Group Size and Behavioural Ecology in the Superfamily Delphinoidea (Delphinidae, Phocoenidae and Monodontidae)

Dissertation
zur
Erlangung der naturwissenschaftlichen Doktorwürde
(Dr. sc. nat.)
vorgelegt der
Mathematisch-naturwissenschaftlichen Fakultät
der
Universität Zürich
von

Lorenz Gygax
von
Seeberg BE

Begutachtet von
Prof. Dr. Andrew D. Barbour
Dr. Werner Stahel
Prof. Dr. Paul Ward

Zürich 2000

Die vorliegende Arbeit wurde von der Mathematisch-naturwissenschaftlichen Fakultät der Universität Zürich auf Antrag von Prof. Dr. Andrew D. Barbour und Prof. Dr. Paul Ward als Dissertation angenommen.

General Acknowledgements

I would like to thank Andrew D. Barbour for the opportunity to conduct the work on this PhD and the invaluable opportunity to learn more about statistics, Lotti and Andres Gygax for their continuing interest in my work, my colleagues Roman Kälin, Christof Luchsinger, Michael Trachsler and many more people from the departments of Mathematics and Zoology for the friendly setting and the many discussions, my friends Erika Bucheli, Gyula Gajdon, Heinz Jufer and Gerulf Rieger for accompanying and supporting me during this rich time, and especially Jeannine Ammann who showed me how to be happy.

For more specific thanks, please refer to the acknowledgement sections of the different chapters.

This PhD thesis has been supported by the Swiss National Science Foundation with grants No. 20-43'453.95 and 20-50'686.97.

Contents

Appendix B: Details and History of the Statistical Evaluations

Prerequisite

The data that is used in these evaluations originate from different studies which differ in their methodology. It is assumed that the variability in the data due to this differences in methodology is rather small because the data used consist of basic biological and physical measurements which should not depend much on the technique of measuring. Additionally, the patterns that are investigated are supposed to be strong such that they would be apparent even if some additional variability was introduced due to different methodologies in different studies.

Statistical evaluations and figures were done on a SuSE 5.2 Linux system with R Versions 0.63.1 and 0.64.1 (see http://www.ci.tuwien.ac.at/R) and its libraries. Most methods can be found in Venables and Ripley (1997) and are discussed there. Where indicated evaluations were conducted using SPSS for Unix on IBM/RS 6000 Release 6.1 or S-PLUS Version 5.0 Release 2 for Sun SPARC.

History of the interspecific comparison

Stepwise backward-forward regression

Variables

The following variables served as response variables in this step of the evaluation (see also the section on interspecific comparisons, page 23 and Appendix A): *mean observed group size*, *standard deviation of observed group size*, *median observed group size* the *inter–quartile range of observed group size*, *mean perceived group size*, *standard deviation of perceived group size*, *median perceived group size*, the *inter–quartile range of perceived group size*, *minimum group size* and *maximum group size*. A further variable gives the *number of observed groups* that are included in the measures of group size.

As explanatory variables the following were considered: the phylogenetic variables *species* and *region*, the variables of the physical environment *habitat*, *habitat structure*, *latitude*, *typical temperature*, *range of temperature*, *variability of temperature*, *typical depth*, *range of depths*, *depth contour index*, the diet variables for cephalopods, fishes and mammals & birds (occurrence, frequency and proportion data for either (1) the unsplit categories [total 9 variables], (2) the food categories split by where they occur in the water column [total 27 variables] or (3) food categories split by the occurrence in the water column and their grouping patterns [total

91 variables]).

The variables on the physical environment were used in the main regressions (see below) and in the analysis of the coefficients of the variable *region*. For the latter which averages of these variables were calculated for each region.

For some variables the information was missing for so many combinations of species and region (= populations) that no attempt was made to incorporate them in the main models. All these variables can be regarded as properties of the species and often there was only one measure available per species. Thus, it was assumed that all these properties are accounted for by the variable *species* in the main models, but additionally, the coefficients of *species* were further analysed in dependence of these variables. Where more than one study reported these measures for a given species an average was calculated weighed by the number of observations in each study.

These species specific variables included the phylogenetic variable *family*, the social variables *group fluidity*, *group structure* and *lifestyle*, the life-history variables on size: *length of females*, *weight of females*, *length of males*, *weight of males*, *length dimorphism*, *weight dimorphism*, on reproduction: *seasonality*, *length of gestation*, *length of lactation*, *calving interval* and on maturation: *age of females at maturation*, *length of females at maturation*, *age of males at maturation*, *length of males at maturation*, *maximum life span of females*, *maximum life span of males*.

Data for these variables was still not available for all species. Thus the coefficients of the main model for the variable *species* were regressed against the *group fluidity*, *group structure*, the *family* and the two representatives for the life-history variables *length of females* and *length dimorphism*.

Statistical Methods

Multiple linear and logistic regressions were used. Variables were log transformed where necessary (basically all response variables and all continuous explanatory variables except *habitat*, *habitat structure* and *latitude*, see also section on interspecific comparison, page 23 and Appendix A). For a given regression problem a backward–forward stepwise procedure was conducted using the BIC criterion (BIC $= -2 \cdot$ maxlog-lh $+ p \cdot \log(n)$, where p is the number of variables and n the number of cases included in a model and maxlog-lh the maximum log-likelihood).

The sample size for a given regression was restricted to the cases where all the necessary variables were known. Model assumptions were checked using graphical methods (see below on the details of the final evaluations, page 109).

Each of the ten measures of group size was used as the response variable in separate regressions. As these measures of group size were highly correlated (see next section), the idea was that one can see these models as a kind of robustness/sensitivity analysis within the location and variability measures. The explanatory variables that are important in many of the models seem to explain the variability and patterns in the response variable most reliably.

The minimum group size could not be subjected to a standard regression analysis because many of these observations were one (see also the next section). An attempt was made to use a logistic regression on whether minimal group size equalled one or not. This was not feasible with the present data set because there were numerical problems in solving the equations as estimated probabilities were too close to zero and one.

All explanatory variables were included in the models as indicated in the section above. For

the diet only the variables of a certain split level (1: by prey type, 2: by prey type and location in water column, 3: by prey type, location in water column and grouping pattern) was included at the same time. Every model was conducted including occurrence, frequency and proportion and including the occurrence variables only. For the latter, the sample size was highest and at the same time the information on prey species was least equivocal. The model selections were conducted in a way that each occurrence category could be included singly but the frequency and proportion categories were either included all or none.

Additionally, based on the residual-plots of initial results, the interaction between *species* and *region* was included in the evaluation of the main models, the squared *length dimorphism* and the interaction between *lifestyle* and *family* in the evaluation of the coefficients of *species* and the squared *typical temperature* in the coefficients of *region*.

Each regression problem was defined by the combination of one of the response variables (10) with one of the three levels of detail in splitting the diet variables, either including occurrence, frequency and proportion or occurrence only. This results in 60 main models. As the sample size was mainly given by the response variable and by whether all types of diet variables or occurrence only were included, I consider 20 "submodels" which in itself contain three models each (the different levels of splits in the diet variables).

Both the coefficients of the variable *species* and *region* were further regressed against life-history variables and variables of the physical environment as described above. The number of models calculated was given by the number of the 60 models that actually included the variables *species* and *region*.

The maximum model with which the backward–forward procedure was started was restricted by the available data and by the number of cases in the data base (total number of cases: 480). In principle a linear main effects model is considered. Non-linear (e. g. quadratic) effects were detected with the RS-plots (residuals plus component effect versus explanatory variables) and tested (see next section).

As residuals were plotted against all combinations of two possible explanatory variables, at least two–way–interactions should have been detected and were tested if necessary (see next section). The interactions between species and all other variables, between food variables and the indicator whether diet was from the same region and between food variables and the number of observations from which diet was calculated were thought to be likely to occur and thus graphically checked for carefully.

Results from the main stepwise backward-forward regression

The response variables were all highly correlated. The correlation coefficients between the response variables (except for the minimum group size) ranged between 0.68 and 0.98 with a median of 0.91. Thus, it seemed reasonable to assume that these response variables basically measure the same parameter, and thus, the different (sub-)models can be seen as a sensitivity analysis for the explanatory variables that are included in the models. Details on the main models are given in Table 14.

Additionally to the variables listed in Table 14 and mentioned in the methods I ran some regressions with other sets of variables: (a) the *average group size* was regressed against all variables except *species* and *region*. The residual error increased about 1.5 times and residuals deviated strongly from a normal distribution. (b) the interactions between *species* and *habitat* and *species* and *habitat structure* were also included in the regressions of *average group size*, though there was no (strong) indication for it in the residuals. As could then be expected, these

Table 14: Number of occurrences of the different explanatory variables in all models. #: Number of occurrences in the 20 submodels. Range: sign and range of values of coefficients.

response variable [a]				A		B		C		D		E		F		G		H		I		J	
diet variables [b]				a	o	a	o	a	o	a	o	a	o	a	o	a	o	a	o	a	o	a	o
explanatory variables	#		range																				
species	16		categorical	3	3	3	3	2	3	2	3	3	2	3		3		3		3	3	2	
region	15		categorical	3	3		2	2	3	2	3	3	2	3		3		3		3	3	2	
habitat	6	+	0.37–0.52					1		1			1		3		3		3				
structure	2	−	0.22–0.23							1										1			
latitude	3	−	0.01–0.02			3		1		1													
typical temp.	4	+	11.9–28.9					1		1			1		3								
range of temp.	0																						
variab. temp	3	+	0.31–0.39						3							1					3		
	1	−	0.40							1													
typical depth	4	+	0.11–0.16									2		2							3		3
range of depth	3	+	0.12–0.18						3				2							2			
contour index	4	+	0.22–0.34									3		2		3		2					
number of groups	2	+	0.31–0.34																	3	3		
o-c[c]	4	+	0.40–0.84								1				1		1		1				
o-f	7	+	0.39–1.41		1	1	1		1		1		1								1		
f-c, f-f, f-m	1		–0.01–0.01							1													
o-c-m	3	−	0.50–0.60	1								1		1									
o-f-e	4	−	0.51–0.78	1										1				1		1			
o-c-b	1	−	0.75			1																	
o-c-e	4	+	0.52–0.68							1					1		1		1				
o-f-m	3	+	0.56–0.70										1				1		1				
o-c-m-i	4	−	0.49–0.60	1						1		1		1									
o-c-b-g	1	+	0.53	1																			
o-c-b-i	1	−	0.69			1																	
o-m-e-g	1	−	0.63–1.98				1			1													
o-c-m-g	1	+	0.76								1												
o-f-e-s	1	+	1.21								1												
o-f-e-i	1	−	0.50																	1			
o-f-b-s	1	+	0.58								1												
o-c-e-i	1	−	0.64													1							
o-c-e-g	1	+	1.01													1							
o-f-m-i	5	+	0.54–0.91												1	1	1		1	1			
o-f-b-i	2	−	0.58–0.71													1		1					

[a] A: *average observed group size* (N: 189/444 [a/o]), B: *median observed group size* (130/298), C: *standard deviation of observed group size* (163/372), D: *inter-quartile range of observed group size* (112/252), E: *average perceived group size* (112/254), F: *median perceived group size* (112/253), G: *standard deviation of perceived group size* (112/251), H: *inter-quartile range of perceived group size* (107/242), I: *maximum observed group size* (162/389), J: *minimum observed group size* (71/153) [b] a: all; o: occurrence only; each with three possible split levels) [c] prey variables: first letter = type of measure o: occurrence, f: frequency, p: proportion; second letter = class of prey species i: invertebrates, c: cephalopods, f: fish, m: mammals and birds; third letter = location of prey in water column e: epipelagic, m: mesopelagic, b: benthic; fourth letter = prey grouping s: single prey, m: mixed, g: grouping prey

occurs only in four submodels and mostly only in one model within those submodels (Table 14). It is surprising that the latitude should have a small (negative) influence on group size, especially as there was no discernible quadratic effect which could be expected as here latitude is measured from 0 (South Pole) to 180 (North Pole).

