



On the Choice of Linear Regression Algorithms for Biological and Ecological Applications

Vasco M. N. C. S. Vieira^{1*}, Joel Creed², Ricardo A. Scrosati³, Anabela Santos⁴, Georg Dutschke⁴, Francisco Leitão⁵, Aschwin H. Engelen⁵, Oscar R. Huanel⁶, Marie-Laure Guillemin^{6,7}, Marcos Mateus¹ and Ramiro Neves¹

¹MARETEC, Instituto Superior Técnico, Universidade Técnica de Lisboa, Av. Rovisco Pais, 1049-001, Lisboa, Portugal.

²Departamento de Ecologia, Instituto de Biologia Roberto Alcântara Gomes, Universidade do Estado do Rio de Janeiro, Rua São Francisco Xavier 524, 20.559-900, Rio de Janeiro, Brazil.

³Department of Biology, Saint Francis Xavier University, Antigonish, Nova Scotia B2G 2W5, Canada.

⁴Universidade Autónoma de Lisboa, Rua de Santa Marta, nº 56 - 1169-023, Lisboa, Portugal.

⁵CCMAR, Center of Marine Science, University of Algarve, Campus Gambelas, 8005-139 Faro, Portugal.

⁶Instituto de Ciencias Ambientales y Evolutivas, Facultad de Ciencias, Universidad Austral de Chile, Casilla 567, Valdivia, Chile.

⁷CNRS, Sorbonne Universités, UPMC University Paris VI, UMI 3614, Evolutionary Biology and Ecology of Algae, Station Biologique de Roscoff, CS 90074, Place G. Tessier, 296888 Roscoff, France.

Authors' contributions

This work was carried out in collaboration between all authors. Authors VMNCSV, MM and RN designed the study, developed the software and analyzed the data. Author AHE analyzed the data. Authors JC, RAS, AS, GD, FL, ORH and MLG anchored the field study, gathered the initial data and performed preliminary data analysis. All authors managed the literature searches, produced the initial draft, read and approved the final manuscript.

Article Information

DOI: 10.9734/ARRB/2016/25219

Editor(s):

(1) George Perry, Dean and Professor of Biology, University of Texas at San Antonio, USA.

Reviewers:

(1) Alejandro Córdova Izquierdo, Universidad Autónoma Metropolitana Unidad Xochimilco, Mexico.

(2) Douglas S. Glazier, Juniata College, Huntingdon, PA, USA.

(3) Kunio Takezawa, National Agriculture and Food Research Organization, Japan.

Complete Peer review History: <http://sciencedomain.org/review-history/14639>

Method Article

Received 23rd February 2016
Accepted 2nd April 2016
Published 16th May 2016

ABSTRACT

Model II regression (i.e. minimizing residuals obliquely) is the adequate alternative to Model I regression by Ordinary Least Squares (i.e. minimizing residuals vertically) given the absence of well-established dependence relationships or x measured with error. Yet, it has no perfect solution. Determining the true slope from errors-in-the-variables models requires the errors in x and y estimated from higher order moments. However, their accurate estimation requires enormous data sets and thus they are not applicable to most ecological problems. The alternative Reduced Major Axis (RMA) is dependent on a strict set of assumptions, hardly met with real data, making it prone to bias, whereas Principal Components Analysis (PCA) becomes less reliable with decreasing correlations while x and y presenting approximate variances. We used artificial data (allowing for the determination of the true slope) to demonstrate when RMA or PCA should be preferred. Consequently, we propose using PCA whenever $r^2 + s_x^2/s_y^2$ is higher than 1.5. Otherwise, we suggest generating artificial data manipulated to match the structure of the original, and to test which method provides closer estimates to the input true slope. We provide a user-friendly script to perform this task. We tested the use of RMA and PCA with real data about intraspecific and interspecific biomass-density relations in algae and seagrass, algae frond growth, crustacean and bird morphometry, sardine fisheries and social sciences data, commonly finding widely divergent slope estimates leading to severely biased parameter estimations and model applications. Their analyses support the suggested approach for method selection summarized above.

Keywords: Model II regression; Principal Components Analysis; Reduced Major Axis.

