

Multidimensional scaling for animal traits in the context of dynamic energy budget theory

Sebastian A. L. M. Kooijman^{*,1}, Konstadia Lika², Starrlight Augustine³ and Nina Marn⁴

¹Department of Theoretical Biology, VU University Amsterdam, de Boelelaan 1087, 1081 HV Amsterdam, The Netherlands

²Department of Biology, University of Crete, 70013, Heraklion, Greece

³Akvaplan-niva AS, Fram High North Research Centre for Climate and the Environment, 9296 Tromsø, Norway

⁴Division for Environmental and Marine Research, Rudjer Bošković Institute, 10000 Zagreb, Croatia

*Corresponding author: Department of Theoretical Biology, VU University Amsterdam, The Netherlands. Email: bas.kooijman@vu.nl

The method of multidimensional scaling (MDS) has long existed, but could only recently be applied to animal traits in the context of dynamic energy budget (DEB) theory. The application became possible because of the following: (i) the Add-my-Pet (AmP) collection of DEB parameters and traits (approximately 280) recently reached 3000 animal species with 45000 data sets of measurements; (ii) we found a natural distance measure for species based on their traits as a side result of our research on parameter estimation in DEB context; and (iii) we developed plotting code for visualization that allows labelling of taxonomic relationships. Traits, here defined as DEB parameters or any function of these parameters, have different dimensions, which hamper application of many popular distance measures since they (implicitly) assume that all traits have the same dimensions. The AmP collection follows the workflow that measured data determine parameters and parameters determine trait values. In this way we could fill up the species traits table completely, which we could not do by using measured values only, as data availability varies considerably between species and is typically poor. The goodness of fit of predictions for all data sets is generally excellent. This paper discusses links between the MDS method and parameter estimation and illustrates the application of MDS for the AmP collection to five taxa, three ectothermic and two endothermic, which we consider to be 'complete', in the sense that we expect that it will be difficult to find more species with data in the open literature. This application of MDS shows links between traits and taxonomy that supplements our efforts to find patterns in the co-variation of parameter values. Knowledge about metabolic performance is key to conservation biology, sustainable management and environmental risk assessment, which are seen as interlinked fields.

Key words: Add-my-Pet collection, animals, classical multidimensional scaling, conservation biology, Dynamic Energy Budgets, environmental risk assessment, life history, loss function, metabolism, parameter estimation, principal coordinate analysis, sustainable management, traits

Editor: Steven Cooke

Received 15 June 2021; Revised 14 October 2021; Editorial Decision 18 October 2021; Accepted 9 December 2021

Cite as: Kooijman SALM, Lika K, Augustine S, Marn N (2021) Multidimensional scaling for animal traits in the context of dynamic energy budget theory. *Conserv Physiol* 9(1): coab086; doi:10.1093/conphys/coab086.

Introduction

Trait databases are proliferating (Kearney *et al.*, 2021, Parra *et al.*, 2015), which calls for methods to analyse them in order to extract knowledge about the underlying metabolism and life history in a form that can readily be applied in practice. This paper discusses the application of classical multidimensional scaling (CMDs) to traits of animals species—as made available in the Add-my-Pet (AmP) collection—to supplement other methods (Augustine *et al.*, 2019a, b, Kooijman, 1986, 2009, 2013, 2014, Lika *et al.*, 2014) for the analysis of patterns in parameter values of dynamic energy budget (DEB) models.

This paper first introduces the AmP collection, then discusses distances between species on the bases of the set of traits and the application of these distances in MDS. The last sections illustrate applications of MDS with numerical examples, illustrate the relationship between MDS and cluster analysis and give a brief discussion.

The AmP database and parameter patterns

The AmP collection is a free online database of referenced data on animal energetics, including life history and aging, and DEB model parameters (AmP, 2021, Marques *et al.*, 2018). It contains eco-physiological data on 3000 animal species (as of 1 June 2021) from all major phyla from start of development to death by ageing. The collection has been set up as a journal, so everybody can contribute, and has a careful curation procedure and version tracking. The use of the collection is supported by free and frequently updated Matlab packages that allow access to the DEB parameters and implied traits (both at individual and population levels) offline in order to analyse eco-evolutionary patterns in parameter values. The ability to compare species in a common framework (i.e. DEBs) supports parameter estimation-in-context and improves insight into eco-evolutionary development. The unique property that the species-trait table is fully filled makes it very suitable for multivariate analysis.

Each species represents an entry in the AmP database; both the measured data and the code used to estimate parameters from data are available. The estimation method is described in (Marques *et al.* 2018, Marques *et al.* 2019, Augustine *et al.* 2019a, Augustine *et al.* 2020 and Lika *et al.* 2020). The database is continuously growing both in species number and in the quality of each species entry because entries are revised and improved with new data. The various versions of each entry are archived and associated with a unique identifier. Over the years, the AmP database became a means to reference parameter values used for analysing and extracting parameter values from new and complex eco-physiological experiments, hindcasting and forecasting simulation studies. AmP grew into a unique platform whereby new insights on

animal metabolism and physiology could be included into various modelling tools used to support decision-making.

We initiated the online database in 2009, and from there felt an urgency to expand the collection rapidly, given the following: (i) the world-wide collapse of species diversity, stressing the need for effective conservation programs; (ii) the need for sustainable management, including optimization of culture and harvesting programs; and (iii) the need for effective environmental risk assessment, e.g. by supporting the analysis of ecotoxicity data to maximize efficiency (and minimize the use of test animals) and the translation of toxic effects on individuals in the laboratory to population and ecosystem levels under field conditions.

All these interlinked activities require detailed knowledge of energetics and the life history of species. This knowledge is rarely sufficient, however, and must be supplemented with knowledge from other fields. DEB theory found already quite a few applications in conservation biology (Arnall *et al.*, 2019, Goedegebuure, 2018, Haberle *et al.*, 2020, Marn *et al.*, 2020, 2017, McKenzie *et al.*, 2016, Molnár *et al.*, 2010, Sarà *et al.*, 2014, Taylor *et al.*, 2019, Teixeira, 2016) and sustainable management/aquaculture (Bertolini *et al.*, 2021, Chary *et al.*, 2020, Jackson, 2018, Jin *et al.*, 2020, Pete *et al.*, 2020, Sangare *et al.*, 2020, Sarà *et al.*, 2018, Stavrakidis-Zachou *et al.*, 2021, Yang *et al.*, 2020), while it was developed since 1979 in the context of ecotoxicity research (Kooijman & Bedaux, 1996, Kooijman & Metz, 1984, and many other papers). Direct links between sensitivity for chemicals and metabolic parameters (Baas & Kooijman, 2015) demonstrate the relevance of knowledge of metabolism for this important set of applications.