The results from the diet variables is somewhat inconclusive. Only variables describing whether a specific food type did occur in the diet were included in the models. Most of these variables only occur in one or a few submodels and the sign of the influence is not consistent over the different levels of detail of the food categories (cephalopods, fish, mammals & birds). Additionally, both the occurrence of fish and occurrence of cephalopods influence the group size positively (though in different submodels). These influences are also comparatively low as they are in the range of a factor of about 1.5 (depending on the presence of absence of a specific food category).

In general, the main models are very consistent in that (with one exception: *variability of temperature*) all the estimated coefficients have the same sign and the variations in the size of the effects is small (Table 14).

Results from the stepwise backward-forward regression of the species effects

In the evaluations of the species effect an attempt is made at explaining the size of these effects by variables that are constant within the species (see also the methods section of the interspecific comparison, page 23).

The sample sizes for these evaluations were given by the number of species that were included in the main models and by the availability of the life-history variables for these species. The submodels that evaluated the coefficients from the main models that used all the prey variables on a given split level had a sample size of about 20 and the ones that evaluated coefficients from the main models that had only used prey occurrence had a sample size of about 30.

In general, the step-wise backwards-forwards method based on the BIC worked well again for the regressions of the coefficients of *species* in that the BIC value was smaller than the ones from the null and the full model in 39 of 44 models (in the other five models its value was close to the one of the null model) and in that most explanatory variables included in the models were significant on the basis of the classical t- or F-tests. The residual errors of these models ranged from 0.40 to 1.60 with a median of 1.03 and the adjusted R^2 from 0.24 to 0.83 with a median of 0.53. Again the results were consistent in that the effects had the same direction and were similar in magnitude.

The variable *family* occurred in all 16 submodels for which coefficients for *species* were available either as a main effect or in an interaction with lifestyle (Table 15). In general, Delphinidae had the biggest coefficients followed by the Monodontidae and the Phocoenidae. If lifestyle was included as a main effect species with a mixed pattern had the biggest coefficients followed by migratory and resident species. This pattern was different in the interactions for the Delphinidae where there was a decrease in coefficients from migratory to mixed and resident species.

The *social structure* was still included in 6 submodels but was actually never significant based on the classical F-test. All the same, the pattern was pretty consistent in that there was a decrease in size of coefficients from matrifocal to segregated to mixed and fission-fusion species.

Length of females was included in 12 submodels and coefficients of *species* decreased with increasing female length. At the same time the coefficients increased with increasing *length*

Table 15: Explanatory variables from the regressions of the coefficients from the variable *species*. #: Number of occurrences in the 20 submodels. Range: sign and range of values of coefficients. # models: number of models for which the coefficients for *species* were available.

response variable [a]			A		B		C		D		E		F		G		H		I		J	
diet variables [b]			a	o	a	o	a	o	a	o	a	o	a	o	a	o	a	o	a	o	a	o
	#	range																				
# models	16		3	3	3	3	2	3	2	3	3	2	3	0	3	0	3	0	3	3	2	0
Categorical variables:																						
family	11		3	2		1	2		2		3	2	3		3		3		3	3		
social structure	6		2		3		2		2	3							1					
lifestyle	6		2	2		1	2		2			2								3		
lifestyle:family	6			1	3	2		3		3											2	
fluidity	1																				1	
Continuous variables:																						
length of female	12	– 1.33–4.36	2		3		2	3	2	3	3		3				2		2	3	2	
dimorphism	9	+ 4.72–16.02	2	3	1	1	2		2	3		2									2	

[a] A: *average observed group size*, B: *median observed group size*, C: *standard deviation of observed group size*, D: *inter-quartile range of observed group size*, E: *average perceived group size*, F: *median perceived group size*, G: *standard deviation of perceived group size*, H: *inter-quartile range of perceived group size*, I: *maximum observed group size*, J: *minimum observed group size* [b] a: all; o: occurrence only; each with three possible split levels)

dimorphism (length of females/length of males) a variable that was included in 9 submodels.

Results from the stepwise backward-forward regression of the region effects

The sample sizes for these evaluations were given by the number of regions that were included in the main models which varied between 11 and 13, a low sample size especially in view of the number of possible explanatory variables.

In general, the step-wise backwards-forwards method based on the BIC did not work well for the regressions of the coefficients of *region* in that the BIC value was often larger than the one of the null model and in that many explanatory variables included in the models were not significant on the basis of the classical t- or F-tests. The residual errors of these models ranged from 0.25 to 1.55 with a median of 0.85 and the adjusted R^2 from –0.42 to 0.92 with a median of 0.31. The results were not consistent in that the effects of one explanatory variable were variable in direction and in magnitude (Table 16).

As could have been suspected from the fact that some of the variables on the physical environment were already included in the main models together with the *region*, the coefficients

Table 16: Explanatory variables from the regressions of the coefficients from the variable *region*. #: Number of occurrences in the 20 submodels. Range: sign and range of values of coefficients. # models: number of models for which the coefficients for *region* were available.

response variable [a]				A		B		C		D		E		F		G		H		I		J	
diet variables [b]				a	o	a	o	a	o	a	o	a	o	a	o	a	o	a	o	a	o	a	o
	#	range																					
# models	15			3	3	2	0	2	3	2	3	3	2	3	0	3	0	3	0	3	3	3	0
Continuous variables:																							
habitat	15	–	0.56–3.76	3	3	2		2	3	2	3	3	2	3		3		3		3	3	2	
structure	15	+	1.55–7.28	2	3	2		2	3	2	3	3	2	3		3		3		3	3	1	
latitude	12	–	0.01–0.05		3	2			3	2	3	3	2	3		3		3			3	3	
variab. temp.	3	+	47.15–105.72	2				2												3			
	6	–	25.41–242.80						3	2		3		3		2		3				2	
contour index	1	+	3.24																			2	
	11	–	1.01–6.03	2	3	2		2		1	3	3	2			3				3	3		
typical temp.	7	+	12.34–123.35	3	3			2			2					1				3	3		
	7	–	62.33–225.26						3	1		1		3		2		3				2	
typical depth	1	+	0.92																			2	
	11	–	0.40–1.56	2		2		2		1	3	3	2			3				3	3	1	
range of temp.	9	+	1.34–210.99			1			3	2	1	3		3		2		3				2	
	3	–	50.01–110.15	2				2												3			
range of depth	14	+	0.57–6.81	3	3	2		2	3	2	3	3	2			3		1		3	3	1	
	1	–	0.82																			2	

[a] A: *average observed group size*, B: *median observed group size*, C: *standard deviation of observed group size*, D: *inter-quartile range of observed group size*, E: *average perceived group size*, F: *median perceived group size*, G: *standard deviation of perceived group size*, H: *inter-quartile range of perceived group size*, I: *maximum observed group size*, J: *minimum observed group size* [b] a: all; o: occurrence only; each with three possible split levels)

of the regions are not well defined by the variables on the physical environment: some of the effects (e. g. from *habitat* and *structure*) were opposite to those in the main models and most explanatory variables were included as positive effects in some and as negative effects in other models and even within a sign showed a big variation in the size of their coefficients.

It should be kept in mind, that in these evaluations the explanatory variables were just the average values over regions and thus might not be a very good characterisation of these. Obviously, the influence of the *region* in the main model is due to the effect of variables that are not well captured with the ones on the physical environment included in the present evaluation and might be rather phylogenetic than environmental.

Conclusions

No detailed (biological) conclusions on these results are given here as they largely coincide with the interpretation of the results in the final evaluations (see the chapter on interspecific comparisons, page 39) though they are less concise here. In this initial evaluations there are several major drawbacks, though:

- It is not always clear, whether and how transformations should be conducted. Thus an attempt was made at using alternating conditional expectation, "ACE" (see below).
- High correlations between the response variables suggest that basically only one measure is available. Thus the different group size measures were subjected to a principal component analysis (PCA, see below). Additionally an attempt was made at using canonical correlation to model the set of group size variables in dependence of some of the explanatory variables (species and region, see below).
- Collinearity among explanatory variables. To avoid the arbitrary exclusion of some explanatory variables that are highly correlated with others and to avoid the conclusion that it is possible to differentiate between the influence of highly correlated explanatory variables, groups of explanatory variables were subjected to a PCA (see below).
- It is not satisfactory to rely solely on R^2 measures for goodness-of-fit judgements. More importantly one should consider the proportion of variability that can be explained given the variability of replicates (see below and the chapter on the interspecific comparisons of group size, page 36).
- The regions are possibly too large and thus too heterogenous, such that their coefficients cannot be subjected to a regression analysis based on the variables of the physical environment.

Additional problems were faced with the variables *minimum* and *maximum group size*. The number of observations is bigger for those populations that have a minimum group size of 1 compared with those where the minimum is bigger than 1 (Welch two sample t-test: $t = 8.40$, $df = 376.026$, $p \ll 0.001$). This indicated that if enough groups are observed there will almost always be one with just one member. On the other hand *maximum group size* increases with increasing number of observations (see above). As this seems to indicate that there is no (strong) selective pressure on short-time minimum and maximum group size, these group measures were not further evaluated in detail.

ACE: alternating conditional expectation

The aim of using ace was to get a better grip on how the variables should best be transformed such that the transformed variables can fit a linear model. Variables were plotted against the transformations suggested by the ACE procedure and inspected visually.

All sixty models that were used as starting points for the step-wise backwards forward selection in the interspecific comparison described above were subjected to an ace analysis. A qualitative summary of these sixty ace analyses is given in Table 17.

The conclusion of this analysis was, that group size measures seem to be reasonably transformed using the logarithm. Diet variables do not seem to need a transformation and the

Table 17: Summary of the used transformations and the ones suggested by the ace analyses.

variable	T[a]	T1[b]	T2[c]	T3[d]
group size measures	l	l,s	l,s	n(p)
species / region	n	n	–	n
habitat / habitat structure	n	n,l,s,p	n,l,p	n,l,p
latitude	n	p,b	n,p,b	n,p,b
typical temperature	l	n,l,p	l,p	p,b
temperature range	l	n,l,p	l,p	n
variability of temperature	l	b	p,b	n
typical depth	l	n,l,b	l,p	n,p,b
depth range	l	n,l,p	n,p	n(p)
depth contour index	l	b	l,p,b	p,b
occurrence of food categories	n	n	n	n
frequency of food categories	n	n(p,b)	n,p	n,p
proportion of food categories	n	n	n	n

[a] transformation used in preliminary models, PCAs and final models; n: none, l: log, s: square root, p: polynomial (quadratic), b: bizarre, i. e. polynomial of high order or non-smooth [b] transformations suggested by ace [c] transformations suggested by ace if species and region were not included in the models [d] transformations suggested by ace if transformation under T was used

variables of the physical environment seem to be related to group size in a possibly non-linear way.

