
Information criteria on multimodel selection of parametric regression: Biological applications¹

Criterios de información en la selección multimodelo de regresiones paramétricas: aplicaciones biológicas

Daniela Moraga López^a
danielamoraga05@gmail.com

Cristian Román P.^b
cromanpa94@gmail.com

Abstract

When carrying out modelling analysis using experimental data is important to obtain a measure of the relative fit of the each model as a primary selection criterion. In this sense, there are few studies based on multi-model selection techniques for the theoretical representation of data sets, so it is common to incur in a misinterpretation of the existing patterns, or even more, the incorrect extrapolation and prediction based on the wrong model. This paper is intended to evaluate in 40 datasets from various ecological published studies the effectiveness of the linear regression model designated by the authors by contrasting with six regression models using the Akaike and Bayesian information criteria, and furthermore to discuss its implications on subsequent interpretations made. It was found that the linear regression model was successful only in 13.35% of the datasets (15% of datasets for AIC and 11.7% of datasets for BIC), but in the other hand, the logarithmic model was the most successful model in 38.5% of the cases (35% of datasets for AIC and 41.1% of the datasets for BIC), casting doubts on the efficiency of the linear regression over other types of regression under biological data.

Keywords: AIC, BIC, least squares, regression.

Resumen

Cuando se lleva a cabo análisis de modelamiento usando datos experimentales es importante obtener una medida de la confiabilidad del ajuste relativo de cada

¹Moraga, D., Román, C. (2015). Information criteria on multimodel selection of parametric regression: Biological applications. *Comunicaciones en Estadística*, **8**(1), 45-57.

^aEstudiante. Departamento de Biología, Facultad de Ciencias Naturales y Exactas. Universidad del Valle. Colombia.

^bEstudiante. Departamento de Biología, Facultad de Ciencias Naturales y Exactas. Universidad del Valle. Colombia.

modelo como un criterio principal de selección. En este sentido, existen pocos estudios basados en técnicas de selección multimodelo para realizar representaciones teóricas de conjuntos de datos, por lo que es común incurrir en una mala interpretación de los patrones existentes, o más aún, extrapolar incorrectamente y basar predicciones en modelos equívocos. Este documento está enfocado en evaluar en 40 conjuntos de datos provenientes de varios estudios ecológicos publicados, la efectividad de la regresión lineal designada por los autores al contrastarla con seis modelos de regresión usando los criterios de información de Akaike y Bayesiano, y además discutir las implicaciones de las interpretaciones subsecuentes de acuerdo al modelo. Se encontró que el modelo de regresión lineal fue exitoso en solo el 13.35% de los conjuntos de datos (15% de los conjuntos de datos para AIC y 11.7% de los datos para BIC), pero por otro lado, el modelo logarítmico fue más exitoso en el 38.5% de los casos (35% de los conjuntos de datos para AIC y 41.1% de los datos para BIC), generando dudas sobre la eficiencia del modelo de regresión lineal sobre los otros tipos de regresión en datos biológicos.

Palabras clave: AIC, BIC, mínimos cuadrados, regresión..

1 Introduction

The studies focused on understanding our world through modeling analysis play a major role in providing new knowledge about patterns of the real (Burnham & Anderson 2004). Models are important by considering in addition to the parameters, the relationships expressed between and within variables (Lahiri 2001, Burnham et al. 2011). In the scientific scope is crucial to execute valid inferences from evidence and therefore it is important to consider alternative hypotheses as models representing each paradigm conceived during the design stage of the investigation or even during the analysis process (Hulme 1992). It is common to find that despite a careful reflection in the selection of an *a priori* model, many mistakes are made by excluding the reflective context of the general aspects on research, or more frequently the philosophical background in which analysis lays (see Milliken (1990) and Spiegelhalter et al. (2002)).

