# Improving priors for the steepness of the stock-recruitment function within the R package fishnets

Unpublished report prepared for Trident Systems

**Authors:**
Philipp Neubauer

**DRAGONFLY**
Data Science

# 1.   INFERRING STEEPNESS

Steepness is generally defined as the fraction of recruitment from an unfished population ($R_0$) when the spawning stock biomass is at 20% of its unfished (equilibrium) level ($B_0$). Given a functional form of the stock-recruit (S-R) function $f$, steepness can be tied to the parameters of that function (at least in the case of the Ricker and Beverton-Holt (B-H) S-R functions), such that parameters of $f$ can be expressed in terms of steepness. However, depending on the form of $f$, the bounds on steepness are different: the Beverton-Holt S-R function admits steepness values between $0.2$ and 1, while the Ricker S-R function admits values between $0.2$ and infinity (i.e., recruitment can increase above $R_0$ at $0.2B_0$).

Theory provides some insight into potential life history predictors for steepness. For example, recruitment variability and natural mortalityi, $M$, set the lower bounds for steepness (He et al. 2006). However, at low recruitment variability and $M$, there is little information about steepness in these parameters alone. Given an age structured model, Mangel et al. (2010; 2013) showed that steepness depends on growth, fecundity and larval and adult mortality rates, and Rose et al. (2001) provided evidence that empirical estimates of steepness follow expectations from life-history theory. However, given the multitude of assumptions that need to be made for a node based on theory alone (i.e., Mangel et al. 2010), from larval mortality schedules to fecundity at age, a theoretical node was not deemed achievable in the context of fishnets. As an aside, the paper by Mangel et al. (2010) should be understood as an exercise that provides a mechanistic underpinning for steepness, not a practically applicable model.

The most promising avenue for a steepness node within fishnets that could be used to predict steepness for (almost) arbitrary stocks, seemed to be an empirical node that related steepness to life-history. The biggest issue with the empirical node, however, was that Myers et al. (1999) remains the only data source that spans a range of families and orders. Shertzer & Conn (2012) obtained priors for demersal fish, but only provided an ad-hoc fit of truncated normal and beta distributions to estimates of steepness from stock assessments, for south-east US stocks only. They did not provide individual estimates for species. Other datasets that provide more formally derived (and unbiased) priors for steepness (e.g., Michielsens & McAllister 2004, Dorn 2002, Forrest et al. 2010) are too narrow to inform over a broad enough taxonomic range for an empirical node.

Given that the range of steepness varies among the B-H and Ricker S-R functions, steepness estimated from one S-R function is not directly applicable to the other. This posed a non-trivial problem for the development of a general purpose prior for steepness. Ideally, the steepness node should have a method for Ricker and B-H steepness. However, empirical approaches to the specification of prior distributions for steepness have generally focused on either one S-R model (with Michielsens & McAllister (2004) and Forrest et al. (2010) being the exceptions). Myers et al. (1999) first provided estimates of B-H steepness for a large number of species and families based on Myers' S-R database. The estimates were

simple transformations of the maximum lifetime reproductive rate, which was estimated from a meta-analysis using the Ricker S-R model across all stocks. Michielsens & McAllister (2004) pointed out that fits of the Ricker and B-H functions lead to different estimates for the slope at the origin and different expectations for steepness. The direct transformation from Ricker model parameters to B-H steepness applied by Myers et al. (1999) therefore gives misleading (i.e., overly conservative) results. Furthermore, there is a potentially strong selection bias in the study (for managed, Virtual Population Analysis assessed stocks), which could have further biased results at the species level by including only a biased subset of populations for any particular species.

Nevertheless, short of doing a new meta-analysis of steepness from the RAM legacy database or other data sources, Myers et al. (1999) was considered the only possible starting point for an empirical node in fishnets. Estimates for US-Canada west-coast rockfish from Forrest et al. (2010) were included to improve estimates of empirical relationships with life-history parameters, as were estimates for tuna by S. Harley (Hoyle et al. 2011). The reported estimates for the maximum lifetime reproductive rate in Myers et al. (1999) made it possible to re-calculate the Ricker steepness (alongside the somewhat biased Beverton-Holt steepness reported in their paper). These estimates were then combined with life-history information in Fishbase (Froese & Pauly 2014) to build a predictive model for steepness within fishnets. Given the complex relationships between life-history parameters and their constraints on steepness, a non-parametric node based on `Brter` was considered the most practical starting point.

## 2.   AN EMPIRICAL STEEPNESS NODE

### 2.1   Data: a collated dataset for steepness

Empirical estimates based on maximum lifetime reproductive rates were extracted from Myers et al. (1999), and both mean B-H and Ricker steepness were recorded (the latter calculated using formulae by Michielsens & McAllister (2004)), along with other parameters reported in Myers et al. (1999). These data were augmented with data from Forrest et al. (2010) for *Sebastes ssp.* and estimates from S. Harley for tuna. The combined dataset is available in the /data/steepness directory as `Myers_et_al_steepness_-extended.csv`. Data at the family level was also extracted and saved into `Myers_et_al_steepness_families_extended.csv`.