The AmP collection is unique in that all DEB parameters have been estimated for all species in the collection, which enables computation of any physiological property that can be written as a function of parameters, for any species.

Five different types of patterns of co-variation of DEB parameter values have been identified so far: the physical co-variation rules (Kooijman, 1986, 2009), metabolic acceleration (Kooijman, 2014), waste-to-hurry (Augustine *et al.*, 2019b, Kooijman, 2013), supply/demand spectra (Lika *et al.*, 2014) and altricial/precocial spectra (Augustine *et al.*, 2019a).

While the physical co-variation rules explain most of what is known as body size scaling relationships (Calder, 1984, Peters, 1983), such as the famous Kleiber's law (Kleiber, 1932), DEB theory shows that body size itself is an emergent property of metabolism (Lika *et al.*, 2019), and suggests that traits can better be linked to processes from which they result, rather than to each other. This enables the understanding of traits, as opposed to merely describing the relationships.

We see MDS, which takes both parameters and functions of parameters as input, as complementary to the identification of patterns in parameter values. A disadvantage of MDS is that the eigenvectors do not have a direct biological interpretation,

but an advantage is that they present a more holistic view on the difference between species by including a set of several traits, rather than a single one.

Traits and distances between species

A trait is here defined, in the context of DEB theory, as a parameter or a function of parameters, which quantifies some eco-physiological properties of a species. In principle, all parameters and any imaginable function of parameters can be used, but the present practical restriction is that they have to occur in the (Matlab) structure allStat, which is used to read trait values for the various species in the AmP collection. As is clear from the definition of traits that we use, trait values are interlinked, and sometimes in complex ways. The strict application of mass and energy conservation rules in DEB theory contributes to this interlinking.

Since some traits are specific for particular DEB models (Marques *et al.*, 2018), which are simple extensions of a common standard model, a second restriction is that these traits must be available for all species in the selected set of species, so the species traits table must be completely filled. This is still some 250 of the 280 traits per species; the number of traits can readily be extended. Notice that measured data has been used to estimate parameters, and parameters are used to compute functions of parameters. So, here, traits are not measured data; it would otherwise be impossible to fill the species traits table completely for a larger set of species.

Even though some traits are available as measured data for some species, the availability of data varies considerably among species and is typically very much restricted. The mean relative error of predictions for all 45000 data sets for all species in AmP collection is 0.05, meaning that the predictions are, on average, very accurate and reliable enough to be used instead of measured data. In other words, we do not see it as a disadvantage that model-derived values for traits are used, rather than measured values. If data come from various sources, which they typically do, the various data sets can easily be inconsistent for various reasons, while this does not apply to the model-derived traits, since the model provides the consistency. This has the drawback that the results are model dependent, but we think that no plausible alternative models will exist with a comparable simplicity and generality that fully specify individuals thermodynamically from start of development to end of life (Kooijman, 2020b).

For the present illustrative purpose, we used the selection of traits as given in Table 1. Most traits speak for themselves and only the specific somatic maintenance counts as parameter in this set, the others are functions of parameters, all at the reference temperature of 20 °C. The precociality coefficient, i.e. maturity at birth as fraction of that at puberty, is discussed in (Augustine *et al.* 2019a), and the supply stress, i.e. maturation maintenance times squared somatic maintenance, divided by cubed assimilation, in Lika *et al.* (2014). The latter

Table 1: A selection of 10 traits (out of 250 or 280 traits depending on the DEB model) that have been used here to illustrate the application of MDS in a DEB context

Symbol	Units	Description
a_m	d	Life span
a_p	d	Age at puberty
a_b	d	Age at birth
W_∞^w	g	Ultimate wet weight
W_p^w	g	Wet weight at puberty
W_b^w	g	Wet weight at birth
\dot{R}_∞	#/d	Ultimate reproduction rate
s_s	-	Supply stress
s_H^{bp}	-	Precociality coefficient
$[p_M]$	J/d.cm ³	Specific somatic maintenance

quantifies where species are in the supply–demand spectrum. Notice that the dimensions of the various traits differ.

By definition, the distance d_{ij} between points i and j is non-negative, while $d_{ij} = d_{ji}$ and $d_{ii} = 0$. Distances can be quantified in many ways, but most of the 13 distance measures that were reviewed by Shirktorshidi *et al.* (2015) cannot be applied to traits of species, since they assume that all traits have the same dimensions. Attempts to remove dimensions, e.g. by dividing trait values through their mean value, or subtracting the mean and dividing by the standard deviation, would distort the distances between species. A natural distance between two species should depend only on their traits, and not on those of other species that happen to be in the comparison set. Moreover, adding species would affect the position of all other ones. We encountered this dimension problem when developing methods to estimate parameters simultaneously for a set of models from a set of data sets (Lika *et al.*, 2020, Marques *et al.*, 2019). Such an exercise makes sense if different models for the different data sets of a species share one or more parameters, which couples the estimation of the parameters: all parameters of data set-specific models for all data sets for a species are simultaneously estimated. This estimation is based on the minimization of a loss function, which essentially quantifies the distance between data sets and predictions for these data sets. So, an intimate relationship exists here between parameter estimation and MDS, since they use the same loss function as a distance measure.

The common basis of parameter estimation and MDS becomes important in future extensions of the concept augmented loss functions (Lika *et al.*, 2020), where terms are added to the loss function that quantify distances to other

species in the AmP collection in terms of traits. We come back to this issue in the discussion.

The loss function that we identified as being most useful for parameter estimation is the symmetric bounded loss function

$$F_{sb} = \sum_{i=1}^n \sum_{j=1}^{n_i} \frac{w_{ij} (d_{ij} - p_{ij})^2}{n_i (d_i^2 + p_i^2)}$$

$$\text{with } d_i = \sum_{j=1}^{n_i} \frac{d_{ij}}{n_i} \quad \text{and } p_i = \sum_{j=1}^{n_i} \frac{p_{ij}}{n_i},$$

where n stands for the number of data sets and data set i has n_i data points, d_{ij} is the (measured) data point j for data set i , p_{ij} its corresponding prediction and w_{ij} its corresponding (dimensionless) weight coefficient (set equal to 1 by default). We call this loss function symmetric, because interchanging data and predictions does not affect its value and bounded, because, if a prediction goes to infinite, its contribution to the loss function remains limited. The symmetry property makes that this loss function is a distance measure. This is of importance for avoiding over-estimation as well as under-estimation, as discussed in Marques *et al.* (2019).

