

A Bayesian approach for estimating length-weight relationships in fishes

Rainer Froese, GEOMAR Helmholtz-Centre for Ocean Research, Düsternbrooker Weg 20,
24105 Kiel, Germany, rfroese@geomar.de (corresponding author)

James T. Thorson, Fisheries Resource Analysis and Monitoring Division, Northwest Fisheries
Science Center, National Marine Fisheries Service, National Oceanic and Atmospheric
Administration, 2725 Montlake Blvd. East, Seattle, WA 98112-2087, USA,
James.Thorson@noaa.gov

Rodolfo B. Reyes Jr., FIN, G.S. Khush Hall, IRRI, Los Baños, Laguna, PH 4031, Philippines,
r.reyes@fin.ph

Summary

We present a Bayesian hierarchical approach to the estimation of length-weight relationships (LWR) in fishes. In particular, we provide prior estimates for the LWR parameters a and b in general and by body shape. We use these priors and existing LWR studies to derive species-specific LWR parameters. In the case of data-poor species, we include in the analysis LWR studies of closely related species with the same body shape. This approach yielded LWR parameter estimates with measure of uncertainty for practically all known 32,000 species of fishes. We provide a large LWR data set extracted from www.fishbase.org, the source code of the respective analyses, and ready-to use tools for practitioners. We present this as an example of a self-learning online database, where the addition of new studies improves the species-specific parameter estimates, and where these parameter estimates inform the analysis of new data.

Keywords

Length-weight relationships and data, Bayesian statistics, ichthyology, data-poor species, FishBase

32 **Introduction**

33 For convenience, size in fishes is often measured in body length. However, management for
34 fisheries or conservation requires information about body weight for regulation of catches and
35 estimation of biomass. Weight (W) can be predicted from length (L) with the help of length-
36 weight relationships (LWR) of the form $W = a L^b$, where parameter b indicates isometric
37 growth in body proportions if $b \sim 3$, and a is a parameter describing body shape and condition
38 if $b \sim 3$ (Froese 2006). FishBase (Froese and Pauly 2012) has compiled LWR parameters for
39 thousands of species of fishes. However, usage of published LWRs brings up three questions:
40 1) If there are many studies for a species, how can this information be meaningfully combined
41 into a joint LWR? 2) If there is only one study for a given species, how well does this study
42 represent the variability that is to be expected? 3) How can existing studies inform a new
43 LWR estimate derived from new data? The aim of this paper is to apply hierarchical Bayesian
44 inference to answer these questions. We present web tools that facilitate the application of the
45 methods by practitioners and that provide the basis for a self-learning online database.

46

47 **Material and Methods**

48 We first describe our general approach to the analysis. We then describe in more detail the
49 data and the statistical models.

50

51 **General approach**

52 Bayesian methods combine existing knowledge (prior probabilities) with additional
53 knowledge derived from new data (the likelihood function). This results in updated
54 knowledge (posterior probabilities), which can be used as priors in subsequent analyses and
55 thus provide learning chains in science (Kuikka et al., 2013). Note that the standard deviation
56 (SD) of a posterior distribution for a parameter represents the uncertainty about the sampling
57 distribution and thus is a standard error (SE) by definition.

58

59 We first established broad overall priors for parameters a and b , based on textbooks and
60 reviews (step 1 below). We then estimated posterior distributions for model parameters for
61 fishes in general by analyzing the distribution of a and b in a large data set of LWR studies
62 (step 2). We further refined the estimated posterior distributions by grouping fish species into
63 body-shape groups, from eel-like to short & deep, and estimating the parameters for each

64 individual group (step 3). We used the body-shape posteriors as priors for the analysis of
65 studies done for a given species (step 4). In data-poor species, we used the model to learn also
66 from studies done on related species with the same body shape, i.e., we applied multivariate
67 hierarchical Bayesian inference, treating each species as its own hierarchical level (step 5). As
68 a result we obtained LWR parameter estimates for practically all fish species, with indication
69 of uncertainty of the parameters and of the weight predicted from length. These species-
70 specific parameters can then be applied directly, or they can serve as priors in the analysis of
71 new weight-at-length data (step 6). FishBase (www.fishbase.org) contains online tools that
72 incorporate these steps and facilitate the analysis of existing parameters and of new weight-at-
73 length data (see also Web Tools section in the Appendix).

74

75 *Step 1: Getting overall priors for LWR parameters a and b , based on the literature:*

76 Parameter b is the slope of a regression line over log-transformed weight-at-length data. It is
77 considered to be normally distributed (Carlander 1969). Parameter b should average
78 approximately 3 in species that do not change body shape as they grow (Spencer 1864-1867)
79 and usually falls between 2.5 and 3.5 (Carlander 1969). This information is interpreted here as
80 a normally distributed prior for b with mean = 3 and SD = 0.5. Parameter a is the intercept of
81 a regression line over log-transformed weight-at-length data. It is considered to be log-
82 normally distributed (Carlander 1977) and reflects the body-shape of the species (Froese
83 2006). With weight in gram and length in centimeter, $a = 0.01$ represents a fusiform fish,
84 bracketed by $a = 0.001$ in eel-like fish and $a = 0.1$ in spherical fish (Froese 2006). This
85 information is here interpreted as a normally distributed prior of $\log_{10}(a)$ with mean = -2 and
86 SD = 1.

87

88 *Step 2: Getting parameter estimates across all available LWR studies*

89 LWR studies compiled in FishBase were used to obtain across-all-studies distributions for
90 parameters a and b . A score reflecting the reliability of a study (see below) was used as
91 weighting factor. The overall priors from step 1 were used in this analysis. For the
92 measurement error in length and weight we assumed an uninformative prior (Gelman 2006).