A data node for steepness was implemented, with the data table returned by:

```
# Source in the package ----

source('collate.R')
require(dplyr)

steep <- Steepness$create('./data/steepness')
```

```
head(tbl_df(steep))
```

```
## Source: local data frame [6 x 11]
##
##                 species populations
## 1     Harpodon nehereus          1
## 2 Alosa pseudoharengus          4
## 3     Alosa sapidissima          1
## 4   Brevoortia tyrannus          1
## 5       Alosa aestivalis          3
## 6   Brevoortia patronus          1
## Variables not shown: log_alpha_tilde (dbl),
##   SE_log_alpha_tilde (dbl), var_log_alpha_tilde
##   (dbl), exp_alpha_hat (dbl), z_20 (dbl), z_50
##   (dbl), z_80 (dbl), mean_R_z (dbl), mean_BH_z
##   (dbl)
```

The table can then be merged with a Fishbase data extract to pair empirical steepness estimates with life-history data. Note that the Fishbase data are not restricted to species with steepness estimates, rather the merge is, in SQL terms, a left join onto the Fishbase data:

```
# Load the Fishbase data
fb <- FishbaseWeb$read('data/fishbase-web')
# Limit to the 7940 records with both k and linf
fb <- subset(fb,!is.na(k) & !is.na(linf))
# Add an id column for indexing later
fb$id <- 1:nrow(fb)
# Add a dummy row for helping with predictor nodes
# that need to have at least two predictors
fb$dummy <- 1.0

# merge with steepness data
steep_merged <- Steepness$merge(steep,fb)

head(tbl_df(steep_merged))
```

```
## Source: local data frame [6 x 46]
##
##                 species populations log_alpha_tilde
## 1 Abbottina rivularis          NA              NA
## 2       Ablennes hians          NA              NA
## 3        Abramis brama          NA              NA
## 4        Abramis brama          NA              NA
## 5        Abramis brama          NA              NA
## 6        Abramis brama          NA              NA
## Variables not shown: SE_log_alpha_tilde (dbl),
##   var_log_alpha_tilde (dbl), exp_alpha_hat (dbl),
```

```
##   z_20 (dbl), z_50 (dbl), z_80 (dbl), mean_R_z
##   (dbl), mean_BH_z (dbl), order (fctr), country
##   (fctr), sex (fctr), genus (fctr), family
##   (fctr), swimmode (fctr), ltype (fctr), linf
##   (dbl), k (dbl), t0 (dbl), m (dbl), temp (dbl),
##   lmat (dbl), lmatmin (dbl), lmatmax (dbl), amat
##   (dbl), amatmin (dbl), amatmax (dbl), a (dbl), b
##   (dbl), trophic (dbl), diet (fctr), feeding
##   (fctr), habit (fctr), migration (fctr),
##   depthmin (dbl), depthmax (dbl), fecundmin
##   (dbl), fecundmax (dbl), fecundity (dbl), class
##   (fctr), lmax (dbl), amax (dbl), id (int), dummy
##   (dbl)
```

In order to set up a fishnet that that extracts empirical relationships between steepness and life-history, the merged Fishbase and steepness data were aggregated at the species level to avoid multiple, possibly contradictory, life-history correlates with steepness. Geometric means were used to aggregate life-history parameters, under the assumption that most follow log-normal distributions.

```
# geometric mean
gmean <- function(x) exp(mean(log(x),na.rm=T))

# reduce dataset; gometric means for paramters by species
steep_reduced <- steep_merged %>%
  select(species,
         genus,
         family,
         class,
         order,
         mean_BH_z,
         linf,
         m,
         fecundity,
         trophic,
         lmat,
         lmax ,
         k,
         amax,
         habit,
         trophic,
         depthmax) %>%
  group_by(order,class,genus,family,species) %>%
  summarise(mean_BH_z = unique(mean_BH_z),
            habit = unique(habit),
            trophic = gmean(trophic),
            linf = gmean(linf),
            m = gmean(m),
```

Improving steepness priors in fishnets

```
            depthmax = gmean(depthmax),
            fecundity = gmean(fecundity),
            trophic = gmean(trophic),
            lmax = gmean(lmax),
            lmat = gmean(lmat),
            k = gmean(k),
            amax = gmean(amax),
            recsigma = NA) %>%
  ungroup() %>%
  data.frame()
```