The selection of an appropriate model tends to avoid underutilization of information from the data and consequently prevent the “discovery” of false positives (Wasserman 2000). Among the methods developed recently as criteria for the selection of a “best model” approached from *a priori* models are the Akaike (AIC) and Bayesian (BIC) information criterion (Sugiura 1978, Hurvich 1991). Both offer a new proposal in the biological sciences with significant computational and philosophical advantages, but they are somewhat different from the usual methods based on p values derived from the theory of least squares (LS) (Forster 2000).

Meanwhile, the minimal implementation of methods based on maximum likelihood or Bayesian approaches in statistical packages often used for analysis of biological data (eg. PAST, STATISTICA and PRIMER, among others), generates a bias towards supports for p values to relatively simple hypothesis, which prevents a

critical and holistic approach on more paradigms in modeling and interpretation of the data (Breiman 1992, Wells 1992). In order to highlight the importance of integrating a multimodel selection approach in the methodological research phase, we reviewed 40 scientific articles to verify the advantage of the author's selected model of regression by comparing it with five other regression models according to the AIC and BIC criteria. The main objective was to emphasize the consequences of the erroneous selection model by confronting deviations and probable errors during the data interpretation (Akaike 1998).

2 Materials and methods

Information criteria allow the identification of models that involves less loss of information based on a balance with respect to the number of parameters to which it is incurred. The analysis where developed in accordance with the Akaike information criterion (AIC) which is defined as,

$$AIC = 2K - 2 \ln L$$

where K is the number of parameters and L is the maximized value of the likelihood of the model. On the other hand, the Bayesian Information Criterion (BIC), which penalizes more rigorously the number of parameters, is defined as,

$$BIC = -2 \ln L + K \ln n$$

where, L denotes the maximum likelihood model, K the number of parameters and n the number of observations. Under these two information criteria we evaluated the relative fit of the linear model selected by the authors of all 40 published papers consulted in the Ecological data wiki (Jones et al. 2009, Table 1) with the quadratic, cubic, polynomial, logarithmic and inverse logarithmic regression models we implemented for comparison purpose. All analysed datasets contained continuous variables, which allowed us to compare the performance of each general regression model in the datasets.

The simple linear regression model is the simpler regression type in which data is modelled using a linear predictor function and the parameters are estimated from the data. It involves an independent or regressor variable (x) and the corresponding response variable (y). The equation of the straight line which relates the two variables is:

$$y = \beta_0 + \beta_1 x + \varepsilon$$

where β_0 is the intercept (if 0 is included in the range), β_1 the slope and ε the associated error. A second-order model regression was also evaluated. The quadratic model involves a parabolic adjustment of the two variables and is described by the equation:

$$y = \beta_0 + \beta_1 x + \beta_2 x^2 + \varepsilon$$

Table 1: *References considered in the analysis of the adjustment for the linear regression models in 40 biological datasets. Source: own elaboration.*

No.	Reference	No.	Reference
1	McArthur et al. (2008)	21	Marshall et al. (2013)
2	Bode et al. (444)	22	Olsen et al. (2008)
3	Eagle et al. (2013)	23	Riget et al. (2007)
4	Kienast et al. (2012)	24	Suzuki et al. (2013)
5	Krebs et al. (2012)	25	Trotter et al. (2011)
6	Lamberson et al. (2011)	26	Zimmermann et al. (2015)
7	Madin et al. (2001)	27	Wheater et al. (2011)
8	Wagner (2013)	28	Miller et al. (2014)
9	Hashizume et al. (2014)	29	Fu et al. (2014)
10	Boyer et al. (2014)	30	Bhattacharyya et al. (2013)
11	Asp et al. (2012)	31	Boisson et al. (2013)
12	Calkins et al. (2013)	32	Hilbe (2011)
13	Hone (2013)	33	Jones (2012)
14	Vitt & Pianka (2014)	34	Quinn et al. (2013)
15	Shero et al. (2014)	35	Tang et al. (2014)
16	Wen et al. (2012)	36	Werner & Griebeler (2014)
17	Zhang et al. (2013)	37	Desapriya et al. (2014)
18	Hiregoudar et al. (2014)	38	Hossieny (2014)
19	Murugkar (2014)	39	Jitnarin et al. (2014)
20	Ogawa et al. (2014)	40	Koike et al. (2014)

