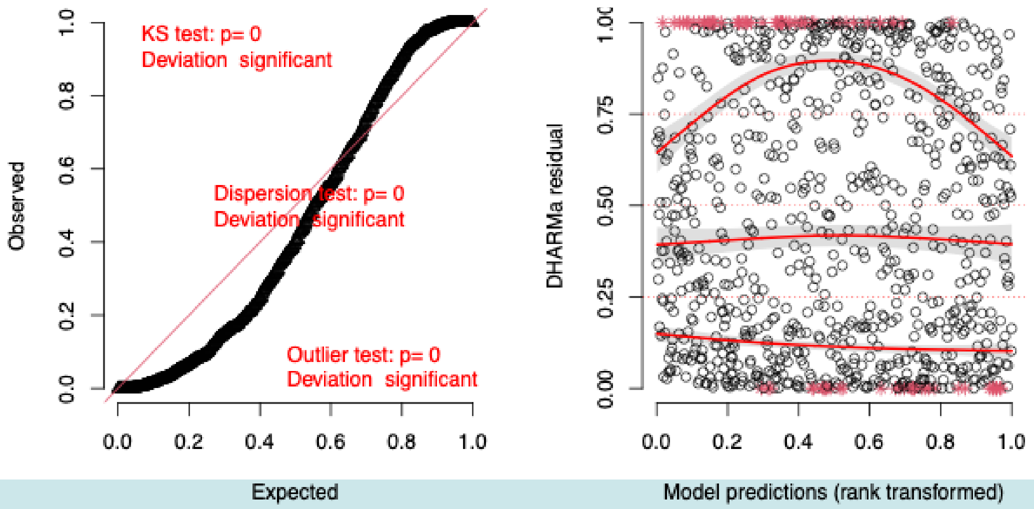




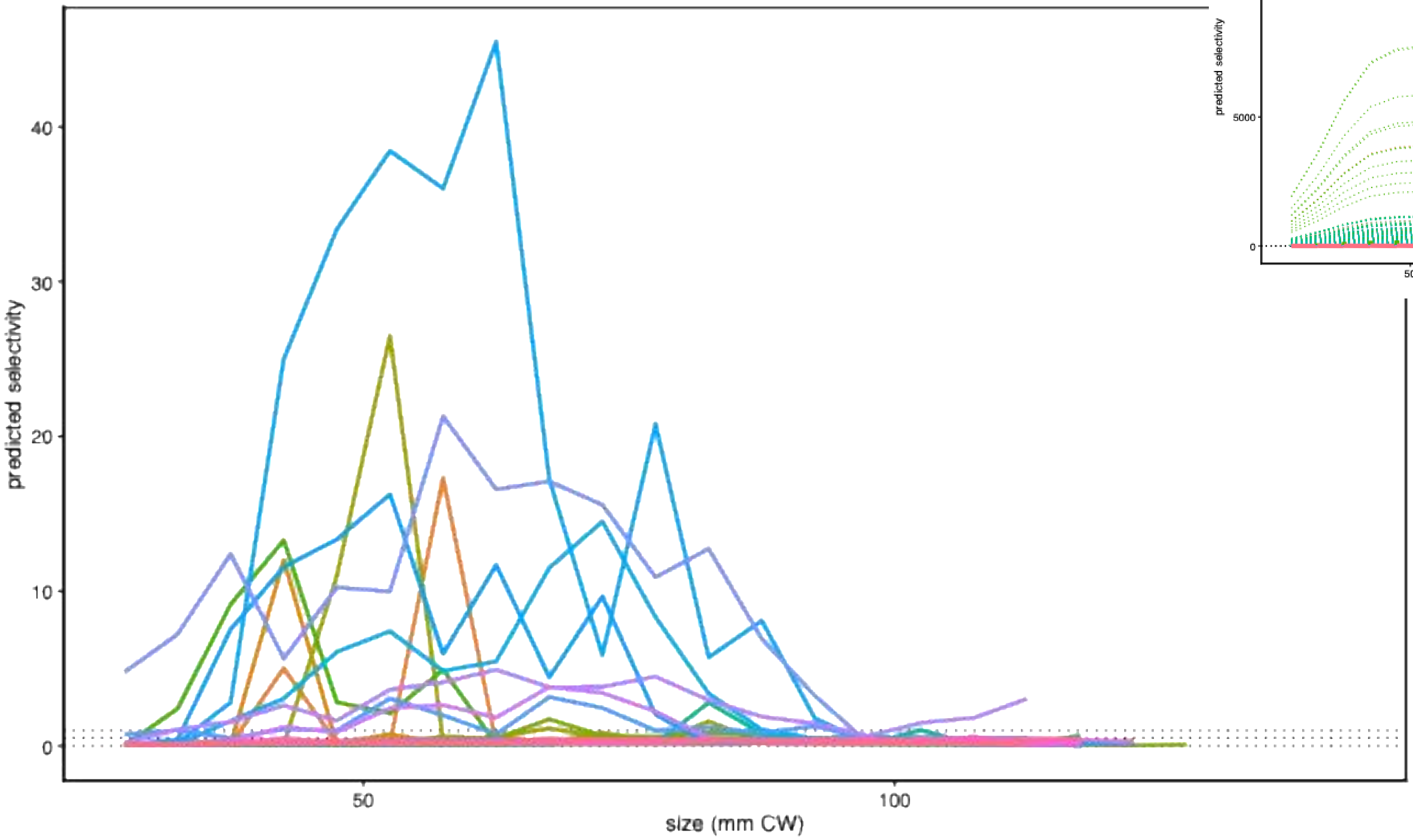
Best binomial model has covariate terms for temperature (*t*) and grain size (*f*)

