DE GRUYTER OPEN

RESEARCH ARTICLE

A generalized, nonlinear regression approach to the length-weight relationship of European perch (*Perca fluviatilis* L.) from the Polish coast of the southern Baltic Sea

Marcello De Giosa, Przemysław Czerniejewski

Abstract. Length-weight relationships in fish are important tools in fisheries management. The aim of this study was to estimate the growth curve $W=aL^b$ for European perch, *Perca fluviatilis* L., from the Polish coast of the southern Baltic Sea (ICES Subdivision 25). The data set comprised the total weights and total lengths of 827 specimens caught during the 2011-2013 period. The mean total length (TL) was 169.4 mm (range 100.0-310.0 mm), and the mean total weight was 83.9 g (range 11.3-553.2 g). First, a multiplicative error term and a linear regression approach to loglog-transformed data was considered. The following estimated values for the parameters were obtained: a = exp(-12.5323), b=3.25, s=0.07862. This approach was not successful in solving the common heterogeneity problem of the length-weight data. A generalized nonlinear regression approach to the original data was more suitable in our case. The estimated model was W = $3.83 \times 10^{-6} \times L^{3.238} + \varepsilon$, with $\varepsilon \sim \text{norm}(0, 0.0281 \times E[W|L]^{2 \times 1.242})$. The estimated 95% confidence interval for b was (3.218, 3.259), and the growth was allometric. The perch from the coastal waters of the Baltic Sea (ICES Subdivision 25) was characterized by slightly better condition than that caught in inland waters and estuaries.

M. De Giosa [✉]
Dipartimento di Matematica, Università di Bari
via Orabona 4, 70125 Bari, Italy
e-mail: marcello.degiosa@uniba.it

P. Czerniejewski
Department of Fisheries Management
West Pomeranian University of Technology in Szczecin
ul. Krolewicza 4, 71-550 Szczecin, Poland

Introduction

The composition of coastal fish communities in different regions of the Baltic Sea varies in relation to the different habitat characteristics of these regions. Salinity, water temperature, and nutrient availability are among the important factors (Olsson et al. 2012). One of the most common freshwater species caught in coastal fish monitoring in the shallow south Baltic Sea is European perch, *Perca fluviatilis* L. It is a widely distributed member of the perch family, widespread throughout Europe and Asia, and successfully introduced in South Africa, Australia, and New Zealand (Thorpe 1977). European perch often plays a key role in aquatic ecosystem functioning via top down control (Eriksson et al. 2011), and it is one of the most valuable commercial and sport fish species in the offshore zone of the Baltic Sea. This species has also been caught in the sea up to 4 km from the coasts.

European perch populations have declined in some coastal areas in recent decades, which has had

negative ecological, commercial, and cultural impacts. The protection and restoration of recruitment areas (perch spawning grounds) are believed to be important aspects of supporting future stocks (Ljunggren et al. 2010, Byström et al. 2015).

The freshwater, shallow-sea fish fauna of the south Baltic Sea has been poorly studied and very little biological information is available. In some areas, perch has been subjected to biological monitoring with the objectives of describing and elucidating long-term trends in fish populations and attempting to explain changes of biological characteristics of this species, e.g., age structure, year-class strength, growth rate (Karås 1990, Böhling at al. 1991, Ložys 2004). A standard weight equation for European perch has only recently been calculated (Giannetto et al. 2012).

The fish length-weight relationship can provide important ecological insights (Froese 2006). It is useful for estimating fish weight or community biomass when only length and species data are available, and it is required for stock assessment, fishery management, and conservation (Froese 1998, Oscoz et al. 2005, Jellyman et al. 2013). It also permits estimating the condition factor and comparing fish life history characteristics and growth trajectories among species, sexes, seasons, and regions (Froese 2006, Jellyman et al. 2013, Lloret et al. 2014). Growth parameter b provides information on the type of fish growth: isometric (b = 3), negative allometric (b<3), positive allometric (b>3) (Cone 1989).