1. INTRODUCTION

The common Ordinary Least Squares (OLS), Iterative Reweighted Least Squares (IRLS) and Maximum Likelihood Estimation (MLE) methods for estimating unknown parameters in linear regressions, also known as model I regression, describe the relation between x and y variables performing a vertical least squares minimization, under the inherent assumption that x is measured in the absence of error. More specifically, Draper [1] identified three situations where this procedure is adequate: (i) when the error in the x_i estimation (δ) is small compared to the error in the y_i estimation (ϵ), (ii) when the x_i are fixed and determined by the experimenter, or (iii) when the experimenter wishes to estimate y_i from an actual x_i observation, irrespective of the error contained within it. Otherwise, least squares should be minimized perpendicularly to the regression line, commonly known as model II regression. Smith [2] extended the list of mostly philosophical arguments in favour of each regression type, with a *a priori* choice taking into consideration the research objectives, the data collection constrains and previous knowledge about the data, but not the goodness-of-fit.

Theoretically, model II regression can estimate the true slope from error-in-the-variables models, for which several solutions are available. The “instrumental variables” alternatives require a subset of data with known errors for training the

model; constrain hardly met by ecological data sets. The Generalized Method of the Moments or Cumulants solves it using lower-order moments provided the errors in the x estimates (σ_x) and y estimates (σ_y) are known; another constrain hardly met. The unknown error structure can be inferred from higher-order moments and cumulants. However, this renders the method ill-conditioned and prone to numerical instability, only yielding accurate results under a set of restrictions and provided very large sample sizes (samples as big as 600 may not be big enough). The “grouping” alternatives, partitioning data into subsets, are problematic because there is no clear consensus on the adequate number of groups. Furthermore, testing with fish biology data, Ricker [3] found RMA being superior to “grouping” methods. For more insights into errors-in-the-variables see the works by Cragg [4], Dagenais and Dagenais [5], Gillard and Iles [6] and Gillard [7].

The classical Principal Components Analysis (PCA) and Reduced Major Axis (RMA) alternatives theoretically optimize their performance in opposing circumstances, with severe impact on data analysis. Both the RMA and PCA intercepts are given by $a = \bar{y} - b\bar{x}$, since regression lines must pass through the bivariate mean [8,9,10]. The problem lies in their slope estimates. PCA was initially developed by Pearson [11] to determine the line of best fit from

the dominant eigenvector. Only later it became successfully applied to the generalized multivariate case [9,12]. The PCA slopes are given by the ratio of the loadings of the dominant eigenvector (u_y/u_x). Usually, this is extracted from the (x,y) covariance matrix, which standardizes the principal components to have zero mean and σ^2 variances [9,12]. However, as two consecutive principal components tend to justify equal amounts of variation (i.e. present approximate eigenvalues) their true axis orientation becomes less reliable [9]. With the bivariate case, this happens with decreasing correlation between x and y while simultaneously exhibiting approximate variances, i.e. the more the dispersion cloud approximates a perfect circumference, in which case the slope estimates are biased [13]. The RMA method estimates the Geometric Mean Slope (GMS) for the vertical regression of y on x and the horizontal regression of x on y , corresponding to the ratio of the standard errors (σ_y/σ_x) [7,8,9,10,14,15]. However, this does not correspond to the line of best fit given by equation (1), where b is the slope, a is the y -intercept, S_{xx} , S_{yy} and S_{xy} the variances-covariances of x and y , and $\lambda = \sigma_y/\sigma_x$. Two properties of equation (1) are fundamental to RMA: (i) when $\lambda = S_{yy}/S_{xx}$ equation (1) becomes $b = (S_{yy}/S_{xx})^{1/2}$, the RMA ratio of the standard errors (σ_y/σ_x), and (ii) the sum of squares of perpendicular deviations is only minimized when $\lambda = 1$. Hence, RMA only yields the best fit upon the unlikely event that $\lambda = \sigma_y/\sigma_x = S_{yy}/S_{xx} = 1$ [1,2,7]. Otherwise, the slope has to be calibrated by $\lambda = \sigma_y/\sigma_x$ accordingly to Draper ([1], equation 12) and Smith ([2], equation 6). However, σ_y is usually unknown in ecological data sets and thus the RMA slope estimates usually come biased. Creasy [16] and Gillard [7] argue that methods relying on the geometric mean slope may also be prone to indefinite axis orientation.

$$b = \frac{S_{yy} - \lambda S_{xx} + \sqrt{(S_{yy} - \lambda S_{xx})^2 + 4\lambda S_{xy}^2}}{2S_{xy}}$$

$$a = \bar{y} - b\bar{x} \quad (1)$$

Although of common use in biology and ecology, the numerical background and constraints of linear regression methods, and in particular of the most useful model II regression algorithms, are seldom well understood by biologists and ecologists. For the above reasons, the objective of this paper is to review briefly the RMA and PCA methods, investigate under which conditions each should be preferred and

demonstrate the consequences for biological and ecological applications of making a poor choice.

