RESEARCH ARTICLE

Information-theory approach to allometric growth of marine organisms

Stelios Katsanevakis · Maria Thessalou-Legaki · Constantina Karlou-Riga · Eugenia Lefkaditou · Evagelos Dimitriou · George Verriopoulos

Abstract Allometric growth investigations are usually conducted by fitting the allometric model (L) $y = ax^b \Leftrightarrow \log y = \log a + b \log x$ (y, x are morphometric characters and b the allometric exponent), which is quite simple both conceptually and mathematically, and its parameters are easy to estimate by linear regression. However b is not necessarily constant and it may change either continuously or abruptly at specific breakpoints; thus, the simple L model quite often fails to describe allometric growth successfully. In the current context, a better alternative is proposed, based on Kullback–Leibler (K-L) information theory and multi-model inference (MMI). Allometric growth was investigated in eight marine species: the bivalves *Pecten jacobaeus* and *Pinna nobilis*, the squids *Todarodes sagittatus* and *Todaropsis eblanae*, the crab *Pachygrapsus marmoratus* (females), the ghost shrimp *Pestarella tyrrhena* (males), and the fishes *Trachurus trachurus* and *Sparus aurata*. In each of the eight species, a pair of body parts was measured and the allometric growth of one body part in relation to the other (reference dimension) was studied, by fitting five different candidate models including: the simple allometric model, two models assuming that b changed continuously and two other assuming that b had a breakpoint. For each species, the 'best' model was selected by minimizing the small-sample, bias-corrected form of the Akaike Information Criterion. To quantify the plausibility of each model, given the data and the set of five models, the 'Akaike weight' w_i of each model was calculated; based on w_i the average model was estimated for each case. MMI is beneficial, more robust, and may reveal more information than the classical approach. As demonstrated with the given examples, estimation of b from the linear model, when it was not supported by the data, revealed some characteristic pitfalls, such as concluding positive allometry when there is actually negative or vice versa, or reporting allometry when the data in reality support isometric growth or vice versa.

S. Katsanevakis (✉) · M. Thessalou-Legaki ·
G. Verriopoulos
Department of Zoology-Marine Biology,
Faculty of Biology, School of Biology,
University of Athens, Panepistimioupolis,
15784 Athens, Greece
e-mail: stelios@katsanevakis.com

C. Karlou-Riga
Ministry of Agriculture, Fisheries Laboratory,
15 Karaoli and Demetriou Street, 18531 Piraeus, Greece

E. Lefkaditou
National Centre for Marine Research, Aghios Kosmas,
16604 Helliniko, Greece

E. Dimitriou
Prefecture of Aitoloakarnania, 30200 Messolonghi, Greece

Introduction

Growth is often accompanied by changes in proportion as well as in size, the phenomenon of relative or allometric growth. The use of the allometric equation (Huxley 1932) is the most extensively used method of analysis for relative growth during ontogeny; the

relationship between the size of a part of the body y and another part x (taken as a reference dimension) has the form $y = ax^b$, where the exponent b is a measure of the difference in the growth rates of the two parts of the body; when $b < 1$ the type of growth is described as negative allometric, when $b > 1$ as positive allometric and when $b = 1$ as isometric. To estimate the coefficients of the allometric equation, the data are usually log-transformed and a linear regression is fitted; b is the slope of the resulting linear equation. Logarithmic transformation is generally appropriate because morphological data tend to have log normal structure, as they are non-negative, with positively-skewed distributions and variances that increase with the mean (Jolicoeur 1990; Ebert and Russell 1994). However, the allometric exponent b is not necessarily constant. The existence of breakpoints (i.e. points of discontinuity in slope b) in allometric data has been recognized since the allometric equation was first proposed (Huxley 1932). Such breakpoints are usually identified visually, when data are obviously separated, and then regression analysis is done on both sides of the selected breakpoint. Estimating breakpoints visually is not, however, an accurate method and the use of segmented regression models is a better alternative (Shea and Vecchione 2002; Hall et al. 2006).

The allometric exponent may also change continuously following a smooth, curvilinear trend without exhibiting marked breakpoints (e.g. Hall et al. 2006). Such a smooth change of the allometric exponent (in contrast to the marked instantaneous change at breakpoints) might occur e.g. when maturity favors the relative growth of reproduction-related body parts but later in ontogeny size constraints reduce the relative growth rates.

When conducting an allometric growth investigation, it is common practice among researchers to simply estimate the slope of the linear regression equation fitted to log-transformed data. In the present paper, it is demonstrated that such an approach may give very poor and misleading results. A better alternative is proposed, based on Kullback–Leibler (K-L) information theory and multi-model inference (MMI) as described by Burnham and Anderson (2002).

Model selection based on information theory is a relatively new paradigm in biological sciences and is quite different from the usual methods that are based on null hypothesis testing. The information theory approach to model selection and inference is based on K-L information $I(f,g) = \int f(x) \log \left(\frac{f(x)}{g(x|\theta)} \right) \mathrm{d}x$ (Kullback and Leibler 1951). $I(f,g)$ is the 'information' lost when model g (with parameters θ) is used to approximate full reality or truth f; equivalently $I(f,g)$ is interpreted as the distance from the approximating model to full reality (Burnham and Anderson 2002). According to information theory, minimization of K-L distance is a fundamental basis for model selection. Full reality f is unknown in real problems and thus $I(f,g)$ may not be computed directly. Akaike (1973) devised a method to approximate K-L distance, based on the empirical log-likelihood function. This is known as Akaike's information criterion (AIC) and is summarized by the formula $\text{AIC} = -2\log(\mathcal{L}(\hat{\theta}|\text{data})) + 2K$, where $\log(\mathcal{L}(\hat{\theta}|\text{data}))$ is the numerical value of the log-likelihood at its maximum point, $\hat{\theta}$ is the vector of the estimated model parameters and K the number of estimable parameters.

According to the information-theoretic approach, data analysis is assumed to be the integrated process of a priori specification of a set of candidate models (based on the science of the problem), model selection based on the principle of parsimony (i.e. selection of a model with the smallest possible number of parameters for adequate representation of the data; a bias versus variance tradeoff) according to AIC, and estimation of parameters and their precision.