Likewise for application in MDS, where $n_i = 1$, and re-using the symbol d_{ij} to denote the distance between species i and j , this distance can be defined as

$$d_{ij} = \sum_{k=1}^p \frac{w_k (\theta_{ik} - \theta_{jk})^2}{p (\theta_{ik}^2 + \theta_{jk}^2)},$$

where θ_{ik} is the value of trait k for species i and the summation is over all p traits. w_k is the weight given to trait k (by default is set equal to 1). Notice that this distance measure is dimensionless.

Classical multidimensional scaling

The CMDS, known also as principal coordinate analysis (PCoA), is a method that allows to position objects, species here, in a space of reduced dimensionality, while preserving the between-object distances as well as possible. PCoA has similarities with principle component analysis (PCA), but operates on dissimilarities, i.e. distances, rather than similarities, i.e. correlations or covariances. CMDS also has similarities with cluster analysis, as will be discussed in a section below. These methods are part of multivariate analysis.

A summary of CMDS is as follows (Legendre & Legendre, 1998, Mardia *et al.*, 1982, for details). For a distance, or dissimilarity, matrix D with elements $\{d_{ij}\}_{i,j=1}^n$, CMDS defines a centered transformed distance matrix $B = CAC$, where the transformed distance matrix A has elements $a_{ij} = -\frac{1}{2}d_{ij}^2$

and the centering matrix is $C = I - n^{-1}E$, where E is a $n \times n$ matrix of all ones. The eigenvalue (or spectral) decomposition of B is now $B = QQ' = Q\Lambda Q^{-1}$, where Q is the matrix of normalized eigenvectors in its columns, also called the configuration matrix, and Λ the diagonal matrix with the eigenvalues of B , which are all real valued, since B is symmetric. This implies that $Q'Q = \Lambda$.

The number of rows of the square eigenmatrix Q equals the number of species, and the number of *relevant* columns depends on the eigenvalues. The leading eigenvalue is typically much larger than the next one and, if the 4th and subsequent eigenvalues are small, the plot of the species (i.e. the rows of Q) in three dimensions (i.e. the first three columns of Q) captures most of the scatter among species. The coordinates of species i in the 3D eigenspace are the i -th elements of the eigenvectors corresponding to the first three eigenvalues of the centered transformed distance matrix. Since the distance matrix is symmetric, we know that all eigenvalues are real, rather than imaginary.

Code implementation in AmPtool

AmPtool (AmPtool, 2021) is a software package (in Matlab) that supports the analysis of the AmP collection; it is interlinked with the more general package DEBtool (DEBtool, 2021) for applications of DEB theory. The taxonomic structure of AmP is detailed, meaning that a large number of taxonomic ranks is recognized; member species of each rank can be selected via the name of the clade. Traits for all species are collected in a large structure allStat, and several functions read from this structure. Function dist_traits takes the name of a clade and a cell string of names of traits as input, to deliver a distance matrix, using the data in structure allStat. The names of the traits are standardized codes, which follow the rules of DEB notation (Kooijman, 2010). Core-Matlab function cmdscale computes the eigenvalues and eigenvectors of the centered transformed distance matrix, using the distance matrix as input. AmPtool function shstat presents the species in the eigenspace graphically, while function connect_subclade connects all subclade members, to recognize the clade more easily; a legend with (user-defined) markers is also presented.

AmPtool function prt_report_my_pet can be used to see the 280 possible traits that are presently included in allStat; this function takes the name of an entry as input and the resulting html page has search options. Some traits depend on the DEB model for that species. The species list gives an overview of the model type that has been used for all entries.

Illustrative examples of application

We illustrate the use of the code for five clades, namely Cephalopoda, Chondrichthyes, Testudines, Austrodyptor-