2. TESTING ALTERNATIVE METHODS USING AN ARTIFICIAL DATA SET

An artificial data set was used to test how PCA and RMA accuracy is affected by the known variances of x , y and their residuals. The observed x was the sum of its true value (ξ) and associated error (u), i.e. $x = \xi + u$. The ξ varied from 0.05 to 5 at 0.05 intervals, whereas u had a distribution $N_u(0, \sigma_u^2)$ with σ_u ranging from 0.5 to 1 at 0.02 intervals. The observed $y = \phi x + \varepsilon$, with the true slope ϕ ranging from 0.1 to 2 at 0.1 intervals, whereas the error ε had a distribution $N_\varepsilon(0, \sigma_\varepsilon^2)$ with σ_ε ranging from 0.5 to 3 at 0.01 intervals. A small Matlab script simulating this data and plotting its analysis is provided as supplementary material, together with a script enabling the user to simulate his bivariate data with normal distribution and run a similar test. Both scripts allow the simulation of artificial data with other distributions besides the Normal.

The distances between the PCA, RMA and true slopes (ϕ) were estimated by the sine of the angle formed ($\sin\theta$) using to the geometric definition of inner product of two vectors, the sine-cosine trigonometric relation and the Euclidean norm. Besides, two other statistics were obtained for each linefit, namely (i) the square of Pearson's correlation coefficient (r^2) between x and y , and (ii) the unevenness of the x and y variances given by σ_y^2/σ_x^2 .

PCA was unable to estimate the true slope when correlations were weak, as expected from Jackson [9], but provided y was not much less scattered than x (Fig. 1). Under these circumstances RMA estimated the true slope conspicuously better than PCA, which should be of restricted relevance as ecologists are seldom interested in slopes relative to weakly correlated variables. For the remaining situations RMA generally performed not better than PCA, often failing tremendously either underestimating or overestimating, as debated by Creasy [16] and Gillard [7]. A helpful rule of thumb is to prefer PCA whenever $r^2 + \sigma_x^2/\sigma_y^2$ is higher than 1.5 (and using s^2 as estimate of σ^2). This way, x and y are well correlated and/or have clearly unbalanced variations. In such cases PCA is guaranteed to provide good approximations, although this comprises situations where RMA is equally well suited. Otherwise, try the Matlab script provided as supplement to generate artificial data similar

to the desired real regarding slope, means and error structure, and check which method is expected to approximate better the true slope under those specific circumstances.

3. TESTING ALTERNATIVE METHODS USING REAL DATA

PCA and RMA were applied to nine factual data sets about intraspecific and interspecific biomass-density relations in algae and seagrass, algae frond growth, crustacean and bird morphometry, sardine fisheries and social sciences data, thus encompassing a wide spectrum of proportionalities among the variances of x , y , δ and ϵ . Where possible, were applied the same metrics used with the artificial data set. The biggest differences between PCA and RMA estimates occurred for irrelevant x-y correlations (i.e, statistically non-significant). Nevertheless, there were still plenty of meaningful correlations for which both methods yielded widely different slopes (Fig. 2). Many cases were identified by the rule of thumb as better suited for PCA. Nevertheless, for most of the topics in ecology it was not possible to guarantee that one particular method was always better. Therefore, it is wise for researchers to test their data using artificial data as a proxy as suggested above, before choosing between PCA or RMA. Below, we present several cases highlighting the impacts of taking the wrong option and demonstrating the benefits of the rule of thumb and the artificial data generator script.

3.1 Fisheries Data

Sardine fisheries data was collected on a monthly basis during 21 years (1989-2009) in 3 regions along the Portuguese coast (North-western, South-western and South-Algarve [17]). We used Sardine Landings (Kg) and Unit Effort (boatxdays) from the south-western region to estimate Surplus Production Models and demonstrate how parameter estimation from inadequate regression models can lead to biased estimates of optimal fishing effort and maximum sustainable yield (Fig. 3). OLS is the commonly used method to regress catch per unit effort (Y/f) on effort (f) when applying the Surplus Production Models, by Schaefer [18] and Fox [19]. We compared it to the use of PCA and RMA. Only when using the Fox model, did PCA produce estimates as well as OLS. But when using the Schaefer model, both RMA and PCA performed terribly.