This may be less of an issue in the final main model of the interspecific comparisons, as the main variable there is the categorical variable *species* and the variables of the physical environment are not used individually but after subjecting them to a PCA (see below).

An ace analysis was also conducted for the evaluations of the coefficients of *species* and *region* with basically the same conclusions.

Natural variability: non-robust approach

In a preliminary attempt to judge how well the models fitted the data, the standard deviation of all group size measures was plotted for a given region against the number of studies of this species in that region and compared with the residual error of the regression resulting from the step-wise backward forward procedure and the residual errors of some of the analyses of variance that were calculated (see below; for the robust analogon see page 36).

Result: Residual errors are very similar to the variability in the group measures for those species-region combinations where at least 5 observations were available. Thus the variability

among replicates was similar to the residual error.

Conclusion: Little of the variability additional to the one among replicates remains unaccounted for and no (gross) overfitting is done.

Simpler Models

Several simple models were conducted to see how much of the variability of group size they can explain and how many variables are needed to explain the variability in the data.

ANOVA: group size versus species

All ten group size measures were subjected to a univariate analysis of variance with the factor species. Graphical checks of residuals were mostly satisfying (regarding outliers and heteroscedascity see the chapters below on robustness and Levene test, respectively), error variance was only slightly higher than the ones from the step-wise backward forward procedure and estimated species coefficients correlated highly with the ones from the preliminary step-wise backward forward procedure.

nested ANOVA: group size versus species within family

An attempt was made to calculate a nested ANOVA with the factor *species* nested in *family*. The data set was too unbalanced (two, six and thirty-one species per family) though to yield satisfactory results.

ANOVA: group size versus species and region

All ten group size measures were subjected to a bivariate analysis of variance including the main effects of the factors *species* and *region*. Graphical checks of residuals were mostly satisfying (regarding outliers and heteroscedascity see the chapters below on robustness and Levene test, respectively), error variance was very slightly higher than the ones from the step-wise backward forward procedure and slightly lower than the ones from the univariate model using species, and estimated species coefficients correlated highly with the ones from the preliminary step-wise backward forward procedure.

Coefficients of *species* and *region* from this analysis were evaluated for their dependence from diet and life-history variables and from the variables of the physical environment, respectively. A model could be built which explained the size of the species coefficients reasonably (and with similar interpretation as in the final model). The coefficients of the *region* could not be modelled by the variables of the physical environment (this could be seen e. g. in that the Tukey-Anscombe plot was made up of almost horizontal lines, i. e. the estimated values for a given coefficient spread over almost the whole range of estimated values). This makes sense in that the regions as defined are large and thus cover a broad range of the values of the variables of the physical environment. From this, it was concluded, that the variable *region* does not reflect the physical environment but can possibly reflect phylogenetic similarity of populations occuring in the same rather than different regions.

Regression: group size versus all variables except species and region

All ten group size measures were subjected to a multiple linear regression including all variables used in the step-wise backward forward procedure except for the two categorical variables species and region. Graphical checks of residuals revealed some change in expected residual value depending on fitted values and some structure in the residual versus explanatory variables and residual plus component effect versus explanatory variables plots. Most residual errors were a bit larger than those of the models including species only.

Conclusions from the simpler models

Species seems to be the single most important variable and also leads to models that follow the statistical assumptions best. *Species* are so unequally distributed over *families*, that a nested procedure is difficult, because the data set is too unbalanced, and thus a two-step evaluation (a main model including *species* and a further model of the species coefficients including *family* seems appropriate). The dependence of the groups size on the continuous variables seems to be highly non-linear or even non-smooth but can be captured in the variables species (and region) in a linear model.

Even with this knowledge it is not clear what the best set or combinations of variables would be. That is why in the final interspecific comparisons allsubsets and allsubsets for a given number of variables were calculated.

Principal component analysis (PCA)

In the present section attempts at principal component analyses not reported in the chapter on interspecific comparison are presented. Variables were log transformed according to the list in Appendix A before being subjected to a PCA.

In general PCA was used as a descriptive tool with the aim of explaining as much variability in as few variables as possible. Additionally, PCs were preferred that allowed an interpretation in biological terms. No formal tests were conducted on whether additional PCs explained "significantly" more of the variability or not.

Group size measures

The first attempt at PCA for the group size measures included all ten available variables (Table 18, top section). One can see that the minimum group size has a loading lower than any other measure in the first PC and has the highest loading in the second PC. As it was known already that minimum group size behaved differently from the other variables (see above) it was excluded from the PCA (see the final PCA reported on page 30).

As it was also known that maximum group size behaved specially, an attempt was made to exclude both, the minimum and the maximum group size from the PCA (Table 18, middle section). As the results were very similar to those that included the maximum group size an exclusion seemed not necessary and a (slight) loss of information.

It can be seen, that in the two previous PCAs the second PCs tend to be a contrast between the location and the variability variables, though not very clearly so. An attempt was made to emphasise this tendency in that the variation variables were exchanged with variables measuring the relative variability (the logarithm of the variability measures minus the logarithm of the location measures). Though the first PC from this analysis was a measure of location mainly

Table 18: Principal component loadings for standardised group size measures.

	PC1	PC2	PC3
ALL GROUP SIZE VARIABLES			
cumulative explained variance	0.86	0.95	0.98
mean observed group size	–0.33	0.16	0.18
standard deviation of observed group size	–0.33	–0.12	0.04
median observed group size	–0.30	0.40	0.40
inter-quartile range of observed group size	–0.32	0.15	0.51
minimum group size	–0.22	0.72	–0.62
maximum group size	–0.32	–0.24	–0.09
mean perceived group size	–0.34	–0.15	–0.08
standard deviation of perceived group size	–0.32	–0.29	–0.18
median perceived group size	–0.33	–0.08	–0.00
inter-quartile range of perceived group size	–0.32	–0.30	–0.33
EXCEPT MINIMUM AND MAXIMUM GROUP SIZE			
cumulative explained variance	0.91	0.97	0.98
mean observed group size	–0.36	0.27	0.19
standard deviation of observed group size	–0.36	–0.12	–0.10
median observed group size	–0.33	0.64	0.26
inter-quartile range of observed group size	–0.35	0.35	–0.74
mean perceived group size	–0.37	–0.18	0.15
standard deviation of perceived group size	–0.35	–0.38	–0.25
median perceived group size	–0.36	–0.08	0.49
inter-quartile range of perceived group size	–0.35	–0.45	–0.01
RELATIVE VARIABILITY			
cumulative explained variance	0.53	0.82	0.90
mean observed group size	0.45	–0.23	–0.13
median observed group size	0.38	–0.35	–0.28
mean perceived group size	0.48	–0.06	–0.09
median perceived group size	0.46	–0.14	0.04
relative standard deviation of observed group size	0.28	0.41	0.20
relative inter-quartile range of observed group size	0.30	0.22	0.75
relative standard deviation of perceived group size	0.16	0.55	–0.22
relative inter-quartile range of perceived group size	0.12	0.53	–0.49

and the second a contrast between the location and the relative variability less of the variability can be explained with the same number of PCs (Table 18, bottom section).

Additionally, PCAs were run using only the location measures of the group size (using all such variables, excluding minimum group size and excluding both, minimum and maximum group size) and using only the variability measures separately. Whereas the first PCs from these evaluations always explained over 90 % of the variability, the PCs from the location measures still correlated with an $r = 0.96$ with those from the variability measures and thus a separate treatment of these two sets of variables did not seem justified.

Variables of the physical environment

Using all variables of the physical environment as is, resulted in a PCA that still needed quite a few variables to explain a reasonable amount of variability (five variables to reach 90 %) and the PCs were difficult to explain. Thus, *latitude* was recoded from a scale of 0 to 180 (from the South to the North Pole) to the usual scale of degrees from the equator.

After this transformation the first few PCs accounted for a somewhat higher proportion of the variability but were still difficult to explain. The exclusion of the *variability of temperature* and the *depth contour index* lead to PCs that were quite easily interpreted and explained even a bit more of the variability (see chapter on interspecific comparisons, page 30).

Diet variables

The main aim of the PCA on the diet variables was to reduce a high number of partly correlated variables. Thus the interpretation was less important for these variables and the initial PCAs were used in the final evaluations (see the chapter on interspecific comparisons, page 30).

Robustness

Even though plots of residuals looked mostly fine, an attempt was made to see whether there were outliers in conducting a robust linear model with species and region as the explanatory variables for each of the group size measures.

In plots of the residuals of these analyses there were no extreme outliers but some suspect data points. Also, the variability of the residuals for the different levels of the two factors seemed quite variable (see below in the section on the Levene test). Residual errors were the smallest of all models conducted thus far.

There was no clear biological problem with the data points that were at the edge of the distribution. In one case the problem might have been that the calculated values only based on very few observations.

If one looked at the residuals in more detail though, more data points lay in the extreme tails of the residual distribution than would have been expected under the assumption of normality:

2.4 to 11.3 % (median: 5.0 %) of the residuals were bigger than 1.96 * estimated deviation of error (expected: 2.5 %).

2.8 to 6.3 % (median: 3.6 %) of of the residuals were smaller than -1.96 * estimated deviation of error (expected: 2.5 %).

0.8 to 4.9 % (median: 2.8 %) of the residuals were bigger than 2.58 * estimated deviation of error (expected: 0.5 %).

Table 19: Loadings for the group size measures (response variable) in a canonical correlation versus species and region.

	loading 1	loading 2	loading 3
GROUP SIZE VARIABLES ALSO USED IN THE FINAL PCA			
correlations	0.878	0.667	0.589
mean observed group size	0.058	0.173	0.003
median observed group size	–0.007	–0.145	0.056
mean perceived group size	–0.017	0.085	0.129
median perceived group size	0.005	0.006	0.028
maximum group size	–0.008	0.005	–0.037
standard deviation of observed group size	0.017	–0.014	–0.049
inter-quartile range of observed group size	–0.001	–0.051	–0.058
standard deviation of perceived group size	0.004	–0.030	–0.022
inter-quartile range of perceived group size	–0.008	–0.039	–0.043
NORMALISED GROUP SIZE VARIABLES			
correlation	0.878	0.667	0.589
mean observed group size	0.081	0.239	0.004
median observed group size	–0.009	–0.189	0.073
mean perceived group size	–0.027	0.137	0.208
median perceived group size	0.009	0.010	0.047
maximum group size	–0.013	0.008	–0.060
standard deviation of observed group size	0.029	–0.023	–0.080
inter-quartile range of observed group size	–0.002	–0.080	–0.091
standard deviation of perceived group size	0.007	–0.053	–0.039
inter-quartile range of perceived group size	–0.015	–0.072	–0.080

0.4 to 2.4 % (median: 1.55 %) of the residuals were smaller than -2.58 * estimated deviation of error (expected: 0.5 %).

Conclusion: There are some outliers and thus robust methods should be applied (see the final evaluations of both the inter- and the intraspecific comparisons).

