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

Introduction

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

Material and Methods

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

General approach

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

We first established broad overall priors for parameters a and b, based on textbooks and reviews (step 1 below). We then estimated posterior distributions for model parameters for fishes in general by analyzing the distribution of a and b in a large data set of LWR studies (step 2). We further refined the estimated posterior distributions by grouping fish species into body-shape groups, from eel-like to short & deep, and estimating the parameters for each individual group (step 3). We used the body-shape posteriors as priors for the analysis of studies done for a given species (step 4). In data-poor species, we used the model to learn also from studies done on related species with the same body shape, i.e., we applied multivariate hierarchical Bayesian inference, treating each species as its own hierarchical level (step 5). As a result we obtained LWR parameter estimates for practically all fish species, with indication of uncertainty of the parameters and of the weight predicted from length. These species-specific parameters can then be applied directly, or they can serve as priors in the analysis of new weight-at-length data (step 6). FishBase (www.fishbase.org) contains online tools that incorporate these steps and facilitate the analysis of existing parameters and of new weight-at-length data (see also Web Tools section in the Appendix).

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

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

Step 2: Getting parameter estimates across all available LWR studies

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

In this analysis, a and b estimates for each individual species were considered as co-varying within the bounds of the species-specific body plan. However, for the across species analysis, a and b were considered as not correlated (see also Discussion). Looking at within and across species variability allowed for decomposing the total variability into measurement error and predictive error, where the latter is a combination of true natural variability and the error resulting from the LWR model only approximating the true relationship between length and weight. The predictive posterior parameter distributions arising from this across-all-studies-and-species analysis can be used as priors in single species analysis where body shape information is missing or does not match any of the shapes defined below.

Step 3: Getting parameter estimates by body shape group

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

Step 4: Getting joint parameter estimates for a species

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

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

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

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

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

Data

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

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

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

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

Statistical models

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

The Body-Shape model

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

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

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

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

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

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

where Σ_s is the measurement error covariance for observation s, which is composed of measurement error variance σ^2_{loga} and σ^2_b for $\log_{10}(a)$ and b, as well as the correlation ρ in measurement errors:

$$\Sigma_s = S_s^{-2} \begin{vmatrix} \sigma^2_{\log a} & \rho\sigma_{\log a}\sigma_b \\ \rho\sigma_{\log a}\sigma_b & \sigma^2_b \end{vmatrix} \qquad (4)$$

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

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

Across-all-Observations-and-Species model

The model for all observations and species but without body-shape is identical to the preceding Body-Shape model, with one exception. Specifically, the vector *bs* is replaced with a dummy vector *1*, which has the value one for all entries. This change implies that all species in this model have the same value for $\log_{10}(\alpha_{bs})$ and β_{bs}. It consequently provides an average value for $\log_{10}(\alpha)$ and β for species for which the body-shape is unknown.

The Few-Studies model

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

The Single-Species model

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

The New Weight-at-Length-Data model

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

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

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

Results and Discussion

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

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

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

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

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

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

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

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

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

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

An example of a species with a single LWR study in FishBase was the Pacific short-finned eel, *Anguilla obscura* (Figure 3). The parameters given were n=145, a = 0.00021, 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 negative parameter co-variation, i.e., the *a* estimate appears too low and *b* too high. In this case, single-species analysis would combine the only study with the information provided by the prior for eel-like species, suggesting $a = 0.00067$ and $b = 3.09$, and thus pulling the parameters suggested by the single study strongly in the direction of the prior. However, other LWR studies for species of the Genus *Anguilla* confirm a deviation from the eel-like prior, although less strongly than suggested by the single study. Including the information from these related species gives $a = 0.00085$ (0.00058 – 0.0013) and $b = 3.17$ (3.07 – 3.26), which appears to be a meaningful summary of the available information, accommodating the single study under the tails of the proposed parameter distributions (see single points in Figure 3).

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

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

It is interesting to compare the results of the Bayesian LWR analysis with those of a regular linear regression. In our example for turbot, the Bayesian analysis included, in a hierarchical process, information from the body-shape group and from other studies done for the species. In contrast, the regular regression only analyzed the data at hand. The prior means for $\log_{10}(a) = -1.83$ and $b = 3.04$ did not differ much from the means of the data, as provided by regular regression with $\log_{10}(a) = -1.81$, $b = 3.06$, and hence the means provided by the Bayesian analysis were identical to those of the regular regression. However, the prior estimates of uncertainty $\mathrm{SD}[\log_{10}(a)] = 0.069$ and $\mathrm{SD}[b] = 0.0486$ were considerably wider than those of the regular regression with $\mathrm{SE}[\log_{10}(a)] = 0.0271$ and $\mathrm{SE}[b] = 0.0187$. In other words, the estimates of uncertainty provided by the regular regression were only representative for the analyzed data, but too narrow if data from other years and areas were considered. The Bayesian analysis incorporated this additional information and provided more realistic estimates of uncertainty that were intermediate between the priors and the data, with SD[$\log_{10}(a)$] = 0.0461 and SD[b] = 0.0317.