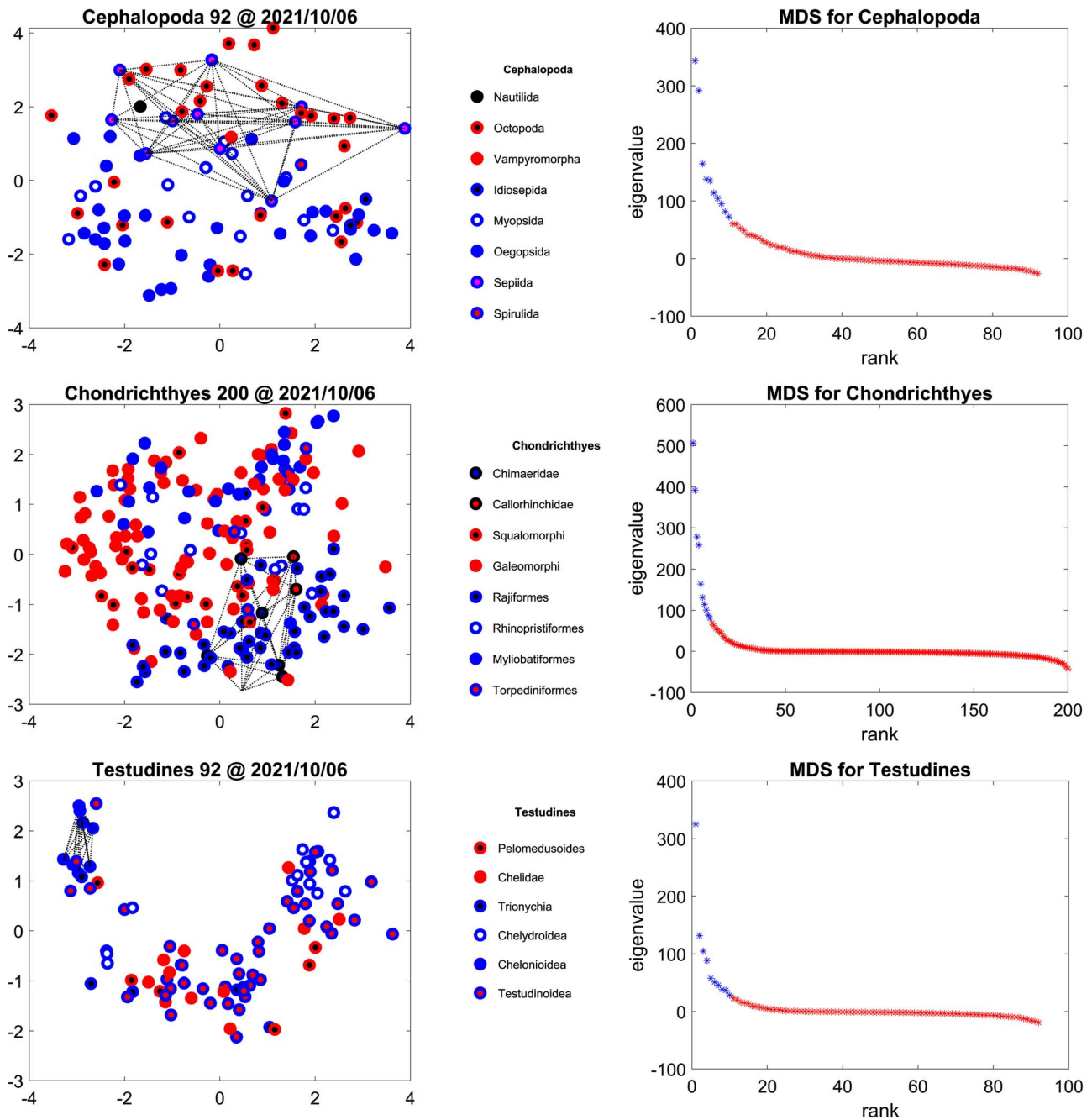


Figure 1: MDS for three ectotherm taxa in the AmP collection, with number of species and the sampling date. Members of the same clade are linked by a black line. The abscissa and ordinate in the left figures represent the first two (dimensionless) eigenvectors of the centered transformed distance matrix, based on the traits of Table 1. The edge color of the markers refers to higher ranked taxa. **Cephalopoda:** Nautiloidea, black; Octopodiformes, red; Decapodiformes, blue. **Chondrichthyes:** Chimaeriformes, black; Selachii, red; Batoidea, blue. **Testudines:** Pleurodira, red; Cryptodira, blue. Members of the clades Sepiida in the Cephalopoda, Chimaeriformes in the Chondrichthyes and Chelonioida in the Testudines are connected. The right figures present all eigenvalues. The first p eigenvalues, p is the number of traits, are presented in blue.

nithes and Carnivora, which we consider to be ‘complete’, meaning that it will not be that easy to find more species with data in the open literature. The five scripts

mydata_msd_taxon, where taxon is one of the five clade names, are available in AmPtool [AmPtool \(2021\)](#) and can easily be adapted for other taxa and/or traits. Possibly with

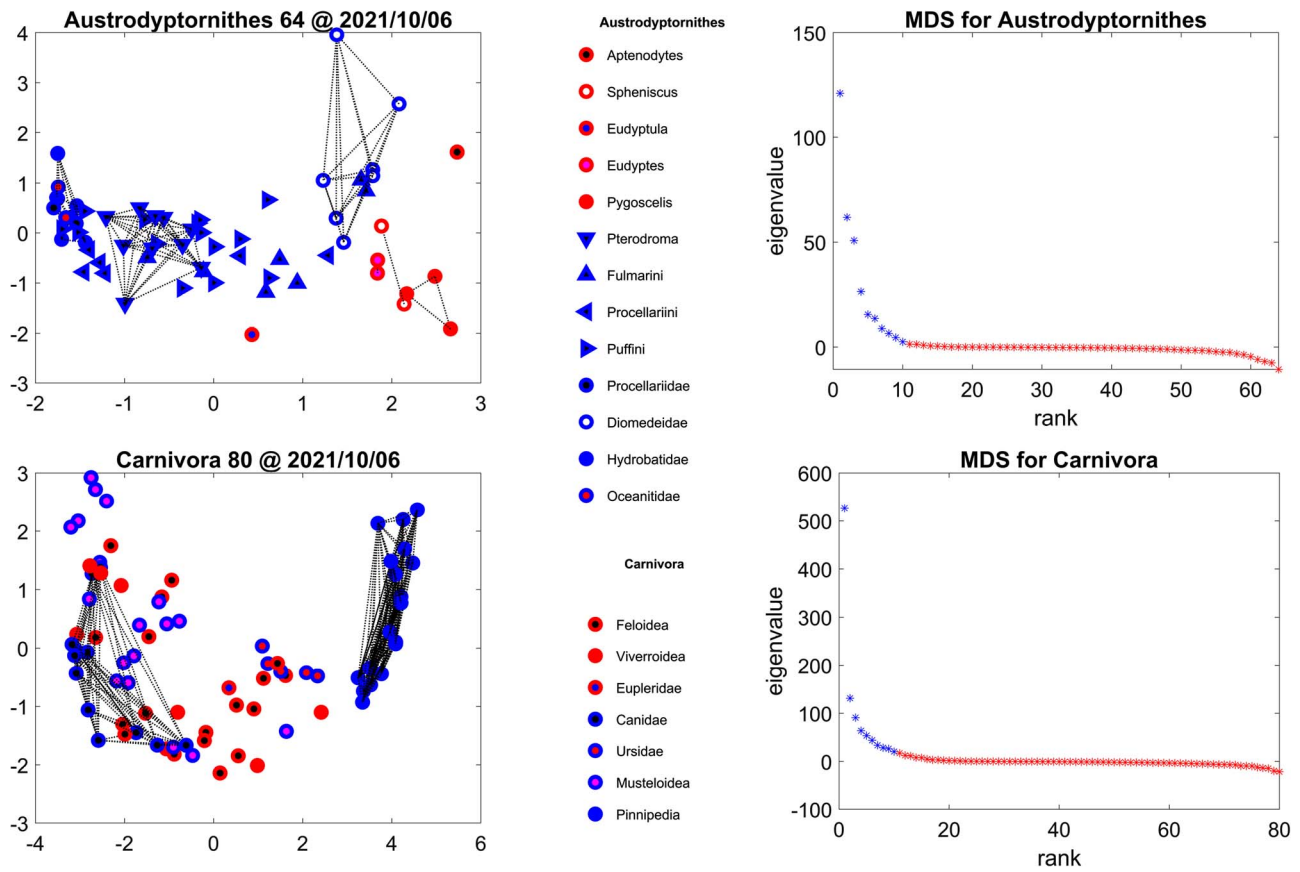


Figure 2: Like Fig. 1 for two endotherm taxa in the AmP collection. The edge color of the markers refer to higher ranked taxa. **Austrodyptornithes:** Sphenisciformes, red; Procellariiformes, blue. **Carnivora:** Feliformia, red; Caniformia, blue. Members of the clades Hydrobatidae, Diomedeidae, Eudyptes, Pygoscelis, Spheniscus and Pterodroma in the Austrodyptornithes and Canidae and Pinnipedia in the Carnivora are connected.

the exception of the cephalopods, most species of the other four clades are considered to be vulnerable, endangered or even critically endangered.