Information-theoretic methods free the researcher from the limiting concept that the proper approximating model is somehow 'given'. When a model is 'picked' in some way, independent of the data, and used to approximate the data as a basis for inference, both the uncertainty associated with model selection and the benefits of selecting a parsimonious model are ignored. This strategy incurs substantial costs in terms of reliable inferences, because uncertainty in model selection is assumed to be zero. If model selection uncertainty is ignored, precision is often overestimated, estimated confidence intervals of the parameters are below the nominal level, and predictions are less accurate than expected (Burnham and Anderson 2002; Katsanevakis 2006). When the data support evidence of more than one model, model-averaging the predicted response variable across models is advantageous in reaching a robust inference that is not conditional on a single model. Rather than estimating parameters from only the 'best' model, parameters can be estimated from several or even all the models considered. This procedure is termed multi-model inference (MMI) and has several theoretical and practical advantages (Burnham and Anderson 2002). Model-averaging ideas are well developed from the Bayesian perspective (e.g. Hoeting et al. 1999) but have not yet been commonly adapted into applied frequentist inferences (Burnham and Anderson 2002).

Materials and methods

Allometric growth was investigated in eight different marine species: the bivalves *Pecten jacobaeus* and *Pinna nobilis*, the squids *Todarodes sagittatus* and *Todaropsis eblanae*, the crab *Pachygrapsus marmoratus* (females), the ghost shrimp *Pestarella tyrrhena* (males), and the fishes *Trachurus trachurus* and *Sparus aurata* (Fig. 1). All specimens were collected from Greek waters. In each of the eight species, a pair of body parts was measured and the allometric growth of one body part in relation to the other was studied (Fig. 1, Table 1).

Five candidate models were fitted to the log-transformed data (natural logarithms) of each dataset, with least squares: the linear (L), quadratic (Q), cubic (C), broken-stick (BS), and two-segment (TS) (Table 2). The L model was fitted with simple linear regression, while polynomial regression was used for the Q and C models. The BS and TS models assume a marked morphological change at a specific size of $X = B$. The BS model represents two straight line segments with different slope that intersect at $X = B$. The TS model represents two straight line segments that do not intersect; thus, there is a point of discontinuity at $X = B$, and the slope of the two segments may or may not be equal. To fit the BS and TS models, the breakpoint $X = B$ was allowed to vary between the minimum and maximum value of the independent variable with a sufficiently small step. For each value of the breakpoint, two separate lines were fitted with linear regression to the data before and after the breakpoint (independent lines in the case of TS or connected lines at the breakpoint in the case of BS)

and the corresponding residual sum of squares (RSS) was calculated as the sum of the two RSS for the two lines; this was done automatically in MsExcel by what-if analysis (one variable data table). The value of the breakpoint that gave the minimum RSS was found and the corresponding model parameters were estimated.

The L model assumes that allometry does not change as body size increases ($b = b_1 = $ const.). In the current context the 'allometric exponent b' of the allometric relationship of the morphometric variable Y in relation to the variable X is generalized and taken to mean the first derivative of $\log Y$ with respect to $\log X$. The Q and C models assume that b changes continuously with increasing body size ($b = b_1 + 2b_2\log X$ and $b = b_1 + 2b_2\log X + 3b_3(\log X)^2$, respectively). The BS and TS models assume that the allometric exponent takes two (generally different) constant values b_1 and b_2, before and after the breakpoint $X = B$.

The small-sample, bias-corrected form AIC_c (Hurvich and Tsai 1989) of the AIC (Akaike 1973; Burnham and Anderson 2002) was used for model selection. Specifically, $AIC_c = AIC + \frac{2k(k+1)}{n-k-1}$, where for least squares $AIC = n\left(\log\left(2\pi\frac{RSS}{n}\right) + 1\right) + 2k$, RSS is the residual sum of squares, n the number of observations, and k is the number of regression parameters plus 1 (normally distributed deviations with constant variance were assumed). The model with the smallest AIC_c value ($AIC_{c,min}$) was selected as the 'best' among the models tested. The AIC_c differences, $\Delta_i = AIC_{c,i} - AIC_{c,min}$ were computed over all candidate models. According to Burnham and Anderson (2002), models with $\Delta_i > 10$ have essentially no support and might be omitted from further consideration, models with $\Delta_i < 2$

Fig. 1 The pairs of morphometric characters measured for each species. For *Todaropsis eblanae* the morphometric characters measured were the same as for *Todarodes sagittatus*, and for *Sparus aurata* the same as for *Trachurus trachurus*. Pictures are not in scale. Abbreviations of morphometric characters as in Table 1
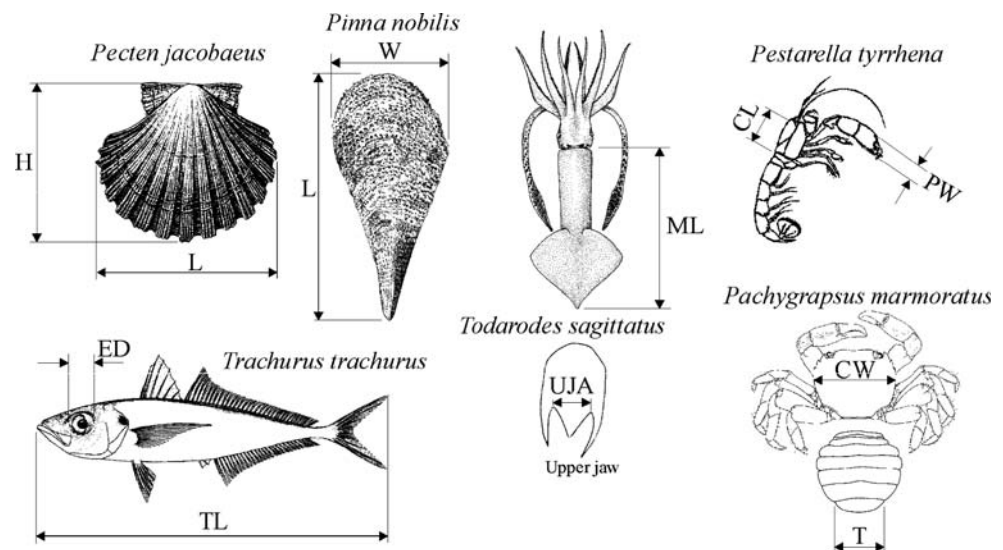
Table 1 The measured body parts and the corresponding size ranges for each species (Y is the dependent and X the independent variable), the corresponding sample size n, and the sources of data. **f** females, **m** males