where β_0 is the mean of y when $x = 0$, β_1 and β_2 are the linear and quadratic effect parameters respectively, and ε the error. The quadratic regression model can be generalized in a K -th order polynomial model in one variable,

$$y = \beta_0 + \beta_1 x + \beta_2 x^2 + \dots + \beta_k x^k + \varepsilon$$

where the parameters are essentially the same as the quadratic regression, and $\beta_k x^k$ represents the k -th term of the k th order polynomial regression. The logarithmic regression involves

$$y = \log_b x$$

where b is the base of the logarithm. Its inverse function is closely related and is expressed as,

$$y = b^x$$

requiring the same parameters that logarithmic model.

Analyses were performed using the package **Stats** in R software based on the original matrices and using the independent and dependent variables in accordance with that defined by the authors. Because all variables analyzed were continuous, we verified the assumptions for a general linear regression analysis in each data set, and also interpreted the relevance and implications of the model selection in analyses performed subsequently.

Table 2: *Best model fits in the different sets of data according to the AIC and BIC criteria. Source: own elaboration.*

Modelo	AIC(%)	BIC(%)
Logarithm	35,0	41,1
Lineal	15,0	11,7
Cuadratic	15,0	11,7
Cubid	15,0	11,7
Polynomial	15,0	11,7
Inverse logarithm	5,0	11,7

3 Results and discussion

AIC values for the data sets examined indicated an incorrect selection of the regression model in 75% of cases and 88.3% according to BIC. The papers consulted considered only linear regressions and therefore directly assume on the data set characteristics of normality and homogeneity of variances that are not corroborated. The Akaike information criteria and Bayesian had a match on the model that best explains the data in 90.9% of the events of selection, therefore, the models specified by each criterion are generally not so dissimilar. It was found that the logarithmic model presented the highest frequency in the fitting for all relationships between variables considered (Table 2), being selected by AIC as correct in 35% of cases and 41.1% for BIC. According to AIC, the linear, quadratic, cubic and polynomial models are also explanatory in 72.7% of the data records and therefore are less explicative than the logarithmic model, which collects the widest part of the information contained in the data.

According to AIC, the inverse logarithmic model was the least favored and therefore presented the smallest explanatory power relative to other models in the same set of data. The BIC criterion indicated only the logarithmic model as superior in explanatory power, but not discerned in capacity among the other models with less explanation. In the dataset and according to the Bayesian approach, the linear, quadratic, cubic, polynomial inverse logarithm models where correct the same number of times (11.7%) and therefore, in some datasets are the most efficient in modeling the dispersion characteristics than in others.

The selection of one of the papers considered for graphically analysis indicated that the scatter of points around the linear regression is much greater than when comparing a logarithmic regression (Figure 1). This is corroborated by each R^2 values, being lower in linear regression with respect to the logarithmic (0.4333 and 0.4352 respectively).

The data analysis methods do not allow the unambiguous description of a system, but the parameters used may be useful towards a better approximation of the patterns. Is considered to be important to find the simplest model that defines best in its equation the characteristics present in the data (Hilborn & Mangel 1997,

