



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

Authors:
Philipp Neubauer



PO Box 27535, Wellington 6141
New Zealand
dragonfly.co.nz



Cover Notes

To be cited as:

Neubauer, P. (2014). Improving priors for the steepness of the stock-recruitment function within the R package *fishnets*, 11 pages. .

1. INFERRING STEEPNESS: AVAILABLE LITERATURE AND ESTIMATION IN THE CONTEXT OF FISHNETS

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 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 in $[0.2; 1]$ while the Ricker S-R function admits values between 0.2 and infinity (i.e., recruitment can increase above R_0 at $0.2B_0$).

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, which 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) being the exception. 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 in Myers et al. 1999 therefore gives misleading (i.e., overly conservative) results.

Short of doing a new meta-analysis of steepness from the RAM legacy DB or other data source, Myers et al. (1999) was nevertheless considered an appropriate starting point for an empirical node as the reported estimates for the maximum lifetime reproductive rate make it possible to re-calculate the Ricker steepness (instead of the Beverton-Holt steepness reported in their paper). However, given the potentially strong selection bias in the study (for managed, VPA assessed stocks), it seemed less than ideal to use species or family level estimates from reported values directly. Rather, I attempted to use these estimates in combination with life-history to build a predictive model for steepness.

Theory suggests that recruitment variability and natural mortality 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 mortality rates, and Rose et al. (2001) provided evidence that empirical estimates of steepness follow expectations from life-history theory. Von Bertalanffy

growth and fecundity were therefore considered in addition to recruitment variability and natural mortality in the development of the empirical steepness node. Furthermore, 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 start.

The biggest issue with an empirical node is that the only data that span a range of families and orders remains the somewhat outdated Myers et al. (1999). The Myers dataset is the largest, but only gives unbiased estimates for the Ricker model steepness (i.e., after back-transforming and using equations in Michielsens & McAllister (2004) to calculate Ricker steepness). Shertzer & Conn (2012) obtain priors for demersal fish, but only provide an ad-hoc fit of truncated normal and beta distributions to estimates of steepness from stock assessments. Other datasets that provide more formally derived priors for steepness (e.g., Michielsens & McAllister (2004), Dorn (2002) and Forrest et al. (2010)) are too narrow to inform over broad enough a taxonomic range for an empirical node. Data from Forrest et al. (2010) were nevertheless included as estimates for US/Canada west-coast rockfish, in hope to obtain more robust estimates of empirical relationships with life-history parameters. However, the fundamental problem of inadequate sources of empirical estimates of steepness for the B-H function steepness means that it was be difficult to construct an empirical node for the B-H.

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 in 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.*. The combined dataset is available in the *datasteepness* directory as *Myers_et_al_steepness_extended.csv*. Data at the family level as also extracted and saved into *Myers_et_al_steepness_families_extended.csv*.

A data node for steepness was implemented, such that

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

source('collate.R')

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

##           species populations log_alpha_tilde SE_log_alpha_tilde
## 1  Harpodon nehereus           1           0.31           0.07
```

```
## 2 Alosa pseudoharengus      4      1.29      0.09
## 3 Alosa sapidissima         1      1.65      0.30
## 4 Brevoortia tyrannus       1      2.20      0.12
## 5 Alosa aestivalis         3      2.60      0.55
## 6 Brevoortia patronus      1      1.25      0.16
## var_log_alpha_tilde exp_alpha_hat z_20 z_50 z_80 mean_R_z mean_BH_z
## 1      NA      2.0 NA 0.34 NA 0.35 0.33
## 2      0.00      5.7 NA 0.59 NA 0.80 0.59
## 3      NA     18.5 NA 0.82 NA 2.06 0.82
## 4      NA     24.8 NA 0.86 NA 2.61 0.86
## 5      0.81     31.9 0.71 0.84 0.92 3.19 0.89
## 6      NA      5.3 NA 0.57 NA 0.76 0.57
```

returns the data table. 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 is not restricted to species with z estimates, rather the merge is, in SQL terms, a left join onto the fishbase data:

```
# Load the Fishbase data
fb <- FishbaseWeb$read('data/fishbase-web')
# Limit to the 7940 with both k and linf
fb <- subset(fb,!is.na(k) & !is.na(linf))
# An id column for indexing later
fb$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(steep_merged)