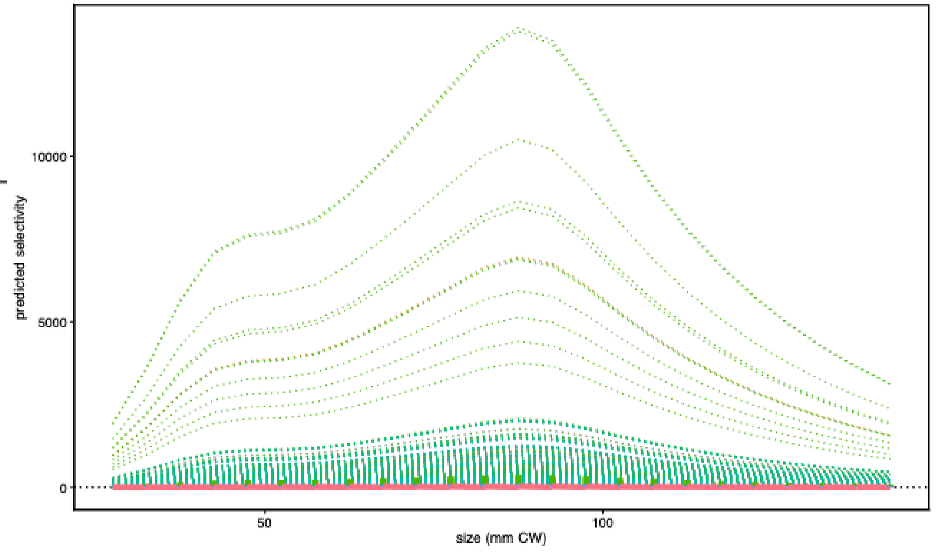
“Best” binomial model for females results





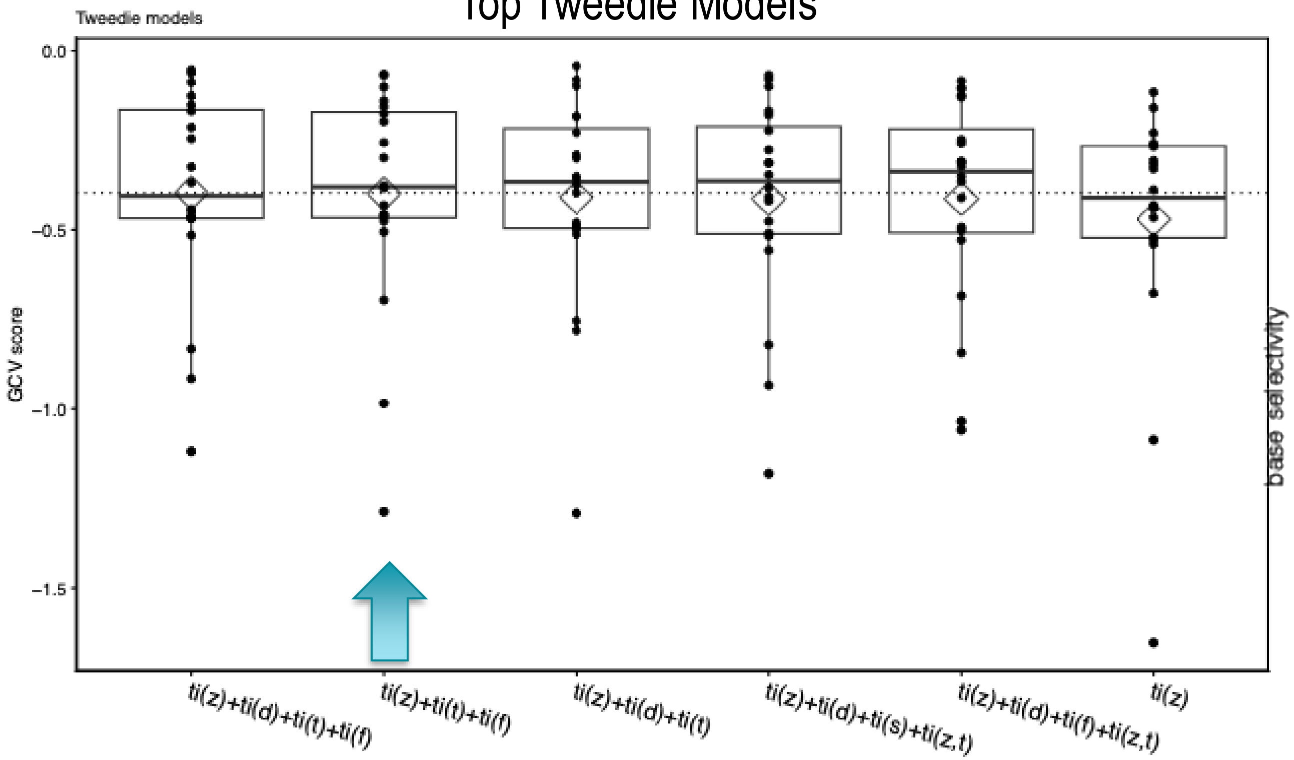
“Best” binomial model for females results