The length-weight relationship in fish is well described by the nonlinear growth model: $W=aL^b$. Length can be measured as total, fork, or standard length, and it plays the role of the explanatory (independent) variable. Weight (W) can be wet or dry, whole or dressed, and is the response (dependent) variable (Jenning et al. 2012).

Historically, the growth curve is mainly estimated using a linear regression approach to the loglog-transformed equation (Huxley 1924, Huxley 1932, Le Cren 1951, Cone 1989). A multiplicative error term is often assumed on the right side of the growth model: $W = aL^b \times exp(\varepsilon)$ (Ogle 2015). The error becomes additive on taking logarithms of both sides: $log(W) = log(a)+b \times log(L)+\varepsilon$. Parameters a and b, and the residual standard error, can then be estimated by the least squares method of linear regression (Guy and Brown 2007, Fox and Weisberg 2011). The loglog-transformation should be the easiest solution to linearize the length-weight relationship and to stabilize the mean-variance relationship when trying to solve the typical heterogeneity problem of length-weight data (the weight of longer fish is more variable than that of shorter fish).

In this paper, the relationship between weight (W) and length (L) for coastal European perch in the southern Baltic Sea was investigated. Both linear regression on loglog-transformed data and generalized nonlinear regression on the original data were used.

Material and methods

Sampling and measurements

The present study involved a total of 831 individuals of perch (*P. fluviatilis*) caught by commercial fishermen in the shallow areas of the Polish coast of the southern Baltic Sea (ICES Subdivision 25) in 2011-2013. To avoid differences between the sexes stemming from the female gonads (Le Cren 1951), the fish were not caught during the spawning season (March-May). Samples from six stations (Świnoujście, Dziwnów, Rewal, Kołobrzeg, Darłowo, Ustka, Jarosławiec) were collected using seine nets, trawl nets, and gillnets. Live specimens were identified and measured for total length (TL) to the nearest mm and total weight (W) to the nearest 0.1 g.

Statistical methods

The length-weight relationship was described by the growth equation: $W = aL^b$. First, a lognormal multiplicative error term in the model was assumed:

$$W = aL^b \times exp(\varepsilon), \ \varepsilon \sim norm(0, s^2).$$

The loglog-transformed equation (with additive normal error):

$$log(W) = log(a)+b\times log(L)+\varepsilon, \; \varepsilon\sim norm(0, s^2)$$

was fitted to log-transformed data by linear regression. Parameters were estimated by the ordinary least squares method. To correct bias, before plotting the estimated model in the original scale, the back-transformed (exponentiated) predicted mean weights and the relative confidence intervals were multiplied by the correction factor for allometric equation $exp(s^2/2)$, where s is the estimated residual standard error (Sprugel 1983). Linear regression assumptions (normality and homogeneity) were checked using residuals plots (Fox and Weisberg 2011).

Then nonlinear regression was considered. In this approach, the minimization of the residual sum of squares is reached via numerical iterative optimization procedures. In ordinary nonlinear regression, constant variance structure is assumed. However, because of heterogeneity, ordinary least squares was not the right fitting method. To include heterogeneity in the model, more complexity and extra mathematical effort were required. We used a generalized nonlinear regression approach with variance structure, which combines nonlinear and heterogeneous variance modeling. The model was considered:

$$W = aL^b + \varepsilon, \; \varepsilon\sim norm(0,var(\varepsilon))$$

consisting of a fixed term, describing the mean growth curve $E(W|L)=aL^b$, and an additive normally distributed random term ε with zero mean and variance structure var(ε).

Four models with the same fixed component and different variance structures were considered. Candidate variance structures were chosen using biological knowledge and suggestions deduced from the plot of the normalized residuals of a model with constant variance. We fitted the nonlinear models to data using generalized least squares estimation methods (Pinheiro and Bates 2000), which is essentially weighted nonlinear regression. With this approach, more accurate estimates of standard errors and statistical significance tests for the fixed part of the model were also obtained (Carroll and Ruppert 1988).