3.2 Intraspecific Macroalgae Biomass-density Data

Macrostages of *Laminaria digitata* Lamouroux were cultivated on 10x15 cm plates with initial densities of 10, 20, 30, 40 and 80 individuals per plate corresponding to 650, 1334, 2000, 2668 and 5186 individual per m² [20]. The experiments were carried in culture tanks at Port Erin Marine Laboratory, Isle of Man, with four replicates were made for each density.. The density and stand biomass on each plate was accessed for 11 sampling times at approximately 19 day intervals, to determine whether the fronds were subject to different competitive stresses reflected by various self-thinning slopes (on a $\log B - \log D$ plot). Only, the first nine sampling times were used because afterwards the fronds stopped self-thinning. Only the four higher densities were used in this study, since the absence of mortality in the lower one yielded an infinite slope. A similar procedure was adopted for experiments with *Fucus serratus* Linnaeus [20]. However, in this case the higher density was absent and only the time instances 2nd to 7th were used, as self-thinning was restricted to this period. Overall, we estimated four slopes for *Laminaria digitata* and three slopes for *Fucus serratus*.

Stand biomass and frond (ramet) density for *Mazzaella parksii* (Setchell & N.L. Gardner) Hughey, P.C. Silva & Hommersand (= *M. cornucopiae* (Postels & Ruprecht) Hommersand) were estimated non-destructively for seven quadrats sized 100 cm² on the west coast of Vancouver Island, Canada [21]. Each quadrat was sampled seven times at approximately bimonthly intervals from 4-6 June 1993 to 11-13 July 1995. Time series of the quadrats were represented in a $\log B - \log D$ plot. All regression lines presented positive slopes. Thus, the fronds of *M. parksii* did not undergo self-thinning on the account of most neighbouring fronds being clonal ramets sharing resources through a common holdfast. For the present analysis, the quadrats were used to estimate seven slopes.

Self-thinning macroalgae often fell in the undetermined zone of the rule of thumb (Fig. 2), requiring dedicated tests to determine the appropriate method. The self-thinning *L. digitata* error structure was approximated by artificial data using its generator script. The best choice between PCA and RMA was case specific. When RMA failed, it always underestimated the

steepness of the slope, thus also underestimating the intercept. This resulted in overestimated intraspecific competitive pressures and underestimated maximum biomasses.

3.3 Interspecific Seagrass Biomass-density Data

Data on above-ground biomass and shoot density were collected for 28 seagrass species

distributed among 11 experimental and 17 descriptive studies [22]. Each species was sampled under natural conditions and nutrient enrichment. Four static interspecific $\log B$ - $\log D$ regression lines were estimated relative to the 2 nutrient levels (high and low) \times 2 study types (Experimental and Descriptive). The data used here was taken from Fig. 2 in Cabaço et al. [22]. It was a clear situation where the data dispersion cloud approximated a circumference, and thus

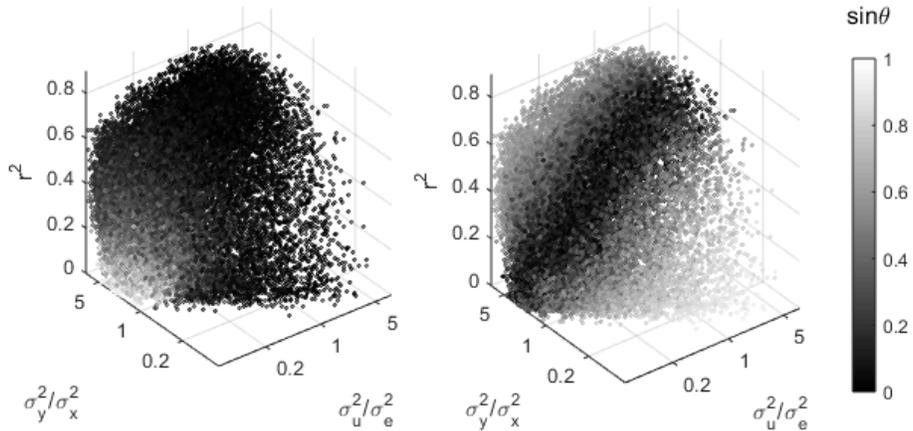


Fig. 1. Performance of PCA and RMA applied to the artificial data. Greyscale – sine of the angle θ formed between the estimated and the true slope