93 In this analysis, a and b estimates for each individual species were considered as co-
94 varying within the bounds of the species-specific body plan. However, for the across species
95 analysis, a and b were considered as not correlated (see also Discussion). Looking at within
96 and across species variability allowed for decomposing the total variability into measurement

97 error and predictive error, where the latter is a combination of true natural variability and the
98 error resulting from the LWR model only approximating the true relationship between length
99 and weight. The predictive posterior parameter distributions arising from this across-all-
100 studies-and-species analysis can be used as priors in single species analysis where body shape
101 information is missing or does not match any of the shapes defined below.

102

103 *Step 3: Getting parameter estimates by body shape group*

104 Based on available drawings, photos or morphometric data, FishBase staff has assigned
105 species to the body shape groups *eel-like*, *elongated*, *fusiform*, and *short & deep*. The
106 approach described in step 2) was used for each of these body shape groups. The
107 measurement and predictive error distributions resulting from this analysis were used as
108 respective priors in the subsequent steps. The parameter and error distributions resulting from
109 this analysis were used as priors for single species analysis within the respective body shape
110 group, see below.

111

112 *Step 4: Getting joint parameter estimates for a species*

113 For species with many available LWR studies, the parameters a and b from these studies were
114 considered as negatively correlated due to well-known correlations between intercept and
115 slope induced by common estimation methods (Peters 1983). The a and b values were
116 analysed together with the priors from the respective body shape group (see Single-Species
117 model below). The resulting species-specific parameter estimates can then either be used
118 directly for predicting weight from length, or they can serve as priors for a new LWR study.

119

120 *Step 5: Getting parameter estimates for species with few available studies*

121 For species with few available studies (e.g. less than 5), information from related species
122 (species in the same Genus, Subfamily or Family and with the same body shape) was used in
123 a hierarchical analysis. First, parameters were derived for every related species, as in step 4).
124 Then these parameters, together with the body shape priors, were used to derive the parameter
125 estimates for the target species (see Few-Studies model below). The resulting species-specific
126 parameter estimates can then either be used directly for predicting weight from length, or they
127 can serve as priors for a new LWR study.

128

129 *Step 6: Using parameter estimates as priors in the analysis of new weight-at-length data*

130 For analysis of new weight-at-length data, the posteriors of the parameter analysis for the
131 respective species (steps 4 or 5) can be used as priors. If no previous LWR study exists for the
132 species, then the body shape priors (from step 3) can be treated as if they were an existing
133 study, and the parameter analysis of step 5 can be run to updated the body shape priors with
134 information from related species. If there are no LWR estimates for related species, the body
135 shape priors can be used instead of species-specific priors. Additionally, if no previous LWR
136 study exists and the body shape does not match the available choices, then generic priors
137 (from step 2) can be used. The analysis of new weight-at-length data is done with a Bayesian
138 linear regression of $\log_{10}(W)$ as a function of $\log_{10}(L)$, weighted by number of individuals,
139 with priors as indicated above. The analysis assumes a raw data set that has been cleansed
140 beforehand of extreme outliers.

141

142 **Data**

143 For steps 2-5, we analyzed LWR parameters compiled in FishBase 12/2012. We only used
144 studies of species that had independently assigned body shapes (eel-like, elongated, fusiform,
145 short & deep) and where length measurements were reported in total length or fork length.
146 Additionally, we only included studies where the parameters were estimated with type-I linear
147 regression of log-transformed weights and lengths. Finally, we excluded studies that were
148 marked by FishBase staff as questionable. This data filtering yielded 5150 studies for 1821
149 species (see Table 1).

150 We assigned scores (S) that represent data quality for each study. These were
151 subsequently used to downweight information from studies that were deemed less reliable
152 than others, and ranged from 0.5 to 1 using the following scoring guide:

- 153 • If a coefficient of determination (r^2) was given by the study, then $S = r^2$
- 154 • Else, if the length range of the raw data was indicated, then $S = 0.7$
- 155 • Else, if the number of measured specimens was > 10 , then $S = 0.6$
- 156 • Else, $S = 0.5$

157 Thus, a high-quality study (i.e. with a high coefficient of determination) received about
158 double the score of a study that just presented the parameters a and b without additional
159 information. This data file is available for download, see Table 5.

160

161 **Statistical models**

162 We used the R statistical package with libraries r2jags (Su & Yajima 2012) and the JAGS
 163 sampler software (Plummer 2003) for conducting the Bayesian analyses, called from the R
 164 Statistical Environment (R Development Core Team 2011). These packages are open source
 165 and freely available on the Internet. The models used in steps 2-6 above are described below
 166 in more detail. Logarithmic transformation of length and weight data can be done with any
 167 base. For convenience, we used natural logarithms in the model description below. In the R-
 168 code and the resulting graphs we used base-10 logarithms, because this facilitates the reading
 169 of log-axes, with $\log_{10}(a) = -3$ giving $a = 0.001$, $\log_{10}(L) = 2$ giving $L = 100$ cm, etc. For
 170 presentation of the models, we also adopted the convention that all parameters are represented
 171 by Greek letters while all data are represented by Latin letters. Thus, in the following section
 172 formally describing the models, a and b from existing LWR studies are considered data,
 173 whereas α and β represent the respective parameters estimated by the models. We additionally
 174 specify that the character i is reserved for indices.