	n	Y	$Y_{\min}$–$Y_{\max}$	X	$X_{\min}$–$X_{\max}$	Source
Pecten jacobaeus	245	Shell length (L, cm)	1.79–16.00	Shell height (H, cm)	1.72–13.85	Katsanevakis (2005a) plus 13 extra measurements at missing sizes
Pinna nobilis	98	Shell width (W, cm)	1.6–25.0	Shell length (L, cm)	6.1–70.1	Katsanevakis (2005b) plus 59 extra measurements
Todarodes sagittatus	76	Upper jaw angles distance (UJA, mm)	1.559–7.070	Mantle length (ML, mm)	121–341	Present study
Todaropsis eblanae	123	Upper jaw angles distance (UJA, mm)	1.186–4.284	Mantle length (ML, mm)	36–170	Present study
Pachygrapsus marmoratus (f)	51	Telson width (T, mm)	0.78–16.20	Carapace width (CW, mm)	4.34–34.90	Protopapas (2006)
Pestarella tyrrhena (m)	114	Propodus width (PW, mm)	0.28–7.39	Carapace length (CL, mm)	1.12–10.87	Present study
Trachurus trachurus	144	Eye diameter (ED, mm)	9.8–18.9	Total length (TL, mm)	160–313	Present study
Sparus aurata	243	Eye diameter (ED, cm)	0.44–2.06	Total length (TL, cm)	6.1–41.9	Present study

Table 2 The candidate models used for allometric growth investigation

Name of model	Abbr.	k	Equation
Linear	L	3	$\log Y = a_1 + b_1\log X$
Quadratic	Q	4	$\log Y = a_1 + b_1\log X + b_2(\log X)^2$
Cubic	C	5	$\log Y = a_1 + b_1\log X + b_2(\log X)^2 + b_3(\log X)^3$
Broken-stick	BS	5	$\log Y = \begin{cases} a_1 + b_1\log X, & X \le B \\ a_1 + (b_1 - b_2)\log B + b_2\log X, & X > B \end{cases}$
Two-segment	TS	6	$\log Y = \begin{cases} a_1 + b_1\log X, & X \le B \\ a_1 + b_2\log X, & X > B \end{cases}$

have substantial support, while there is considerably less support for models with $4 < \Delta_i < 7$. To quantify the plausibility of each model, given the data and the set of five models, the 'Akaike weight' w_i of each model was calculated, where $w_i = \frac{\exp(-0.5\Delta_i)}{\sum_{k=1}^{5}\exp-(0.5\Delta_k)}$. The 'Akaike weight' is considered as the weight of evidence in favor of model i being the actual best model of the available set of models (Akaike 1983; Buckland et al. 1997; Burnham and Anderson 2002). 'Average' models were estimated by averaging the predicted response variable across models, using the corresponding w_i as weights (Burnham and Anderson 2002). Absolute residuals were plotted against $\log X$ for the linear and the average model, as a diagnostic tool to check model assumptions and especially to check for curvature in the pattern of residuals; the display was enhanced by a

smooth loess (local regression) curve fitted to the residuals (first degree loess with a span of 0.6).

Results

The regression parameters and the corresponding AIC_c, Δ_i, and w_i of the five candidate models were estimated for each dataset (Tables 3, 4). For *P. jacobaeus*, L was the best model, but the rest of the models were also supported by the data, especially Q and TS. For *P. nobilis* TS was the best model with BS also having some support. For both squids *T. sagittatus* and *T. eblanae*, TS was the best model and no other models had substantial support. For *P. marmoratus*, C was the best model with TS having considerably less support. For *P. tyrrhena*, Q was the best model,

Table 3 The parameters and the residual sum of squares *RSS* of the regression equations between the logarithms of the measured biometric variables *Y* and *X*. Model abbreviations as in Table 2

Model		P. jacobaeus	P. nobilis	T. sagittatus	T. eblanae	P. marmoratus	P. tyrrhena	T. trachurus	S. aurata
L	a_1	0.0647	–0.7701	–6.4375	–2.2883	–2.7302	–1.4652	–1.9447	–1.8995
	b_1	1.0365	0.9581	1.4430	0.7199	1.5787	1.4556	0.8517	0.6945
	RSS	0.15696	2.27883	1.38386	1.92287	0.74479	0.89374	0.90759	1.96447
Q	a_1	0.0293	–3.6164	16.3485	–5.2748	–2.2304	–1.5791	–25.5047	–2.9811
	b_1	1.0779	2.8374	–7.1119	2.0841	1.1472	1.6598	9.5435	1.5640
	b_2	–0.0110	–0.2967	0.8011	–0.1548	0.0864	–0.0733	–0.8011	–0.1655
	RSS	0.15601	0.70981	1.18107	1.86484	0.71253	0.83594	0.85692	1.73616
C	a_1	–0.0070	–2.6807	158.6050	–5.1108	3.7138	–1.5201	–125.8580	–6.1495
	b_1	1.1513	1.8896	–87.7376	1.9713	–6.4521	1.4612	65.3062	5.2098
	b_2	–0.0568	0.0125	16.0093	–0.1291	3.1842	0.0936	–11.1238	–1.5197
	b_3	0.0089	–0.0326	–0.9547	–0.0019	–0.4064	–0.0406	0.6366	0.1636
	RSS	0.15577	0.70270	1.15845	1.86484	0.46855	0.82926	0.85578	1.68902
BS	a_1	0.0611	–1.6809	–3.3510	–2.6832	–1.824	–1.514	–4.930	–2.387
	b_1	1.0387	1.3029	0.8384	0.8141	1.067	1.499	1.418	0.933
	b_2	0.9681	0.5691	1.8273	0.4748	1.700	1.290	0.721	0.499
	logB	2.451	3.227	5.274	4.625	2.025	1.851	5.303	2.542
	RSS	0.15584	0.63666	1.06263	1.81874	0.59432	0.82561	0.83890	1.63651
TS	a_1	0.0743	–1.5715	–4.1557	–2.2700	–2.324	–1.499	–1.742	–1.424
	b_1	1.0288	1.2554	0.9970	0.7119	1.354	1.480	0.799	0.432
	a_2	0.1781	0.6436	–3.3953	–0.4620	–1.987	–1.036	–1.071	–1.218
	b_2	0.9905	0.5801	0.9160	0.3405	1.351	1.248	0.693	0.482
	logB	2.090	3.015	5.491	4.516	2.807	1.767	5.260	2.363
	RSS	0.15327	0.60761	0.97099	1.68396	0.47868	0.80928	0.77574	1.52479