Canonical correlations

To see whether the group size measures are combined similarly in the context of species and region as with the PCA, a canonical correlation was conducted with the group size measures

(except *minimum group size*) as the response and the species and region as the explanatory variables (Table 19, top). Additionally, the same analysis was conducted starting from the normalised group size measures (Table 19, bottom).

In both canonical correlation analyses the emerging pattern is very similar but rather difficult to understand. All loadings are very small. The first set of loadings is difficult to interpret and the second and third seem to be slightly different versions of contrasts between location and variability measurements but not purely so. It might be possible that this result is due to the fact that the explanatory variables only consist of a set of 47 dummy variables.

Levene test

As there was some suspicion that residuals might be heteroscedastic over the levels of the factors *species* and *region* a robust linear model was calculated with mean observed group size or the first PC of group size (see above and the chapter on the interspecific comparison, page 33) as the response (three outliers were excluded in this analysis) and species and region as the explanatory variable to make preliminary estimates of homo-/heteroscedascity.

In the model describing average observed group size heteroscedascity was indicated for the factor species (Levene-test $p = 0.04$) but not for region ($p = 0.11$). If only those species with more than three observations were included in the Levene test (26 instead of 39 species) there was no indication of heteroscedascity ($p = 0.45$).

A similar picture was visible if the first PC of group size was the response variable. Species and region showed some heteroscedascity (Levene $p = 0.01$ for both variables). Again, if only those species or regions were included that had more than three observations (16 instead of 36 species and 12 instead of 13 regions), there was no indication of strong heteroscedascity ($p = 0.96$ and 0.09 for species and region, respectively).

Conclusion: Considering the number of observations per species and/or per region there was no indication of heteroscedascity (see also below for the Levene tests of the final model).

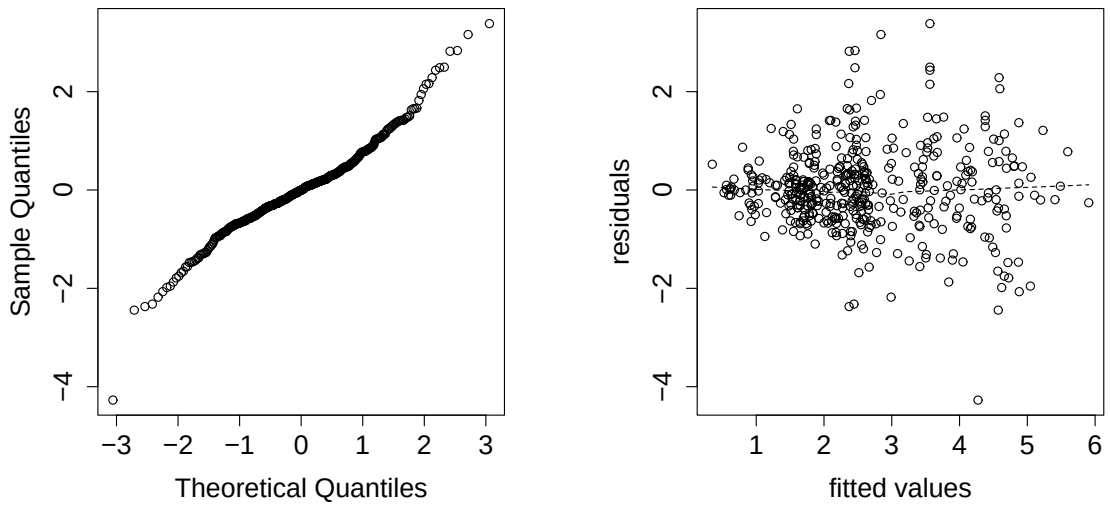


Figure 13: Quantile-quantile and Tukey-Anscombe plots of the residuals from the main interspecific model. Dashed line is a lowess smoother.

Statistical details of the interspecific comparison

Macroscopic view: Residuals

The details in this section are on the model given in formula (1), i. e. the model explaining the average group size with the variable species and the PCs of the physical environment.

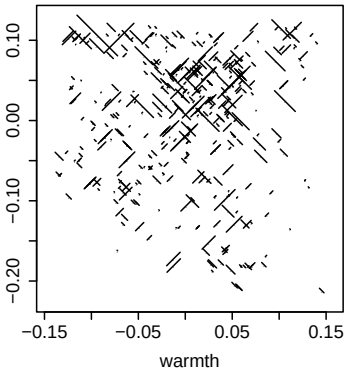
Quantile-quantile plots (Fig. 13) In the quantile-quantile plots the sample quantiles of the residuals are plotted against the theoretical quantiles. If the resiudals are normally distributed they lie on a straight line. Outliers can be seen as well as shapes that imply a necessary transformation of the (response) variable(s). The quantile-quantile plot of the final main model shows one outlier that should be well handled by the robust methods. (Fig. 13).

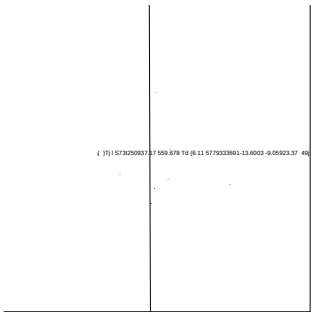
Tukey-Anscombe plots (Fig. 13) Residuals are plotted against the fitted values in the Tukey-Anscombe plot. Variability in the positive direction is expected to be similar to the one in the negative direction. The variability of the residuals is also supposed to remain constant over the range of fitted values and the expected value for the residuals over the whole range (visualised by a smoother) is expected to be zero. Except for the outliers mentioned in the last section the Tukey-Anscombe plot looks pretty much as expected.

Residuals versus explanatory variables (Fig. 14) Plotting the residuals versus the explanatory variables is analogous to the Tukey-Anscombe plot. The same points are expected, i. e. no structure in the residuals is supposed to be visible. The residual versus explanatory variables plots look quite random for the main model and the expected value visualised by a lowess smoother follows the zero line quite closely. The variability of the residuals is somewhat smaller in regions of the explanatory variables where fewer data points are available. Some structure seems to be present in the plot of the squared effect of openness and the two plots of the specificity. These effects would be included in the model if one followed the allsubset analysis, i. e. including these variables would lead to a model with a smaller residual error. But the inclusion of these variables does not lead to a statistically significant improvement of the model.

Residuals plus component effect versus explanatory variables (Fig. 4) These plots are similar to the residuals versus explanatory variables plots. Some deviations from a straight line (in a linear model) are seen more easily in these plots, especially if one adds the modelled line plus a smoother to the plot. I. e. if the dependence between the response and the explanatory variables is not linear the shape of the dependence can be seen in these plots. As suggested the model prediction and a lowess smoother are added to these plots (Fig. 4). The two lines are almost identical and thus there is no apparent unmodelled shape of dependence hidden in the residuals.

Residual-interaction plots (Fig. 15) In the Residual-interaction plots residuals are plotted against two (continuous) variables. Their sign is visualised by a positive or negative slope of one and their size by the length of the dash. No clusters of identical sign and big size are supposed to occur, otherwise it is likely that existing two-way interactions are not considered. As an example the residual-interaction plots are shown for openness against all other PCs of the physical environment. No pattern(s) are apparent from these plots (Fig. 15).


0.10
0.00
−0.10
−0.20
−0.15
−0.05
0.05
0.15
warmth



Explaining the coefficients of species

In a further evaluation of the dependence of the species coefficients on the *family*, *social structure*, the *residency pattern*, *length*, *length dimorphism* and the PCs of the diet variables,the former dependent variable was replaced with the estimated species effect. As the shape of dependence in this model is heavily influenced by the different number of cases for the different species this approach has later been abandoned because it is not biologically appropriate. Nevertheless, there were some interesting methodological observations connected with this evaluation. That is why it is presented here.

Model selection

This evaluation was conducted three times for each of the sets of the diet PCs. As results of the allsubset evaluations suggested similar models, only the results of the prey occurrence PCs were inspected more closely and presented because they are based on the largest data set (N= 477 instead of 198). All continuous explanatory variables, (*length* and *length dimorphism* and the PCs of the diet variables), were included linearly and quadratically, as an ace analysis suggested, that the influence of (some of) these variables might be more complex than linear.

It was not possible to conduct a complete all subsets analysis in this case, because there were too many possible combinations in each of the models (over half a million). Instead, all subsets of a given number of explanatory variables were conducted for zero to four variables. In the sets of models including one, two and four variables the best model always had a clearly lower residual error than the other models. The best model in each step also reflected a continuous step in the inclusion of variables. Thus, the model with the lowest residual error and four variables was taken as a starting point. The inclusions of all these variables were significant and thus the procedure had to be continued. This was done in a kind of step-wise procedure. All models were calculated that included one more variable and their residual errors were compared. If the model with the lowest error was clearly better than the others it was accepted. If it was only the best without being clearly so, all possible additions of two variables were evaluated for the last accepted model. This procedure was followed until the addition of a variable did not yield a significant improvement of the model using robust analysis of variance (Table 20).

No comparison of the residual error with the "natural variability" is possible here, as no estimate of the latter is available (there are no replicates showing variability).

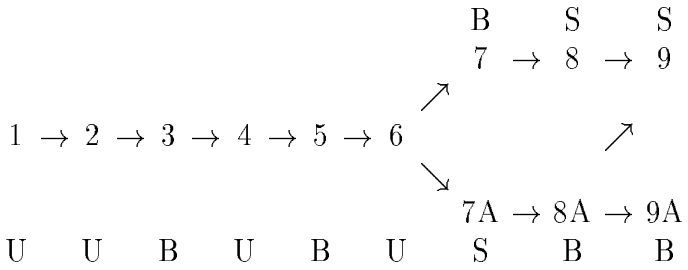
Residuals

The distribution of the model on the species coefficients deviates some more from the expected straight line in a qq-plot (Fig.16, left) and thus the results from this evaluation should be interpreted more cautiously. The flat part in the middle section of this curve might be due to cases that are fitted well in the model and the cases that lie at both ends are not fitted equally well. This may be due to the fact that the number of cases is unbalanced between the different species, so that the species with large numbers of cases are well fitted but the species with few number of cases do not fit the model equally well.

In the Tukey-Anscombe plot for the model on the species coefficients straight lines are visible which are due to the fact that all cases of one species have the same value of the response variable in this model (Fig. 16, right) and variability increases somewhat with increasing fitted values.