175

176 **The Body-Shape model**

177 The Body-Shape model uses the species-specific measure of a_s and b_s for each available study
 178 i_s , as well as the associated quality score S_s and binomial genus-species gs_s (the subscript s
 179 stands for ‘study’, and each variable with subscript s has an individual value for each
 180 observation in the database). Each scientific name is associated with a body-shape, bs_{gs} ,
 181 where i_{gs} is an index associated with each unique species (the subscript gs standards for
 182 ‘genus-species’, and each variable with subscript gs has an individual value for each unique
 183 species in the database). The model estimates a ‘true’ but unobserved value for each species in
 184 the dataset, $\log_{10}(\alpha_{gs})$ and β_{gs} . These vary around their average value for a given body-shape,
 185 α_{bs} and β_{bs} , where i_{bs} is an index associated with each of four body-shape types (the variable
 186 bs standards for ‘body-shape’ and each variable with subscript bs has an individual value for
 187 each unique body-shape in the database). Parameters $\log_{10}(\alpha_{gs})$ and β_{gs} for each species vary
 188 around the average value for their body shape according to a normal distribution, with a
 189 separate variance $\tau_{\log\alpha}^2$ and τ_{β}^2 for $\log_{10}(\alpha)$ and β :

$$190 \quad \log_{10}(\alpha_{gs}) \sim \text{Normal} \left(\sum_{i_{bs}=1}^4 \log_{10}(\alpha_{bs}) \cdot I(bs_{gs} = i_{bs}), \tau_{\log\alpha}^2 \right) \quad (1)$$

$$191 \quad \beta_{gs} \sim \text{Normal} \left(\sum_{i_{bs}=1}^4 \beta_{bs} \cdot I(bs_{gs} = i_{bs}), \tau_{\beta}^2 \right) \quad (2)$$

192 where $I(bs_{gs}=i_{bs})$ is an indicator function that equals one when bs_{gs} equals i_{bs} and zero
 193 otherwise, and $\text{Normal}\left(\sum_{i_{bs}}^{n_{bs}} \log_{10}(\alpha_{bs}) \cdot I(bs_{gs} = i_{bs}), \tau_{\log \alpha}^2\right)$ is normal distribution with mean
 194 $\sum_{i_{bs}}^{n_{bs}} \log_{10}(\alpha_{bs}) \cdot I(bs_{gs} = i_{bs})$ and variance $\tau_{\log \alpha}^2$ (we define other normal distributions
 195 similarly).

196 LWR parameter estimates are known to be negative correlated (Froese 2006), i.e., in a
 197 log-log plot of weight over length for a given species, an increase in the slope of the
 198 regression line will result in a decrease of the intercept on the weight axis, and vice-versa. We
 199 accounted for this correlation between $\log_{10}(a)$ and b within each study by specifying that
 200 study-specific observations vary around the 'true' but unobserved species-specific value
 201 according to a multivariate normal distribution.

$$202 \langle \log_{10}(a_s), b_s \rangle \sim \text{MVN}\left(\left\langle \sum_{i_{gs}=1}^{n_{species}} \log_{10}(\alpha_{gs}) I(gs_s = i_{gs}), \sum_{i_{gs}=1}^{n_{species}} \beta_{gs} I(gs_s = i_{gs}) \right\rangle, \Sigma_s\right) \quad (3)$$

203 where Σ_s is the measurement error covariance for observation s , which is composed of
 204 measurement error variance $\sigma_{\log a}^2$ and σ_b^2 for $\log_{10}(a)$ and b , as well as the correlation ρ in
 205 measurement errors:

$$206 \Sigma_s = S_s^{-2} \begin{vmatrix} \sigma_{\log a}^2 & \rho \sigma_{\log a} \sigma_b \\ \rho \sigma_{\log a} \sigma_b & \sigma_b^2 \end{vmatrix} \quad (4)$$

207 This measurement error covariance varies among studies such that measurement errors are
 208 greater for low-scoring studies. Using a multivariate distribution has previously been shown
 209 to reduce the uncertainty of the parameter estimates (Pulkkinen et al. 2011).

210 Parameters are given priors, as is necessary for any Bayesian analysis. Specifically,
 211 standard deviation parameters $\tau_{\log a}$, τ_b , $\sigma_{\log a}$, and σ_b , were given initially broad inverse-gamma
 212 (0.001, 0.001) priors, and measurement error correlation ρ was given a uniform negative prior
 213 from -0.99 to 0. Prior distributions for each body shape α_{bs} and β_{bs} were defined as described
 214 previously.

215

216 **Across-all-Observations-and-Species model**

217 The model for all observations and species but without body-shape is identical to the
 218 preceding Body-Shape model, with one exception. Specifically, the vector bs is replaced with
 219 a dummy vector I , which has the value one for all entries. This change implies that all

220 species in this model have the same value for $\log_{10}(\alpha_{bs})$ and β_{bs} . It consequently provides an
221 average value for $\log_{10}(\alpha)$ and β for species for which the body-shape is unknown.

222

223 **The Few-Studies model**

224 The Few-Studies model uses the same set of equations (Eq. 1-4) as the Body-Shape model,
225 but incorporates the following changes. First, it replaces the broad priors for $\log_{10}(\alpha_{gs})$ and β_{gs}
226 with more informative priors estimated from the previous Body-Shape analysis. Second, it
227 replaces the uninformative priors for between-species ($\tau^2_{\log\alpha}$ and τ^2_{β}) and measurement error
228 variance ($\sigma^2_{\log\alpha}$ and σ^2_{β}) with informative priors. Specifically, it specifies a gamma
229 distribution for the standard deviation of between-species and measurement error variability,
230 and parameterizes it such that the mean and standard deviation of this gamma distribution
231 match the posterior mean and standard deviation from the Body-Shape model.

232

233 **The Single-Species model**

234 The Single-Species model uses a reduced set of equations (Eq. 3-4) from the Body-Shape
235 Model. It assumes that previous LWR studies for the species are sufficiently numerous and
236 informative so that no inclusion of data from other related species is needed. It uses priors for
237 $\log_{10}(\alpha)$ and β and for the standard deviation of measurement errors based on the Body-Shape
238 model.