## 2.2   A Bayesian empirical node to avoid over-fitting

An initial attempt at defining a fishnet based on previously available
nodes (`Brter`: boosted regression trees; `Svmer`: support vector machines;
`Glmer`: generalized linear models) suggested that over-fitting of empirical
relationships might be interfering with predictive power of the nodes. To
investigate this problem, I implemented a Bayesian node based on the
INLA package (Ru et al. 2009). Fully Bayesian methods are less prone to
over-fitting, and the INLA approach, based on Laplacian approximations,
allows for fast parameter estimation in linear and linear mixed models. The
resulting node, called `Bayser`, uses the same formula syntax as previous
empirical nodes, with an additional option to estimate random effects with
the `f(effect, 'iid')` syntax (e.g., `f(family, model='iid')` for a family
level random effect). This option should facilitate the use of taxonomic
information in empirical nodes, whereas the use of taxonomy as a fixed
effect frequently leads to over-fitting. The `Bayser` node comes with `fit`,
`sample`, `predict` and `tune` methods:

```
linf = Bayser(log(linf) ~ f(family,model="iid") +
                f(order,model="iid") +
                log(lmax) +
                habit,exp)

# fitting to data
linf$fit(steep_reduced[1:1500,])
psamp <- 1501:nrow(steep_reduced)

# predicting onto new data
example_predictions <-
    linf$predict(steep_reduced[psamp,])

nna_data <- !is.na(steep_reduced$lmax[psamp])

predict_df <- data.frame(bayser = example_predictions[nna_data],
    observed = steep_reduced[psamp,'linf'][nna_data])

linf_brter = Brter(log(linf) ~ family +
```

```
       order + log(lmax) + habit, exp)

# fitting to data
linf_brter$fit(steep_reduced[1:1500, ])
```

```
## Error in linf_brter$fit(steep_reduced[1:1500, ]): could not
find function "gbm.fixed"
```

```
# predicting onto new data
example_predictions_brter <-
    linf_brter$predict(steep_reduced[psamp,])
```

```
## Error in linf_brter$predict(steep_reduced[psamp, ]):  could
not find function "predict.gbm"
```

```
predict_df <- reshape2::melt(
        data.frame(brter = example_predictions_brter[nna_data],
        predict_df), id.vars = c('observed')
    )
```

```
## Error in data.frame(brter = example_predictions_brter[nna_-
data], predict_df): object 'example_predictions_brter' not found
```

```
require(ggplot2)

ggplot(predict_df) +
  geom_point(aes(y = value, x = observed, col = variable)) +
  geom_abline(slope=1, intercept=0) +
  theme_bw() +
  xlab('Data') +
  ylab('Prediction') +
  scale_colour_discrete('Method')
```

```
## Error in eval(expr, envir, enclos): object 'value' not found
```

### 2.3   Sequential imputing of data in fishnets

In the Bayser example above, linf depends on lmax in an empirical node. The relationship can only be estimated for data rows (i.e., species) with data for both traits. For more complex life history interactions, missing data for the combination of covariates in the model can strongly influence the total number of species for which the relationship can be estimated. For example, one species may not have a reported von Bertalanffy $k$, while another species may have $k$ but no estimate of fecundity. Neither row could be included

in a node like steepness, which may depend on both factors. In the case of steepness, only 46 empirical estimates at the species level are available from the dataset described above, and missing values dramatically reduce the number of covariates that could be used to construct an empirical node for steepness.

To obtain more data for complex empirical relationships, I implemented an option within the generic `fit` method of a fishnet, which allows data to be sequentially imputed. Thus, if `linf` depends on `lmax`, and `k` in turn depends on `linf`, then `linf` could be imputed for species with missing `linf` based on the node relating it to `lmax`. This allows $k$ to be estimated for species with data for `lmax` and missing data for `linf`. The imputation naturally relies on a linearly evolving net, where relationships depend on values imputed at lower level nodes. The option can be turned on using `impute=TRUE` when fitting a fishnet (see the steepness node fitting below).

### 2.4 The empirical steepness node: fishing for predictive power

The empirical fishnet node to estimate steepness was embedded in a larger fishnet that aimed to build a covariate set (using the `impute=TRUE` option). All empirical nodes leading up to steepness node itself used the `Bayser` node, and included taxonomic information as random effects. The actual steepness node was initially specified as a `Brter` node, in order to capture potentially non-linear relationships between steepness and covariates. After some discussion, the B-H estimates by Myers et al. (1999) were used as these were deemed most useful in the context of NZ stock assessments, which typically use B-H S-R functions.

```
# Build a net for steepness.
# Use Bayesian nodes in an attempt to not over-fit
# Transformations are defined for Beverton-Holt S-R
# parameters to map steepness onto the real line and back.

logit     <- function(x) log(x/(1-x))
logit_inv <- function(xt) 1/(1 + exp(-xt))
BH_tr     <- function(h) h/0.8-0.25
BH_tr_inv <- function(ht) (ht+0.25)*0.8
logit_BH  <- function(h) logit(BH_tr(h))
logit_BH_inv <- function(ht) BH_tr_inv(logit_inv(ht))

BH_net <- Fishnet(
  species  = SpeciesRandom(),
  genus    = GenusParser(),
  family   = FamilyLookupper(),
  order    = OrderLookupper(),
  class    = ClassLookupper(),

  habit    = TaxonomicImputer('habit'),
  depthmax = TaxonomicImputer('depthmax',c(log,exp),5),
```

```
  trophic   = TaxonomicImputer('trophic',c(log,exp),3),
  lmax      = TaxonomicImputer('lmax',c(log,exp),5),
  amax      = TaxonomicImputer('amax',c(log,exp),5),

  linf      = Bayser(log(linf) ~ f(family,model="iid") +
                    f(class,model="iid") +
                    log(lmax),exp),

  fecundity = Bayser(log(fecundity) ~ f(family,model="iid") +
                    f(class,model="iid") +
                    log(linf) +
                    log(depthmax),exp),

  k         = Bayser(log(k) ~ f(family,model="iid") +
                    log(linf) +
                    f(habit,model="iid") +
                    log(depthmax),exp),

  m         = Bayser(log(m) ~ f(family,model="iid") +
                    f(class,model="iid") +
                    log(k) +
                    log(linf) +
                    f(habit,model="iid") +
                    log(depthmax) +
                    trophic,exp),

  lmat      = Bayser(log(lmat) ~ f(family,model="iid") +
                    log(k) +
                    log(linf) +
                    f(habit,model="iid") +
                    log(depthmax),exp),

  recsigma  = RecsigmaThorsonEtAl2014(),

  mean_BH_z = Brter(logit_BH(mean_BH_z) ~  habit +
                    log(linf) +
                    log(k) +
                    log(m) +
                    log(fecundity) +
                    recsigma +
                    trophic +
                    log(depthmax),
                  transform = logit_BH_inv,
                  ntrees=3500,
                  bag.fraction=0.9)

)

# fit the BH_net to the summarised Fishbase data
```