If residuals are plotted against the explanatory variable family some heteroscedasticity is visible (Fig. 17 top, see also below). This is not obvious if residuals are plotted against the

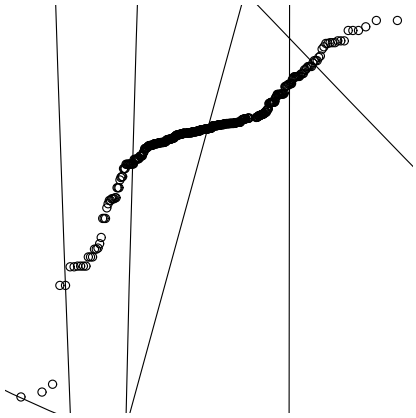
Table 20: ANOVA table comparing models for the species effect with the lowest residual error for a given number of variables (for further information on the procedure see text). The sequence of models is given by (U: unique best model, i. e. one for which the residual error is clearly lower as for the others with the same number of variables, B: best model for a given number of variables, S: included such that a continuous inclusion of additional variables is possible):



row [a]	included variables	cpt [b]	rdf [c]	df [d]	Wald	p-value
0	intercept		476			
1	+ social structure	0	472	4	9682.06	$\ll 0.001$
2	+ family	1	470	2	25935.89	$\ll 0.001$
3	+ (length dimorphism)^2	2	469	1	8468.48	$\ll 0.001$
4	+ length dimorphism	3	468	1	54.44	$\ll 0.001$
5	+ PC5	4	467	1	14.26	< 0.001
6	+ PC1	5	466	1	35.83	$\ll 0.001$
7	+ (PC5)^2	6	465	1	6.47	0.011
7A	+ PC4	6	465	1	3.77	0.052
8	+ PC4	7	464	1	6.85	0.009
8A	+ (PC6)^2	7A	464	1	0.99	0.319
9	+ (PC6)^2	8	363	1	1.93	0.164
9A	+ (PC1)^2	8A	363	1	1.71	0.191
9	+ (PC5)^2	8A	363	1	2.76	0.097

[a] also equal to number of variables added to intercept [b] compared to row no.
[c] residual degree of freedom [d] degree of freedom

continuous variables. No structure is clearly visible in the plots of the residuals plus component effects against the variables nor for the plots of residuals against two variables, either (not shown).



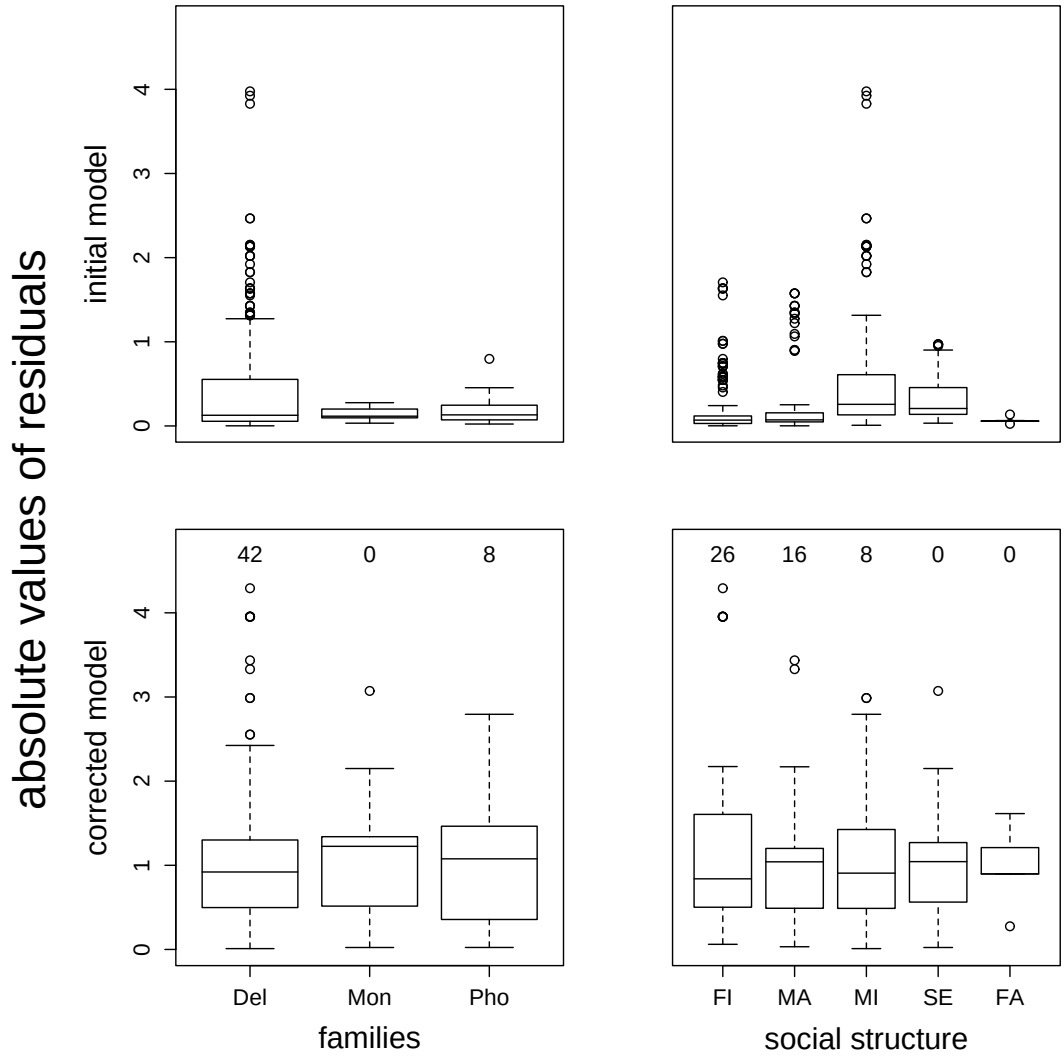


Figure 17: Absolute values of residuals from the model on the species coefficients split according to family (DELphinidae, MONodontidae, PHOcoenidae) and social structure (FIssion-fusion, MAtrifocal, MIxed, SEgregated, FAmilial) in the initial (heteroscedastic) model and the corrected (homoscedastic) model. In the corrected model the number of observations with absolute residuals larger than 4.5 is indicated (these points are not plotted).

where E_X is the effect of X, ssgs the predicted species specific group size, D the length dimorphism and the PCs the principal components of the diet variables. If parameters are normalised (see statistical methods on page 29) one gets:

$$\begin{aligned} \text{ssgs} = e^{\text{se}} &= 6.43 \cdot e^{E_{\text{family}}} \cdot e^{E_{\text{social structure}}} \\ &\cdot e^{-0.17D+0.10D^2} \cdot e^{0.11PC1} \cdot e^{-0.01PC4} \cdot e^{0.04PC5-0.03PC5^2}. \end{aligned}$$

Delphinidae have on average larger groups than Phocoenidae and Monodontidae (the factors that go into formula 4 are 1.00, 0.09 and 0.05 for these three families, respectively). The factors for the social structure are 1.00 for fission fusion, 0.42 for matrifocal, 4.10 for mixed, 2.66 for segregated and 0.34 for familial groups. Familial groups include only few individuals in Delphinoidea and thus this factor level also reflects group size. There are only 7 cases of groups organised as families, though and even if they were included in the analysis as matrifocal, the

most similar social organisation, the model would not change.

The *family* reflects the common evolutionary heritage of the different species resulting from evolutionary processes, that took place before the radiation, and can account for a twenty fold increase of group size and the *social structure* for a twelve fold increase, whereas the continuous variables *length dimorphism*, and the first, fourth and fifth principal components of the diet variables can only account for a 3, 1.6, 1.04 and 1.3 fold change in expected group size, respectively.

Correcting for the heteroscedasticity the estimates of the parameters are:

$$\begin{aligned} \text{ssgs} = e^{\text{se}} &= 1.67 \cdot e^{E_{\text{family}}} \cdot e^{E_{\text{social structure}}} \\ &\cdot e^{-4.10D-0.36D^2} \cdot e^{0.73PC1} \cdot e^{0.08PC4} \cdot e^{0.31PC5+1.70PC5^2}, \end{aligned} \tag{5}$$

and if normalised:

$$\begin{aligned} \text{ssgs} = e^{\text{se}} &= 5.32 \cdot e^{E_{\text{family}}} \cdot e^{E_{\text{social structure}}} \\ &\cdot e^{-0.43D-0.01D^2} \cdot e^{0.08PC1} \cdot e^{0.01PC4} \cdot e^{0.02PC5+0.01PC5^2}. \end{aligned}$$

The effects estimated in the latter two models slightly differ from the original model (formula 5 versus formula 4). The estimated influence of the family and social structure is very similar though (only the importance of mixed and segregated is approximately reversed). The differences are mainly in the quadratic effects and in the diet variables which are of little explanatory power, i. e. they are of little biological relevance compared to the other variables.

Explaining the coefficients of species: Final model

These details on the residuals concern the model that was formulated in formula 3. In the qq-plots (Fig. 18, left) some single outliers are visible for the different models using the different categorisations of subfamily. These are not problematic as a robust regression method was used. Variability of the residuals in the different subfamilies looks very similar (Fig. 18, right). It looks most variable in the model where the species coefficients were explained by family and residency pattern, but here the Levene was not significant. This second series of plots can also represent the Tukey-Anscombe plot and the component plus residual plot because only one (or two) factor variable(s) were included in these models.

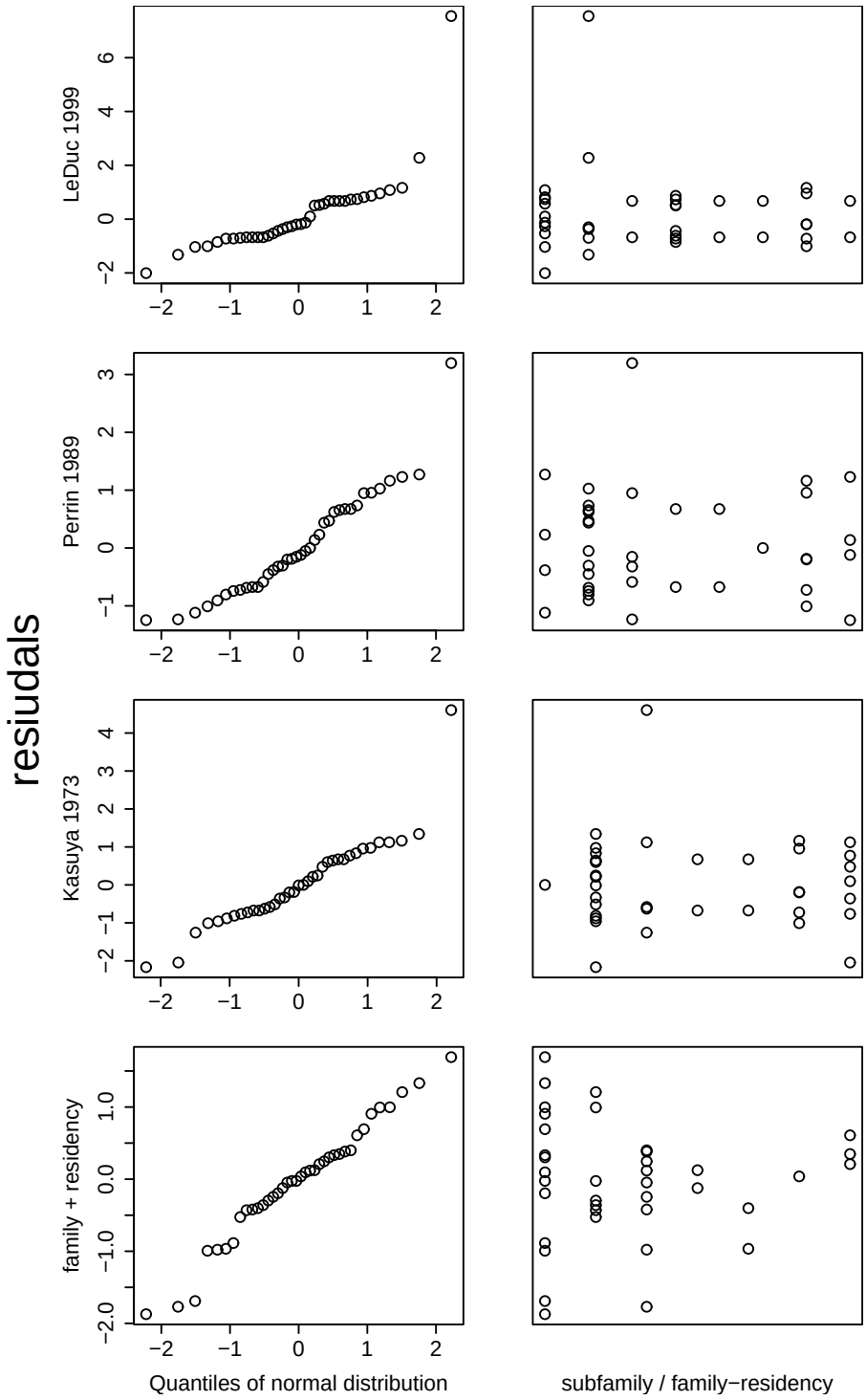


Figure 18: QQ-plots (left) and residuals versus subfamily categories (right). For the model including family and residency pattern residuals are plotted against all combinations of these two factors.