Preliminary LWR parameters for all species of fishes

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

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

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

Conclusion

We present 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. We used a Bayesian approach to the estimation of length-weight relationships for practically all species of fishes. We show how the use of all available prior information can improve parameter estimates. The increased uncertainty in species with little available data is expressed in wider respective parameter distributions. We make a large standardized data set available for further research. We hope our read-to-use tools will help in spreading the application of Bayesian methods in fisheries.

Acknowledgements

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

References

Carlander, K. D., 1969: Handbook of freshwater fishery biology, Vol. 1. Ames, IA., The Iowa State University Press, 752 pp.

Carlander, K. D., 1977: Handbook of freshwater fishery biology, Vol. 2. Ames, IA. The Iowa State University Press, 431 pp.

Clark, J.S., 2003: Uncertainty and variability in demography and population growth: a hierarchical approach. Ecology 84(6), 1370-1381.

Froese, R., 2006: Cube law, condition factor, and weight-length relationships: history, meta-analysis and recommendations. Journal of Applied Ichthyology 22(4), 241-253.

Froese, R.; Pauly, D. Editors. 2012: FishBase. World Wide Web electronic publication, version (12/2012). Accessed December 1, 2012 at http://www.fishbase.org

Gelman, A., 2006: Prior distributions for variance parameters in hierarchical models. Bayesian Analysis, **1**, 515–533.

Gelman, A.; Hill, J., 2007: Data analysis using regression and multilevel/hierarchical models. Cambridge, UK, Cambridge University Press, 648 p.

Jellyman, D.J., 1991: Biology of the shortfinned eel *Anguilla obscura* in Lake Te Rotonui, Mitiaro, Cook Islands. Pac. Sci.45(4), 362-373.

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

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

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

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

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

Spencer, H., 1864‑1867: The Principles of Biology, 2 volumes. Williams & Norgate, London. 678, pp.

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

Figures

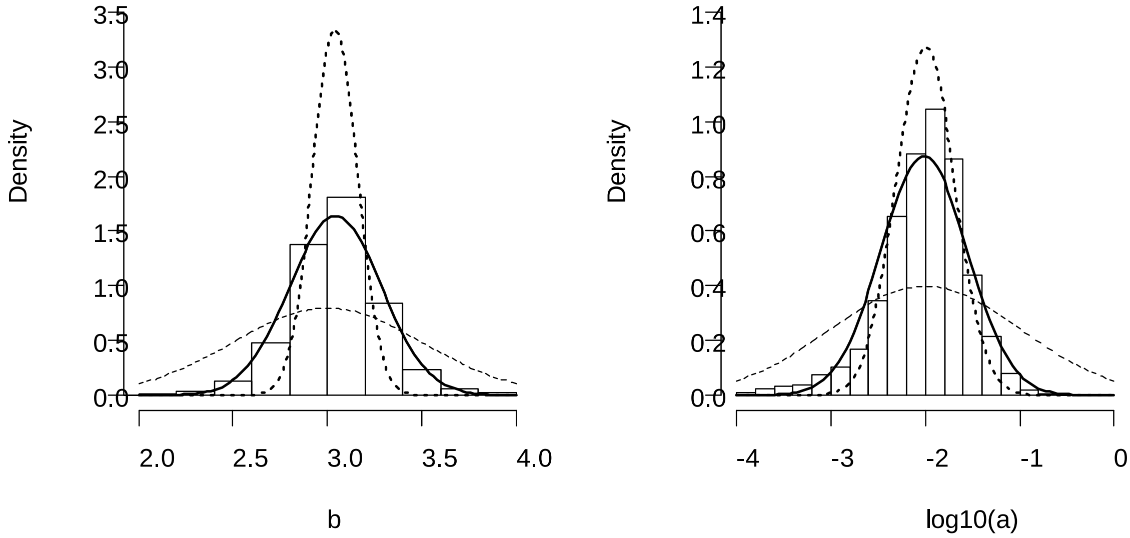

Figure 1. Weighted distribution of parameters *b* and *a* in 5150 LWR studies for 1821 species of fishes. The overlaid curves are normal density functions, i.e. the areas under the histograms and under the curves are identical and equal to 1. The bold normal curves use mean and standard deviation of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