Figure 1 shows three examples of application of MDS to ectotherm taxa of the AmP collection; Fig. 2 does this for two endotherm taxa. The number of eigenvalues equals the number of rows (or columns) of the square symmetric distance matrix between species, i.e. the number of species. Although the code in AmPtool/mydata_msd_* generates 3D mds-plots, the figures in this paper show only the first two. The screen plots are interactive; you can rotate the figure, and click on markers to see the species name. Species that cluster in the eigenspace can be connected to enhance their clustering. The traits of the Austrodyptornithes (i.e. penguins and petrels) are discussed in Kooijman (2020a), in the context of their natural history. The two outliers of the Procellariini belong to the genus *Procellaria*; the three other genera do cluster.

The cephalopods have the most complex configuration (i.e. highest λ_{10}/λ_1 ratio); they are the only one of these five taxa that sport metabolic acceleration, they have the largest range

in (maximum) body sizes and the size of neonates as fraction of size at death is the smallest, so they grow most during their (short) life span.

The Pinnipedia (seals) among carnivora show a clear separation from Canidae (dogs) with respect to the first axis (Fig. 2). To determine which traits contribute most to the observed pattern among species, we correlated the eigenvectors from the MDS with each trait. It turns out that the ultimate reproduction rate (-ve), precociality coefficient (+ve) and age traits (+ve) correlate the highest with the first axis (correlation coefficients larger than 0.7). Getting a few large pups doubtlessly relates to their aquatic life style, minimizing the time they need to spend on the beach.

Like the seals among the carnivores, the petrels separate from the penguins among the Austrodyptornithes, see the blue versus red points in Fig. 2, and possibly for the same reason: petrels have large eggs for their adult weight, while the penguins have egg sizes that are typical for birds of that weight (Kooijman, 2020a). This trait is intimately linked to quite a few other traits.

Table 2: The 10th eigenvalue of the centered transformed distance matrix as fraction of the first one, with 10 traits per species, see Table 1

Cephalopoda	0.214
Chondrichthyes	0.160
Testudines	0.088
Austrodyptornithes	0.022
Carnivora	0.039

The sea turtles (Chelonioidae) cluster among the turtles (Fig. 1). Although the mean egg weight of sea turtles is twice that of turtles, their relative egg weight is much smaller, since their adult weight is much larger than the mean for turtles.

The right-hand sub-figures in Figs. 1 and 2 show that the subsequent eigenvalues rapidly go to zero and hardly become negative, meaning that the distances between species behave like Euclidean distances. Notice that the number of eigenvalues equals the number of species (so the size of the distance-matrix). The number of eigenvalues colored in blue equals the number of traits that has been used, see Table 1. This was motivated by the idea that the more traits are included the more complex the distance structure between species becomes, with the consequence that more eigenvalues will be large. Table 2 shows the 10th eigenvalues as fraction of the first (i.e. largest) ones for the five examples. The lower this fraction, the fewer dimensions are required to capture most scatter in position of the species in the eigenspace.

These illustrative examples show that, in quite a few cases, related species appear clustered in the eigenspace. This points to links between parameter values and taxonomic position of a species.

Cladograms versus MDS

This section compares a data presentation in the form of a cladogram with that in MDS to illustrate the relationship between these methods. For clarity, we will use a simple example of a distance matrix that can be presented explicitly. A cladogram presentation is a special case of the larger class of cluster techniques, as are nowadays frequently used to present DNA, RNA or amino acid sequences.

The distance between two taxa can be expressed in terms of years since divergence of these taxa during evolution. For vertebrate taxa, divergence timing is given in Table 3, and this table is used to compose the cladogram in Fig. 3 and the CMDS plot in Fig. 4, two alternative representations of the same thing. The cladogram can be produced in Matlab with function seqlinkage, method average, in toolbox bioinformatics; the CMDS plot with function cmdscale in toolbox statistics.

Table 3: The timing of divergences from Hedges & Kumar (2021), quantified in MYA

	Cyclostomata	Holocephali	Galeomorphi	Squalomorphi	Batoidea	Cladistii	Chondrostei	Holostei	Teleostei	Actinistia	Dipnoi	Tetrapoda
Cyclostomata	0	615	615	615	615	615	615	615	615	615	615	615
Holocephali		0	399	399	399	473	473	473	473	473	473	473
Galeomorphi			0	199	265	473	473	473	473	473	473	473
Squalomorphi				0	265	473	473	473	473	473	473	473
Batoidea					0	344	344	344	435	435	435	435
Cladistii						0	386	386	386	435	435	435
Chondrostei							0	344	344	435	435	435
Holostei								0	315	435	435	435
Teleostei									0	435	435	435
Actinistia										0	413	413
Dipnoi											0	395
Tetrapoda												0