History of the intraspecific comparison

Stepwise backward-forward regression

Variables

The following measures of group size were considered as response variables: *mean observed group size*, *standard deviation of observed group size*, *median observed group size*, the *inter–quartile range of observed group size*, *mean perceived group size*, *standard deviation of perceived group size*, *median perceived group size*, *inter–quartile range of perceived group size*, *minimum group size* and *maximum group size*.

As data sets were rather small and because there was only one value per species for most life-history variables (especially for those regarding maturation and reproduction) only three of these variables were included as explanatory variables in the analysis: *lifestyle*, *length of females* and *length dimorphism*.

To describe the physical environment the variables *habitat*, *habitat structure*, *latitude*, *typical temperature*, *variability of temperature*, *typical depth* and *depth contour index* were included as explanatory variables. Additionally the variables on the occurrence of the food categories cephalopods, fish and mammals & birds were included. No finer split of these variables was included in these evaluations as sample sizes (number of cases) were rather small.

Methods

Basic multiple linear and logistic regressions were used (methods "lm" and "glm"). Variables were log transformed where necessary (basically all response variables and all continuous explanatory variables except *habitat*, *habitat structure* and *latitude*; see also Appendix A). For a given regression problem a backward-forward stepwise procedure was conducted using the BIC criterion.

Model assumptions were checked using graphical methods (see above for the details on which plots were used). The sample size for a given regression was restricted to the cases where all the necessary variables were known.

Each of the ten measures of group size was used as the response variable in separate regressions. As these measures of group size were highly correlated (see below and chapter on interspecific comparison, page 30), one can view these models together as a kind of robustness/sensitivity analysis. The explanatory variables that are important in many of the models seem to explain the variability and patterns in the response variable most reliably.

Minimum group size could not be subjected to a standard regression analysis in all species because many of these observations were one and thus residuals were no longer normally distributed. In these cases a logistic regression was used on whether minimal group size equalled to one or not.

In principal a linear main effects model is considered. Non-linear (e. g. quadratic) effects were detectable with the RS-plots (residuals plus component effects versus explanatory variables; some such effects seemed to occur but did not reach statistical significance). As residuals were plotted against all combinations of two possible explanatory variables (residual-interaction plots), at least two–way–interactions should have been possible to detect.

Table 21: Overview of the directions (+, –) of the effects from the regression models of group size for all species. Oo = Orcinus orca, Gs = Globicephala spp., Gg = Grampus griseus, Tt = Tursiops truncatus, Dd = Delphinus delphis, Sc = Stenella coeruleoalba, Pp = Phocoena phocoena, Pd = Phocoenoides dalli.

	Oo	Gs	Gg	Tt	Dd	Sc	Pp	Pd
length of females			–	–	+	–	+	
length dimorphism			+		–			
habitat	+	–		+				
habitat structure			+	–	–	+		
latitude	–		+		+	+	–	
variability of temperature		+			+			
contour index			–	–	–	+	+	+
typical temperature	–	–	+	+	–			
typical depth	+	–	+		+	+	+	
occurrence of cephalopods	+				–		+	
occurrence of fish	+	+						
lifestyle	X			X				

Results

Group size seems not to be determined by the same explanatory variables in all the eight species that were considered in the present evaluations and thus some results are presented by species.

Killer whale (*Orcinus orca*): The measures of group size were highly correlated for the killer whale ranging from 0.25 to 0.98 with a median of 0.81 (excluding minimal group size). All models have a satisfactory number of cases (55–75) and model selection seems reasonable in that the BIC values are all lower than the one of the null model and lower or equal to the one of the full model and models are also significant on the basis of the classical F-values though the R^2 values are only moderately high (0.17–0.49).

Group size in killer whales decreases with *typical temperature* and less clearly with *latitude*. It increases with *habitat*, *typical depth* and the occurrence of *cephalopods* and *fish* in the diet. Additionally, resident populations have a bigger group size than migratory populations or those with a mixed pattern (Table 21).

Pilot whales (*Globicephala spp.*): The group size measures correlated between 0.39 and 0.99 with a median of 0.82 (excluding minimal group size) in pilot whales. Though sample size is rather small (13–25) the variable selection is reasonable in that the BIC values are smaller than the ones of the null models and smaller or equal to the ones of the full model and classical F-values reach significance (with one exception). R^2 values are variable but mostly

high (0.13–0.98).

Group size increases in pilot whales with *variability in temperature* and the occurrence of *fish* in the diet and decreases with *habitat*, *typical temperature* and *typical depth* (Table 21).

Risso's dolphin (*Grampus griseus*): Correlations of measures of group size correlate between 0.07 and 0.99 with a median of 0.84 for Risso's dolphin (excluding minimal group size). Sample size is small (11–27) and model selection based on BIC might be more problematic in these cases because two BIC values are actually bigger than the ones of the null model and most p-values based on the classical F-test are not significant. It is not so surprising then that the R^2 values are rather low (0.12–0.83) though the influence of the variables included in the evaluations are pretty consistent.

Group size in Risso's dolphin decreases with *length* and *contour index* and increases with *length dimorphism*, *structure*, *latitude*, *typical temperature* and *typical depth* (Table 21).

Bottlenose dolphin (*Tursiops truncatus*): Measures of group size are highly correlated in the bottlenose dolphin ranging from 0.46 to 0.98 with a median of 0.86 (excluding minimal group size). Sample size in these models is adequate (33–77) and model selection based on BIC seems to be very reasonable in that the BIC values are all lower than the ones from the null and the full model and all classical F-tests are significant. R^2 values are consistent but not very high (0.12–0.68).

Group size in bottlenose dolphin decreases with *length of females*, *habitat structure* and *contour index* and increases with *habitat* and *typical temperature*. Additionally, both resident and migratory populations seem to have bigger groups than those with a mixed or unknown pattern (Table 21).

Common dolphin (*Delphinus delphis*): Group size measures correlated highly between 0.58 and 0.95 with a median of 0.91 for the common dolphin (excluding minimal group size). Sample size is rather small and thus four BIC values are equal to the one of the full model whereas the others are smaller than both the one from the null and the one from the full model. But still, all (except one) classical F-tests are significant. The R^2 values are very high (0.25–0.99).

Group size of common dolphin increases with *length*, *latitude*, *variability of temperature* and *typical depth* whereas it decreases with *length dimorphism*, *habitat structure*, *contour index*, *typical temperature* and the occurrence of *cephalopods in the diet* (Table 21).

Striped dolphin (*Stenella coeruleoalba*): All group measures are highly correlated in the striped dolphin ranging from 0.66 to 0.99 with a median of 0.97 (excluding minimal group size). Sample sizes are rather small (13–24) but variable selection based on the BIC seems reasonable in that the BIC values are smaller than those of the null models and smaller (or, in two cases, equal) to the value of the full model and all classical F-tests are also significant. The R^2 values are extremely high (0.71–1.00).

In the striped dolphin group size decreases with the *length* and increases with *habitat structure*, *latitude*, *contour index* and *typical depth* (Table 21).

Harbour porpoise (*Phocoena phocoena*): Measures of group size correlate between 0.29 and 0.98 with a median of 0.80 for the harbour porpoise (excluding minimal group size). Sample

size is small (13–22) and BIC values are all smaller than the ones from the null model but four are only equal to those of the full model and thus there was no reduction in explanatory variables. In the same cases the classical F-tests were not significant, either. Still, R^2 values are moderate to high (0.25–1.00).

The interpretation of the effects of the explanatory variables is more difficult here, as the effects seem to be either large and statistically unsignificant or small and highly significant but in the other direction. Nevertheless, it seems that group size in harbour porpoise increases with *length*, *contour index*, *typical depth* and the occurrence of *cephalopods* in the diet and decreases with *latitude* (Table 21).

Dall's porpoise (*Phocoenoides dalli*): In Dall's porpoise the measures of group size correlate between 0.42 and 0.99 with a median of 0.85 (excluding minimal group size). Sample size is moderate (24–31) and the BIC values are minimal in six cases whereas in the other four cases the null model resulted in the minimal BIC. Also the classical F-tests reach (almost) significance. The R^2 are rather low (0.08–0.36).

Group size in Dall's porpoise increases with *contour index* (Table 21).

Conclusions

Again, the interpretation of the results from these preliminary analyses are very similar to the final model (see the section on the intraspecific comparison, page 50) but had some serious drawbacks:

- It is difficult to compare the models of the different species as they do not include the same variables. Thus a set of variables should be defined that is included in the models of all species.
- As in the preliminary evaluations of the interspecific comparison it is not satisfactory to have highly correlated response variables in different models and highly correlated explanatory variables (collinearity) in the same model. Thus the same PCAs were used for the intraspecific evaluation as for the interspecific evaluation (see above and the chapter on interspecific comparisons, page 30).

Statistical details of the intraspecific comparison

Residuals

For the explanation of the different plots see the section on residuals from the interspecific comparison, page 109.

Quantile-quantile plots (Fig. 19)

Common dolphin and Risso's dolphin both show quite a few outliers (six and five data points, respectively) and striped dolphins and pilot whales have each one data point that seems to be an outlier. The models for the other species seem to yield quite normally distributed residuals. In this light it seems highly justified to use robust methods for this intraspecific comparison.

Tukey-Anscombe plots (Fig. 20)

The same outliers are visible in the Tukey-Anscombe plots. Otherwise the variablity seems to be similar in the positive and negative direction and more or less constant over the range of the fitted values. The smoother indicates that in general expected values are close to zero.

Residuals versus explanatory variables

No obvious structure is visible in the residuals except for the outliers already mentioned above.

Residuals plus component effect versus explanatory variables (Fig. 5)

Almost all of the modelled effects coincide exactly with the smoother in these plots. Thus the shape of the dependence seems to be modelled well. Again the outliers are visible.

Residual-interaction plots

No clusters of big residuals of the same sign seem to occur. Thus, there is no (obvious) indication that existing two-way interactions were omitted in the models.