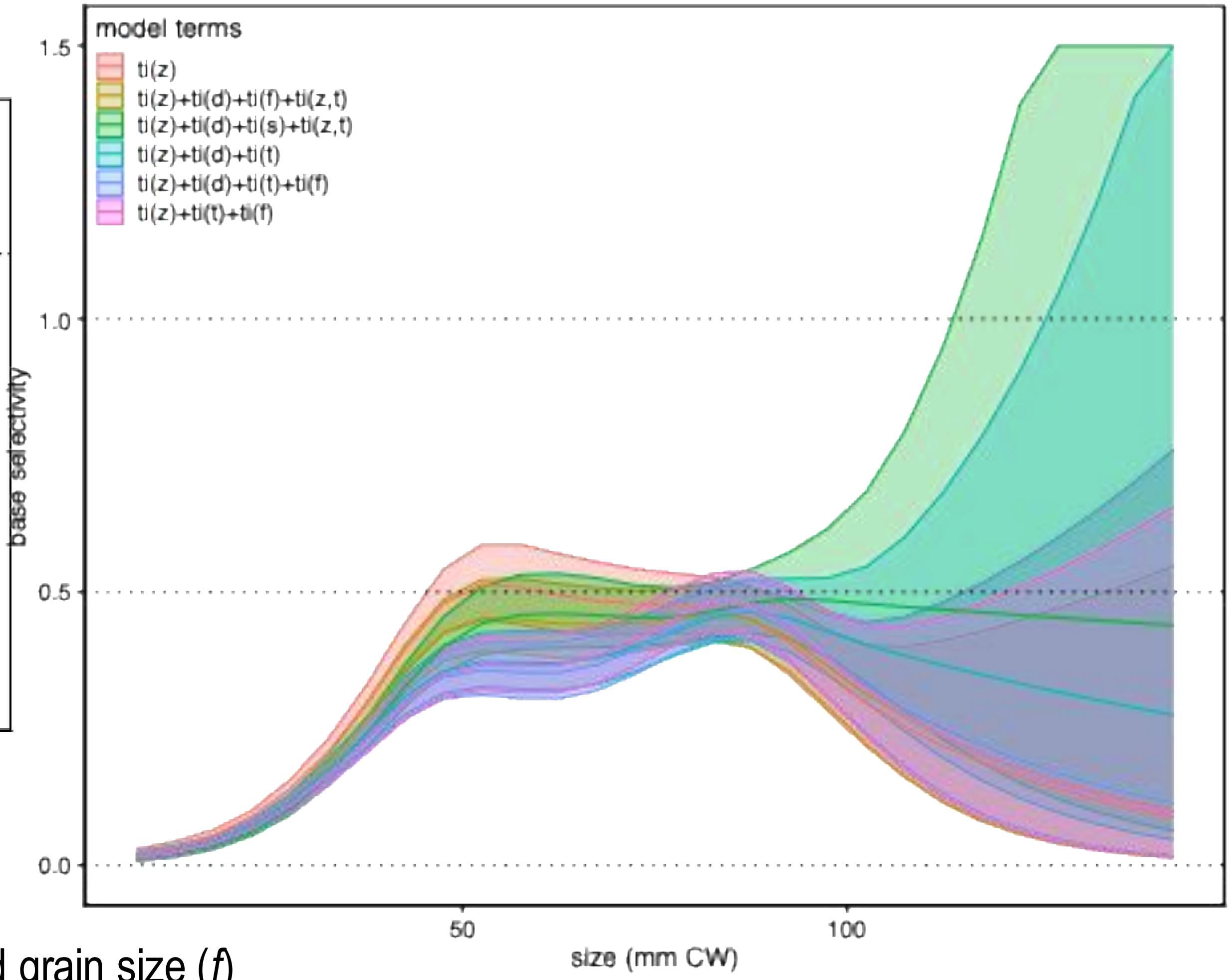




Tweedie model results for females

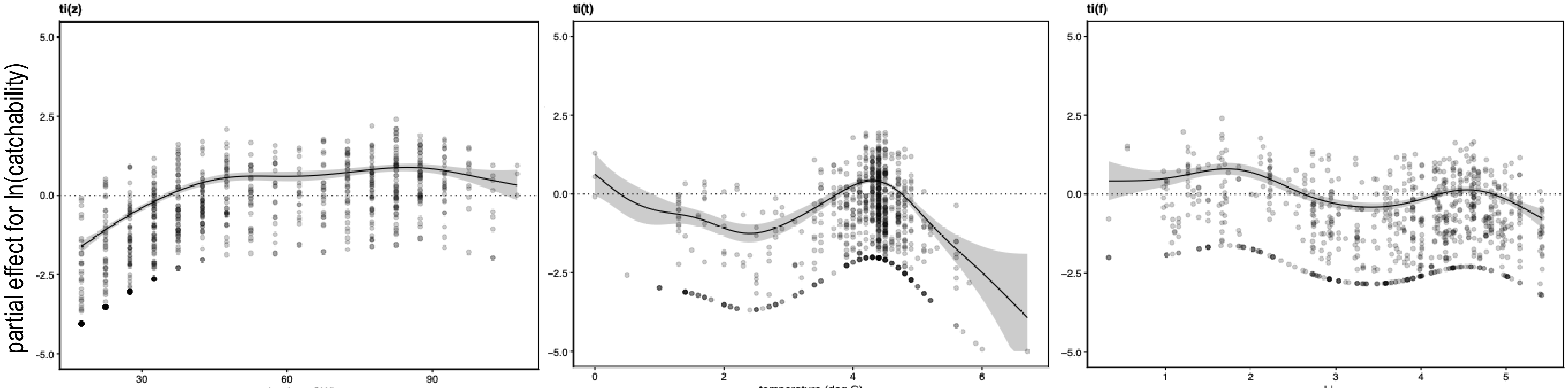
Top Tweedie Models

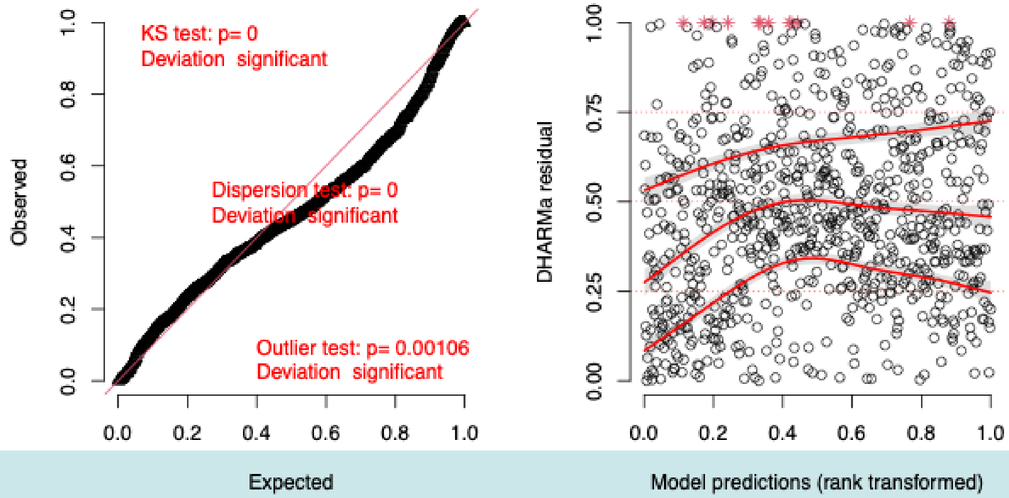