Improving steepness priors in fishnets

```
BH_net$fit(steep_reduced,impute = T)
```

```
## Error in self$nodes[[name]]$fit(data):   could not find
function "gbm.fixed"
```

The predictive power of the steepness node was assessed using jackknifing, and a newly implemented tune method for `Brter`. The tune formulation included 6 formulae with increasing amounts of life-history information.

```
# Function to make a test set for cross validation
make_testset <- function(net, org_data, name){
  testset <- data.frame(net$data[, -which(colnames(net$data) == name)],
                        name = org_data[name])
  testset
}

testset <- make_testset(BH_net, steep_reduced, 'mean_BH_z')
```

```
## Error in data.frame(net$data[, -which(colnames(net$data) ==
name)], name = org_data[name]): arguments imply differing number
of rows: 0, 1876
```

```
# jacknifing - could be done with $cross,
# but is implemented here to produce
# a vector of observed and predicted outcomes.

jackknife_cv <- function(data, net, node){
  testnet <- net
  data <- data[!is.na(data[[node]]),]

  pred <- vector(,nrow(data))
  for (i in 1:nrow(data)){
    cat('CV for observation ', i, '\n')
    train <- data[-i, ]
    test <- data[i, ]
    test[[node]] <- NA
    testnet$nodes[[node]]$fit(train)
    pred[i] <- testnet$nodes[[node]]$predict(test)

  }
  data.frame(Predicted=pred, Observed=data[[node]])
}

covars <- c('log(linf)',
    'log(k)',
    'log(m)',
    'log(fecundity)',
```

```
    'log(recsigma)',
    'log(trophic)',
    'log(depthmax)')

formulae <- vector(, 6)
formulae[1] <- paste('logit_BH(mean_BH_z) ~ ',
    paste(covars[c(3, 4)], collapse = ' + '))
formulae[2] <- paste('logit_BH(mean_BH_z) ~ ',
    paste(covars[c(3, 4, 7)], collapse = ' + '))
formulae[3] <- paste('logit_BH(mean_BH_z) ~ ',
    paste(covars[c(2, 4, 7)], collapse = ' + '))
formulae[4] <- paste('logit_BH(mean_BH_z) ~ ',
    paste(covars[c(2, 3, 4, 7)], collapse = ' + '))
formulae[5] <- paste('logit_BH(mean_BH_z) ~ ',
    paste(covars[c(1, 2, 3, 4, 5, 7)], collapse = ' + '))
formulae[6] <- paste('logit_BH(mean_BH_z) ~ ',
                     paste(covars, collapse = ' + '))
```

```
tune <- BH_net$nodes$mean_BH_z$tune(testset, formulae,
    folds=sum(!is.na(testset[['mean_BH_z']])))
```

```
## Error in t(self$cross(data, ...)$summary$mean):  error in
evaluating the argument 'x' in selecting a method for function
't': Error in is.data.frame(data) : object 'testset' not found
## Calls: <Anonymous> ... model.frame -> model.frame.default ->
is.data.frame
```

```
steep_cv <- jackknife_cv(testset,BH_net, 'mean_BH_z')
```

```
## Error in jackknife_cv(testset, BH_net, "mean_BH_z"): object
'testset' not found
```

```
tune$best
```

```
## Error in eval(expr, envir, enclos): object 'tune' not found
```

```
lm_pred_steep <- lm(Observed ~ Predicted, data = steep_cv)
```

```
## Error in is.data.frame(data): object 'steep_cv' not found
```

```
summary(lm_pred_steep)
```

```
## Error in summary(lm_pred_steep):  error in evaluating the
argument 'object' in selecting a method for function 'summary':
Error: object 'lm_pred_steep' not found
```

```
plot(steep_cv, pch = 16, xlim = c(0.2, 1), ylim = c(0.2,
    1))

## Error in plot(steep_cv, pch = 16, xlim = c(0.2, 1), ylim =
c(0.2, 1)): error in evaluating the argument 'x' in selecting a
method for function 'plot': Error: object 'steep_cv' not found

abline(lm_pred_steep$coeff[1], lm_pred_steep$coeff[2],
    col = 2, lwd = 2)

## Error in abline(lm_pred_steep$coeff[1], lm_pred_steep$coeff[2],
col = 2, : object 'lm_pred_steep' not found

abline(0, 1, lwd = 2)

## Error in int_abline(a = a, b = b, h = h, v = v, untf = untf,
...): plot.new has not been called yet
```