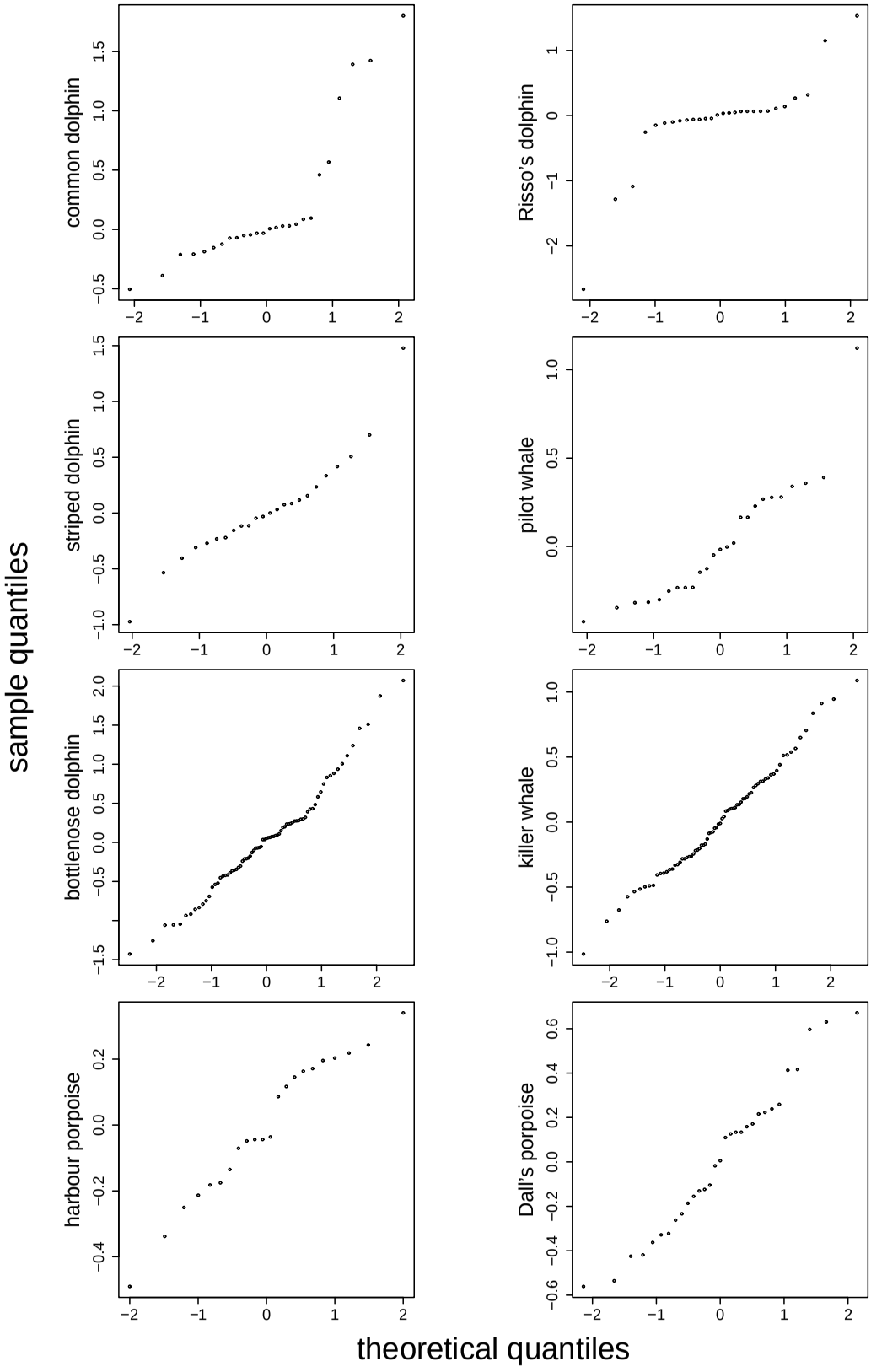
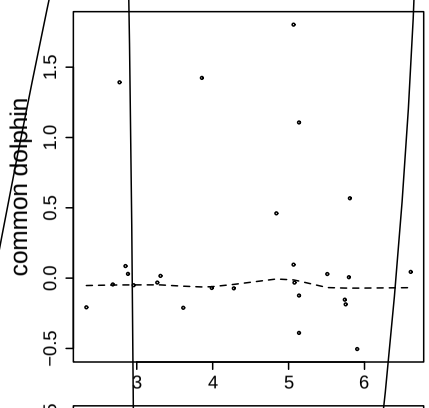
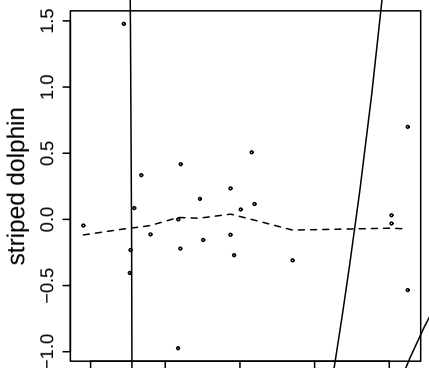


Figure 19: Quantile-quantile plots of the resiudals from the intraspecific models.


common dolphin
1.5
1.0
0.5
0.0
−0.5
3
4
5
6


striped dolphin
1.5
1.0
0.5
0.0
−0.5
−1.0
2
3
4
5
6

History of the classification

Variables

The following variables were used in the classification attempts of the Delphinoidea species: the group size measures *mean observed group size*, *standard deviation of observed group size*, *median observed group size*, the *inter–quartile range of observed group size*, *minimum group size*, *mean perceived group size*, *standard deviation of perceived group size*, *median perceived group size*, the *inter–quartile range of perceived group size* and *maximum group size*; the variables of the physical environment *habitat*, *habitat structure*, *latitude*, *typical temperature*, *range of temperature*, *variability of temperature*, *typical depth*, *range of depths* and *depth contour index*; the life history variables *length of females*, *length dimorphism*, *seasonality* of reproduction, *calving interval*, *length of females at maturation* and *length of males at maturation*; and the diet variables.

The latter were split as usual into the food categories cephalopods, fishes and mammals & birds. For each of these categories there is information on whether they occurred in the diet of a species at all (a yes-no variable), how frequent they were (in how many of the investigated cases/stomachs) and in what proportion they were found. The categories are additionally divided in those species that occur at the surface (epipelagic), in the water column (mesopelagic) or on the ground (benthic). An even finer split is achieved by considering whether the food organisms occur singly, both singly and in swarms or only in swarms (groups with more than one individual).

All continuous variables, i. e. all variables except for the ones indicating presence or absence of specific food types, the *habitat*, the *habitat structure* and *latitude*, were log transformed.

The number of cases in each analysis was given by the availability of data for the variables that were included in the analysis (see below).

General Methods: misclassification

Usually misclassification rates are calculated by dividing the number of misclassified cases by the total number of cases. This measure might be overly optimistic if one of the classified groups is much larger than the others and is well classified.

That is why I evaluated an additional "misclassification index": in principal, it calcualted a missclassification index per column of a usual missclassification table and then averaged over all columns. The results from this index are basically identical to the classical misclassification index and are not further discussed.

Cluster analysis

Methods

The R library "mva" was used for this analysis. Cluster analysis (Venables and Ripley, 1997, chapt. 13.2) was conducted on several subsets of the variables to see which variables might be the most important in clustering the Delphinoidea species into their families. The following groups of variables were each either included or not (two variants each): group sizes, physical environment, size, life-history. Additionally information on diet was either included or not. If included this was done on a specific level of detail and either all variables or only those on occurrence were included (seven variants). As the case where no variables are included

is uninteresting, this gives a total of $2^4 * 7 - 1 = 111$ possible combinations. Further, the combinations using only *average group size* instead of all group measures and the occurrence variables of the diet only were evaluated (additional 24 combinations). For each model single, average and complete clustering was conducted.

Cluster analysis is a descriptive method of "unsupervised" classification, i. e. no prior knowledge of group membership is used. In the present example it could be assumed that the first three splits would divide the species into their families if the variables were closely correlated to the phylogeny of the species.

To calculate a measure of misclassification, I split the calculated tree into the three most dissimilar classes and used the misclassification index defined above.

Results

None of the cluster analyses evaluated in this study clustered the species according to their families. Species of the different families were mixed even in the smallest clusters, i. e. those with the highest similarity. This is also reflected in the misclassification index which was high in general (0.6 to 1). Neither the number of cases, the number of variables nor the number of species included in a model decreased the misclassification index and thus did not lead to a better model, i. e. that would categorise species into families more accurately.

Supervised categorisation

After these exploratory attempts of phylogenetic classification some methods of "supervised" categorisations were used: Firstly a (black-box) logistic discriminant analyses to see how well the Delphinoidea species can be categorised at all, and secondly, tree-based models which give good information on which and how variables are used in the categorisation process (Ripley, 1996, Introduction). As most life-history variables included only one value per species they were not included in these two supervised methods.

Discriminant analysis: Methods

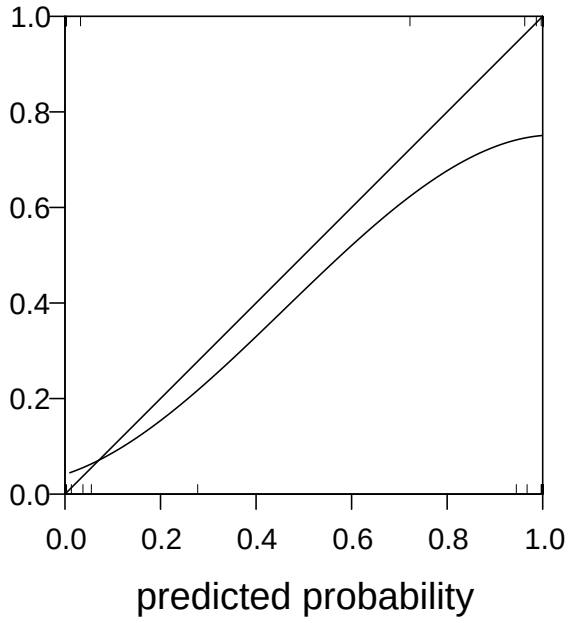
The R libraries "MASS", "nnet", "mva" and "modreg" were used for this analysis. Logistic discriminant analyses (Venables and Ripley, 1997, chapt. 13.3 and 17) were calculated using a neural net (function multinom() in the package "nnet"). In such an evaluation it is possible to conduct a variable reduction based on the AIC criterion. The discriminant analyses were conducted, and their performance tested, by a 10-fold cross-validation (Venables and Ripley, 1997, p. 493).

To judge the suitability of such a model it is useful to check (graphically) whether the predicted probabilities are well calibrated (Venables and Ripley, 1997, p. 495f and Fig. 21).

24 models were conducted including different combinations of variables. The *average group size* and the measures of the physical environment were always included. The other measures of group size and the *length of females* were included or not and one of six diet variable sets was included (on a certain split level and either all types or occurrence only).