239

240 **The New Weight-at-Length-Data model**

241 The model for new weight-at-length data uses the individual observations of length l_j and w_j
242 for n_{obs} fish observations. Specifically, it specifies the base-10 logarithm of weight as a
243 function of the base-10 logarithm of length:

$$244 \log_{10}(w_j) \sim \text{Normal}\left(\log_{10}(\alpha_{gs}) + \beta_{gs} \log_{10}(l_j), \sigma^2_{\log w}\right) \quad (5)$$

245 where $\sigma^2_{\log w}$ is the residual log-normal variance in the LWR. We additionally specify that the
246 priors for α_{gs} and β_{gs} match the estimated posteriors from the Few-Studies or Single-Species
247 models.

248

249 **Results and Discussion**

250 We sought to estimate LWR parameter distributions for $\log_{10}(a)$ and b for a hypothetical
251 species of a given body-shape, while accounting for correlations between $\log_{10}(a)$ and b for

252 observations within a given species, but not between species. We made this distinction
253 because clearly, for a species with a given body shape (which determines a) and a given life
254 history strategy how much this shape changes as the fish grows (which determines b),
255 intercept $\log_{10}(a)$ and slope b cannot but co-vary within the narrow bounds of log-transformed
256 weight-at-length data. Accounting for this negative correlation reduces the uncertainty of the
257 parameter estimates (Pulkkinen et al. 2011). However, other species may have different body
258 shapes but the same growth strategy. For example, an eel will have a thin, long body which
259 fills only a small fraction ($= a$) of a cube with a length equal to the eel's body length. In
260 comparison, a box fish is likely to fill a substantial fraction of its respective cube, resulting in
261 a much higher value of a . This high a , however, does not mean that the boxfish will have a
262 lower b than the eel. This reasoning is confirmed by the results of the body shape analysis
263 shown in Table 1, where the 95% ranges of a values are far apart between eel-like and short &
264 deep body shapes, but the 95% b ranges are nearly identical.

265 We used a hierarchical model that estimates mean and between-species variability in
266 $\log_{10}(a)$ and b for each body-shape. The model then estimates $\log_{10}(a)$ and b for each species
267 with the respective body shape, while shrinking estimates for poorly-estimated species
268 towards their body-shape mean (Gelman and Hill 2007). Essentially, the model uses multiple
269 observations within each species to estimate the 'measurement errors' for the average LWR
270 study. Variability between-species in excess of these 'measurement errors' is then attributed to
271 a 'process error' that arises due to natural between-species variability in $\log_{10}(a)$ and b (Clark
272 2003). Additionally, systematic differences in $\log_{10}(a)$ and b between body-shapes were
273 ultimately attributed to effects stemming from different body plans.

274 Figure 1 shows histograms of parameters a and b across all studies. The overlaid bold
275 normal probability density curves use mean and standard deviation of the data and confirm
276 that $\log_{10}(a)$ and b are approximately normally distributed. Figure 1 also shows nicely the
277 updating of prior beliefs from the initial broad estimates derived from textbooks (dashed
278 curve), to the observed variability in 5150 data sets (bold curve), to the predictive distribution
279 (dotted curve) which excludes measurement errors. The narrower posterior distribution
280 especially for parameter b confirms observations by Carlander (1977) and Froese (2006) that
281 strong deviations from $b=3.0$ often stem from questionable studies with few specimens,
282 narrow length ranges, or low explained variability.

283

284 Table 1 shows weighted means and standard deviations by body-shape group for the LWR
 285 studies compiled in FishBase 12/2012. For all body-shape groups, mean b values were close
 286 to 3, confirming that most fish do not change their body shape as adults (Froese 2006).
 287 However, geometric mean a values clearly differed between body-shape groups, from $a =$
 288 0.001 in eel-like fishes to $a = 0.02$ in short & deep fishes, confirming the pattern proposed by
 289 Froese (2006). Table 2 gives the measurement and process errors, respectively.

290 For the estimation of parameter distributions by species we used the weighted means and
 291 standard deviations of the respective body-shape group as priors. We assumed that differences
 292 in parameter estimates between different studies for a given species were mostly caused by
 293 different sample size structure or season rather than by different localities (Froese 2006).
 294 Therefore we treated all populations of a species as being of the same hierarchical level with
 295 respect to LWR. We applied this approach to 48 weighted LWR studies of the European
 296 Anchovy *Engraulis encrasicolus*. The resulting joint parameters had reasonably narrow
 297 distributions shown in Figure 2, with means (peak of continuous curve) that did not deviate
 298 significantly from the means of the data (indicated by the single points).

299 Note that the posterior standard deviation of $\log_{10}(a)$ is also the standard error of body
 300 weight predicted from length. For example, using the parameters estimated for European
 301 anchovy in Figure 2, the mean weight predicted for 12 cm total length is given by

$$302 W_{mean} = 10^{-2.26 + 3.04 \log_{10}(12)} = 10.5$$

303 and the range that is likely to contain 95% of the variability in weight is given by

$$304 W_{range} = 10^{(-2.26 + 3.04 \log_{10}(12) \pm 1.96 \times 0.0399)} = 8.8 - 12.6$$

305 For the estimation of parameter distributions by species and related species (congeners or
 306 Family members with the same body-shape), we applied multivariate hierarchical Bayesian
 307 inference, treating each species as its own hierarchical level. In other words, we did not use
 308 hierarchical levels for Genus- or Family-groups, because we considered the deviation of the
 309 body shape of a species from the mean shape of its Genus or Family-group not as an error but
 310 as a true manifestation of differences between species. Again, we assumed a correlation
 311 between parameters a and b within species, but we treated these parameters as independent
 312 when summarizing across species.