Even for the best `Brter` model, the jackknifing suggests very low predictive power (**??**), possibly due to over-fitting. To investigate this possibility, the node was re-implemented as a `Bayser` node, and a tune step was used to define a best model among a set of 8 candidate models, which included taxonomic levels as predictors:

```
BH_net$nodes$mean_BH_z <-
    Bayser(logit_BH(mean_BH_z) ~
    f(family,model='iid') +
    f(habit,model='iid') +
    log(linf) +
    log(k) +
    log(m) +
    log(fecundity) +
    log(recsigma) +
    log(m)*log(recsigma) +
    log(trophic) +
    log(depthmax), transform=logit_BH_inv)

BH_net$nodes$mean_BH_z$fit(testset)

##  Error  in  BH_net$nodes$mean_BH_z$fit(testset):    object
'testset' not found

formulae <- vector(, 6)
formulae[1] <- paste('logit_BH(mean_BH_z) ~ ',
    paste(covars[7], collapse = ' + '))
formulae[2] <- paste('logit_BH(mean_BH_z) ~ ',
```

```
    "f(family, model='iid')")
formulae[3] <- paste('logit_BH(mean_BH_z) ~ ',
    "f(family, model='iid') +",
    paste(covars[c(3, 7)], collapse = ' + '))
formulae[4] <- paste('logit_BH(mean_BH_z) ~ ',
    paste(covars[c(2, 3, 4, 7)], collapse = ' + '))
formulae[5] <- paste('logit_BH(mean_BH_z) ~ ',
    "f(family, model='iid') +",
    paste(covars[c(2, 3, 4, 7)], collapse = ' + '))
formulae[6] <- paste('logit_BH(mean_BH_z) ~ ',
    "f(family, model='iid') +",
    "f(habit, model='iid') +",
    paste(covars[c(2, 3, 4, 7)], collapse = ' + '))
formulae[7] <- paste('logit_BH(mean_BH_z) ~ ',
    "f(family, model='iid') +",
    paste(covars, collapse = ' + '))
formulae[8] <- paste('logit_BH(mean_BH_z) ~ ',
    "f(family, model='iid') +",
    "f(habit, model='iid') +",
    "log(m)*log(recsigma) +",
    paste(covars, collapse = ' + '))
```

```
tune <- BH_net$nodes$mean_BH_z$tune(testset, formulae,
    folds=sum(!is.na(testset[['mean_BH_z']])))
```

```
## Error in t(self$cross(data, ...)$summary$mean):  error in
evaluating the argument 'x' in selecting a method for function
't': Error in is.data.frame(data) : object 'testset' not found
## Calls: <Anonymous> ... model.frame -> model.frame.default ->
is.data.frame
```

```
steep_cv_bayes <- jackknife_cv(testset, BH_net, 'mean_BH_z')
```

```
## Error in jackknife_cv(testset, BH_net, "mean_BH_z"): object
'testset' not found
```

```
tune$best
```

```
## Error in eval(expr, envir, enclos): object 'tune' not found
```

```
lm_pred_steep <- lm(Observed ~ Predicted, data = steep_cv_bayes)
```

```
## Error in is.data.frame(data):  object 'steep_cv_bayes' not
found
```

Improving steepness priors in fishnets

```
summary(lm_pred_steep)

## Error in summary(lm_pred_steep):  error in evaluating the
argument 'object' in selecting a method for function 'summary':
Error: object 'lm_pred_steep' not found

plot(steep_cv_bayes, pch = 16, xlim = c(0.2, 1), ylim = c(0.2,
    1))

## Error in plot(steep_cv_bayes, pch = 16, xlim = c(0.2, 1), ylim
= c(0.2, : error in evaluating the argument 'x' in selecting a
method for function 'plot': Error: object 'steep_cv_bayes' not
found

abline(lm_pred_steep$coeff[1], lm_pred_steep$coeff[2],
    col = 2, lwd = 2)

## Error in abline(lm_pred_steep$coeff[1], lm_pred_steep$coeff[2],
col = 2, : object 'lm_pred_steep' not found

abline(0, 1, lwd = 2)

## Error in int_abline(a = a, b = b, h = h, v = v, untf = untf,
...): plot.new has not been called yet
```

The Bayesian node does not improve predictions (**??**); neither node thus seems to produce particularly good point predictions for steepness. However no clear bias seems to be present in the predictions. Predictive error may be in part due to the somewhat biased estimates of steepness in the input data, with potentially low-quality estimates for various life-history parameters in Fishbase further obscuring their relationship with steepness.

## 3.  EXAMPLES

The examples follow some examples used in the bea2014 node, with additional life history data collected for hoki and snapper from the MPI plenary for Primary Industries (2014).