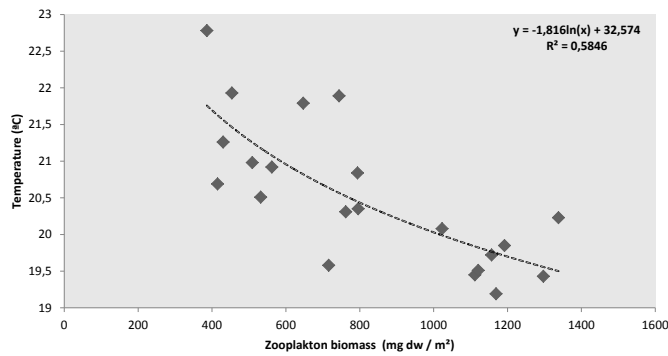
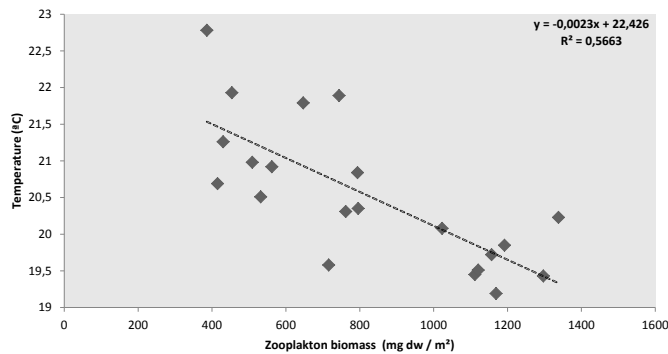


Figure 1: *Differences between two regression models applied to the same set of data. The regression equation and the value of R^2 is indicated for each relation. A. linear regression, B. logarithmic regression. Source: own elaboration.*

Stephens et al. 2007). In this sense, the logarithmic model presents, along with the inverse logarithmic, fewer parameters among the tested models, implying minor penalty by the Akaike and Bayesian criteria. The distinction and decision rule between these two models, varies therefore by the value of maximum likelihood in each dataset. This implies therefore that the logarithmic model represents better

the characteristics of the data than the inverse logarithm, since it presents the optimum combination of the number of parameters and maximum likelihood in the different datasets (Motulsky & Christopoulos 2004).

The logarithmic model is usually used as an efficient solution to various common aspects in biological systems: (i) where the scale of the data is sufficiently large, the logarithmic standardization reduces the spectrum, (ii) for data where the residuals increase with higher values in the dependent variable or (iii) if there are multiplicative effects (interaction) between the variables. On the other hand, the inverse logarithmic model which also has the lowest number of parameters, was expected to be one of which presents a low value for AIC and BIC, but despite this, it was the least representative among the models tested (5% and 11.7% respectively). In this case, the likelihood discerns between this and the logarithmic model that fits better to the data set (Draper & Smith 1998). On the other hand, the polynomial model does not balance model complexity with respect to their explanatory power and is only recommended when a good curve is obtained for graphing or interpolate unknown values. This is one of the reasons why the polynomial model does not show the best fits (Brand & Johnson 2004).

It is clear the high concordance between the decisions of fit between the two criteria (90.1%), so that the differences are due to the strict penalties made by BIC (Symonds & Moussalli 2011), but both criteria reflects fidelity to the selection of the model with the best balance in the adjustment. Analyses indicated that the linear model was not always the best descriptor of the datasets, because between 75% and 83.3% of the times the model that best fit the data was a linealizable model. This is a consequence of subjective selection of a linear regression model for being the simpler and easy to perform than other methods, without regard the assumptions of the function (Gaussian data distribution and homogeneity of variances (Draper & Smith 1998)).

Some clear grounds for not selecting linear regressions as descriptors of complexity in datasets is defined within the same analysis (R^2 value), which at low values implies a low explanatory power of the observed variance in the dependent variable by the dependent (Bevington et al. 1993, Varin & Vidoni 2005). Another reason why linear regression cannot fits the data significantly corresponds to the assumption that the independent variable has been properly defined and that the experimental error and biological variability affects only the values of the dependent variable, which in a pragmatic sense never occurs (Bates & Watts 1988). In biological systems the variance explained by a variable in relation to another is usually results of the synergistic effects (Gabrielsson & Weiner 2001).