313 An example of a species with a single LWR study in FishBase was the Pacific short-
 314 finned eel, *Anguilla obscura* (Figure 3). The parameters given were $n=145$, $a = 0.00021$,
 315 $b=3.38$, $r^2=0.99$ (Jellyman 1991), which represents a considerable deviation from the body
 shape means for eel-like fishes of $a = 0.001$ and $b = 3.06$ (Table 3), probably as a case of

316 negative parameter co-variation, i.e., the a estimate appears too low and b too high. In this
317 case, single-species analysis would combine the only study with the information provided by
318 the prior for eel-like species, suggesting $a = 0.00067$ and $b = 3.09$, and thus pulling the
319 parameters suggested by the single study strongly in the direction of the prior. However, other
320 LWR studies for species of the Genus *Anguilla* confirm a deviation from the eel-like prior,
321 although less strongly than suggested by the single study. Including the information from
322 these related species gives $a = 0.00085$ ($0.00058 - 0.0013$) and $b = 3.17$ ($3.07 - 3.26$), which
323 appears to be a meaningful summary of the available information, accommodating the single
324 study under the tails of the proposed parameter distributions (see single points in Figure 3).

325 Finally, we wanted to inform a new analysis of weight-at-length data with parameter
326 estimates from existing studies. If no previous study existed for the target species, then the
327 body shape priors in Tables 1 and 2 would represent the existing knowledge. Otherwise, a
328 parameter analysis as described above was first conducted on the existing studies for the
329 target species, including related species if necessary. This analysis then provided the priors for
330 the new study.

331 For example, we used weight-at-length data for North Sea turbot (*Scophthalmus maximus*)
332 extracted in November 2012 from the DATRAS database (<http://datras.ices.dk>) for the years
333 2010-2012. A plot of $\log_{10}(W)$ over $\log_{10}(L)$ showed one extreme outlier, which we removed.
334 We run a parameter analysis across the 10 existing studies for the species. We used the
335 resulting means and standard deviations for $\log_{10}(a)$, b , and measurement error of $\log_{10}(a)$ as
336 priors for the new analysis. The results are presented in Table 4, which can serve as a model
337 for meaningful reporting of Bayesian LWR analyses in publications.

338 It is interesting to compare the results of the Bayesian LWR analysis with those of a
339 regular linear regression. In our example for turbot, the Bayesian analysis included, in a
340 hierarchical process, information from the body-shape group and from other studies done for
341 the species. In contrast, the regular regression only analyzed the data at hand. The prior means
342 for $\log_{10}(a) = -1.83$ and $b = 3.04$ did not differ much from the means of the data, as provided
343 by regular regression with $\log_{10}(a) = -1.81$, $b = 3.06$, and hence the means provided by the
344 Bayesian analysis were identical to those of the regular regression. However, the prior
345 estimates of uncertainty $SD[\log_{10}(a)] = 0.069$ and $SD[b] = 0.0486$ were considerably wider
346 than those of the regular regression with $SE[\log_{10}(a)] = 0.0271$ and $SE[b] = 0.0187$. In other
347 words, the estimates of uncertainty provided by the regular regression were only
348 representative for the analyzed data, but too narrow if data from other years and areas were

349 considered. The Bayesian analysis incorporated this additional information and provided more
350 realistic estimates of uncertainty that were intermediate between the priors and the data, with
351 $SD[\log_{10}(a)] = 0.0461$ and $SD[b] = 0.0317$.

352

353 **Preliminary LWR parameters for all species of fishes**

354 FishBase 12/12 contained 32,470 species of fishes in 554 Families. However, LWR studies
355 were only available for 3,587 species in 357 Families. Based on the results of this study, the
356 FishBase team assigned preliminary LWR parameters as follows:

- 357 • For the over 2,500 species in the 197 Families without LWR studies, the respective
358 body shape priors (step 3 above) were assigned. If no matching body shape
359 information was available, the overall priors (step 2 above) were assigned.
- 360 • For the over 26,000 species without specific LWR studies but with studies for other
361 species in their Families, the respective body shape priors were treated as if they
362 were an existing study and the parameter analysis of step 5 above was run to
363 updated the body shape priors with information from related species.
- 364 • For the over 3,500 species with existing LWR studies, steps 4 or 5 above were used
365 to estimate representative parameters.

366

367 This approach assigned preliminary LWR parameters to practically all species of fishes,
368 summarizing the best available information. These parameters will be updated whenever new
369 studies are added to FishBase.

370

371 **Conclusion**

372 We present an example of a self-learning online database, where the addition of new studies
373 improves the species-specific parameter estimates, and where these parameter estimates
374 inform the analysis of new data. We used a Bayesian approach to the estimation of length-
375 weight relationships for practically all species of fishes. We show how the use of all available
376 prior information can improve parameter estimates. The increased uncertainty in species with
377 little available data is expressed in wider respective parameter distributions. We make a large
378 standardized data set available for further research. We hope our read-to-use tools will help in
379 spreading the application of Bayesian methods in fisheries.

380

381 **Acknowledgements**

382 We thank Crispina Binohlan for compiling most of the LWR studies used in this study. We
383 thank Josephine Barile and Kimberly Banasihan for implementing the web tools in FishBase.
384 We thank Sakari Kuikka for useful comments. Rainer Froese acknowledges support by the
385 Future Ocean Excellence Cluster 80, funded by the German Research Foundation on behalf of
386 the German Federal State and State Governments. James Thorson acknowledges supportive
387 discussions with J. Cope and W. Patrick regarding the model design. The authors would like
388 to thank the i-Marine project (FP7 of the European Commission, FP7-INFRASTRUCTURES-
389 2011-2, Contract No. 283644) for making available the computational infrastructure that
390 facilitated the computation of LWR estimates for all species in FishBase. Rainer Froese and
391 Rodolfo B. Reyes Jr. acknowledge support from the European Union's Seventh Framework
392 Programme (FP7/2007-2013) under grant agreement no. 244706/ECOKNOWS project.
393 However, the paper does not necessarily reflect the views of the European Commission (EC),
394 and in no way anticipates the Commission's future policy in the area. This is FIN
395 Contribution number 139.