##          species populations log_alpha_tilde SE_log_alpha_tilde
## 1 Abbottina rivularis      NA      NA      NA
## 2 Ablennes hians      NA      NA      NA
## 3 Abramis brama      NA      NA      NA
## 4 Abramis brama      NA      NA      NA
## 5 Abramis brama      NA      NA      NA
## 6 Abramis brama      NA      NA      NA
## var_log_alpha_tilde exp_alpha_hat z_20 z_50 z_80 mean_R_z mean_BH_z
## 1      NA      NA NA NA NA NA NA NA
## 2      NA      NA NA NA NA NA NA NA
## 3      NA      NA NA NA NA NA NA NA
## 4      NA      NA NA NA NA NA NA NA
## 5      NA      NA NA NA NA NA NA NA
## 6      NA      NA NA NA NA NA NA NA
##          order  country sex  genus  family  swimmode ltype
```

```

## 1 Cypriniformes Russian Fed U Abbottina Cyprinidae na NG
## 2 Beloniformes India U Ablennes Belonidae na FL
## 3 Cypriniformes Poland U Abramis Cyprinidae subcarangiform TL
## 4 Cypriniformes Former USSR U Abramis Cyprinidae subcarangiform TL
## 5 Cypriniformes Russian Fed U Abramis Cyprinidae subcarangiform TL
## 6 Cypriniformes Germany U Abramis Cyprinidae subcarangiform TL
## linf k t0 m temp lmat lmatmin lmatmax amat amatmin amatmax
## 1 15.8 0.30 0 NA NA NA 4 5.0 1 NA NA
## 2 123.0 0.61 0 NA 27.0 NA NA NA NA NA NA
## 3 91.4 0.09 0 NA NA NA 25 32.5 NA NA NA
## 4 109.0 0.10 0 NA 6.5 NA NA NA NA 3 5
## 5 42.0 0.23 0 0.17 8.0 NA NA NA NA NA NA
## 6 61.8 0.07 0 NA 8.3 NA NA NA NA NA NA
## a b trophic diet feeding
## 1 NA NA NA <NA> <NA>
## 2 0.00036 3.3220 4.50 secondary hunting macrofauna (predator)
## 3 0.01763 3.0825 3.15 secondary hunting macrofauna (predator)
## 4 0.02500 3.0000 3.15 secondary hunting macrofauna (predator)
## 5 0.01940 3.0365 3.15 secondary hunting macrofauna (predator)
## 6 NA NA 3.15 secondary hunting macrofauna (predator)
## habit migration depthmin depthmax fecundmin fecundmax
## 1 benthopelagic freshwater NA NA NA NA
## 2 reef-associated oceanodromous 0 3 NA NA
## 3 benthopelagic potamodromous 1 NA 20800 309600
## 4 benthopelagic potamodromous 1 NA 20800 309600
## 5 benthopelagic potamodromous 1 NA 20800 309600
## 6 benthopelagic potamodromous 1 NA 20800 309600
## fecundity class lmax amax id dummy
## 1 NA Actinopterygii NA NA 7728 1
## 2 NA Actinopterygii NA NA 3229 1
## 3 759847 Actinopterygii 48.33333 NA 1813 1
## 4 759847 Actinopterygii 48.33333 NA 1814 1
## 5 759847 Actinopterygii 48.33333 NA 1815 1
## 6 759847 Actinopterygii 48.33333 NA 1816 1

```

In order to set up a fishnet that that extracts empirical relationships between steepness and life-history, the merged fishbase/steepness data was aggregated at 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.

```

require(dplyr)

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

# reduce dataset; geometric means for parameters by species
steep_reduced <- steep_merged %>%

```

```

select(species, genus, family, class, order, mean_BH_z, linf, m, fecundity, trophic)
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),
           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 overfitting

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 overfitting of empirical relationships might be interfering with predictive power of the nodes. To investigate this potential problem, I implemented a Bayesian node based on the INLA package in R. Fully Bayesian methods are less prone to overfitting, 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 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 overfitting. 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(steepest_reduced[1:1500,])