4 Conclusion

The incorrect selection of a predictive model for a biological system generates definitively the establishment of erroneous conclusions, leading to mistaken explanation of important events and sacrificing the pragmatic applications of the new

paradigms. It is crucial then, to carry out an objective and clear selection of the models, allowing the matching between the correct predictions with the observed data.

Aknowledgements

We wish tank to Ranulfo Gonzalez-Obando and GIE for the continuous support. CRP thanks YPC, CRV and CaRP.

Recibido: 02 de febrero de 2015

Aceptado: 04 de mayo de 2015

References

- Akaike, H. (1998), *Information theory and an extension of the maximum likelihood principle. Selected Papers of Hirotugu Akaike*, Springer.
- Asp, M., Richardson, J., Collene, A., Droll, K. & Belury, M. (2012), ‘Dietary protein and beef consumption predict for markers of muscle mass and nutrition status in older adults’, *The journal of nutrition, health & aging* **16**(9), 784–790.
- Bates, D. M. & Watts, D. G. (1988), *Nonlinear regression: iterative estimation and linear approximations*, Wiley Online Library.
- Bevington, P. R., Robinson, D. K., Blair, J. M., Mallinckrodt, A. J. & McKay, S. (1993), ‘Data reduction and error analysis for the physical sciences’, *Computers in Physics* **7**(4), 415–416.
- Bhattacharyya, S., Adhikari, B. & Rawat, G. (2013), ‘Forage selection by Royle’s pika (*Ochotona roylei*) in the western Himalaya, India’, *Zoology (Jena)* **116**(5), 300–306.
- Bode, M., Schukat, A., Hagen, W. & Auel, H. (444), ‘Predicting metabolic rates of calanoid copepods’, *Journal of Experimental Marine Biology and Ecology* (1-7).
- Boisson, S., Stevenson, M., Shapiro, L., Kumar, V., Singh, L. P., Ward, D. & Clasen, T. (2013), ‘Effect of household-based drinking water chlorination on diarrhoea among children under five in Orissa, India: a double-blind randomised placebo-controlled trial’, *PLoS medicine* **10**(8), e1001497.
- Boyer, D., Winchester, J. & Kay, R. (2014), ‘The effect of differences in methodology among some recent applications of shearing quotients’, *American journal of physical anthropology* **156**(1), 166–178.

- Brand, L. & Johnson, M. L. (2004), *Numerical computer methods*, Academic Press.
- Breiman, L. (1992), 'The little bootstrap and other methods for dimensionality selection in regression: X-fixed prediction error', *Journal of the American Statistical Association* **87**(419), 738–754.
- Burnham, K. & Anderson, D. (2004), 'Multimodel inference understanding AIC and BIC in model selection', *Sociological methods & research* **33**(2), 261–304.
- Burnham, K., Anderson, D. & Huyvaert, K. (2011), 'AIC model selection and multimodel inference in behavioral ecology: some background, observations, and comparisons', *Behavioral Ecology and Sociobiology* **65**(1), 23–35.
- Calkins, E. S., Fuller, T. K., Asa, C. S., Sievert, P. R. & Coonan, T. J. (2013), 'Factors influencing reproductive success and litter size in captive island foxes', *The Journal of Wildlife Management* **77**(2), 346–351.
- Desapriya, E., Giulia, S., Subzwari, S., Peiris, D. C., Turcotte, K., Pike, I., Sages, D. & Hewapathirane, D. S. (2014), 'Does Obesity Increase the Risk of Injury or Mortality in Motor Vehicle Crashes? A systematic review and meta-analysis', *Asia-Pacific journal of public health* **26**(5), 447–460.
- Draper, N. R. & Smith, H. (1998), *Applied regression analysis*, John Wiley & Sons.
- Eagle, R., Eiler, J., Tripathi, A., Ries, J., Freitas, P., Hiebenthal, C., Wanamaker, A., Taviani, M., Elliot, M., Marensi, S., Nakamura, K., Ramírez, P. & Roy, K. (2013), 'The influence of temperature and seawater carbonate saturation state on 13C–18O bond ordering in bivalve mollusks', *Biogeosciences* **10**(7), 4591–4606.
- Forster, M. R. (2000), 'Key concepts in model selection: Performance and generalizability', *Journal of mathematical psychology* **44**(1), 205–231.
- Fu, Y., Wang, T., Fu, Q., Wang, P. & Lu, Y. (2014), 'Associations between serum concentrations of perfluoroalkyl acids and serum lipid levels in a Chinese population', *Ecotoxicology and environmental safety* **106**, 246–252.
- Gabrielsson, J. & Weiner, D. (2001), *Pharmacokinetic and pharmacodynamic data analysis: concepts and applications*, Vol. 2, CRC Press.
- Hashizume, N., Tanabe, A., Inoue, Y., Sawada, T., Murakami, H., Suzuki, Y., Sumi, S., Tsubokura, Y., Yoshida, T., Ajimi, S., Tsuji, T. & Furukawa, K. (2014), 'Prediction of the bioconcentration factor in common carp (*Cyprinus carpio* L.) using data from the dietary exposure bioaccumulation fish test', *Environmental toxicology and chemistry / SETAC* **33**(6), 1406–1414.
- Hilbe, J. M. (2011), Modeling Count Data, in 'International Encyclopedia of Statistical Science', Springer, pp. 836–839.
- Hilborn, R. & Mangel, M. (1997), 'The ecological detective: Confronting models with data Princeton Univ'.