Table 4 Values of AIC$_c$, AIC$_c$ differences (Δ_i) and of the Akaike weights w_i for the five models of the measured morphometric variables, for each species. For each species, values corresponding to the best models are in *bold* characters

	P. jacobaeus	P. nobilis	T. sagittatus	T. eblanae	P. marmoratus	P. tyrrhena	T. trachurus	S. aurata
AIC$_c$								
L	**–1100.11**	–84.24	–82.43	–156.22	–64.31	–223.00	–314.79	–475.03
Q	–1099.53	–196.37	–92.25	–157.85	–64.21	**–228.47**	–320.95	–502.98
C	–1097.82	–195.14	–91.42	–155.68	**–83.12**	–227.20	–318.99	–507.59
BS	–1097.72	–204.81	–97.98	–158.76	–71.00	–227.70	–321.86	–515.26
TS	–1099.68	**–207.12**	**–102.48**	**–166.01**	–79.45	–227.75	**–330.95**	**–530.34**
Δ_i								
L	**0.00**	122.88	20.04	9.80	18.81	5.47	16.16	55.31
Q	0.58	10.74	10.23	8.16	18.91	**0.00**	10.01	27.36
C	2.29	11.98	11.06	10.34	**0.00**	1.27	11.96	22.75
BS	2.40	2.30	4.49	7.26	12.13	0.77	9.09	15.08
TS	0.43	**0.00**	**0.00**	**0.00**	3.67	0.72	**0.00**	**0.00**
w_i (%)								
L	**31.5**	0.0	0.0	0.7	0.0	2.2	0.0	0.0
Q	23.5	0.4	0.5	1.6	0.0	**33.7**	0.7	0.0
C	10.0	0.2	0.4	0.5	**86.0**	17.8	0.2	0.0
BS	9.5	23.9	9.5	2.5	0.2	22.9	1.0	0.1
TS	25.5	**75.6**	**89.6**	**94.6**	13.8	23.4	**98.0**	**99.9**

while C, BS, and TS also had substantial support. For *T. trachurus* and *S. aurata*, TS was the best model with essentially no support for any other model. The L model had essentially no support in most cases (*P. nobilis, T. sagittatus, T. eblanae, P. marmoratus, T. trachurus*, and *S. aurata*), except for *P. jacobaeus* where it was the best model and *P. tyrrhena* where it had considerably less support than the best model.

From the loess fit of the residuals, it is clear that in all cases except for *P. jacobaeus* the assumption of linearity is more or less violated as the residuals of the L model do not have a random distribution around zero but exhibit evident curvature (Fig. 2). The situation clearly improved with the average model, with the loess fit of its residuals much closer to the horizontal axis (Fig. 2). In the case of *P. jacobaeus*, the residuals of the L model were randomly distributed around the horizontal axis and the average model did not make any essential difference in residual distribution.

Discussion

When investigating allometric growth, researchers almost always 'pick' the linear model for log-transformed data, which is quite simple both conceptually and mathematically and its parameters are easy to estimate by linear regression. Researchers also check for breakpoints using some version of broken stick or segmented model (e.g. Shea and Vecchione 2002), but this is not common practice. Examples of model selection for allometric growth studies are rare (e.g. Hall et al. 2006). In the examples of the present study, a large part of information would have been lost had we arbitrarily chosen the classic allometric model (L model). In fact, with the exception of the *P. jacobaeus* dataset, the L model had essentially no support or considerably less support in all other cases, gave residuals that were not randomly distributed but exhibited non-linear patterns, and gave estimations of the allometric exponent that actually 'smoothed' the true picture or even contradicted the outcomes of the more accurate average models.

During ontogeny, some somatic parts grow with a constant allometric exponent, others exhibit a change of their allometric exponent, while others show a discontinuity of their allometric exponent at a breakpoint. Thus, a set of candidate models including the simple linear, models that assume a continuous change in allometry, and models that assume discontinuity at breakpoints should be considered in allometric growth studies.

In the past twenty years, modern statistical science has been moving away from traditional formal methodologies based on statistical hypothesis testing. In particular, traditional approaches of hypothesis testing when applied to model selection (such as stepup, stepdown, and stepwise selection) have been often found to be poor and of limited value (Akaike 1981) and it is suggested that their application will diminish in future (Burnham and Anderson 2002). As the latter authors point out, hypothesis testing schemes are based on arbitrary levels of type I error probabilities (commonly 0.05 or 0.01), while multiple testing and tests between models that are not nested are problematic. The adjusted coefficient of multiple determination (R^2) is also often used in model selection, but this approach was found to be very poor (McQuarrie and Tsai 1998). Model selection based on K-L information theory is a relatively new and promising approach in biological sciences.

In the present study, the comparison of the estimated b-values from both the linear and the average model revealed some characteristic pitfalls in investigating allometry using the L model when it actually has no support from the data. The most serious risk is to judge the type of allometry wrongly, i.e. concluding positive allometry when there is actually negative or vice versa, or reporting allometry when the data support isometric growth or vice versa (Fig. 2). For example, in *P. nobilis* there is a marked change in the relative growth of width in relation to length; initially there is strong positive allometry which after a length of ~20 cm becomes strongly negative. The linear model 'smoothes' this picture and derives an allometric exponent with a 95% confidence interval between 0.90 and 1.03, supporting isometric growth during ontogeny and thus reaching a quite different conclusion. In *T. sagittatus* positive allometry of UJA in relation to ML was supported by the L model, while piecewise isometry was supported by the average model (Fig. 2). In the case of female *P. marmoratus* the allometry for the growth of the telson in relation to carapace width was initially negative (for small sizes), but the allometric exponent increased continuously reaching values > 1 (positive allometry), attained a maximum of 1.86 and decreased again in larger individuals. This pattern is probably related to maturity, when the size of the abdominal segments in females has to grow in a much higher rate ($b >> 1$) in order to increase the size of the abdomen for the accommodation of incubated egg mass, while later in ontogeny the relative growth rate may decline as the abdomen size is constrained by the width of the carapace. Constant positive allometry during ontogeny with a high and constant $b = 1.58$ was concluded by the L model, missing the true pattern.