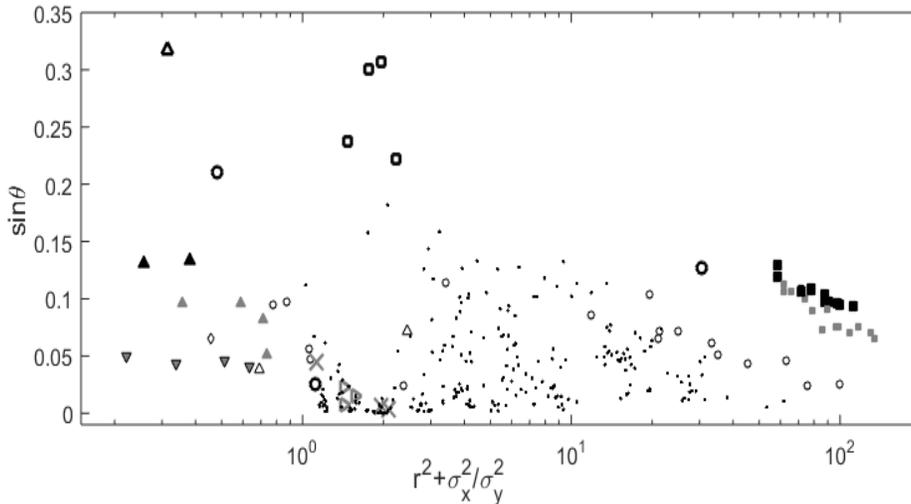


Fig. 2. Divergency between slopes estimated by PCA and RMA

Markers legend: Large black symbols – statistically non-significant correlations, small grey symbols– statistically significant correlations, full squares - culture-to-felicity, full inverted triangles - self-thinning *Laminaria digitata*, open triangles - self-thinning *Fucus serratus*, full triangles – *Mazzaella parksii* biomass-density relationship, open squares - seagrass interspecific biomass-density relation, big open laid triangles – *Gracilaria chilensis* growth, full circles – *Nephrops norvegicus* morphometry, open circles - sparrow morphometry, diamonds – Sardine Surplus Production Models, big crosses – Sardine landings per fishing effort

PCA could not give trustable results [13]. In this situation, RMA provides better estimates. Not surprisingly, the highest mismatches between PCA and RMA slopes occurred with this data set (Fig. 2). The dispersion cloud will always resemble a circumference if x and y are weakly correlated and have been scaled. When PCA is performed using the correlation matrix, this scaling is automatic. In the example below with the social science data, x and y were weakly correlated but presented widely unequal variances. In this case, performing PCA using the covariance matrix gave better results than RMA.

3.4 Macroalgae frond Growth Data

Tetrasporophytes, as well as male and female gametophytes of *Gracilaria chilensis*, an isomorphic biphasic life-cycle red algae, were followed in two Chilean localities during 2 years at 4 month intervals, having their volume estimated (cm^3), among other metrics (unpublished data). The demographic census followed the protocol described by Engel et al. [23] while phases and sexes were defined using the sex markers described by Guillemín et al. [24]. Positive changes in volume with time (t) corresponded to frond growth whereas negative changes, corresponding to frond breakage, were discarded. The growth model best fitting

the data was a modification of the Gompertz growth model where the linear fit estimating model parameters also required the logarithm of x . Then, the linear relation became $\log(\Delta V/\Delta t \cdot 1/t) = a + b \cdot \log(V)$. Assuming that $V_0=0$, the frond growth dynamic model became $V_t = -bct^{1/b}$, with $c=e^a$ and b always exhibiting negative values.

OLS is the popular choice for parameter estimation when modelling ecology and evolution and/or fitting the Logistic or the Gompertz density-dependent sigmoidal growth models [25,26]. However, Evans [27] argues OLS is unlikely to estimate econometric growth regressions consistently, proposing a Model II regression alternative by Instrumental Errors-in-Variables Model. *Sensu strictu* seems intuitive that growth is dependent on present size. However, future size is dependent on current growth and hence, *sensu lato* there should not be a hierarchical relation between them. The *Gracilaria chilensis* frond growth data (unpublished) demonstrated that OLS estimates regression coefficients widely different from PCA and RMA with a severe impact on growth model performance (Fig. 3). In this case PCA and RMA match and it is OLS underestimating frond size at earlier ages. Moreover, because fecundity is size dependent, this bias propagates to biased estimates of population growth, generation times and reproductive value.