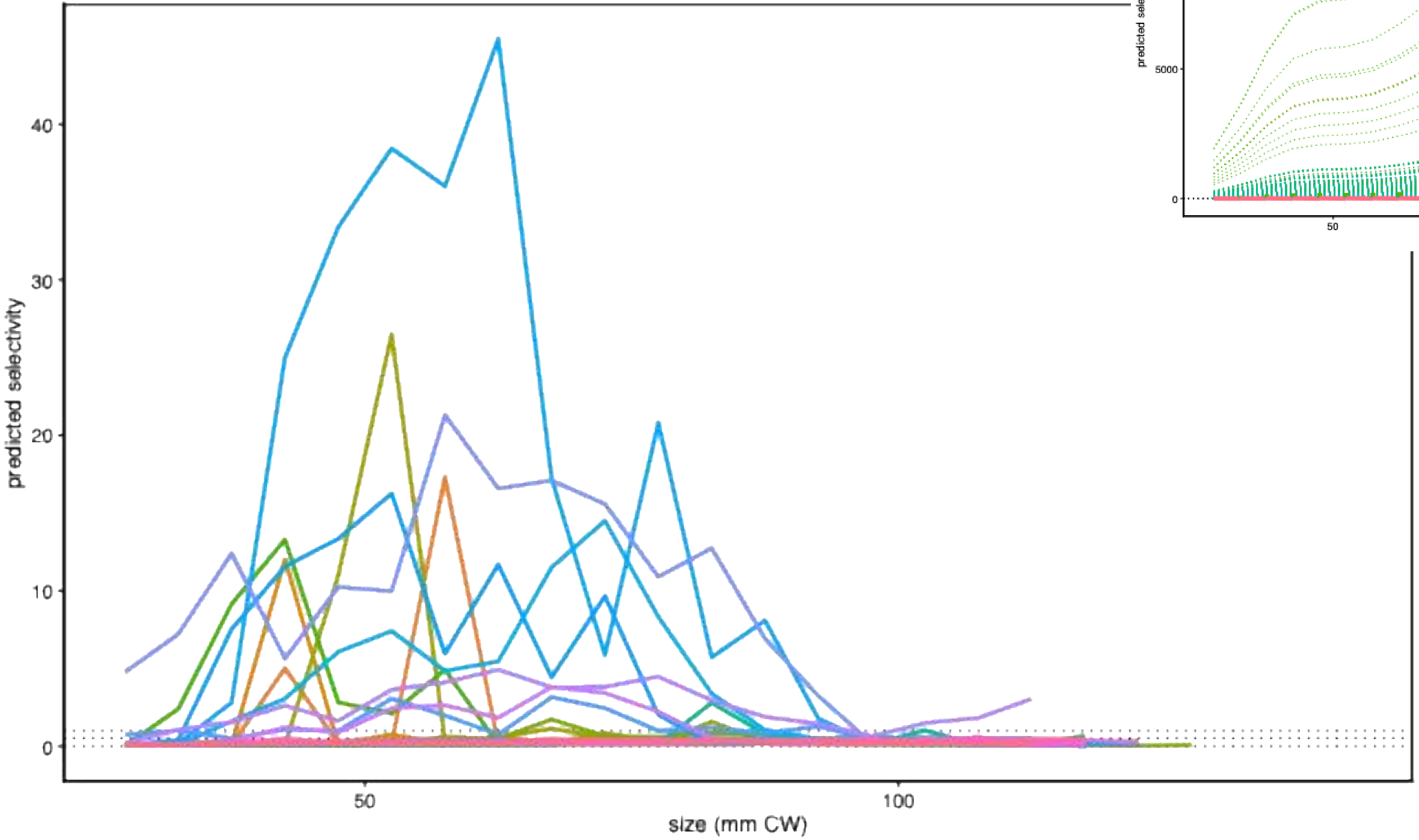
Best Tweedie model also has covariate terms for temperature (*t*) and grain size (*f*)

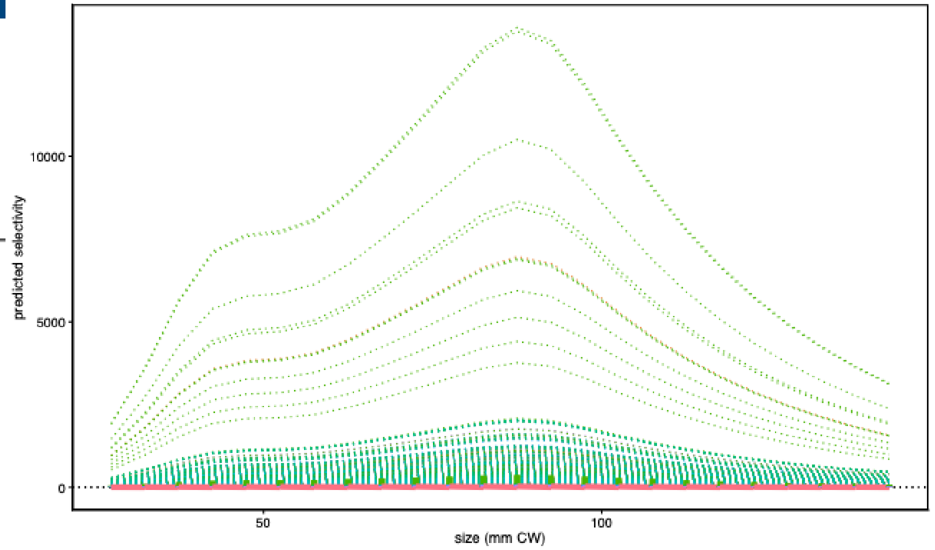
“Best” Tweedie model for females results