### 3.1  Information gained from life history: Bluenose

This example shows how the steepness predictive distribution is shrunk when life-history information is added to the predictions (**??**). The life-history data particularly truncates the upper end of the distribution, suggesting that steepness values above 0.9 are not consistent with life-history.

```r
bwa <- BH_net$sample(list(
  species = 'Hyperoglyphe antarctica',
  # Maximum length, temperature and
  # maximum depth from Fishbase
  lmax = 140,
  temp = 11,
  depthmax = 1500,
  # Female growth and max age from
  # Horn et al 2010
  linf = 92.5,
  k = 0.071,
  amax = 71
),samples = 1000)


## Error in eval(expr, envir, enclos): object 'mean_BH_z' not
found


# how much information is gained from life-history

bwa.org <- BH_net$sample(list(
  species = 'Hyperoglyphe antarctica'),samples = 1000)


## Error in eval(expr, envir, enclos): object 'mean_BH_z' not
found


ggplot(bwa.org) +
  geom_bar(aes(x=mean_BH_z,y=..density..),fill='grey40') +
  scale_x_continuous(limits=c(0.2,1)) +
  labs(x='Steepness (z)',y='Density')


## Error in ggplot(bwa.org): object 'bwa.org' not found

ggplot(bwa) +
  geom_bar(aes(x=mean_BH_z,y=..density..),fill='grey40') +
  scale_x_continuous(limits=c(0.2,1)) +
  labs(x='Steepness (z)',y='Density')


## Error in ggplot(bwa): object 'bwa' not found
```

### 3.2 Comparing against data

This section compares steepness estimates from various levels of life-history information to available data in Fishbase and steepness estimates by (Myers et al. 1999). The first two examples are taken from the bea2014 fishnet, the other two examples treat NZ stocks specifically, namely hoki and snapper.

```
BH_net_test <- BH_net

#' Plot density histograms
plot_samples <- function(samples,
                         inp_data,
                      species_,pars=c('linf','k','m','mean_BH_z')){
  datas = subset(inp_data,species==species_)

  melted <- melt(samples[,pars])
  data_melted <- melt(datas[,pars])
  ggplot(melted,aes(x=value)) +
    geom_histogram(aes(y = ..density..),data=data_melted) +
    geom_density(fill=hsv(0,0.7,0.7),alpha=0.5) +
    facet_wrap(~variable,scales='free') +
    labs(x='',y='Density') +
    theme(strip.text.x=element_text(size=10))
}
```

### 3.2.1 Atlantic Cod

Steepness estimates for Atlantic Cod are consistently below the value
estimated in Myers et al., and do not seem to fluctuate much with
increasingly detailed life-history information (**??,??,??**).

```
# fit test net
BH_net_test$fit(subset(steep_reduced,species!='Gadus morhua'),impute = T)


## NOTE: Imputing data values successively in the order of the specified nodes
## <environment: 0x8c5a818>
## attr(,"class")
## [1] "Fishnet"


# predictions

preds.nlh <- BH_net_test$sample(list(
  species = 'Gadus morhua'
),samples=1000)

plot_samples(preds.nlh,steep_merged,'Gadus morhua')


## Error in plot_samples(preds.nlh, steep_merged, "Gadus
morhua"): could not find function "melt"


preds.slh <- BH_net_test$sample(list(
  species = 'Gadus morhua',
```

```
  swimmode = 'subcarangiform',
  habit = 'benthopelagic',
  depthmax = 600,
  lmax = 132
),samples = 1000)

plot_samples(preds.slh,steep_merged,
  'Gadus morhua'
)
```

```
## Error in plot_samples(preds.slh, steep_merged, "Gadus
morhua"): could not find function "melt"
```

```
m.k <- mean(as.numeric(subset(steep_merged,
                              subset = species == 'Gadus morhua',
                              select = 'k')$k))

sd.k <- sqrt(var(as.numeric(subset(steep_merged,
                              subset = species == 'Gadus morhua',
                              select = 'k')$k)))

preds.lh <- BH_net_test$sample(dists(
  species = Fixed('Gadus morhua'),
  swimmode = Fixed('subcarangiform'),
  habit = Fixed('benthopelagic'),
  depthmax = Fixed(600),
  lmax = Fixed(132),
  linf = Normal(110,20),
  k = Normal(m.k ,sd.k),
  amax=Fixed(20)
),1000)


plot_samples(preds.lh,steep_merged,'Gadus morhua')
```

```
## Error in plot_samples(preds.lh, steep_merged, "Gadus
morhua"): could not find function "melt"
```