The optimal variance structure was chosen using Akaike information criteria (AIC; Akaike 1973). AIC is defined as

$$AIC = -2log(L)+2p$$

where L is the maximum likelihood and p is the number of parameters. It measures goodness of fit and model complexity. The lower the AIC, the better the model fits the data.

Graphical procedures (residual plots) were used to check and validate the assumptions underlying the final model: correct mean function, normality, and independence with respect to sampling sites and sampling time (Ritz and Streibig 2008). A plot of the standardized (normalized) residuals, obtained from the ordinary residuals divided by the square root of the variance, was used to check if any heterogeneity was still present.

Approximate Wald 95% confidence intervals for the parameters were also obtained, using a normal approximation to the distribution of the (restricted) maximum likelihood estimators. The confidence interval for b was also used to test for the isometry or allometry of growth; the isometry null hypothesis (H0: b=3) was rejected at the 5% significance level if the 95% confidence interval for b did not contain the value 3.

For a detailed description of the statistical methods, readers should refer to Pinheiro and Bates (2000). All graphs and statistical analysis in this paper were produced using the software R (R Core Team 2014) and the R package NLME (Pinheiro et al. 2015).

Results

In a preliminary explorative data analysis, four outliers stemming from measurement errors were identified by graphical procedures and removed from the sample. The final sample comprised 827 specimens. The mean total length was 169.4 mm (min=100, max=310) with a standard deviation of 47.3 mm. The mean total weight was 83.9 g (min=11.3 g,
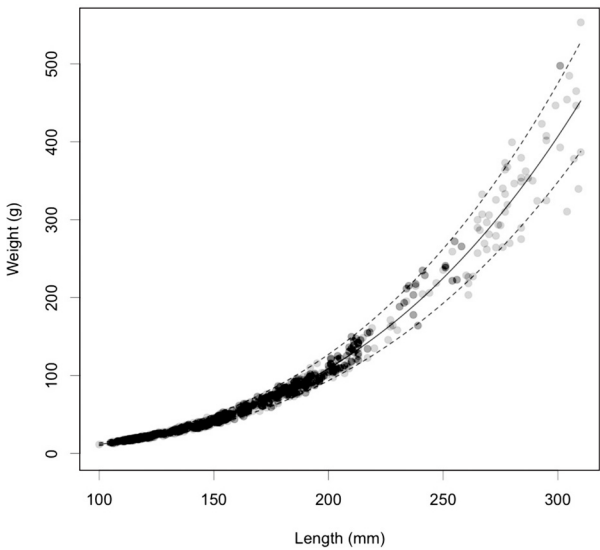
Figure 1. Weight-length data for European perch from south Baltic Sea. The superimposed fitted model (solid line) and 95% prediction bands (dashed lines) have been obtained by a linear regression approach to the loglog-transformed data. Predicted weights are the back transformed (exponentiated) predicted log weights multiplied by Sprugel's correction factor. Prediction bands are obtained in the same way.
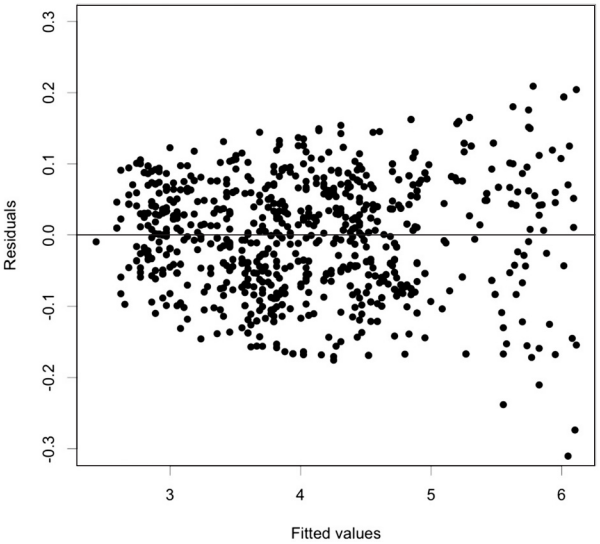