- Hiregoudar, S., Sandeep, T., Nidoni, U., Shrestha, B. & Meda, V. (2014), 'Studies on dhal recovery from pre-treated pigeon pea (*Cajanus cajan* L.) cultivars', *Journal of food science and technology* **51**(5), 922–928.
- Hone, J. (2013), 'Diminishing returns in bovine tuberculosis control', *Epidemiology and infection* **141**(07), 1382–1389.
- Hossieny, P. (2014), 'Ain't no mountain high enough', *Journal of paediatrics and child health* **50**(10), E1–E2.
- Hulme, E. C. (1992), *Receptor-ligand interactions: A practical approach, Practical approach series.*, IRL Press at Oxford University Press, Oxford, New York.
- Hurvich, C. M. a. C. L. (1991), 'Bias of the corrected AIC criterion for underfitted regression and time series models', *Biometrika* **78**(3), 499–509.
- Jitnarin, N., Kosulwat, V., Rojroongwasinkul, N., Boonpraderm, A., Haddock, C. K. & Poston, W. S. (2014), 'The Relationship Between Smoking, Body Weight, Body Mass Index, and Dietary Intake Among Thai Adults Results of the National Thai Food Consumption Survey', *Asia-Pacific Journal of Public Health* **26**(5), 481–493.
- Jones, K., Bielby, J., Cardillo, M., Fritz, S., O'Dell, J., Orme, C., L., D., Safi, K., Sechrest, W., Boakes, E., Carbone, C., Connolly, C., Cutts, M., Foster, J., Grenyer, R., Habib, M., Plaster, C., Price, S., Rigby, E., Rist, J., Teacher, A., Bininda-Emonds, P., O. R., Gittleman, J., Mace, G. & Purvis, A. (2009), 'PanTHERIA: A Species-Level Database of Life History, Ecology, and Geography of Extant and Recently Extinct Mammals', *Ecology* **90**(9), 2648–2648.
- Jones, K. D. (2012), Factors affecting establishment and germination of upland prairie species of conservation concern in the Willamette Valley, Oregon, Master's thesis, Oregon State University, Corvallis.
- Kienast, M., MacIntyre, G., Dubois, N., Higginson, S., Normandeau, C., Chazen, C. & Herbert, T. D. (2012), 'Alkenone unsaturation in surface sediments from the eastern equatorial Pacific: Implications for SST reconstructions', *Paleoceanography* **27**(1), PA1210.
- Koike, K., Kitahara, K., Higaki, M., Urata, M., Yamazaki, F. & Noshiro, H. (2014), 'Clinicopathological features of gastric metastasis from breast cancer in three cases', *Breast Cancer* **21**(5), 629–634.
- Krebs, C., Bilodeau, F., Reid, D., Gauthier, G., Kenney, A., Gilbert, S., Duchesne, D. & Wilson, D. (2012), 'Are lemming winter nest counts a good index of population density', *Journal of Mammalogy* **93**(1), 87–92.
- Lahiri, P. (2001), *Model selection*, Institute of Mathematical Statistics Lecture Notes - Monograph Series.