Tree based analysis: Methods

The R library "tree" was used for this analysis. Tree-based models (Venables and Ripley, 1997, chapt. 14) were conducted to see how variables were used in categorisation. The same 24


1.0
0.8
0.6
0.4
0.2
0.0
0.0
0.2
0.4
0.6
0.8
1.0
predicted probability

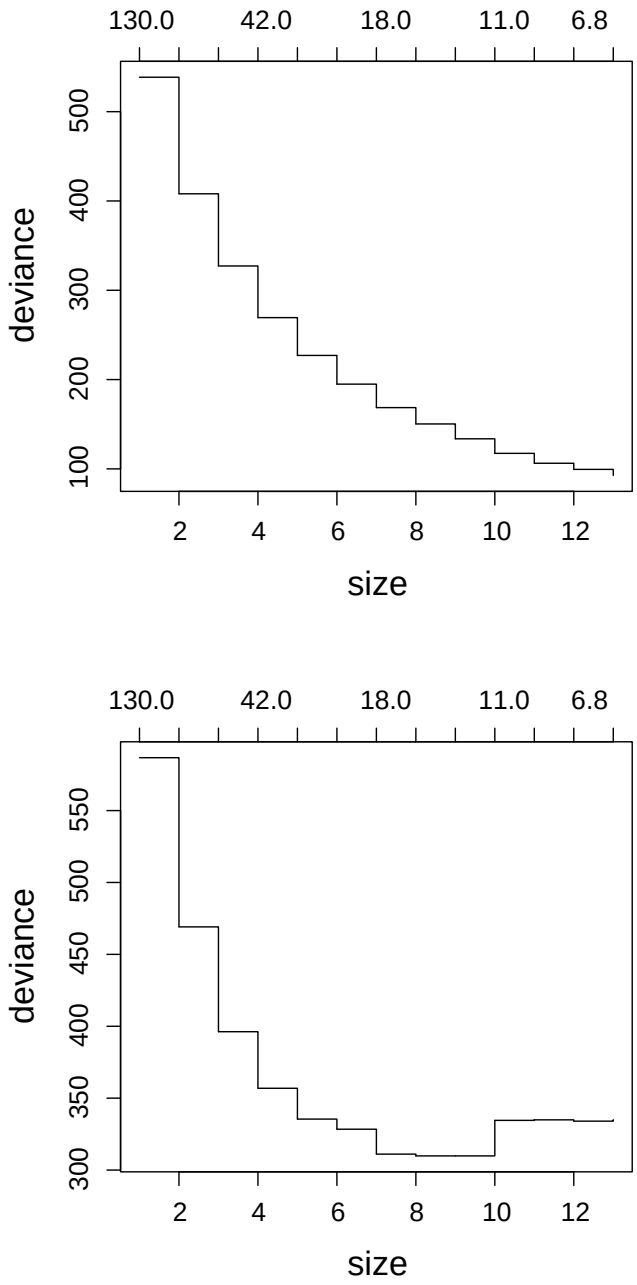


Figure 22: Typical examples of a pruning plot (top) and a cross-validation plot for pruning (bottom) in the tree-based models.

Conclusions from the three preliminary methods

Delphinoidea species seem to be difficult to categorise on the basis of variables from their behavioural ecology (group size, physical environment, life-history and diet). It seems equally difficult to judge on what variables the success is achieved as little as there is. As species are known it was decided to restrict further attempt to the more powerful supervised methods.

Canonical discriminant analysis

A further series of canonical discriminant analysis attempted to model the different species based on average group size (always included), all other group size variables (included or not), the physical environment (always included), the size of the animals (included or not) and one out of six sets of diet variables (using all variables on the three different split levels and using only the occurrence variables only on the three different split levels). All these combinations resulted in 24 models that were run.

Plotting the data on the first two discriminant axes revealed that neither species nor families were well differentiated. Additionally, no cross-validation was conducted for these models and thus overfitting of the data was likely.

Canonical discriminant analysis using crossvalidation and PCs

Using a ten-fold cross-validation species were categorised based on a group size measure (either none, average group size or first PC of group size, for the PCs see the chapter on the inter-specific comparisons, page 30), the physical environment (the four first PCs of the physical environment were either included or not), the size of the animals (either included or not), the length dimorphism (either included or not) and one of three sets of PCs from the diet variables (either included or not). This resulted in 95 possible combinations.

Still, the categorisation based on the plots on the two principal discriminant axes was not overwhelming. Within species variability seemed to be very large. Models including size seemed to fare somewhat better. It was still likely that models including many of the variables could offer more information than was actually contained in the data (overfitting).

Variable selection

One suspicion in the above methods was that some variables only added redundant information and thus, that not all the variables are necessary to achieve the smallest possible misclassification error.

Thus, estimation samples from a ten-fold cross-validation for each of the sets of diet PCs were exported and read into SPSS. There, a step-wise discriminant analysis was conducted (Table 22).

Based on the test available in SPSS all variables improved the models significantly. This was not trusted and thus the sequence of inclusion over the ten-fold cross-validation sample was averaged. The following variables were considered (in parentheses the mean rank of inclusion is given for the diet PCs on all diet variables; using the occurrence variables only, also presented in more detail in Table 22; and using only split level 1): average group size (3, 3, 2), physical PCs (openness: 11, 12, 9; warmth: 4, 2, 6; specificity: 13, 9, 10; smoothness: 2, 8, 4), length of females (1, 1, 1), length dimorphism (5, 4, 3), prey PCs (PC1: 5, 4, 3; PC2: 9, 6, 5; PC3: 6, 13, 12; PC4: 11, 7, 7: PC5: 10, 11, 11; PC6: 8, 10, –).

It is important to note that these are average ranks. Some variables (as e. g. some of the diet PCs, Table 22) varied greatly in their rank of inclusion among the 10 cross-validation sets whereas others were chosen quite consistently (such as e. g. length of females, Table 22).

Based on this average sequence of inclusion discriminant analyses were run in R with an increasing number of variables. Again a ten-fold cross validation procedure was used and repeated ten times. Then, the number of included variables were plotted against misclassification

Table 22: Ranking of inclusion of variables (small rank means early inclusion) in the ten-fold crossvalidation using the set of diet PCs based on the occurrence variables. Total sample size is 406 and includes 23 species. The crossvalidations include between 356 and 389 cases each.

crossvalidation	1	2	3	4	5	6	7	8	9	10	average rank	new rank
average group size	4	3	4	5	6	4	5	4	6	3	4.4	3
openness	11	9	10	11	12	13	10	9	13	11	10.9	12
warmth	1	6	6	1	1	8	7	6	1	6	4.3	2
specificity	9	10	8	10	5	10	11	8	7	5	8.7	9
smoothness	3	7	1	9	9	9	8	7	3	8	6.4	8
length of females	2	1	2	2	2	2	6	1	2	2	2.2	1
size dimorphism	8	4	7	4	4	6	2	3	4	5	4.7	4
diet PC1	7	8	5	8	8	1	4	5	5	7	5.8	6
diet PC2	12	2	3	3	3	12	3	2	9	4	5.3	5
diet PC3	13	13	13	13	13	11	12	12	12	12	12.4	13
diet PC4	5	5	9	6	7	3	9	10	8	1	6.3	7
diet PC5	6	11	12	12	11	5	13	13	11	13	10.7	11
diet PC6	10	12	11	7	10	7	1	11	10	10	8.9	10

errors. From these plots it was obvious that misclassification rates reached a minimum at about four to six included variables (Fig. 23).

From all models it was obvious that the most important variable was size. Thus, this was taken as the starting point for the final evaluation (see the chapter on the phylogenetic classification, page 57).

Statistical details of the classification comparison

No thorough check on the assumptions of the classification analysis were conducted as it was assumed that deviations from these assumptions would lead to models that predicted the species less well. If a model is found that manages to predict species accurately the aim of the analysis is reached.

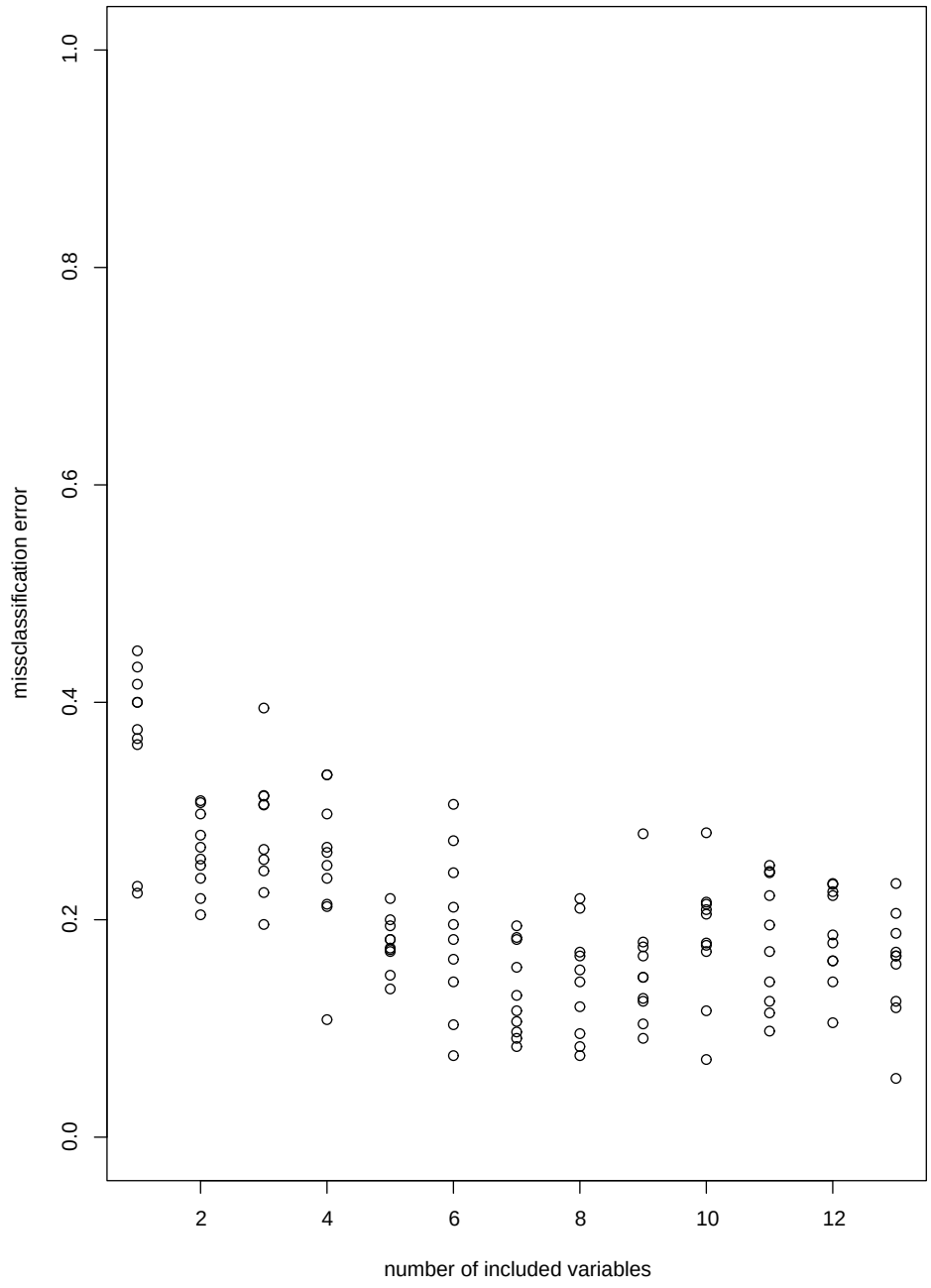


Figure 23: Misclassification errors versus number of variables. Sequence of variables as in Table 22: 'new rank'. Each dot is the result of a ten-fold crossvalidation.