Figure 2. Residuals versus fitted values for the linear regression model fitted to loglog-transformed data. There is a clear violation of the homogeneity (constant variance) assumption of linear regression.

max=553.2 g) with a standard deviation of 87.9 g. The plot of W against total length (TL) showed variance heterogeneity for these data, with a clear tendency for the variability of W: the larger the mean value, the larger the variability around the mean value. From the linear regression analysis of the log-transformed data the following estimated values were obtained for the parameters: a=exp(-12.5323), b=3.25, s=0.07862.

Figure 1 shows a plot of the estimated model in the original scale. Predicted weights, calculated with bias correction, are plotted against total lengths.

Pointwise 95% confidence bands are also shown. Figure 2 shows a plot of the residuals versus the fitted values for the loglog-linear model. There is a clear violation of the variance homogeneity assumption of linear regression.

Table 1 presents the variance structure, the degree of freedom, and the AIC for each of the four generalized nonlinear models considered. The model with the lowest AIC (=4727.7) was that with the so-called *power of the mean* variance structure:

$$var(\varepsilon) = s^2 E[W|L]^{2d}.$$

This was chosen as the best model. Table 2 presents estimates and 95% confidence intervals for each model parameter.

Table 1
Variance structure, degree of freedom and AIC for each of the four considered generalized nonlinear model

Variance structure	Degree of freedom	AIC
$s^2 L$	3	6273.0
$s^2 E[W|L]^{2d}$	4	4727.7
$s^2 (\varepsilon + E[W|L])^2$	5	4729.7
$s^2 \exp(E[W|L])^{2d}$	4	4971.0

Table 2
Estimates and 95% confidence intervals for the parameters of the best nonlinear model

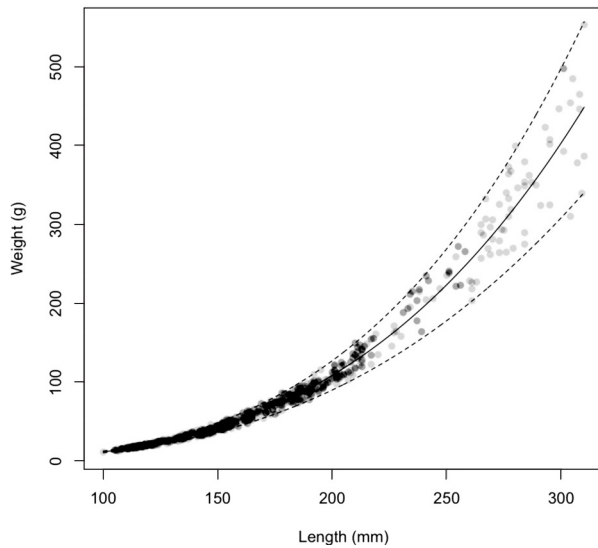| Parameter	Estimate	95% C.I.
a	3.83×10^{-6}	$3.44 \times 10^{-6}, 4.23 \times 10^{-6}$
b	3.238	3.218, 3.259
s	0.0281	0.0224, 0.0353
d	1.242	1.187, 1.297