- Lamberson, J., Frazier, M., Nelson, W. & Clinton, P. (2011), *Utilization patterns of intertidal habitats by birds in Yaquina Estuary, Oregon*, Technical report, U.S. Environmental Protection Agency.
- Madin, L., Horgan, E. & Steinberg, D. (2001), 'Zooplankton at the Bermuda Atlantic Time-series Study (BATS) station: diel, seasonal and interannual variation in biomass, 1994–1998', *Deep Sea Research Part II: Topical Studies in Oceanography* **48**(8-9), 2063–2082.
- Marshall, B., Thunell, R., Henehan, M., Astor, Y. & Wejnert, K. (2013), 'Planktonic foraminiferal area density as a proxy for carbonate ion concentration: A calibration study using the Cariaco Basin ocean time series', *Paleoceanography* **28**(2), 363–376.
- McArthur, J. M., Algeo, T. J., van de Schootbrugge, B., Li, Q. & Howarth, R. J. (2008), 'Basinal restriction, black shales, Re-Os dating, and the Early Toarcian (Jurassic) oceanic anoxic event', *Paleoceanography* **23**(4), PA4217.
- Miller, S., Richardson, K. & Edwards, P. (2014), 'The effect of suspended sediment on fertilization success in the urchin *Evechinus chloroticus*: Analysis of experimental data using hierarchical Bayesian methods', *Marine pollution bulletin* **88**(1-2), 28–33.
- Milliken, G. A. (1990), 'Nonlinear Regression Analysis and Its Applications', *Technometrics* **32**(2), 219–220.
- Motulsky, H. & Christopoulos, A. (2004), *Fitting models to biological data using linear and nonlinear regression: a practical guide to curve fitting*, Oxford University Press.
- Murugkar, D. A. (2014), 'Effect of sprouting of soybean on the chemical composition and quality of soymilk and tofu', *Journal of food science and technology* **51**(5), 915–921.
- Ogawa, T., Hanamura, N., Yamashita, M., Kimura, H. & Kashikura, Y. (2014), 'Long-term results of breast volume replacement using an inframammary adipofascial flap after breast-conserving surgery', *Breast Cancer* **21**(5), 635–640.
- Olsen, A., Brown, K., Chierici, M., Johannessen, T. & Neill, C. (2008), 'Sea-surface CO₂ fugacity in the subpolar North Atlantic', *Biogeosciences* **5**(2), 535–547.
- Quinn, A., Williams, D. M. & Porter, W. F. (2013), 'Landscape structure influences space use by white-tailed deer', *Journal of Mammalogy* **94**(2), 398–407.
- Riget, F., Moller, P., Dietz, R., Nielsen, T., Asmund, G., Strand, J., Larsen, M. & Hobson, K. (2007), 'Transfer of mercury in the marine food web of West Greenland', *Journal of environmental monitoring* **9**(8), 877–883.