The main outcome of allometric growth studies under the classic approach (i.e. using the L model) is to define the allometric exponent and consequently the type of allometry. Such an approach fails to locate the possible existence of discontinuities in relative growth curves. Such changes in the growth trajectories of morphological characters during ontogeny are a potentially useful source of information as they may be caused by marked events in the life history of the species or fast ecological change, and should not be overlooked. Quite often maturity might be the cause of a distinct change in morphology. The attainment of morphometric maturity in crabs is identified by finding such breakpoints in the allometry of certain body parts at the puberty moult, usually chelae in male crabs and abdominal segments in females (Somerton 1981; Hall et al. 2006; Protopapas 2006). Shea and Vecchione (2002) using a variation of a BS model and a quite different approach than K-L information theory for deciding on the significance of breakpoints, found discontinuities in several morphological measurements

Fig. 2 For each species: (Top) the relative growth patterns of the measured pair of morphometric characters, according to the L model (light line) and the corresponding average model (dark line), (Middle) plot of the absolute residuals against log$X$ for the linear and the average model, with a loess curve fitted to the residuals (first degree loess with a span of 0.6), (Bottom) the allometric exponent b for the relative growth of the measured pair of morphometric characters, according to the L model (light line) and the corresponding 'average' model (dark line); the bootstrap 95% confidence interval of b, according to L model is given with dotted lines. Abbreviations of morphometric characters as in Table 1
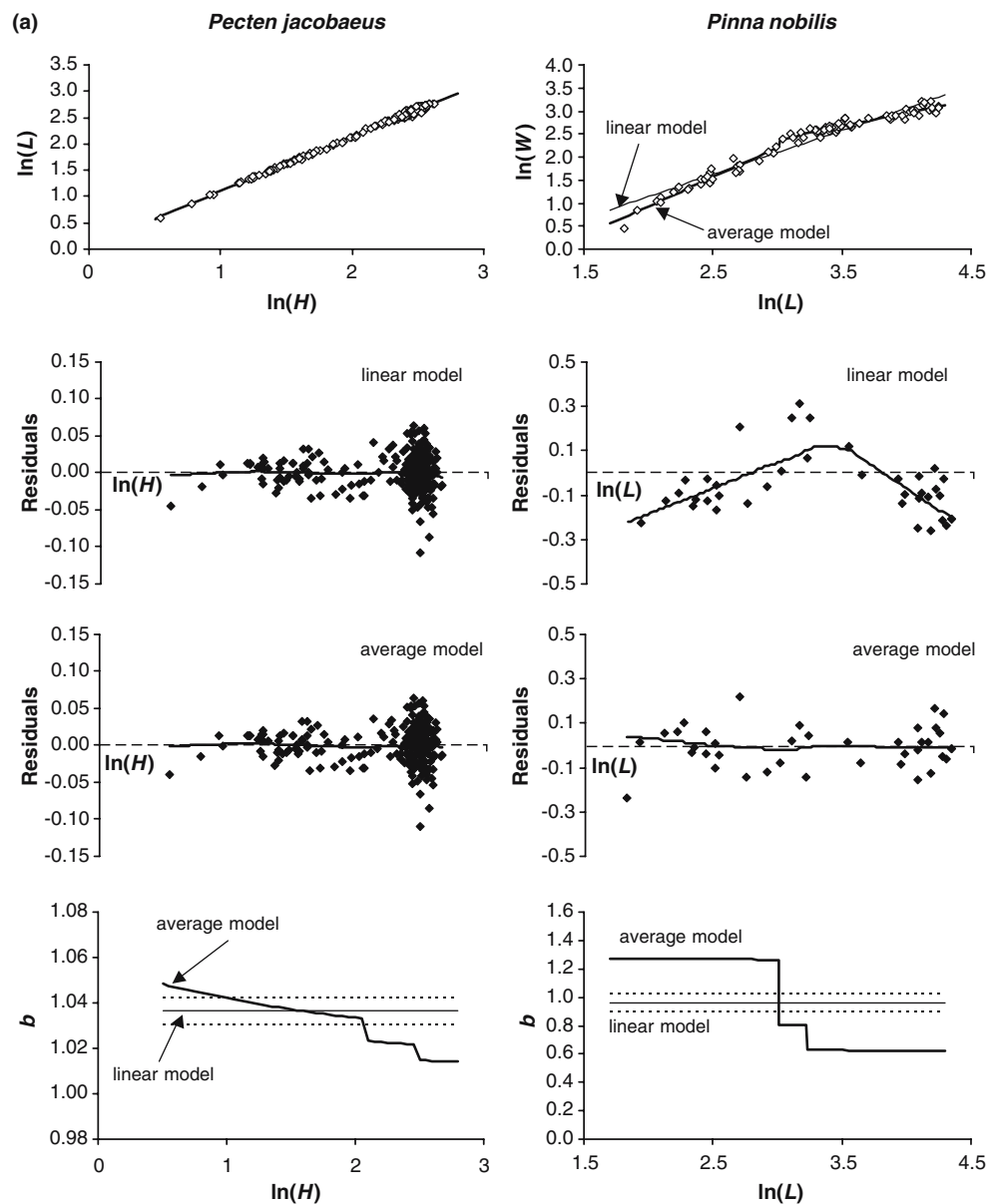

of three species of squid; for each species, there was a clustering of breakpoints into discrete size ranges and it was considered as an 'allomorphosis', which may correlate with rapid ecological change. The information theoretic approach is quite effective in locating breakpoints and thus providing more information than the classic approach, which only states (often wrongly) the type of allometry.