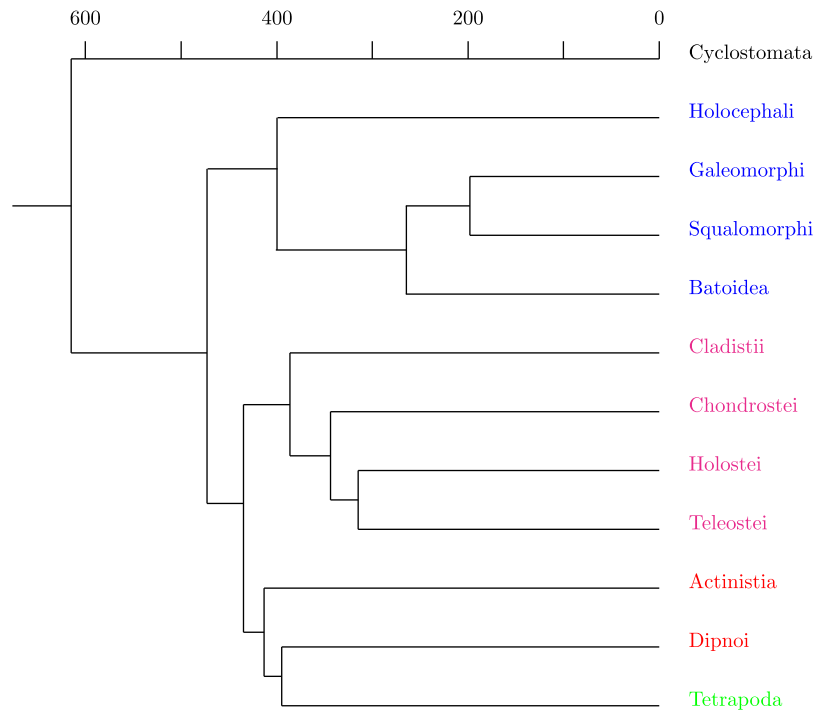


Figure 3: A cladogram of the evolutionary relationships between vertebrate taxa on the basis of the distance matrix of Table 3. The time since divergence is given on top, in MYA.

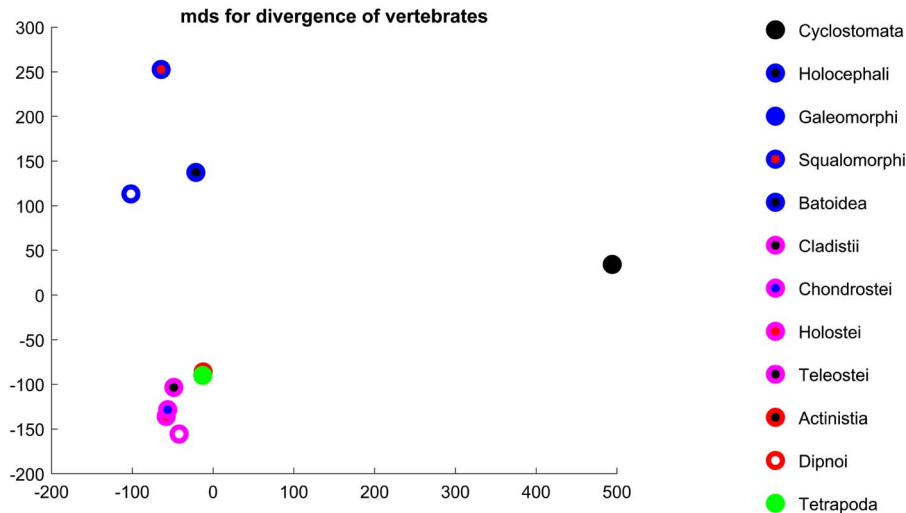


Figure 4: CMDS applied to the distance matrix of Table 3. The Actinistia almost hide behind the Tetrapoda, while the Dipnoi fully hide behind the Tetrapoda and the Galeomorphi behind the Squalomorphi. The edge color of the markers relates to the classification: blue for Chondrichthyes, magenta for Actinopterygii and red and green for the Sarcopterygii.

The divergence between the Cyclostomata (lampreys and hagfishes) and the Gnathostomata (the other vertebrates) is estimated at 797 MYA (Hedges & Kumar, 2021), but the origin of the Vertebrata at 615 MYA; we used the latter value in the Table 3. One does not need a cluster program in this case to construct the cladogram from the distance

table. Although the biological interpretation of this cladogram is presently less relevant, the discussion whether the coelacanth-lungfish group (CL-group) is a monophyletic sister group of the closely related tetrapods (Shan & Gras, 2011, Zardoya & Meyer, 1996), or the prevailing view that lungfish-tetrapod group (Rhipidistia) is monophyletic (Forey, 1986),

is still open and possibly cannot be resolved (Takezaki *et al.*, 2004).

The lampreys, cyclostomes, come out as very distant from the other vertebrates (gnathostomes) in both presentations. The rays and sharks (chondrichthians) cluster separately from the ray-finned fish (actinopterygians) in the CMDS plot while the three sarcopterygian taxa cluster very tightly. This is also visible in the cladogram. The lungfish, Dipnoi, have the same first two coordinates in the eigenspace as the tetrapods in the CMDS plot. This also applies to the Galeomorphi, which have the same first two coordinates as the Squalomorphi. They only differ in the seventh coordinate, where they have opposite signs. A full presentation of the CMDS plot requires 11 dimensions, while the figure shows only the first 2.

Discussion

The strength of MDS is that it can reveal relationships between taxa based on combinations of trait values, in a way that is not that easy to see directly in the values of traits, since it is the combination of traits that matters. The Pinnipedia among the Carnivora and the Sphenisciformes among the Austrodytornithes are striking examples in Fig. 2: they appear fully separated from the other carnivores and petrels, respectively, in the eigenspace. This invites for a closer analysis of such taxa, to better understand what makes them that different, leading to improved eco-evolutionary insight. It turned out that the distances, as quantified by the symmetric bounded loss function, are almost Euclidean.

Another application of MDS is in supporting parameter estimation. To what extent measured data determine parameters accurately is always an issue, which directly affects values of traits and, therefore, distances between species. Concerns might be (i) the limited availability of data, (ii) uncertainty in measurements (in many cases food and temperature conditions had to be guessed and biological data does have scatter) or (iii) differences in geographic races and populations that have been used for measurements. To reduce this problem, we developed several methods to estimate parameters in context (Lika *et al.*, 2020, Marques *et al.*, 2019), and have written code to compare species and use the recognized patterns in parameters to judge to what extent some parameters are outliers. MDS is a valuable additional tool in this respect. If we identify outliers, we revisit the parameter estimation and study the possibility to get the parameters more in concert with that of related taxa, without affecting the goodness of fit. Although the parameter combination for which the loss function has a global minimum is the combination we are looking for, the loss function might have other local minima that are hardly higher than the global one and might be biologically more realistic.

CMDS is just one of a set of related methods aiming to produce ordination of data points in few dimensions, while preserving, as well as possible, the distances among them.

We briefly discussed links with PCA and cluster techniques. The latter methods aim to cluster similar data points into the same clusters, while dissimilar or distant data points are placed into different clusters. Other MDS methods involve additional steps, beyond the initial scaling used by the CMDS. Extended MDS methods iteratively reposition the objects in the configuration using an algorithm that improves the fit between the dissimilarities and the inter-object distances. The Matlab function `mdscale` performs non-classical MDS with choices to perform metric or non-metric scaling.