"Best" Tweedie model for females results



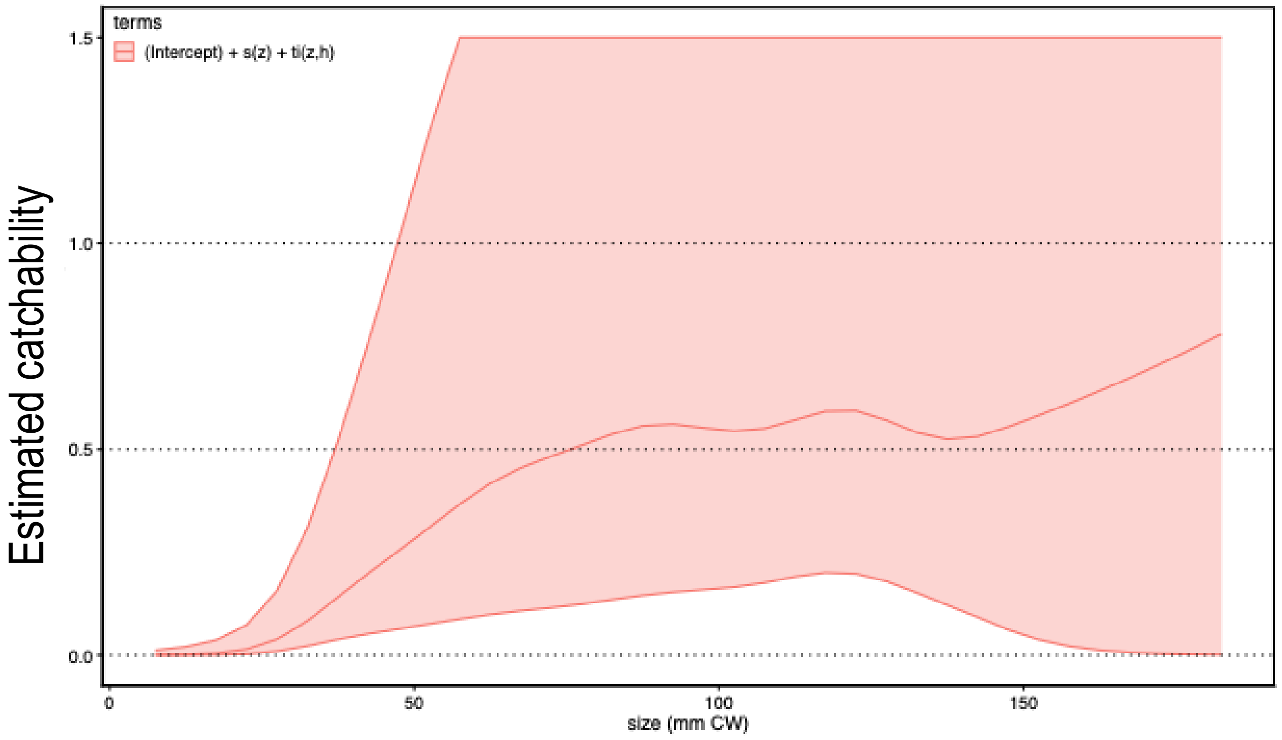


Models for males with haul-level random effects

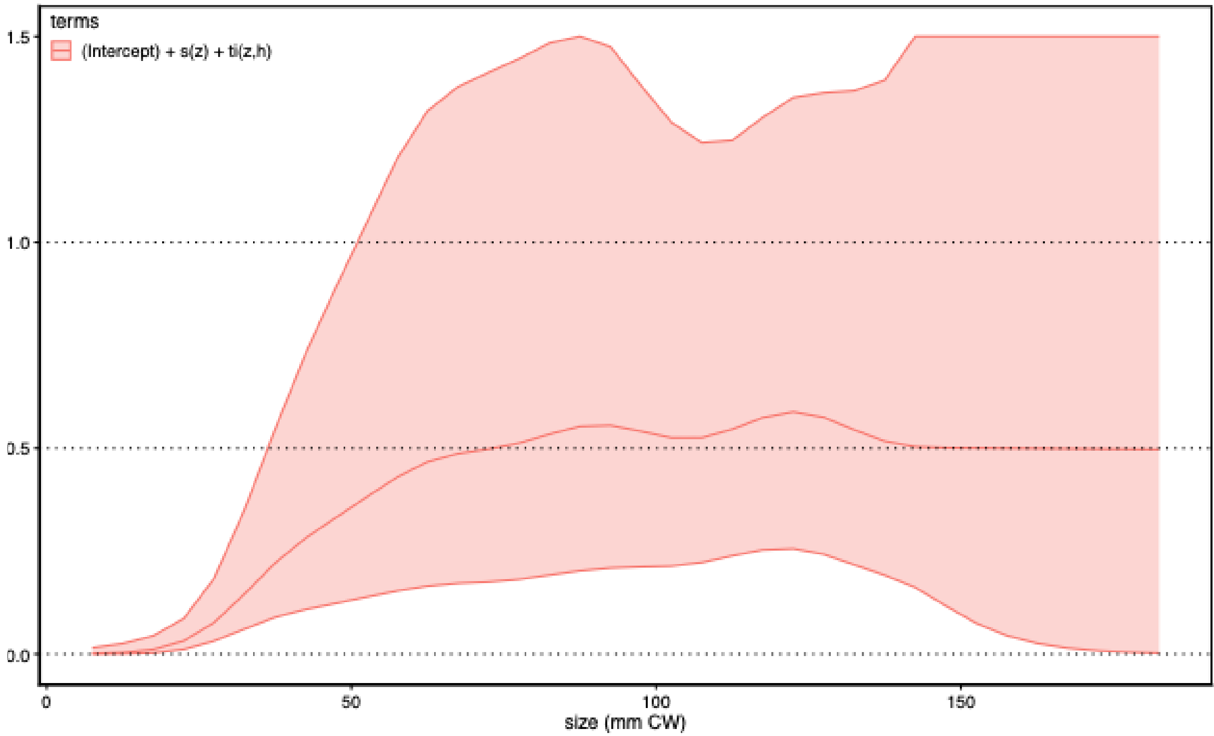
haul-level deviations from smooth curve treated as random effects:

response = intercept + s(z) + ti(z,h,bs="fs")