Figure 3. Weight-length data for European perch from south Baltic Sea. The superimposed fitted model (solid line) and 95% prediction bands (dashed lines) are obtained by a generalized nonlinear regression approach to the original data.
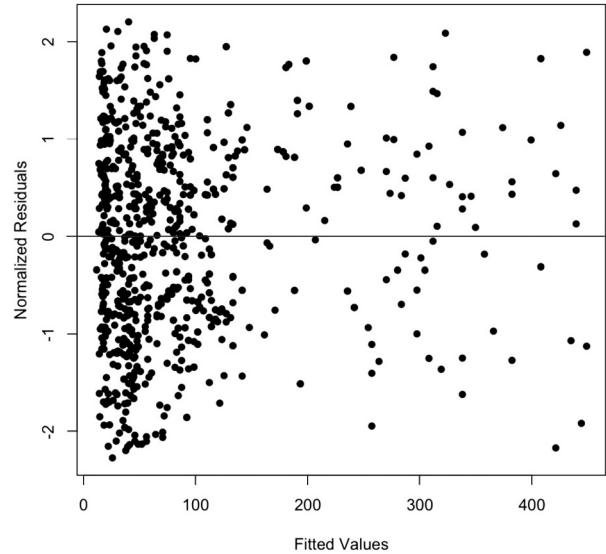

Figure 4. Normalized residuals versus fitted value for the generalized nonlinear regression model fitted to the original data. There is no evidence of a still present heterogeneity.

The estimated model was:

$$W = 3.83*10^{-6}*L^{3.238} + \varepsilon,$$
$$\varepsilon \sim norm(0, 0.0281*E[W|L]^{2*1.242}).$$

The estimated 95% confidence interval for b was (3.218, 3.259), which means that growth was allometric (H0: b=3 rejected at the 0.05 significance level. Residual plots (not shown here to save space) did not show any evident departure from the model assumptions. Figure 4 is a plot of the normalized residuals versus the explanatory variable length. There is no evidence of that heterogeneity is still present.

Discussion

Mathematical models are very useful in describing the great variety of possible individual growth patterns. When, in the three-dimensional growth process, individual length, width, and depth change in proportion to one another, growth is referred to as isometric. In this case, individual weight (W) is proportional to the cube of its length (L): $W = aL^3$. In fishes the allometric model of $W = aL^b$, a two parameter generalization of the isometric cube law, often

provides a more reasonable fit to the data (Quinn and Deriso 1999).

The main consideration in model fitting is that the correct error structure is chosen. The allometric model can either be fitted to data by linear regression on the loglog-transformed variables or by nonlinear regression of non-transformed variables. The two methods differ in their assumptions on the error structure, and the appropriate model fitting procedure depends on the method chosen. The choice should take into account the heteroscedastic nature of allometric data and should be checked by investigating the appropriate residuals plots.

The loglog-transformation of the allometric equation was first used by Huxley (1924, 1932) and some other investigators of the period. Huxley fitted a straight line to data plotted on a double-logarithmic paper by eye and then estimated the coefficients with elementary trigonometry. Other scientists (Galstoff 1931, Feldstein and Hersh 1935) introduced the least squares method of linear regression later.

The linear regression on the loglog-transformed variables approach assumes a multiplicative error structure: $W = aL^b \times e^\varepsilon$, with Σ normally distributed with constant variance (Quinn and Deriso 1999).

A consequence of the multiplicative error structure is that an increase in the mean value of the response variable is accompanied by an increase in its variance. However, the multiplicative error assumption is primarily made only for mathematical convenience and proper justification for it is hardly provided. For decades it was the only way to fit the allometric model to growth data while simultaneously addressing the intrinsic heteroscedasticity of the data. For this reason, several authors have strongly supported this approach (Ogle 2015) despite its several critical issues. Among them the linear regression approach results in bias from the logarithmic transformation (the model passes through the geometric mean, rather than the arithmetic mean), but this can be adjusted using a simple correction factor (Sprugel 1983) (1); growth is interpreted in terms of percentage change (one percentage change in length results in $100 \times (1.01 \times b-1)$ percentage change in weight), so it is easy to get confused about the interpretation of growth parameter b (2); data transformation can change the error distribution – normally distributed errors can became non-normal and only log-normal errors became normal after the transformation (3); data transformation can hide interesting biological and ecological information (4). Some of these issues were already identified by Jerrold Zar (Zar 1968) and more recently by other authors (Hui and Jackson 2007). Nevertheless, this approach is very common in the biological literature (Ogle 2015), mainly because many consider it the only way to fit a model to growth data with intrinsic heteroscedasticity.