One of the assumptions of the ordinary least squares (OLS) regression (or Model I regression) is that the independent variable is under the control of the investigator or known without error. However, with morphological measurements, when we want to find the functional relationship between two (usually log-transformed) morphological variables y and x, both

$\ln(x)$ and $\ln(y)$ are subject to natural variability and measurement errors and, thus, Model I regression may not be appropriate. The Model I regression coefficient (the allometric exponent in our case) is expected to be lower in absolute value than the true slope of the functional relationship (Laws and Archie 1981; Sokal and Rohlf 1995; Prairie et al. 1995). Several authors have recommended the replacement of OLS regression with other slope estimators (subsumed under Model II regression), such as the ordinary major axis or the reduced major axis (e.g. Ricker 1973; Jolicoeur 1990; Sokal and Rohlf 1995; Ebert and Russel 1994) when the observations are believed to be uncertain due either to measurement errors or natural error variability. The appropriateness, though, of any of these slope

Fig. 2 continued
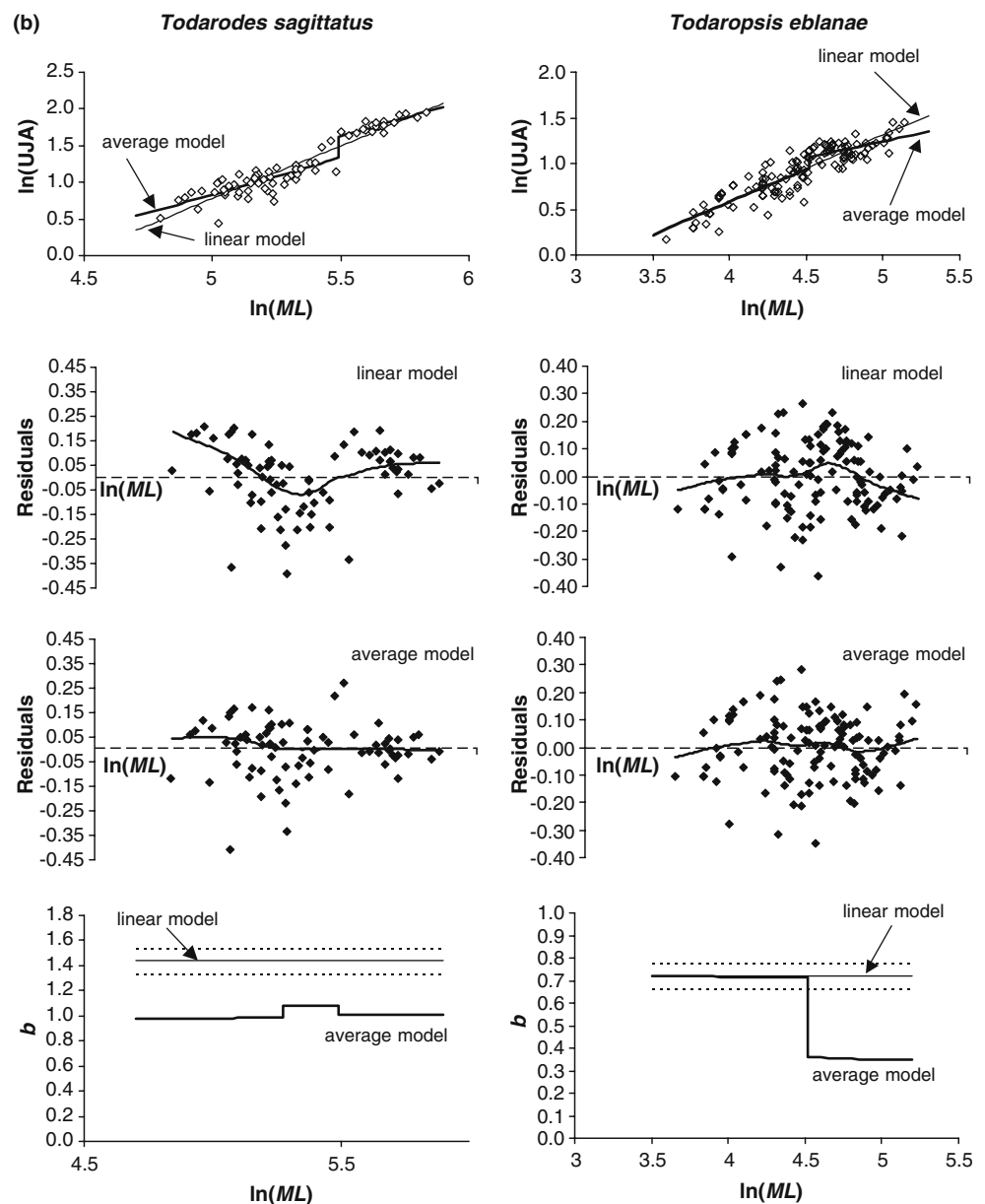
estimators depends on whether the specific data at hand conform to the very restrictive assumptions of each of these techniques. According to Prairie et al. (1995), "the advice often found in the ecological literature regarding the use of these alternative regression techniques should be tempered with the realization that, without an estimate of the natural error variability component (not just measurement errors), these estimators are equally likely to be worse than they are to be better than the OLS estimates". Furthermore, regarding Model II regression almost all published work deal with the simple linear case (of log-transformed data). Only a few publications deal with specific non-linear Model II regressions (e.g. Ebert and

Russell 1994), whereas a unified non-linear Model II approach is lacking. In addition, standard statistical software does not incorporate Model II regression, making its use non-trivial. Calculation of AIC's through maximum likelihood estimation under a Model II non-linear approach might also be challenging (in relation to the simplified AIC calculation with OLS).

Several other methods have been proposed to estimate the functional relationship between two morphological variables (e.g. Prairie et al. 1995; Shafer and Purdy 1996) but they are either mathematically complicated or restricted to the simple linear model. The main target of the present study was to propose a