### 3.2.2  Skipjack tuna

The estimated steepness values are clearly too low given the productivity of
skipjack that is assumed in assessments (**??,??,??**).

```
# predictions
BH_net_test$fit(subset(steep_reduced,
                       species!='Katsuwonus pelamis'),
```

```
                  impute=T)


## NOTE: Imputing data values successively in the order of the specified nodes
## <environment: 0x8c5a818>
## attr(,"class")
## [1] "Fishnet"


Kp_pred <- BH_net_test$sample(list(
  species = 'Katsuwonus pelamis',
  family = 'Scombridae'
), samples=1000)

plot_samples(Kp_pred,steep_merged,
  'Katsuwonus pelamis'
)


## Error in plot_samples(Kp_pred, steep_merged, "Katsuwonus
pelamis"): could not find function "melt"


Kp_pred.slh <- BH_net_test$sample(dists(
  species = Fixed('Katsuwonus pelamis'),
  family = Fixed('Scombridae'),
  depthmax = Fixed(260),
  lmax = Fixed(90.5)
),1000)

plot_samples(Kp_pred.slh,steep_merged,
  'Katsuwonus pelamis'
)


## Error in plot_samples(Kp_pred.slh, steep_merged, "Katsuwonus
pelamis"): could not find function "melt"


m.k <- mean(as.numeric(subset(steep_merged,
                    subset = species == 'Katsuwonus pelamis',
                      select = 'k')$k))

sd.k <- sqrt(var(as.numeric(subset(steep_merged,
                    subset = species == 'Katsuwonus pelamis',
                        select = 'k')$k)))

Kp_pred.lh <- BH_net_test$sample(dists(
  species = Fixed('Katsuwonus pelamis'),
  family = Fixed('Scombridae'),
  depthmax = Fixed(260),
```

```
  lmax = Fixed(90.5),
  linf = Normal(80,10),
  k = Normal(m.k ,sd.k)
),1000)

plot_samples(Kp_pred.lh,steep_merged,
  'Katsuwonus pelamis'
)


## Error in plot_samples(Kp_pred.lh, steep_merged, "Katsuwonus
pelamis"): could not find function "melt"
```

### 3.2.3 Snapper

Life history data were taken from the stock assessment plenary, Fishbase, as well as individual Snapper assessments. The assumed values of steepness in assessments are 0.85 and 0.9 for SNA1 and SNA7, respectively. Steepness is fixed at 1 for SNA2, which assumes that recruitment at 20% of virgin biomass is (with 100% certainty) equivalent to unfished recruitment (Mangel et al. 2013).

```
BH_net_test$fit(subset(steep_reduced,
                       species!='Pagrus auratus'),
                impute = T)


## NOTE: Imputing data values successively in the order of the specified nodes
## <environment: 0x8c5a818>
## attr(,"class")
## [1] "Fishnet"


preds.snapper <- BH_net_test$sample(dists(
  species=Fixed('Pagrus auratus'),
  family=Fixed('Sparidae'),
  habit=Fixed('benthopelagic'),
  depthmax=Fixed(200),
  m=Fixed(0.075),
  amat=Fixed(4),
  lmat=Fixed(24),
  amax=Fixed(60),
  linf=Normal(65,5),
  k=Normal(0.1, 0.05)),
  samples=1000)

plot_samples(preds.snapper,steep_merged,
    'Pagrus auratus')
```

```
## Error in plot_samples(preds.snapper, steep_merged, "Pagrus
auratus"): could not find function "melt"
```

The snapper predictions from the Beverton-Holt steepness node seem to provide sensible estimates (**??**), with a distribution centered around 0.73, close to estimates used in the assessments for SNA1 and SNA7. However, the density near 1 is very low, suggesting that the SNA7 assumption of $z = 1$ is not consistent with life-history traits.

### 3.2.4 Hoki

Hoki life-history parameters were extracted from the 2012 assessment of HOK1 (**mpi_2014_fisheriesi**). The steepness value in the assessment is 0.75, reduced from 0.9 (used prior to 2008). Estimates from the steepness node suggest that even 0.75 is at the upper limit of what life history correlates would suggest (**??**).

```
BH_net_test$fit(subset(steep_reduced,
    species!='Macruronus novaezelandiae'),
    impute = T)


## NOTE: Imputing data values successively in the order of the specified nodes
## <environment: 0x8c5a818>
## attr(,"class")
## [1] "Fishnet"


preds.hoki <- BH_net_test$sample(dists(
    species=Fixed('Macruronus novaezelandiae'),
    family=Fixed('Merlucciidae'),
    swimmode=Fixed('subcarangiform'),
    habit=Fixed('benthopelagic'),
    depthmax=Fixed(1000),
    m=Lognormal(log(0.298)-0.5*log((0.153^2)+1),
        sqrt(log((0.153^2)+1))),
    lmat=Fixed(70),
    amax=Fixed(25),
    linf=Normal(100,10),
    k=Normal(0.2,0.06)),
    samples=1000)

plot_samples(preds.hoki,steep_merged,
    'Macruronus novaezelandiae')


## Error in plot_samples(preds.hoki, steep_merged, "Macruronus
novaezelandiae"): could not find function "melt"
```

## 4. DISCUSSION

Examples for B-H steepness for individual species gave reasonable results for New Zealand stocks, but gave questionable results in the case of Atlantic Cod and especially skipjack tuna. The latter is especially surprising seeing that the dataset includes data from an analysis of tuna stocks. Since the steepness nodes did not produce high predictive power in jackknife tests, the accuracy of the estimates in these examples is unlikely to be good.

This document analysed data using biased B-H estimates; a companion document shows that the unbiased Ricker estimates provide a slightly better basis for a predictive node. In that case, the `Brter` node suggested some predictive power, suggesting that the bias in B-H estimates contributed to obscuring of relationships with life-history parameters. Estimates for skipjack were also notably improved with the Ricker S-R data.

### 4.1 Can the steepness node be improved?

The biggest improvement for a steepness node would most likely come from a new meta-analysis of steepness values, using methods outlined in Michielsens & McAllister (2004). Since both the data and methods are available, this avenue holds the most promise for improving steepness estimates. This could have a substantial benefits for stock assessments, not just in New Zealand, but elsewhere as well. Collecting available S-R data over a range of species from primary and grey-literature, in addition to data available in the RAM legacy database would be an important part of such an analysis.