396
397

398 **References**

- 399 Carlander, K. D., 1969: Handbook of freshwater fishery biology, Vol. 1. Ames, IA., The Iowa
400 State University Press, 752 pp.
- 401 Carlander, K. D., 1977: Handbook of freshwater fishery biology, Vol. 2. Ames, IA. The Iowa
402 State University Press, 431 pp.
- 403 Clark, J.S., 2003: Uncertainty and variability in demography and population growth: a
404 hierarchical approach. *Ecology* 84(6), 1370-1381.
- 405 Froese, R., 2006: Cube law, condition factor, and weight-length relationships: history, meta-
406 analysis and recommendations. *Journal of Applied Ichthyology* 22(4), 241-253.
- 407 Froese, R.; Pauly, D. Editors. 2012: FishBase. World Wide Web electronic publication,
408 version (12/2012). Accessed December 1, 2012 at <http://www.fishbase.org>
- 409 Gelman, A., 2006: Prior distributions for variance parameters in hierarchical models.
410 *Bayesian Analysis*, **1**, 515–533.
- 411 Gelman, A.; Hill, J., 2007: Data analysis using regression and multilevel/hierarchical models.
412 Cambridge, UK, Cambridge University Press, 648 p.
- 413 Jellyman, D.J., 1991: Biology of the shortfinned eel *Anguilla obscura* in Lake Te Rotonui,
414 Mitiaro, Cook Islands. *Pac. Sci.*45(4), 362-373.

415 Kuikka, S. ; Vanhatalo, J.; Pulkkinen, H.; Mäntyniemi, S.; Corander, J., 2013: Experiences in
416 Bayesian Inference in Baltic Salmon management. *Statistical Science*, (accepted).

417 Peters, R.H., 1983: *The ecological implications of body size*. Cambridge University Press,
418 New York, 329 p.

419 Plummer, M., 2003: JAGS: A program for analysis of Bayesian graphical models using Gibbs
420 sampling. *Proceedings of the 3rd International Workshop on Distributed Statistical*
421 *Computing (DSC 2003)*, pp 20–22.

422 Pulkkinen, H.; Mäntyniemi, S.; Kuikka, S.; Levontin, P., 2011: More knowledge with the
423 same amount of data: advantage of accounting for parameter correlations in hierarchical
424 meta-analyses. *Marine Ecology Progress Series* 443, 29-37.

425 R Development Core Team, 2011: *R: A language and environment for statistical computing*.
426 Vienna, Austria. Accessed November 1, 2012, <http://www.R-project.org/>.

427 Spencer, H., 1864-1867: *The Principles of Biology*, 2 volumes. Williams & Norgate, London.
428 678, pp.

429 Su, Y.-S.; Yajima, M., 2012: R2jags: A package for running Jags from R. Accessed
430 November 1, 2012, <http://CRAN.R-project.org/package=R2jags>.