Binomial distribution

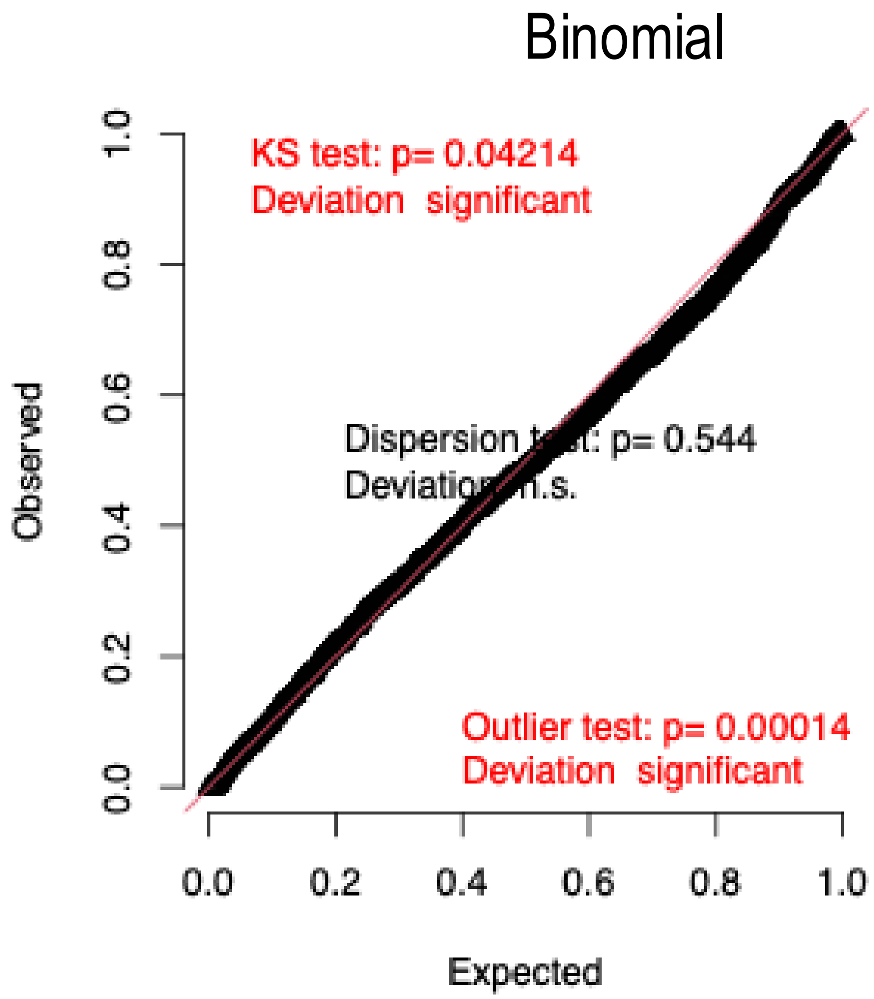


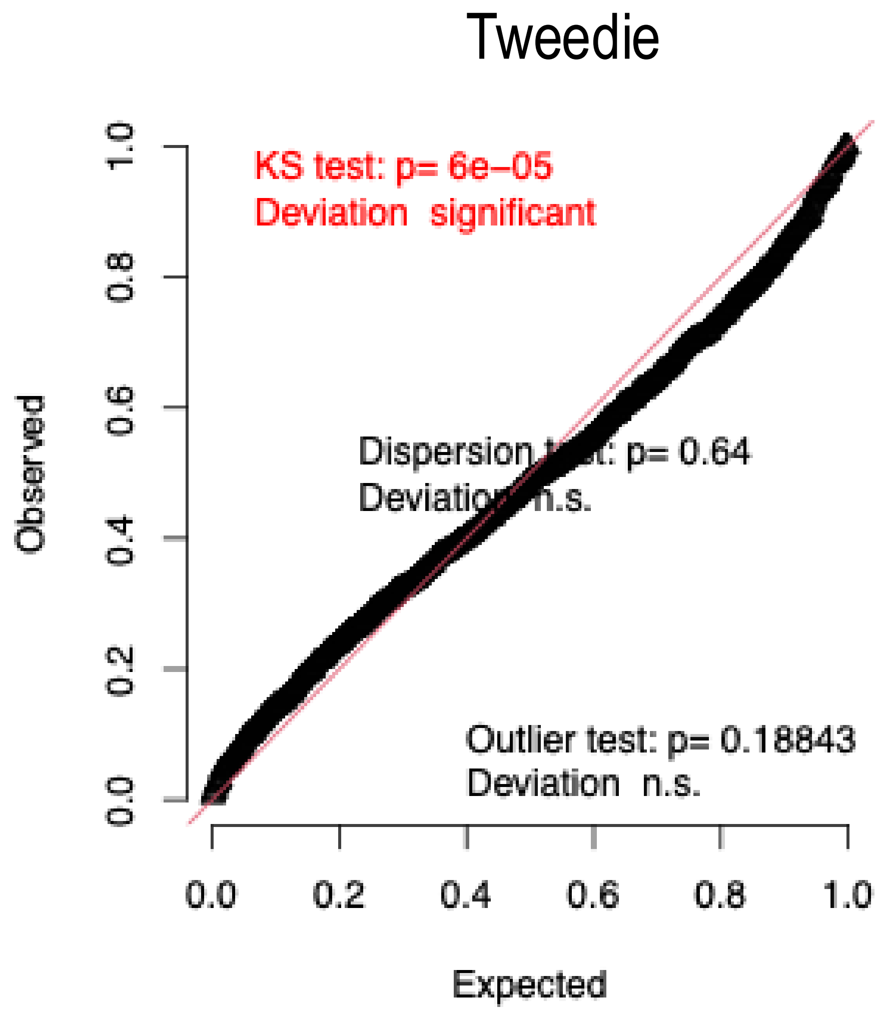
Tweedie distribution



NOAA FISHERIES

Models for males with haul-level random effects



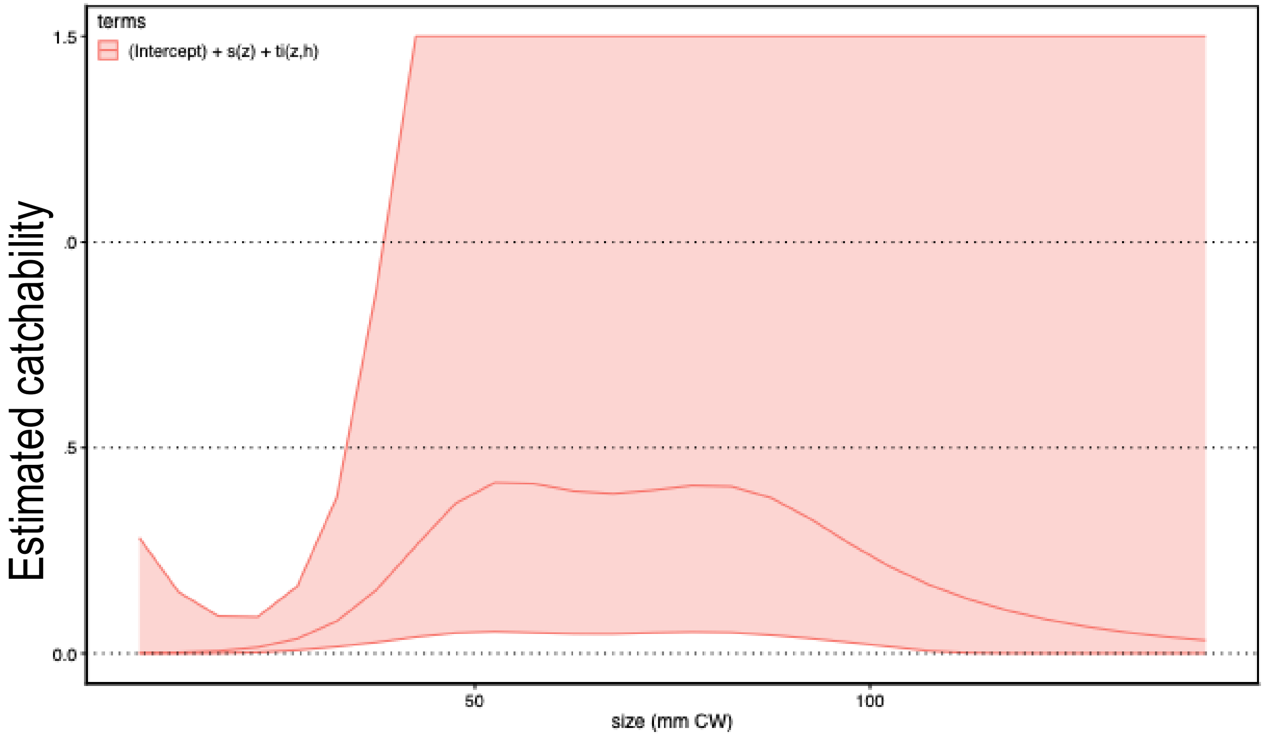