# predicting onto new data
example_predictions <- linf$predict(steepest_reduced[1501:nrow(steepest_reduced),])

predict_df <- data.frame(bayser = example_predictions[!is.na(steepest_reduced$lmax),])

linf_brter = Brter(log(linf) ~ family + order + log(lmax) + habit, exp)
# fitting to data
linf_brter$fit(steepest_reduced[1:1500,])

```

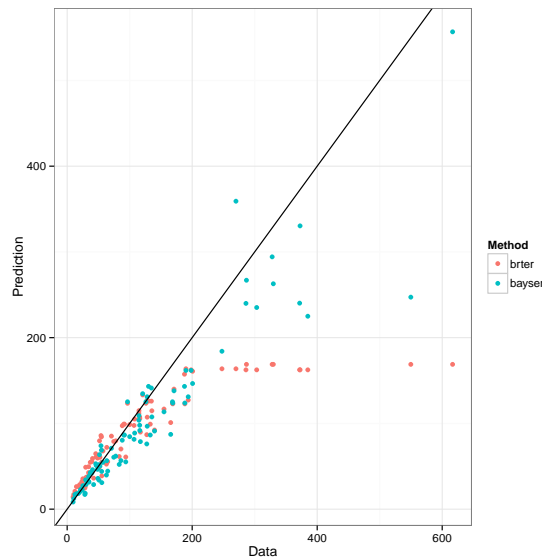


Figure 1: Example predictions for $\ln f$ as a function of $\ln \max$ and taxonomy on a subset of fishbase data, showing that the predictive power in Bayser can be superior to that of Brter when taxonomy is used as a predictor.

```
## [1] fitting gbm model with a fixed number of 2000 trees for log(linf)
## [1] total deviance = 228.52
## [1] residual deviance = 19.65
## predictors: family order log(lmax) habit

# predicting onto new data
example_predictions_brter <- linf_brter$predict(steepest_reduced[1501:nrow(steepest_reduced),])

predict_df <- reshape2::melt(data.frame(brter = example_predictions_brter[!is.na(example_predictions_brter)]))
```

```
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')
```

2.3 Sequential imputing of data in fishnets

In the *Bayser* example above, $\ln f$ depends on $\ln \max$ 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 $\ln f$ depends on $\ln \max$, and k in turn depends on $\ln f$, then $\ln f$ could be imputed for species with missing $\ln f$ based on the node relating it to $\ln \max$. This allows k to be estimated for species with data for $\ln \max$ and missing data for $\ln f$. 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* = *T* option).

```
# build a net for steepness. Use Bayesian nodes in an attempt to not overfit (i.e.,
# transformations are defined for Beverton-Holt S-R parameters to map steepness on

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),

  ln f = Bayser(log(ln f) ~ f(family,model="iid")+f(class,model="iid")+log(lma
```

```

fecundity = Bayser(log(fecundity) ~ f(family,model="iid")+f(class,model="iid") +
k         = Bayser(log(k) ~ f(family,model="iid") + log(linf) + f(habit,model="iid")
m         = Bayser(log(m) ~ f(family,model="iid")+f(class,model="iid")+log(k)+log(l
lmat      = Bayser(log(lmat) ~ f(family,model="iid")+log(k)+log(linf)+f(habit,mod
recsigma  = RecsigmaThorsonEtAl2014(),
mean_BH_z = Brter(logit_BH(mean_BH_z) ~ habit + log(linf) + log(k) + log(m)+ log
)

# fit the BH_et to the summarised fishbase data
BH_net$fit(steepest_reduced,impute = T)

## NOTE: Imputing data values successively in the order of the specified nodes
## [1] fitting gbm model with a fixed number of 3500 trees for logit_BH(mean_BH_z)
## [1] total deviance = 65.82
## [1] residual deviance = 19.09
## predictors: habit log(linf) log(k) log(m) log(fecundity) recsigma trophic log(d
## <environment: 0xe73f380>
## attr(,"class")
## [1] "Fishnet"

```

3. POTENTIAL FURTHER DEVELOPMENTS

3.1 Can the steepness node be improved?

An alternative to using meta-analysis estimates to inform an empirical node would be to repeat simulations based on theory (Mangel et al. 2010, 2013) with species for which sufficient data is available. This subset is probably small, and the simulations would require assumptions about larval/early-juvenile survival. This would be a much trickier start but give consistent priors for both Ricker and B-H. A regression against life history parameters from fish-base could then make it possible to get a predicted distribution from the fishnet.

3.2 Suggested developments for fishnets

4. 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.
- 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.

- 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.
- 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.
- 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.