- Shero, M. R., Pearson, L. E., Costa, D. P. & Burns, J. M. (2014), 'Improving the Precision of Our Ecosystem Calipers: A Modified Morphometric Technique for Estimating Marine Mammal Mass and Body Composition', *PloS one* **9**(3), e91233.
- Spiegelhalter, D., Best, N., Carlin, B. & Van Der Linde, A. (2002), 'Bayesian measures of model complexity and fit', *Journal of the Royal Statistical Society: Series B. Statistical Methodology* **64**(4), 583–639.
- Stephens, P. A., Buskirk, S. W., Hayward, G. D. & Martínez, C. (2007), 'A call for statistical pluralism answered', *Journal of Applied Ecology* **44**(2), 461–463.
- Sugiura, N. (1978), 'Further analysts of the data by akaike's information criterion and the finite corrections: Further analysts of the data by akaike's', *Communications in Statistics: Theory and Methods* **7**(1), 13–26.
- Suzuki, F., Garcia, D. A. Z. & Orsi, M. (2013), 'Reproductive Dynamics of *Iheringichthys labrosus* (Lütken, 1874) (Teleostei, Pimelodidae), in the Capivara Reservoir, Paranapanema River, Parana, Brazil', *Dataset Papers in Biology* **1**, 1–6.
- Symonds, M. R. & Moussalli, A. (2011), 'A brief guide to model selection, multimodel inference and model averaging in behavioural ecology using Akaike's information criterion', *Behavioral Ecology and Sociobiology* **65**(1), 13–21.
- Tang, L., Li, T., Li, D. & Meng, X. (2014), 'Elevational patterns of plant richness in the Taibai Mountain, China', *The Scientific World Journal* **2014**.
- Trotter, J., Montagna, P., McCulloch, M., Silenzi, S., Reynaud, S., Mortimer, G., Martin, S., Ferrier-Pagés, C., Gattuso, J. & Rodolfo-Metalpa, R. (2011), 'Quantifying the pH 'vital effect' in the temperate zooxanthellate coral *Cladocora caespitosa*: Validation of the boron seawater pH proxy', *Earth and Planetary Science Letters* **303**(3–4), 163–173.
- Varin, C. & Vidoni, P. (2005), 'A note on composite likelihood inference and model selection', *Biometrika* **92**(3), 519–528.
- Vitt, L. J. & Pianka, E. R. (2014), *Lizard ecology: historical and experimental perspectives*, Princeton University Press.
- Wagner, H. H. (2013), 'Rethinking the linear regression model for spatial ecological data', *Ecology* **11**, 2381–2391.
- Wasserman, L. (2000), 'Bayesian model selection and model averaging', *Journal of mathematical psychology* **44**(1), 92–107.
- Wells, J. W. (1992), *Analysis and interpretation of binding at equilibrium. Receptor-ligand interactions*, Hulme EC, editor. A practical approach. Oxford: Oxford University Press.

- Wen, Y., Su, L., Qin, W., He, J., Fu, L., Zhang, X. & Zhao, Y. (2012), 'Linear and non-linear relationships between soil sorption and hydrophobicity', *SAR and QSAR in Environmental Research* **23**(1-2), 111–123.
- Werner, J. & Griebeler, E. M. (2014), 'Allometries of Maximum Growth Rate versus Body Mass at Maximum Growth Indicate That Non-Avian Dinosaurs Had Growth Rates Typical of Fast Growing Ectothermic Sauropsids', *PLoS one* **9**(2), e88834.
- Wheater, C., Bell, J. & Cook, P. (2011), *Practical Field Ecology: A Project Guide*, John Wiley & Sons.
- Zhang, H., Liu, H., Li, Y., An, J., Xue, X. & Hou, M. (2013), 'Spatial variation of soil moisture/salinity and the relationship with vegetation under natural conditions in Yancheng coastal wetland', *Huan jing ke xue* **34**(2), 540–546.
- Zimmermann, B., Sand, H., Wabakken, P., Liberg, O. & Andreassen, H. (2015), 'Predator-dependent functional response in wolves: from food limitation to surplus killing', *The Journal of animal ecology* **84**, 102–112.