Models for females with haul-level random effects

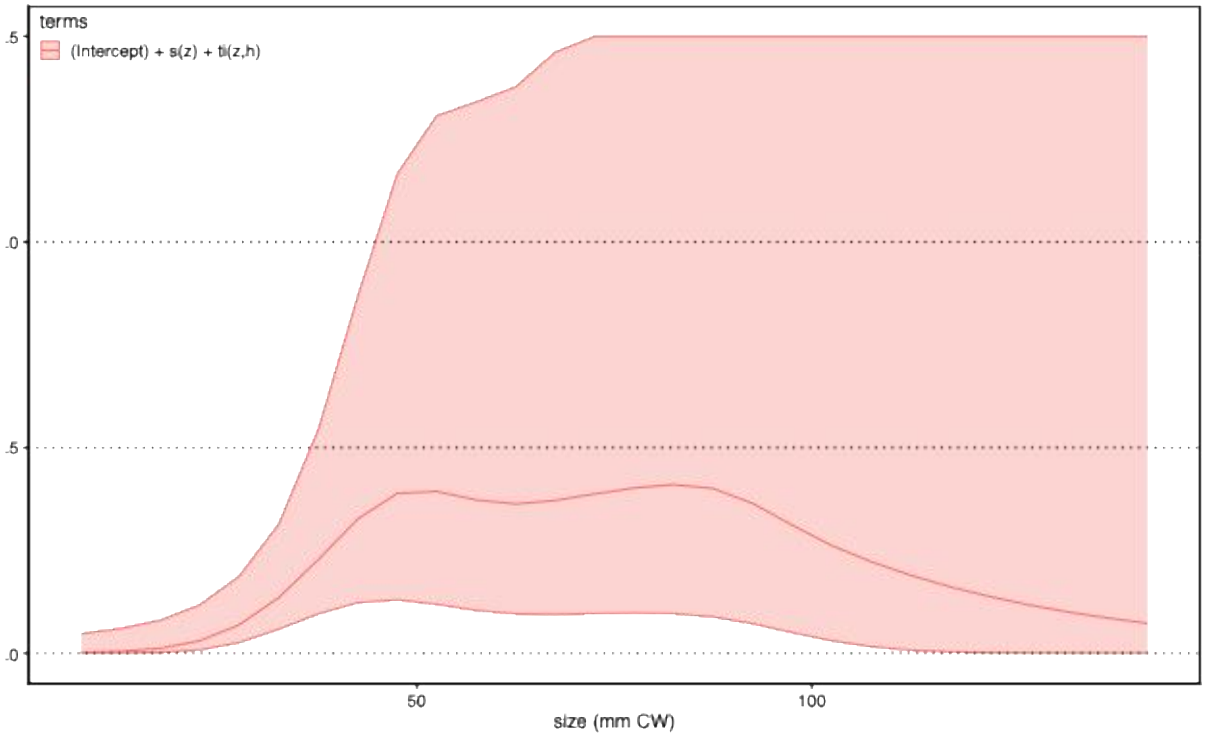
haul-level deviations from smooth curve treated as random effects:

response = intercept + s(z) + ti(z,h,bs="fs")