An alternative to using meta-analysis estimates to inform an empirical node would be to perform simulations based on theory (Mangel et al. 2010, 2013), with species for which sufficient data are available. This subset is probably small, and the simulations would require assumptions about larval and early-juvenile survival, however they could give consistent priors for both Ricker and B-H S-R functions.

### 4.2 Suggested developments for fishnets

Suggestions relating to packaging and ensuring code integrity in the open source context:

- The code is still somewhat cryptic, and the proto-type object structure will be difficult to digest by potential users that are more familiar with the classical R object systems such as S3 and S4. The proto approach has its advantages, and writing some good documentation, along with a vignette that explains the object system, would be a good way to make the package more accessible. The documentation would, I imagine, be somewhat more difficult given that packages like `roxygen2` or others used for R documentation do not integrate automatically with this structure.

- Distributed development can easily lead to modifications by one developer breaking code that others are working on. Having a testing and deployment routine that involves tests over all main features in fishnets would be extremely helpful and could be considered part of the package, so package builds fail if tests fail (and pull requests should not be accepted if the package does not build successfully). The `testthat` package would be an ideal way to achieve this.

Other suggestions:

- Fully Bayesian outputs from the prediction process (i.e., posterior predictive distributions as priors from the `Bayser` node), which would provide more consistent priors in a strict Bayesian sense.

- A hierarchical model parser, which would take a linear fishnet (i.e., the graph) and transform it into a Bayesian hierarchical model. The fully Bayesian approach has the advantage that estimation and predictions would be consistent over the whole graph, with likelihoods at one level of the graph conditional on information at other levels.

# 5. REFERENCES

Dorn, M.W. (2002). Advice on west coast rockfish harvest rates from bayesian meta-analysis of stock- recruit relationships. *North American Journal of Fisheries Management 22(1)*: 280–300.

for Primary Industries, M. (2014). Fisheries assessment plenary, may 2014: stock assessments and stock status. Compiled by the Fisheries Science Group, Ministry for Primary Industries, Wellington, New Zealand. 1381 p.

Forrest, R.E.; McAllister, M.K.; Dorn, M.W.; Martell, S.J.; Stanley, R.D. (2010). Hierarchical bayesian estimation of recruitment parameters and reference points for pacific rockfishes (sebastes spp.) under alternative assumptions about the stock–recruit function. *Canadian Journal of Fisheries and Aquatic Sciences 67(10)*: 1611–1634.

Froese, R.; Pauly, D. (Eds.). (2014). FishBase. Retrieved 9 December, 2014, from http://www.fishbase.org

He, X.; Mangel, M.; MacCall, A. (2006). A prior for steepness in stock-recruitment relationships, based on an evolutionary persistence principle. *Fishery Bulletin 104(3)*: 428–433.

Hoyle, S.; Kleiber, P.; Davies, N.; Langley, A.; Hampton, J. (2011). Stock assessment of skipjack tuna in the Western and Central Pacific Ocean. Report of the Western and Central Pacific Fisheries Commision, Number WCPFC-SC7-2011/SA-WP-04, 133 p. Retrieved 9 December, 2014, from https://www.wcpfc.int/system/files/SC-7-SA-WP-04%20[SKJ%20Assessment-rev1].pdf

Mangel, M.; Brodziak, J.; DiNardo, G. (2010). Reproductive ecology and scientific inference of steepness: a fundamental metric of population dynamics and strategic fisheries management. *Fish and Fisheries 11(1)*: 89–104.

Mangel, M.; MacCall, A.D.; Brodziak, J.; Dick, E.J.; Forrest, R.E.; Pourzand, R.; Ralston, S.; Rose, K. (2013). A perspective on steepness, reference points, and stock assessment. *Canadian Journal of Fisheries and Aquatic Sciences 70(6)*: 930–940.

Michielsens, C.G.; McAllister, M.K. (2004). A bayesian hierarchical analysis of stock recruit data: quantifying structural and parameter uncertainties. *Canadian Journal of Fisheries and Aquatic Sciences 61(6)*: 1032–1047.

Myers, R.A.; Bowen, K.G.; Barrowman, N.J. (1999). Maximum reproductive rate of fish at low population sizes. *Canadian Journal of Fisheries and Aquatic Sciences 56(12)*: 2404–2419.

Rose, K.A.; Cowan, J.H.; Winemiller, K.O.; Myers, R.A.; Hilborn, R. (2001). Compensatory density dependence in fish populations: importance, controversy, understanding and prognosis. *Fish and Fisheries 2(4)*: 293–327.

Ru, H.; Martino, S.; Chopin, N. (2009). Approximate Bayesian inference for latent Gaussian models using integrated nested Laplace approximations (with discussion). *Journal of the Royal Statistical Society, Series B 71*: 319–392. Retrieved 9 December, 2014, from http://www.r-inla.org/

Shertzer, K.W.; Conn, P.B. (2012). Spawner-recruit relationships of demersal marine fishes: prior distribution of steepness. *Bulletin of Marine Science 88(1)*: 39–50.