Fig. 2 continued

(c) *Pachygrapsus marmoratus*   *Pestarella tyrrhena*
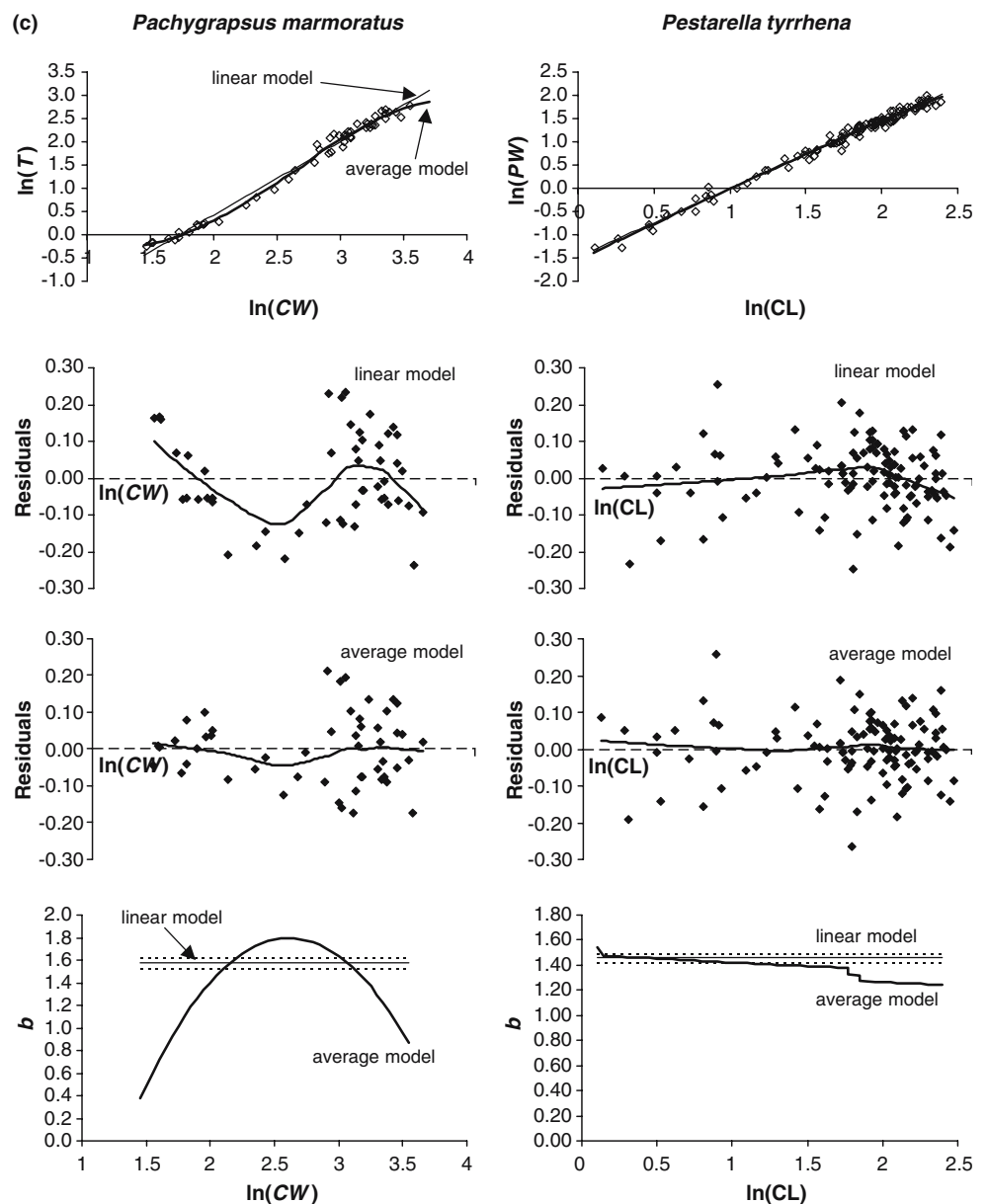
simple and easy to implement (using standard statistical software) method to deal with the allometry problem (incorporating information theory and multi-model inference), capable to extract additional information such as the existence of discontinuities in relative growth curves that are related to episodes of biological importance (such as maturity).

For all the above reasons we kept the simple OLS approach, bearing in mind that our regression results may be slightly biased and are susceptible of improvement. According to Prairie et al. (1995), if both y and x are subject to random error such that $Y = y + \varepsilon_y$ and $X = x + \varepsilon_x$ (where X, Y are the observed values, x, y the true values, $y = \alpha + \beta_{yx}x$ is the true relationship, and ε_x, ε_y the random errors of x and y respectively) the OLS slope estimator β_{YX} is given by the relationship: $\beta_{YX} = \beta_{yx}\ [1-\mathrm{var}(\varepsilon_x)/\mathrm{var}(X)]$. One may sample as wide a range in the predictor variable (x) as possible so that $\mathrm{var}(\varepsilon_x)$ will be small relative to $\mathrm{var}(X)$ and consequently $\mathrm{var}(\varepsilon_x)/\mathrm{var}(X) \rightarrow 0$, thereby making $\beta_{YX} \approx \beta_{yx}$; in this case, OLS is sufficient to estimate the true functional relationship. In our datasets, we tried to sample the predictor variable in the widest possible range; nevertheless it was difficult to know how wide a range is wide enough to ensure that $\beta_{YX} \approx \beta_{yx}$. To check the appropriateness of OLS regression for our datasets (based on the linear model) we applied the 'slope-range' method of Prairie et al.
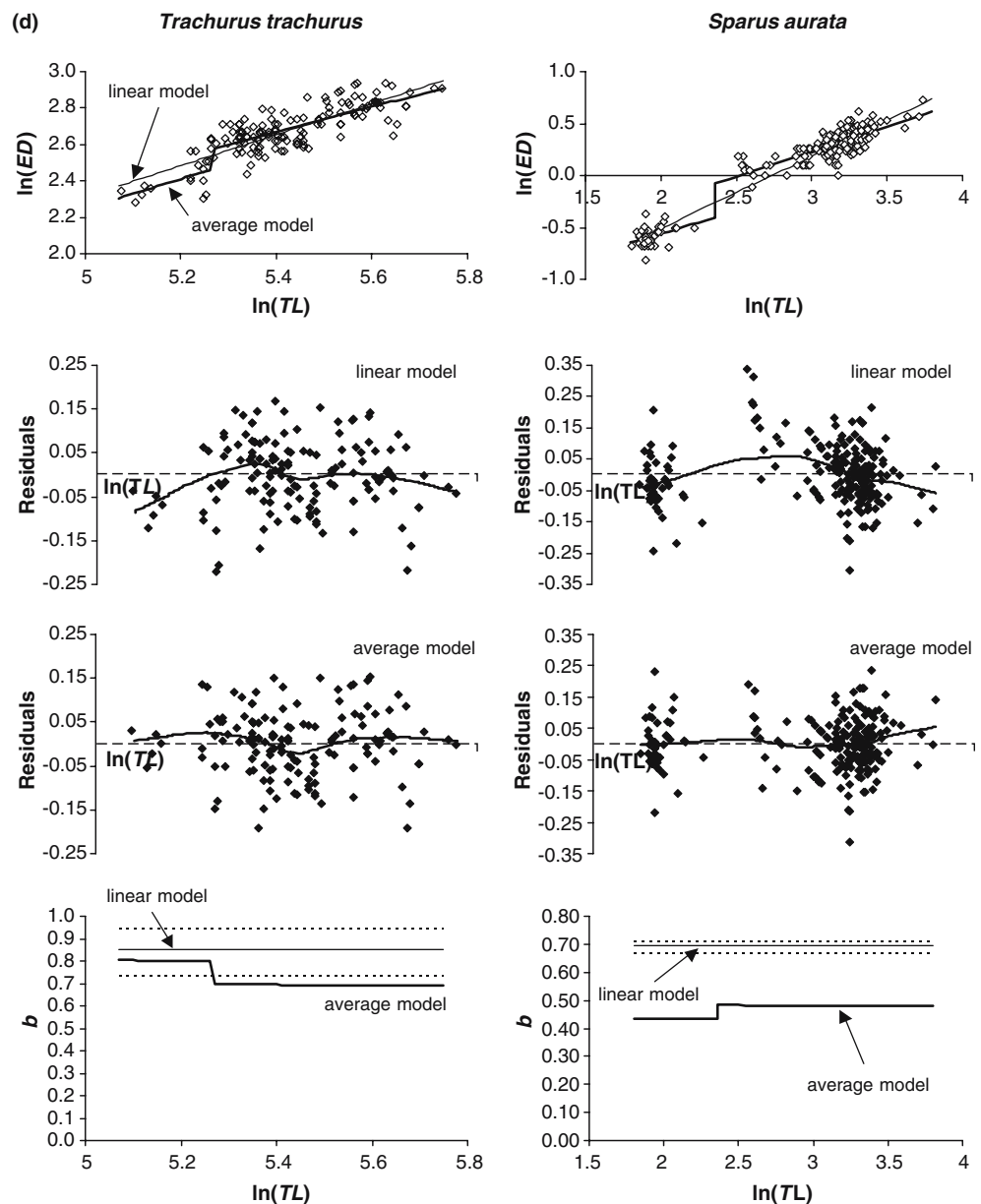
Fig. 2 continued