In addition to the symmetric bounded loss function that we used in this analysis, we have tried several other distance measures that allow traits with different dimensions, e.g. Canberra metric and coefficient of divergence, see Shirikhorshidi *et al.* (2015), with classical and non-classical MDS algorithms but did not find one with better results, i.e. clustering related species better. The symmetric unbounded loss function, as proposed by Marques *et al.* (2019) for parameter estimation as alternative for the symmetric bounded one, seemed to perform poorly in this context.

We also tried several selections of traits and have the impression that a much shorter list performed less well, while a larger list did not change the result much, but we cannot claim that the list as mentioned in Table 1 is optimal in any respect and the best selection might very well depend on the taxon.

We see our discovery that the loss function that we use in parameter estimation and now also in MDS as an important new insight that links the two fields in a deep way. Some long-known distance measures, such as the Euclidean distance, strongly remind of least-squares regression problems and so of parameter estimation, but they are the ones that give dimension problems, both in multi-model parameter estimation and in MDS with traits (Marques *et al.*, 2019). Parameter estimation and MDS are, therefore, only linked using a proper choice for the distance measure. To see that the link is deeper than just using the same loss function, it is possible to include trait distances between species in an augmented term in the loss function that is used in multi-species parameter estimation, like we did for minimizing the variation coefficients for selected parameters (Lika *et al.*, 2020) that is currently implemented in DEBtool. The importance of simultaneous minimization of trait distances in parameter estimation is then controlled by weight coefficients. By step-wise increasing the weight coefficients and monitoring the effect on the mean relative error of predictions for observations, it becomes possible to distinguish random differences in trait estimates from data-supported differences. By imposing these constraints, the total number of parameters to be estimated is reduced, so are the associated ill-posedness problems. We see this as a further step in parameter estimation in context.

The present examples only focus on taxonomic position, but the AmP collection also has identifiers for climate, ecozone, habitat, embryo-environment, migration, diet, gender

and reproduction type for all entries, see https://bio.vu.nl/thb/deb/deblab/add_my_pet/AmPeco.html. It might very well be that, if these identifiers are taken into account as well, the clustering in species positions in the eigenspace improves considerably. We are only at the start of a whole new direction in the analysis of the AmP database, but we can conclude that these first MDS applications look promising.

Acknowledgements

We want to thank all who contributed to AmP collection and Michael Kearney for comments on the manuscript.

Funding

This work was supported by the Norwegian Science Council [NFR 255295 to S.A.] and by the Croatian science foundation [project AqADAPT no. IP-2018-01-3150 to N.M.].

References

- AmP (2021) *AmP collection*. Add-my-Pet collection, online database of DEB parameters, implied properties and referenced underlying data. http://www.bio.vu.nl/thb/deb/deblab/add_my_pet/.
- AmPtool (2021) *Software package AmPtool*. <https://github.com/add-my-pet/AmPtool>.
- Arnall SG, Mitchell NJ, Kuchling G, Durell B, Kooijman SALM, Kearney MR (2019) Life in the slow lane? A dynamic energy budget model for the western swamp turtle, *Pseudemys umbrina*. *J Sea Res* 143: 89–99.
- Augustine S, Lika K, Kooijman SALM (2019a) Altricial-precocial spectra in animal kingdom. *J Sea Res* 143: 27–34.
- Augustine S, Lika K, Kooijman SALM (2019b) Why big-bodied animal species cannot evolve a waste-to-hurry strategy. *J Sea Res* 143: 18–26.
- Augustine S, Lika K, Kooijman SALM (2020) Comparing loss functions and interval estimates for survival data. *Ecol Model* 430: 109077.
- Baas J, Kooijman SALM (2015) Sensitivity of animals to chemical compounds links to metabolic rate. *Ecotoxicology* 24: 657–663.
- Bertolini C, Brigolin D, Porporato EMD, Hattab J, Pastres R, Tiscar PG (2021) Testing a model of pacific oysters' (*Crassostrea gigas*) growth in the Adriatic Sea: implications for aquaculture spatial planning. *Sustainability* 13: 3309.
- Calder WAIII (1984) *Size, Function and Life History*. Harvard University Press, Cambridge, MA, USA.
- Chary K, Aubin J, Sadoul B, Fiandrino A, Covès D, Callier MD (2020) Integrated multi-trophic aquaculture of red drum (*Sciaenops ocellatus*) and sea cucumber (*Holothuria scabra*): assessing bioremediation and life-cycle impacts. *Aquaculture* 516: 734621.
- DEBtool (2021) Software package DEBtool_M. https://github.com/add-my-pet/DEBtool_M.
- Forey PL (1986) Relationships of lungfishes. *J Morphol* 1: 75–91.
- Goedegebuure M (2018) *Improving representations of higher trophic-level species in models: using individual-based modelling and dynamic energy budget theory to project population trajectories of southern elephant seals*. PhD thesis. University of Tasmania, Hobart, Australia.
- Haberle I, Marn N, Geček S, Klanjšček T (2020) Dynamic energy budget of endemic and critically endangered bivalve *Pinna nobilis*: a mechanistic model for informed conservation. *Ecol Model* 434: 109207.
- Hedges, B. and Kumar, S. (2021) TimeTree. <http://www.timetree.org/>.
- Jackson B (2018) *The GREAT-ER model as a tool for chemical risk assessment and management for Chinese river catchments*. PhD thesis. Lancaster University, UK.
- Jin L, Jiang G, Li X (2020) Transforming environmental chemistry and toxicology to meet the anthropocene sustainability challenges beyond silent spring. In GB Jiang, XD Li, eds, *A New Paradigm for Environmental Chemistry and Toxicology: From Concepts to Insights*. Springer Singapore, Malaysia.
- Kearney MR, Jusup M, McGeoch MA, Kooijman SALM, Chown SL (2021) Where do functional traits come from? The role of theory and models. *Funct Ecol* 35: 1385–1396.
- Kleiber M (1932) Body size and metabolism. *Hilgardia* 6: 315–353.
- Kooijman SALM (1986) Energy budgets can explain body size relations. *J Theor Biol* 121: 269–282.
- Kooijman SALM (2009) *Dynamic Energy Budget theory for metabolic organisation*. Cambridge University Press, Cambridge.
- Kooijman, S. A. L. M. (2010) Notation for dynamic energy budget theory. https://www.bio.vu.nl/thb/deb/deblab/bib/Kooy2010_n.pdf.
- Kooijman SALM (2013) Waste to hurry: dynamic energy budgets explain the need of wasting to fully exploit blooming resources. *Oikos* 122: 348–357.
- Kooijman SALM (2014) Metabolic acceleration in animal ontogeny: an evolutionary perspective. *J Sea Res* 94: 128–137.
- Kooijman SALM (2020a) The comparative energetics of petrels and penguins. *Ecol Model* 427: 109052.
- Kooijman SALM (2020b) The standard dynamic energy budget model has no plausible alternatives. *Ecol Model* 428: 109106.
- Kooijman SALM, Bedaux JJM (1996) *The Analysis of Aquatic Toxicity Data*. VU University Press, Amsterdam.
- Kooijman SALM, Metz JAJ (1984) On the dynamics of chemically stressed populations; the deduction of population consequences from effects on individuals. *Ecotoxicol Environ Saf* 8: 254–274.
- Legendre P, Legendre L (1998) *Numerical Ecology*. Elsevier, Amsterdam, the Netherlands.
- Lika K, Augustine S, Kooijman SALM (2019) Body size as emergent property of metabolism. *J Sea Res* 143: 8–17.
- Lika K, Augustine S, Kooijman SALM (2020) The use of augmented loss functions for estimating dynamic energy budget parameters. *Ecol Model* 428: 109110.