431

432

433

434

435

436

437

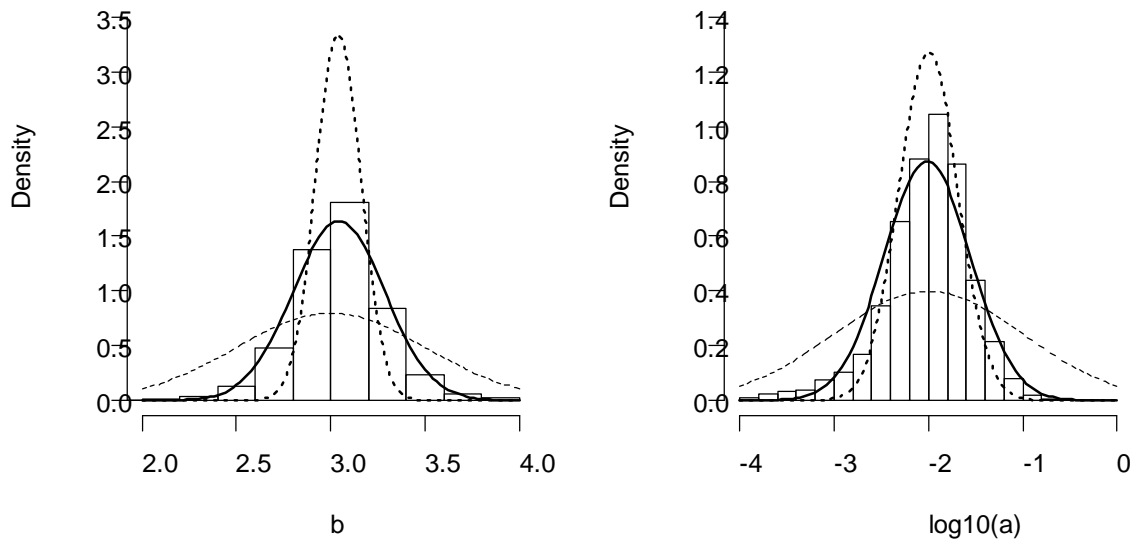
438

439

440

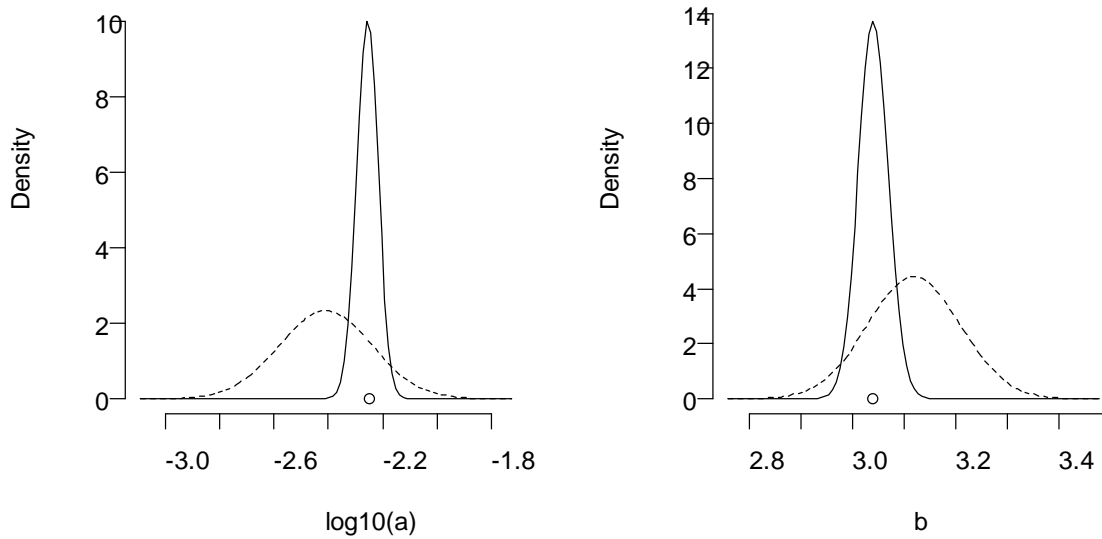
441

442

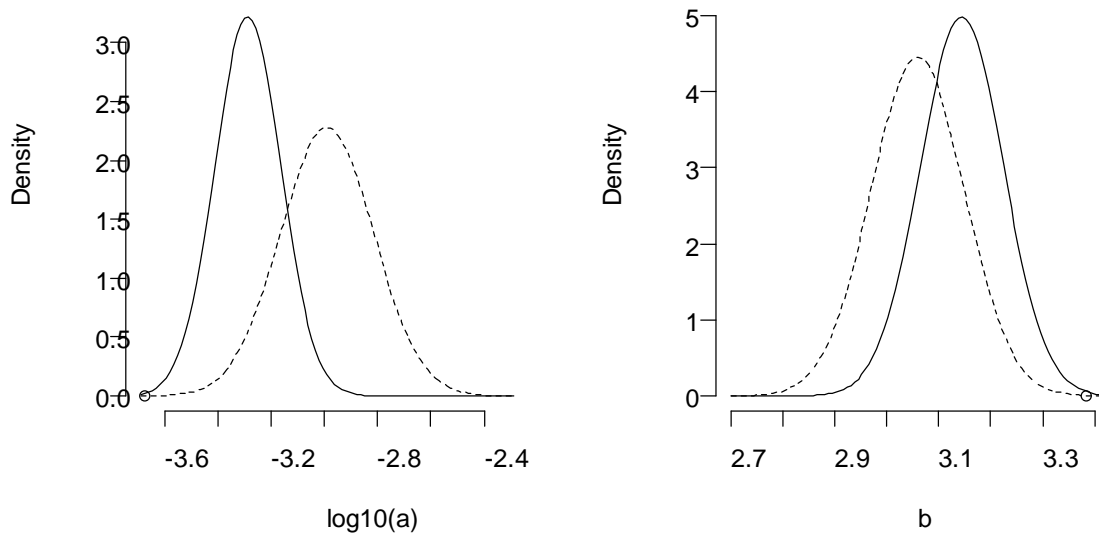


444
 445 Figure 1. Weighted distribution of parameters b and a in 5150 LWR studies for 1821 species
 446 of fishes. The overlaid curves are normal density functions, i.e. the areas under the histograms
 447 and under the curves are identical and equal to 1. The bold normal curves use mean and
 448 standard deviation of the data. They confirm that b and $\log_{10}(a)$ are approximately normally
 449 distributed. The dashed curves represent the overall priors derived from the literature. The
 450 dotted curves represent the predictive posterior distributions. They are narrower because they
 451 represent only the errors in parameter estimation and between-species variability, excluding
 452 measurement errors.

453
 454
 455
 456
 457
 458
 459
 460
 461
 462



463
 464 Figure 2. Distribution of parameters a and b for 48 LWR studies of the European anchovy
 465 *Engraulis encrasicolus*. The single points present the mean values of the data. The dashed
 466 lines indicate the prior distributions for elongated fishes. Mean $\log_{10}(a) = -2.26$, SD of
 467 $\log_{10}(a)$ and $\log_{10}(W) = 0.0397$, geometric mean $a = 0.00554$, 95% range $a = 0.00464 -$
 468 0.00662 , for total length, and mean $b = 3.04$, SD $b = 0.0291$, and 95% credible interval $b =$
 469 $2.98 - 3.1$. The measurement error σ of $\log_{10}(a)$ was mean = 0.255, SD = 0.00319, and of b
 470 was mean = 0.188, SD = 0.00224.
 471
 472



473
 474 Figure 3. Distribution of parameters a and b for one study with $a=0.00021$ and $b=3.38$ for the
 475 Pacific short-finned eel, *Anguilla obscura* (indicated by single points) and 33 LWR studies of
 476 four species of the Genus *Anguilla*. The dashed curves indicate the prior distributions for eel-
 477 like fishes. Resulting mean $\log_{10}(a) = -3.28$, SD of $\log_{10}(a)$ and $\log_{10}(W) = 0.123$, geometric
 478 mean $a = 0.000519$, 95% range $a = 0.000293 - 0.000907$, and mean $b = 3.14$, SD $b = 0.0790$,
 479 and 95% range $b = 2.99 - 3.30$. The measurement error of $\log_{10}(a)$ was mean = 0.264,
 480 SD=0.00324, and for b it was mean = 0.182, SD=0.0225.