Nevertheless, for the data set considered in this paper, linear regression on loglog-transformed data was not a suitable or successful procedure in the face of the heterogeneity problem. A plot of residuals versus fitted values showed that the constant variance assumption was not met by the log-transformed data. Some heterogeneity was still present after the transformation. Ignoring this problem could result in a regression parameter with incorrect standard error and prediction intervals for new observations with the incorrect confidence level.

The advent of generalized nonlinear regression introduced new methods for fitting a nonlinear model with non-constant variance and normally distributed additive error to the original data. Several variance structures can be compared. For the data in this study, a generalized nonlinear fitting of the growth curve to the original data, combined with variance structures accounting for heterogeneity (Pinheiro and Bates 2000, Ritz and Streibig 2008), was a better fitting procedure than linear regression on loglog-transformed data, and it produced more satisfactory results.

For the European perch from south Baltic Sea, the estimated growth parameter is b=3.238 with a 95% confidence interval (3.218, 3.259). This means that this species grows allometrically. However, the weight-length relationship in European perch varies widely among locations and years. Generally, the perch from the coastal waters of the Baltic Sea from 2011-2013 were characterized by slightly better condition than those caught in inland waters and estuaries (Lozys 2004).

References

Akaike H. 1973 – Information theory as an extension of the maximum likelihood principle – In: Second International Symposium on Information Theory (Eds) B.N. Petrov, F. Csaki, Akademiai Kiado, Budapest, Hungary: 267-281.

Böhling P., Hudd R., Lehtonen H., Karås P., Neuman E., Thoresson G. 1991 – Variations in year-class strength of different perch (*Perca fluviatilis*) populations in the Baltic Sea with special to temperature and pollution – Can. J. Fish. Aquat. Sci. 48: 1181-1187.

Byström P., Bergström U., Hjälten A., Ståhl S., Jonsson D., Olsson J. 2015 – Declining coastal piscivore populations in the Baltic Sea: Where and when do sticklebacks matter? – Ambio, 44: 462-471.

Carroll R.J., Ruppert D. 1984 – Power transformations when fitting theoretical models to data – J. Am. Stat. Assoc. 79: 321-328.

Cone R.S. 1989 – The need to reconsider the use of condition indices in fishery science – Trans. Am. Fish. Soc. 118: 510-514.

Eriksson B.K., Sieben S., Eklöf J., Ljunggren L., Olsson J., Casini M., Bergström U. 2011 – Effects of altered offshore food webs on coastal ecosystems emphasizes the need for cross ecosystem management – Ambio 40: 786-797.

Feldstien M.J., Hersch A.H. 1935 – The 242 determination of genetic constants of relative growth – Am. Nat. 69: 344-353.

Fox J., Weisberg S. 2011 – An R companion to applied regression – Thousand Oaks, CA, Sage Publications, 449 p.

Froese R. 1998 – Length-weight relationships for 18 less-studied fish species – J. Appl. Ichthyol. 14: 117-118.

Froese R. 2006 – Cube law, condition factor and weight–length relationships: history, meta analysis and recommendations – J. Appl. Ichthyol. 22: 241-253.

Galstoff P.S. 1931 – The weight-length relationship of the shells of the Hawaiian pearl oyster, *Pinctada* sp. – Am. Nat. 65: 423-433.

Giannetto D., Carosi A., Franchi E., La Porta G., Lorenzoni M. 2012 – Proposed standard weight (*Ws*) equation for European perch (*Perca fluviatilis* Linnaeus, 1758) – J. Appl. Ichthyol. 28: 34-39.

Guy C.S., Brown M.L. 2007 – Analysis and interpretation of freshwater fisheries data – Am. Fish. Soc., Bethesda, Maryland, 996 p.