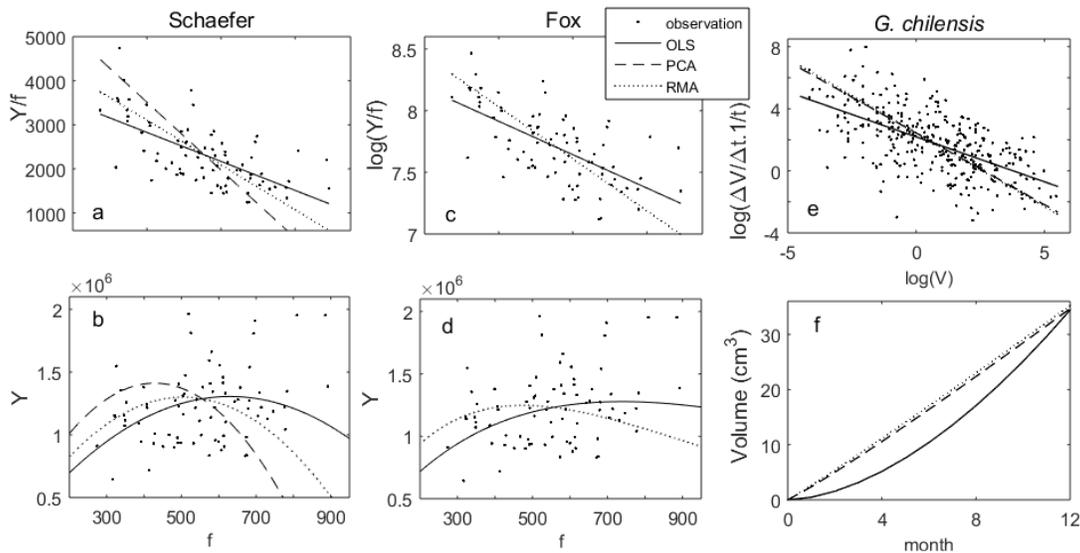


Fig. 3. PCA, RMA and OLS estimations of Surplus production models for Sardine fisheries along the Portuguese south-western coast and of frond growth of *Gracilaria chilensis*. Frond volume increases with time following the relation $V_t = -bct^{1/b}$

3.5 Lobster and Sparrow Morphometry Data

Males of the lobster *Nephrops norvegicus* Linnaeus were sampled at seven sites from the south coast of Portugal (Algarve) in the Atlantic and across the Mediterranean in the Alboran Sea (Malaga), the Catalan Sea (Barcelona), the Ligurian Sea, the Tyrrhenian Sea, the Adriatic Sea and the Gulf of Euboikos (Greece). Many morphometric variables were originally measured by Castro et al. [28]. Presently, the carapace standard length was compared to the carapace total length, carapace posterior length, carapace lateral left and right lengths, carapace width, carapace height, and antenna scaphocerite left and right lengths. These comparisons were performed for each population yielding 8 pairs of variables \times 7 populations, accounting for a total of 56 slopes.

Several sparrows (*Passer domesticus* Linnaeus) were collected after a severe storm on the 1st February 1898 and taken into the biological laboratory at Brown University, Rhode Island, where Hermon Bumpus took five morphometric measurements, namely the total length, alar extent, length of beak and head, humerus length and keel of sternum length. The data set is available in the book by Manly [12] for the 49 female sparrows only, of which 21 survived whereas the remaining 28 died. Regression lines were estimated for all possible pair-wised variables, separately for survivors and deceased, yielding 20 slopes.

Most of the morphometry correlations were situations for PCA, although often RMA estimated similar slopes (Fig. 2). The *N. norvegicus* morphometry showed that when RMA overestimated the slopes, it miss-estimated body shapes leading to a false identification of distinct morphotypes (Fig. 4). Comparison among the seven RMA slopes yielded eight significant differences (estimated from permutations tests following Vieira and Creed (2013a)) whereas based on PCA yielded none. The false results from RMA application were (i) ML significantly different from all the others except GR, (ii) AD significantly different from ML and GR, and (iii) LI significantly different from ML, GR and AD.