Table 5 Estimation of the random error ε_x associated with the 'independent' variable x

Var(X) is the observed variance of the 'independent' variable, and the ratio var(ε_x)/var(X) is an index of the bias of slope estimations based on OLS regression

Dataset	var(ε_x)	var(X)	var(ε_x)/var(X) (%)
P. jacobaeus	3.84E-05	0.194	0.02
P. nobilis	9.54E-04	0.422	0.23
T. sagittatus	1.97E-03	0.066	3.00
T. eblanae	3.44E-03	0.121	2.84
P. marmoratus (f)	1.45E-03	0.409	0.36
P. tyrrhena (m)	1.26E-03	0.274	0.46
T. trachurus	4.26E-04	0.019	2.30
S. aurata	3.78E-04	0.263	0.14

(1995) to estimate ε_x and then the ratio var(ε_x)/var(X) (Table 5). In most datasets the ratio var(ε_x)/var(X) was < 0.5% and thus it seems that OLS is sufficient, while for three datasets (*T. sagittatus, T. eblanae* and

T. trachurus) it was between 2 and 3%, indicating that results might be improved if more complicated errors-in-variables models were incorporated. Further research is necessary in this direction.

References

Akaike H (1973) Information theory and an extension of the maximum likelihood principle. In: Petrov BN, Csaki F (eds) Second international symposium on information theory. Akademiai Kiado, Budapest, pp 267–281

Akaike H (1981) Likelihood of a model and information criteria. J Econom 16:3–14

Akaike H (1983) Information measures and model selection. Int Stat Inst 44:277–291

Buckland ST, Burnham KP, Augustin NH (1997) Model selection: an integral part of inference. Biometrics 53:603–618

Burnham KP, Anderson DR (2002) Model selection and multimodel inference: a practical information-theoretic approach, 2nd edn. Springer, Berlin Heidelberg New York

Ebert TA, Russell MP (1994) Allometry and Model II non-linear regression. J Theor Biol 168:367–372

Hall NG, Smith KD, de Lestang S, Potter IC (2006) Does the largest chela of the males of three crab species undergo an allometric change that can be used to determine morphometric maturity? ICES J Mar Sci 63:140–150

Hoeting JA, Madigan D, Raftery AE, Volinsky CT (1999) Bayesian model averaging: a tutorial (with discussion). Stat Sci 14:382–417

Hurvich CM, Tsai CL (1989). Regression and time series model selection in small samples. Biometrika 76:297–307

Huxley JS (1932) Problems of relative growth. Methuen, London

Jolicoeur P (1990) Bivariate allometry: interval estimation of the slopes of the ordinary and standardized major axes and structural relationship. J Theor Biol 144:275–285

Katsanevakis S (2005a) Abundance and spatial distribution of the Mediterranean scallop, *Pecten jacobaeus*, in a marine lake. Fish Res 76:417–429

Katsanevakis S (2005b) Population ecology of the endangered fan mussel *Pinna nobilis* in a marine lake. Endanger Species Res 1:1–9

Katsanevakis S (2006) Modelling fish growth: model selection, multi-model inference and model selection uncertainty. Fish Res 81:229–235

Kullback S, Leibler RA (1951) On information and sufficiency. Anal Math Stat 22:79–86

Laws EA, Archie JW (1981) Appropriate use of regression analysis in marine biology. Mar Biol 65:13–16

McQuarrie ADR, Tsai CL (1998) Regression and time series model selection. World scientific publishing company, Singapore

Prairie YT, Peters RH, Bird DF (1995) Natural variability and the estimation of empirical relationships: a reassessment of regression methods. Can J Fish Aquat Sci 52:788–798

Protopapas N (2006) Biology of the grapsid crab *Pachygrapsus marmoratus*. MSc thesis. National and Kapodistrian University of Athens, Athens

Ricker WE (1973) Linear regression in fishery research. J Fish Res Board Can 30:409–434

Schafer DW, Purdy KG (1996) Likelihood analysis for errors-in-variables regression with replicate measurements. Biometrika 83:813–824

Shea EK, Vecchione M (2002) Quantification of ontogenetic discontinuities in three species of oegopsid squids using model II piecewise linear regression. Mar Biol 140:971–979

Sokal RR, Rohlf FJ (1995) Biometry. Freeman, New York

Somerton DA (1981) Regional variation in the size of maturity of two species of tanner crab (*Chionoecetes bairdi* and *C. opilio*) in the eastern Bering Sea, and its use in defining management subareas. Can J Fish Aquat Sci 38:163–174