Huxley J.S. 1924 – Constant differential growth-ratios and their significance – Nature 114: 895-896.

Huxley J.S. 1932 – Problems of relative growth – Metheun, London, UK, 319 p.

Hui D., Jackson R.B. 2007 – Uncertainty in allometric exponent estimation: a case study in scaling metabolic rate with body mass – J. Theor. Biol. 249: 168-177.

Jellyman P.G., Booker D.J., Crow S.K., Jellyman D.J. 2013 – Does one size fit all? An evaluation of length-weight relationships for New Zealand's freshwater fish species – New Zeal. J. Mar. Fresh. 47: 450-468.

Jenning C.A., Sloss B.L., Lasee B.A., Burtle G.J., Moyer G.R. 2012 – Care, handling, and examination of sampled organism – In: Fisheries Techniques, 3nd edn (Eds) A.V. Zale, D.L. Parrish, T.M. Sutton, American Fisheries Society, Bethesda, Meryland: 163-221.

Karås P. 1990 – Seasonal changes in growth and standard metabolic rate of juvenile perch, *Perca fluviatilis* L. – J. Fish Biol. 37: 913-920.

Le Cren E.D. 1951 – The length–weight relationship and seasonal cycle in gonad weight and condition in the perch (*Perca fluviatilis*) – J. Anim. Ecol. 20: 201-219.

Ljunggren L., Sandström A., Bergström U., Mattila J., Lappalainen A., Johansson G., Sundblad G., Casini M., Kaljuste O., Eriksson B. K. 2010 – Recruitment failure of coastal predatory fish in the Baltic Sea coincident with an offshore ecosystem regime shift – ICES J. Mar. Sci. 67: 1587-1595

Lloret J., Shulman G., Love R.M. 2014 – Condition and health indicators of exploited marine fishes – Wiley Blackwell, West Sussex, UK, 262 p.

Ložys L. 2004 – The growth of pikeperch (*Sander lucioperca* L.) and perch (*Perca fluviatilis* L.) under different water temperature and salinity conditions in the Curonian Lagoon and Lithuanian coastal waters of the Baltic Sea – Hydrobiologia 514: 105-113.

Ogle D.H. 2015 – Introductory fisheries analysis with R – Chapman and Hall/CRC, Boca Raton, USA, 317 p.

Olsson J., Bergström L., Gårdmark A. 2012 – Abiotic drivers of coastal fish community change during four decades in the Baltic Sea – ICES J. Mar. Sci. 69: 961-970.

Oscoz J., Campos F., Escala M.C. 2005 – Weight–length relationships of some fish species of the Iberian Peninsula – J. Appl. Ichthyol. 21: 73-74.

Pinheiro J., Bates D. 2000 – Mixed effects models in S and S-Plus – Springer, New York, USA, 528 p.

Pinheiro J., Bates D., DebRoy S., Sarkar D. 2015 – NLME: Linear and nonlinear mixed effects models – R package version 3.1-122, URL http://CRAN.R289 project.org/package=nlme.

Quinn T.J., Deriso R.B. 1999 – Quantitative fish dynamics – Oxford University Press, New York, Oxford, 542 p.

R Core Team 2014 – R: A language and environment for statistical computing – R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/.

Ritz C., Streibig J.C. 2008 – Nonlinear regression with R – Springer, Berlin, Germany, 150 p.

Sprugel D.G. 1983 – Correcting for bias in log-transformed allometric equations – Ecology 64: 209-210.

Thorpe J.E. 1977 – Synopsis of biological data on the perch *Perca fluviatilis* Linnaeus, 1758 and *Perca flavescens* Mitchill, 1804 – FAO Fisheries Synopsis 113, 138 p.

Zar J.H. 1968 – Calculation and miscalculation of the allometric equation as a model in biological data – BioScience 18: 1118-1120.