3.6 Social Science Happiness Data

Portugal and Germany were compared relatively to the cultural influence on happiness with 338 Portuguese and 302 Germans responding to the

same questionnaires. The culture questionnaire comprised seven cultural dimensions, namely the power distance, individualism, masculinity, uncertainty avoidance, long-term orientation, indulgence versus restraint, and monumentalism. The happiness was measured with a questionnaire for each of two dimensions, namely the life satisfaction scale and the general health questionnaire. We estimated the regression lines for all pair-wised culture-happiness combinations and separately for Portugal and Germany. The culture questionnaire consisted of the Values Surveys Module 2008 manual [29]. The Life Satisfaction Scale is one of the most used questionnaires developed by Diener et al. [30] and posteriorly subject to several revisions. The General Health Questionnaire is used by the British Household Panel Study. We used version GHQ-12 by Goldberg and Huxley [31]. The social-sciences correlations were very weak and thus, when x and y were scaled, their scatter plot was a circumference.

However, this situation did not occur when using their original units. The social-sciences data set undoubtedly required linefits estimated by PCA using the covariance matrix, and it was fundamental to perform this correct choice as RMA slopes were always widely different. This was a typical situation were, although correlations were weak, PCA performed much better than RMA because y was much less scattered than x .

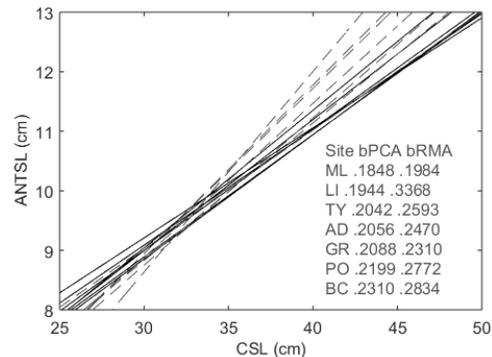


Fig. 4. Morphometry relations estimated by (solid black) PCA and (dashed grey) RMA

4. CONCLUSIONS

When both x and y are measured with error, model II regression is the adequate regression type. However, it does not have a globally better algorithm. RMA provides reasonable estimates

only when x and y are measured with similar error, which is a basic assumption of this method. Furthermore, our tests demonstrated that its inadequate use can easily lead to severely biased data analysis and conclusions about the biology and ecology of living beings. Although PCA often provides much better estimates than RMA, it still is very fallible when x and y have approximate variances and are weakly correlated. Based on our tests, we propose choosing PCA whenever $r^2 + s_x^2/s_y^2$ is higher than 1.5. In the remaining cases where it is difficult to decide which the best method is, a case-by-case analysis is required. We suggest for the researcher to generate artificial data with slope and error structure manipulated to match the original data, and test which method approximates better the true slope. We provide a Matlab script performing this task assuming a Normal distribution.

ACKNOWLEDGEMENTS

Work funded by ERDF Funds of the Competitiveness Factors Operational Programme - COMPETE and by national funds from the FCT - Foundation for Science and Technology project UID/EEA/50009/2013. Research on *Gracilaria chilensis* was supported by CONICYT (Fondo Nacional de Desarrollo Científico y Tecnológico FONDECYT) under grant number 1090360. The authors are grateful to Margarida Castro for providing the Norway lobster biometry data from the project NEMED (EU-DG XIV, MED/92/008, coordinated by ICM, CSIC, Barcelona), and to Carlos Moura (DGPA) for providing fisheries sardine data.

COMPETING INTERESTS

Authors have declared that no competing interests exist.