481
 482
 483
 484
 485
 486
 487

488 **Tables**

489

490 Table 1. Weighted means and standard deviations of parameters a and b from 5150 LWR studies for 1821 species of fishes, by body shape. *Geom.*
 491 *mean* stands for geometric mean and the *95% range* includes about 95% of the observations.

Body shape	Mean $\log_{10}(a)$	SD $\log_{10}(a)$	Geom. mean a	95% range a	Mean b	SD b	95% range b	n
eel-like	-2.99	0.175	0.00102	0.000464 – 0.00225	3.06	0.0896	2.88 – 3.24	162
elongated	-2.41	0.171	0.00389	0.00180 – 0.00842	3.12	0.0900	2.94 – 3.30	712
fusiform	-1.95	0.173	0.0112	0.00514 – 0.0245	3.04	0.0857	2.87 – 3.21	3478
short & deep	-1.70	0.175	0.0200	0.0182 – 0.0218	3.01	0.0905	2.83 – 3.19	798
all	-2.00	0.313	0.0100	0.00244 – 0.0411	3.04	0.119	2.81 – 3.27	5150

492

493

494

495

496

497 Table 2. Measurement and process errors derived from 5150 LWR studies for 1821 species.
 498 For convenience, the parameters are also given as shape and rate, ready for use with a
 499 gamma distribution.

Type of error	mean σ	sd σ	shape	rate
Measurement error $\log_{10}(a)$	0.260	0.00322	25076	6520
Measurement error b	0.184	0.00223	37001	6808
Process error $\log_{10}(a)$	0.173	0.00467	7933	1372
Process error b	0.088	0.00368	6498	572

500
 501
 502

503 Table 3. Demonstration of how parameter estimates from a single LWR study (for *Anguilla*
 504 *obscura*), which deviated strongly from the means for eel-like fishes, were made more
 505 realistic by inclusion of prior information, first for eel-like fishes, and then for eel-like fishes
 506 and related species in the Genus *Anguilla*. The relatively wide standard deviations (also
 507 shown in Figure 3) account for the remaining uncertainty in the estimates.

Data sources	<i>a</i>	$\log_{10}(a)$	sd	<i>b</i>	sd
eel-like prior	0.00102	-2.99	0.175	3.06	0.0896
single study	0.00021	-3.68	-	3.38	-
study + prior	0.000665	-3.18	0.131	3.09	0.0785
33 Genus studies	0.000853	-3.07	0.086	3.17	0.0484
study + prior + Genus	0.000519	-3.28	0.123	3.14	0.0790

508
 509
 510
 511
 512

513 Table 4. Analysis of weight-at-length data for North Sea turbot for the years 2010 - 2012. Priors were derived from parameter analysis of existing
 514 studies in FishBase 12/2012. The analysis used total lengths in cm and whole body weight in g.

Species	n	Length (cm)	Weight (g)	$\log_{10}(a)$	sd	<i>a</i>	95% range	<i>b</i>	sd	95% range	r^2
<i>Scophthalmus maximus</i>	742	9 – 52	15 – 3252	-1.81	0.0467	0.0155	0.0126 – 0.0192	3.06	0.0322	2.99 – 3.12	0.972

515
 516
 517
 518
 519
 520
 521

522 **Appendix: Web tools**

523

524 The Bayesian approaches described in this study have been implemented in web tools
525 available from www.fishbase.org. On a FishBase species summary page, go to the ‘More
526 information’ section and select the link ‘Length-weight’. This opens a new page with a table
527 of available LWR studies, and a plot of $\log_{10}(a)$ over b values, which should typically cluster
528 around a line with a negative slope. This graph is meant to help identification of studies that
529 deviate from the others, often because they used a different type of length measurement. The
530 default scores used for weighting are shown for each study and can be modified by the user.
531 The available studies can then be analysed, with inclusion of other species from the same
532 Genus or Family in cases where, e.g., fewer than 5 studies are available for the target species.
533 The respective priors shown in Tables 1 and 2 are used automatically by the web tools.

534

535 A successful analysis will present the parameter estimates as well as the measurement
536 error, together with standard deviations and 95% ranges. There is also an option to analyze
537 new weight-at-length data, using the results from the available studies as priors. Alternatively,
538 users can download data and R-code and perform the analyses locally. The analyses described
539 above can also be done by life stage or sex or for a certain region, simply by only including
540 the respective studies in the parameter analysis.

541

542 The preliminary LWR parameter estimates assigned to all species in FishBase are available
543 from the bottom of the FishBase species summary page, in the section entitled: *Estimation of*
544 *some characteristics with mathematical models.*

545

546 The R-code and the data used in the Figures and Tables can be downloaded as indicated in
547 Table 5.

548 **Table 5. R-code and data files used for graphs and tables can be downloaded from**
549 **<http://oceanrep.geomar.de/21875/>**

Figure / Table	R-code	Data source
Figure 1	LWR_Stats_3.R	BodyShape_3.csv, also data from Table 1
Figure 2	SingleSpecies LWR_7.R	BodyShape_3.csv
Figure 3	RelativesLWR_4.R	BodyShape_3.csv
Table 1+2	BodyShapePar_v5.R	BodyShape_3.csv
Table 3	RelativesLWR_4.R	BodyShape_3.csv
Table 4	LW_data_v6.R	Scophthalmus_maximus_LW.csv

550