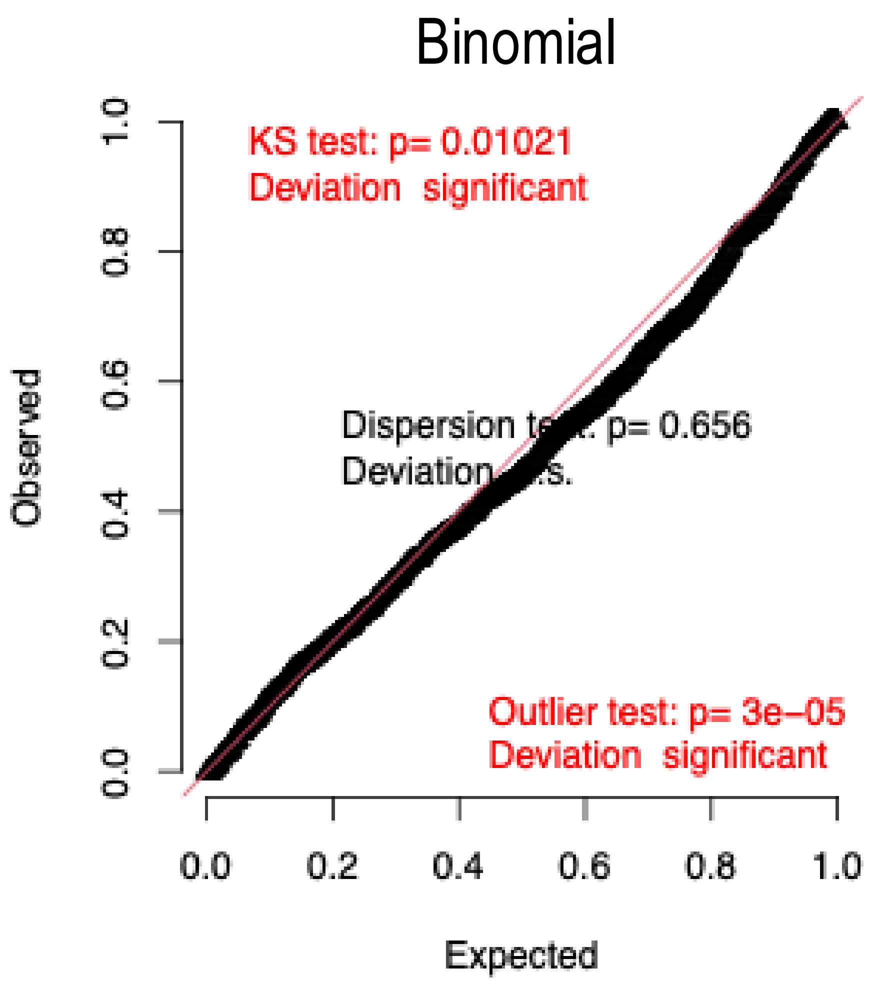
Binomial distribution

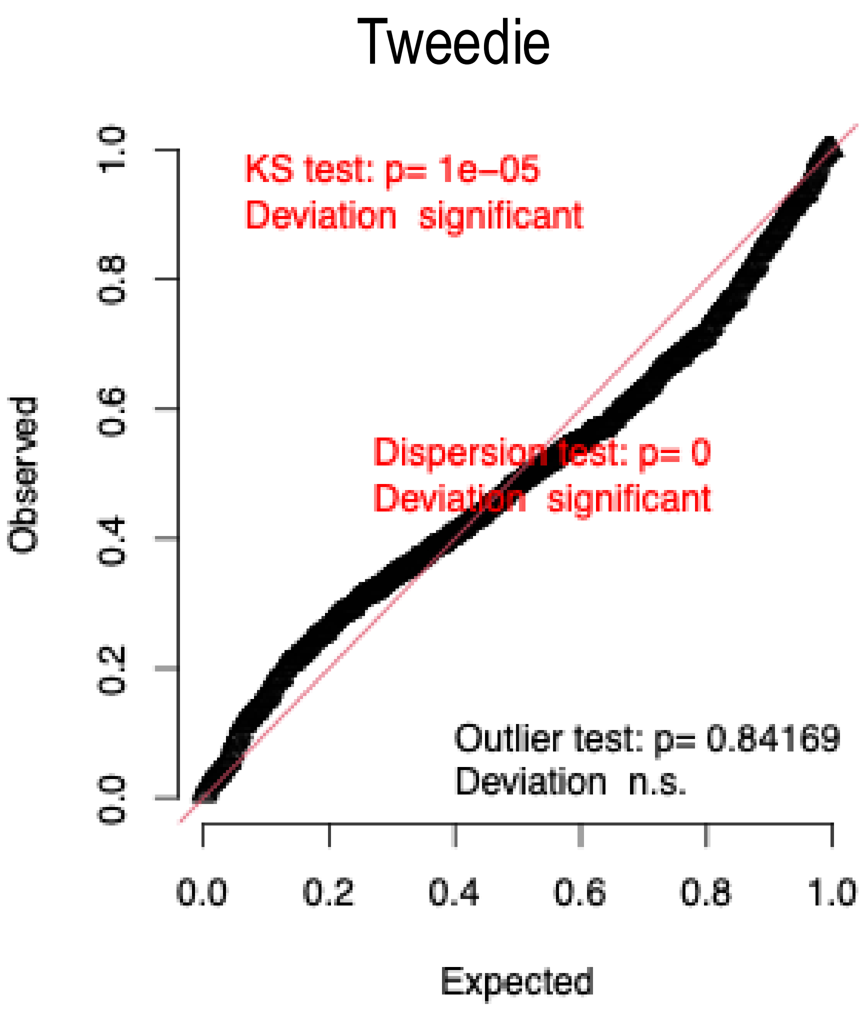


Tweedie distribution



Models for females with haul-level random effects





Wrap-up

Further work

- incorporate results from “best” models into Tanner crab assessment
- finish similar analysis for BBRKC (2013-2016 data)
- revisit snow crab (add 2017, 2018 data)

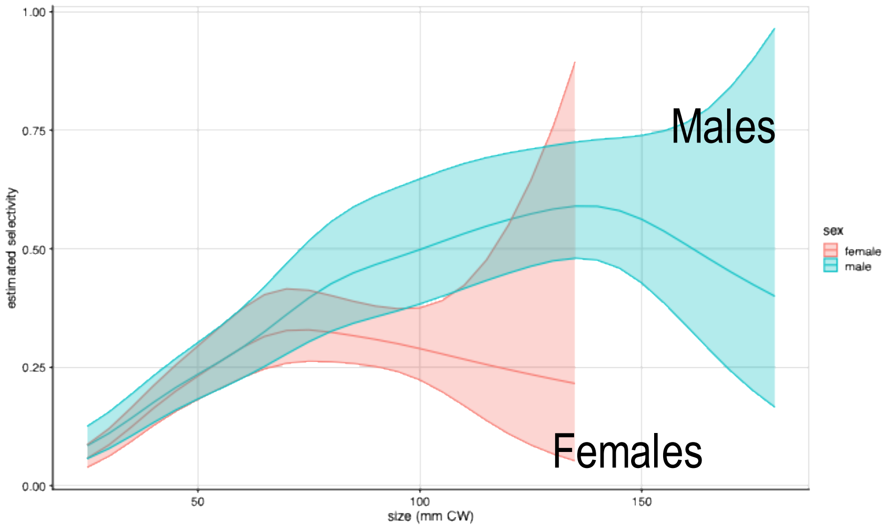
Acknowledgments

- NMFS EBS Shelf Survey crews & vessels
- BSFRF SBS Studies crews & vessels
- Bob McConnaughey and the EBSSED-2 sediment database

Discussion

- “best” estimates of size-specific catchability?
- other model distributions to try?
- mechanisms for low catchability?

Survey-level estimates: Tweedie REs



Haul-level estimates: binomial REs