REFERENCES

1. Draper NR. Straight line regression when both variables are subject to error. In Proceedings of the 1991 Kansas State University Conference on Applied Statistics in Agriculture. 1992;1-18.
2. Smith RJ. Use and misuse of Reduced Major Axis for line-fitting. *Am J Phys Anthropol.* 2009;140:476-486.
3. Ricker WE. Linear regressions in fishery research. *J Fish Res Board Can.* 1973;30:409-434.
4. Cragg JG. Using higher moments to estimate the simple errors-in-variables model. *RAND J Econ.* 1997;28(0):71-91.
5. Dagenais MG, Dagenais DL. Higher moment estimators for linear regression models with errors in the variables. *J Econometrics.* 1997;76:193-221.
6. Gillard JW, Iles TC. Method of moments estimation in linear regression with errors in both variables. Cardiff University School of Mathematics Technical Paper, Cardiff, UK; 2015.
7. Gillard J. An overview of linear structural models in errors in variables regression. *REVSTAT Stat J.* 2010;8(1):57-80.
8. Sokal RR, Rohlf FJ. *Biometry: The principles and practice of statistics in biological research*, 2nd ed. W.H. Freeman and company. New York, USA; 1981.
9. Jackson JE. *A user's guide to principal components*. John Wiley & Sons, New York, USA; 1991.
10. Claude J. *Morphometrics* with R. Springer, New York, USA; 2008.
11. Pearson K. On lines and planes of closest fit to systems of points in space. *Philos Mag.* 1901;2:559-572.
12. Manly B.J.F. *Multivariate methods*. Chapman & Hall, London, United Kingdom; 1986.
13. Vieira VMNCS, Leitão Fand Mateus M. Biomass-density data analysis: A comment on Cabaço et al. (2013). *J Ecol.* 2015;103:537-540. DOI: 10.1111/1365-2745.12294
14. Vieira VMNCS, Creed J. Estimating significances of differences between slopes: A new methodology and software. *Computational Ecology and Software.* 2013a;3(3):44-52.
15. Vieira VMNCS, Creed J. Significances of differences between slopes: An upgrade for replicated time series. *Computational Ecology and Software.* 2013b;3(4):102-109.
16. Creasy MA. Confidence limits for the gradient in the linear functional relationship. *J. Roy. Statist. Soc. Ser. B.* 1956;18:65-69.
17. Leitão F, Alms V, Erzini K. A multi-model approach to evaluate the role of environmental variability and fishing pressure in sardine fisheries. *J Mar Sys.* 2014;139:128-138.
18. Schaefer MB. Some aspects of the dynamics of populations important to the

- management of commercial marine fisheries. IATTC Bull. 1954;1:25-56.
19. Fox WW Jr. An exponential surplus-yield model for optimizing exploited fish populations. Trans. Am. Fish. Soc. 1970; 99:80-88.
 20. Creed JC, Kain JM, Norton TA. An experimental evaluation of density and plant size in two large brown seaweeds. J Phycol. 1998;34:39-52.
 21. Scrosati R, DeWreede RE. Dynamics of the biomass-density relationship and frond biomass inequality for *Mazzaella cornucopiae* (Gigartinales, Rhodophyta): implications for the understanding of frond interactions. Phycologia. 1997;36:506-516.
 22. Cabaço S, Apostolaki ET, García-Maín P, Gruber R, Hernández I, Martínez-Crego B, Mascaró O, Pérez M, Prathep A, Robinson C, Romero J, Schmidt AL, Short FT, Tussenbroek BI, Santos R. Effects of nutrient enrichment on seagrass population dynamics: Evidence and synthesis from the biomass–density relationships. J Ecol. 2013;101:1552-1562.
 23. Engel C, Aberg P, Gaggiotti OE, Destombe C, Valero M. Population dynamics and stage structure in a haploid- diploid red seaweed, *Gracilaria gracilis*. J Ecol. 2001; 89:436- 450.
 24. Guillemin ML, Oscar RH, Martínez EA. Characterization of genetic markers linked to sex determination in the haploid-diploid red alga *Gracilaria chilensis*. J Phycol. 2012;48:365-372.
 25. Skrobachi Z.. Selected methods for the estimation of the logistic function parameters. Maintenance and reliability. 2007;3(35):52-56.
 26. Gillman M. An introduction to mathematical models in ecology and evolution: Time and space, 2nd edition. Willey-Blackwell, Oxford, UK; 2009.
 27. Evans P. How to estimate growth regressions consistently. Mimeo, Department of Economics, the Ohio State University; 1994.
 28. Castro M, Gancho P, Henriques P. Comparison of several populations of Norway lobster, *Nephrops norvegicus* (L.), from the Mediterranean and the adjacent Atlantic. A biometrics study. Sci Mar. 1998; 62(1):71-79.
 29. Hofstede G, Hofstede GJ, Monkov M, Vinken H. Announcing a new version of the values survey module: The VSM 08; 2008. (Accessed 5 April 2016)
Available:<http://www.geerthofstede.nl/vsm-08>
 30. Diener E. Subjective well-being. Psychol Bull. 1984;95:542-575.
 31. Goldberg D, Huxley P. Mental illness in the community: The pathway to psychiatric care. Tavistock Publications, London, UK; 1980.

© 2016 Vieira et al.; This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Peer-review history:

The peer review history for this paper can be accessed here:
<http://sciencedomain.org/review-history/14639>