- Lika K, Augustine S, Pecquerie L, Kooijman SALM (2014) The bijection from data to parameter space with the standard deb model quantifies the supply-demand spectrum. *J Theor Biol* 354: 35–47.
- Mardia KV, Kent JT, Bibby JM (1982) *Multivariate Analysis*. Academic Press, London.
- Marn, N., Jusup, M., Kooijman, S. A. L. M., and Klanjscek, T. (2020) Quantifying impacts of plastic debris on marine wildlife identifies ecological breakpoints. *Ecol Lett* 23:1479–1487.
- Marn N, Jusup M, Legović T, Kooijman SALM, Klanjšček T (2017) Environmental effects on growth, reproduction, and life-history traits of loggerhead turtles. *Ecol Model* 360: 163–178.
- Marques, G. M., Augustine, S., Lika, K., Pecquerie, L., Domingos, T., and Kooijman, S. (2018) The AmP project: comparing species on the basis of Dynamic Energy Budget parameters. *PLoS Comput Biol* 14: e1006100.
- Marques GM, Lika K, Augustine S, Pecquerie L, Kooijman SALM (2019) Fitting multiple models to multiple data sets. *J Sea Res* 143: 48–56.
- McKenzie1 DJ, Axelsson M, Chabot D, Claireaux G, Cooke SJ, Corner RA, Boeck GD, Domenici P, Guerreiro PM, Hamer Bet al. (2016) Conservation physiology of marine fishes: state of the art and prospects for policy. *Conserv Physiol* 4: 1–20.
- Molnár PK, Derocher AE, Thiemann GW, Lewis MA (2010) Predicting survival, reproduction and abundance of polar bears under climate change. *Biol Conserv* 143: 1612–1622.
- Parra CS, Schulz KS, Hammocka J, Wilson N, Leary P, Rice J, Corrigan RJ (2015) Traitbank: practical semantics for organism attribute data. *Semantic Web* 7.
- Pete R, Guyondet T, Bec B, Derolez V, Cesmat L, Lagarde F, Pouvreau S, Fiandrino A, Richard M (2020) A box-model of carrying capacity of the Thau lagoon in the context of ecological status regulations and sustainable shellfish cultures. *Ecol Model* 426: 109049.
- Peters RH (1983) *The Ecological Implications of Body Size*. Cambridge University Press, Cambridge, UK.
- Sangare N, Lo-Yat A, Moullac GL, Pecquerie L, Thomas Y, Lefebvre S, Gendre RL, Beliaeff B, Andréfouët S (2020) Impact of environmental variability on *Pinctada margaritifera* life-history traits: a full life cycle DEB modeling approach. *Ecol Model* 423: 109006.
- Sarà G, Gouhier TC, Brigolin D, Porporato EMD, Mangano MC, Mirto S, Mazzola A, Pastres R (2018) Predicting shifting sustainability trade-offs in marine finfish aquaculture under climate change. *Glob Change Biol* 3654–3665.
- Sarà G, Rinaldi A, Montalto V (2014) Thinking beyond organism energy use: a trait-based bioenergetic mechanistic approach for predictions of life history traits in marine organisms. *Mar Ecol* 35: 254–274.
- Shan Y, Gras R (2011) 43 genes support the lungfish-coelacanth grouping related to the closest living relative of tetrapods with the Bayesian method under the coalescence model. *BMC Res Notes*. 4: 49.
- Shirkhorshidi, A. S., Aghabozorgi, S., and Wah, T. Y. (2015) A comparison study on similarity and dissimilarity measures in clustering continuous data. *PLoS One* 10(12):e0144059.
- Stavrakidis-Zachou, O., Lika, K., Anastasiadis, P., and Papandroulakis, N. (2021) Projecting climate change impacts on mediterranean finfish production: a case study in greece. *Clim Change* 165: 67.
- Takezaki, N., Figueroa, F., Zaleska-Rutczynska, Z., Takahata, N., and Klein, J. (2004) The phylogenetic relationship of tetrapod, coelacanth, and lungfish revealed by the sequences of forty-four nuclear genes. *Mol Biol Evol* 21: 1512–1524.
- Taylor CA, DiStefano RJ, Larson ER, Stoeckel J (2019) Towards a cohesive strategy for the conservation of the united states' diverse and highly endemic crayfish fauna. *Hydrobiologia* 846: 39–58.
- Teixeira CMGL (2016) *Application of dynamic energy budget theory for conservation relevant modelling of bird life histories*. PhD thesis. VU University Amsterdam, The Netherlands & Lisbon University, Portugal.
- Yang T, Ren J, Kooijman S, Shan X, Gorfine H (2020) A dynamic energy budget model of *Fenneropenaeus chinensis* with applications for aquaculture and stock enhancement. *Ecol Model* 431: 109186.
- Zardoya R, Meyer A (1996) Evolutionary relationships of the coelacanth, lungfishes, and tetrapods based on the 28s ribosomal RNA gene. *Proc Natl Acad Sci U S A* 93: 5449–5454.