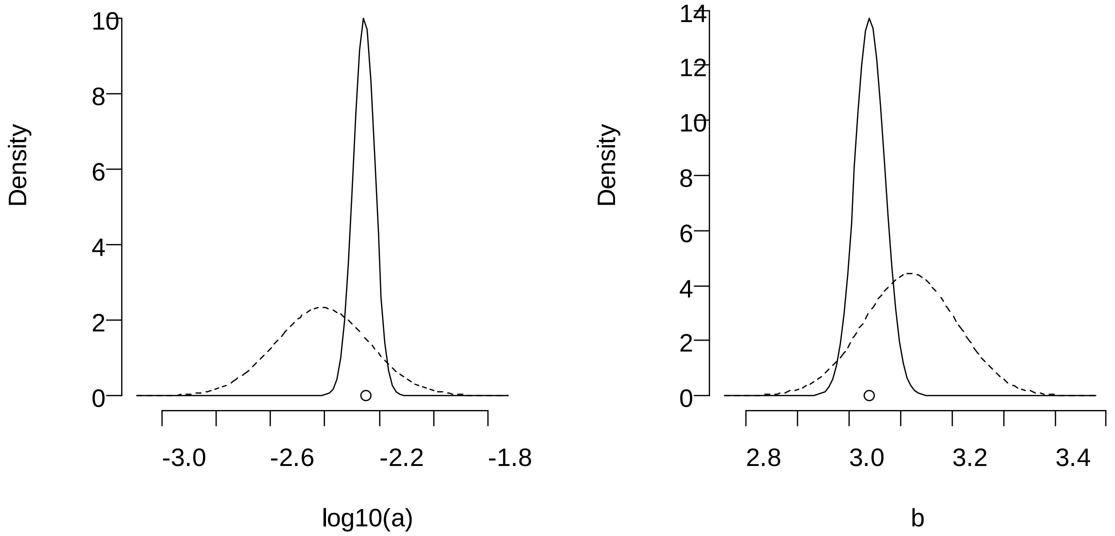

Figure 2. Distribution of parameters *a* and *b* for 48 LWR studies of the European anchovy *Engraulis encrasicolus*. The single points present the mean values of the data. The dashed lines indicate the prior distributions for elongated fishes. Mean $\log_{10}(a)$ = -2.26, SD of $\log_{10}(a)$ and $\log_{10}(W)$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

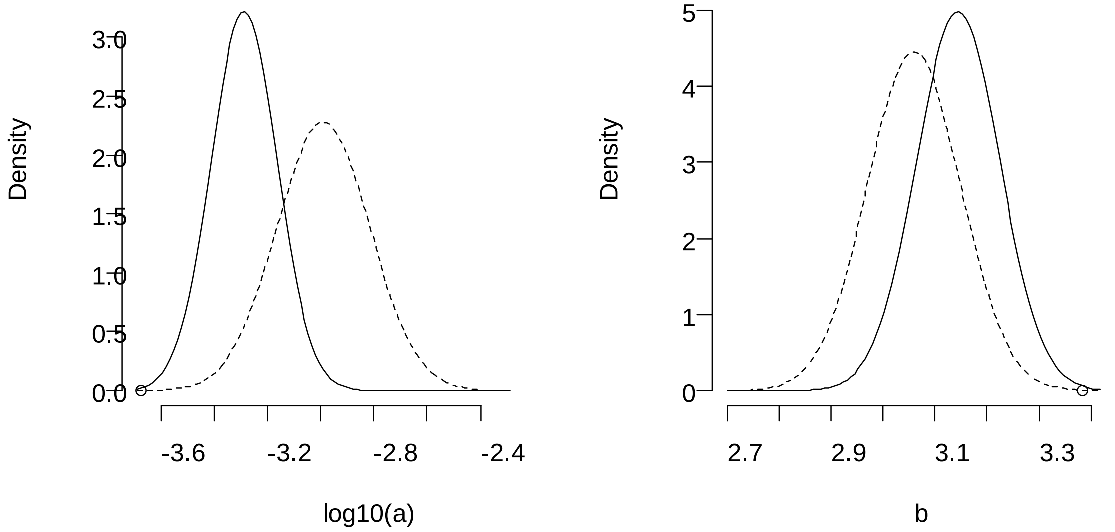

Figure 3. Distribution of parameters a and b for one study with a=0.00021 and b=3.38 for the Pacific short-finned eel, *Anguilla obscura* (indicated by single points) and 33 LWR studies of four species of the Genus *Anguilla*. The dashed curves indicate the prior distributions for eel-like fishes. Resulting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

Tables

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. mean* stands for geometric mean and the *95% range* includes about 95% of the observations.

Body shape	**Mean log10(*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

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

Type of error	**mean σ**	**sd σ**	**shape**	**rate**
Measurement error log10(*a*)	0.260	0.00322	25076	6520
Measurement error *b*	0.184	0.00223	37001	6808
Process error log10(*a*)	0.173	0.00467	7933	1372
Process error *b*	0.088	0.00368	6498	572

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

Data sources	***a***	**$\log_{10}(a)$**	**sd**	***b***	**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

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 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**	***a***	**95% range**	***b***	**sd**	**95% range**	**r^2**
Scophthalmus maximus	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

Appendix: Web tools

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

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

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

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

Table 5. R-code and data files used for graphs and tables can be downloaded from 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