

APPLICATION

rptR: repeatability estimation and variance decomposition by generalized linear mixed-effects models**Martin A. Stoffel^{1,2}, Shinichi Nakagawa³  and Holger Schielzeth^{*4,5} **

¹Department of Animal Behaviour, Bielefeld University, Morgenbreede 45, 33615 Bielefeld, Germany; ²Faculty of Science, School of Natural Sciences and Psychology, Liverpool John Moores University, Liverpool L3 3AF, UK; ³Evolution and Ecology Research Centre and School of Biological, Earth and Environmental Sciences, University of New South Wales, Sydney, NSW 2052, Australia; ⁴Department of Evolutionary Biology, Bielefeld University, Morgenbreede 45, 33615 Bielefeld, Germany; and ⁵Population Ecology Group, Institute of Ecology, Friedrich Schiller University Jena, Dornburger Str. 159, 07743 Jena, Germany

Summary

1. Intra-class correlations (*ICC*) and repeatabilities (*R*) are fundamental statistics for quantifying the reproducibility of measurements and for understanding the structure of biological variation. Linear mixed effects models offer a versatile framework for estimating *ICC* and *R*. However, while point estimation and significance testing by likelihood ratio tests is straightforward, the quantification of uncertainty is not as easily achieved.

2. A further complication arises when the analysis is conducted on data with non-Gaussian distributions because the separation of the mean and the variance is less clear-cut for non-Gaussian than for Gaussian models. Nonetheless, there are solutions to approximate repeatability for the most widely used families of generalized linear mixed models (GLMMs).

3. Here, we introduce the R package rptR for the estimation of *ICC* and *R* for Gaussian, binomial and Poisson-distributed data. Uncertainty in estimators is quantified by parametric bootstrapping and significance testing is implemented by likelihood ratio tests and through permutation of residuals. The package allows control for fixed effects and thus the estimation of adjusted repeatabilities (that remove fixed effect variance from the estimate) and enhanced agreement repeatabilities (that add fixed effect variance to the denominator). Furthermore, repeatability can be estimated from random-slope models. The package features convenient summary and plotting functions.

4. Besides repeatabilities, the package also allows the quantification of coefficients of determination R^2 as well as of raw variance components. We present an example analysis to demonstrate the core features and discuss some of the limitations of rptR.

Key-words: animal personality, coefficient of determination, generalized linear mixed-effects models, nonparametric bootstrapping, software application, variance component analysis

Introduction

Whenever quantitative measurements are hierarchically organized at multiple levels, intra-class correlations *ICC* can be used to express the average correlation among measurements taken from the same hierarchical level (McGraw & Wong 1996). A classical application is the quantification of the reproducibility of measurements from the same study object, and the *ICC* is therefore also known as the repeatability *R*. In the context of ecology and evolution, the repeatability typically represents the fraction of the total phenotypic variance V_P in the population of interest that can be attributed to variation among groups V_G . The ratio of these variance components can be written

as $R = \frac{V_G}{V_G + V_R}$, where V_R represents the within-group (residual) variance and $V_G + V_R = V_P$. The term ‘group’ is used here in the statistical sense and could represent various biological grouping, e.g. within individuals, families, social groups, plots or years. In the study of animal behaviour, for example, repeatabilities are frequently used as a measure of individual consistency with individual identities being the grouping factor (Boake 1989; Bell, Hankison & Laskowski 2009).

Notably, there can be repeatabilities at multiple levels in the same dataset: for example, the repeatability of phenotypes in different years of measurements and the repeatability of phenotypes within individuals. The coefficient of determination R^2 is a similar statistic that quantifies the proportion of variance explained by fixed effects (marginal R^2 sensu Nakagawa & Schielzeth 2013). R^2 is therefore a complementary statistic for decomposing the phenotypic variance. Both *R* and R^2

*Correspondence author. E-mail: holger.schielzeth@uni-jena.de

represent standardized statistics in the sense that they are variance components divided by the total phenotypic variance and are thus expressed as proportions of the phenotypic variance. Sometimes, however, it is relevant to estimate the variance components as they are, without standardization by the phenotypic variance. A convenient software tool for decomposing phenotypic variances will thus allow for the estimating of repeatabilities at multiple levels, offer flexibility in controlling for fixed effects, estimate the marginal R^2 for fixed effects and allow for the estimation of raw (unstandardized) variance components.

Repeatabilities can be estimated from a variety of sampling designs by a variety of statistical tools. The most widely used statistical framework for analyzing various designs are mixed effects models that allow the estimation of the relevant variance components with the possibility to control for fixed effect covariates (Nakagawa & Schielzeth 2010). Non-normal error distributions have represented a challenge for estimating repeatabilities, because of the nonlinearity induced by the link function (de Villemereuil *et al.* 2016). However, we have previously reviewed the equations for estimating repeatabilities and R^2 from generalized linear mixed effects models (GLMMs) (Nakagawa & Schielzeth 2010, 2013). The challenge here is that not all variance components necessary for calculating repeatabilities are part of the model output. Another challenge is the quantification of uncertainty of variance components and ratios of variance components (such as repeatabilities and R^2). We have suggested the use of parametric bootstrapping for quantifying uncertainty of repeatabilities (Nakagawa & Schielzeth 2010) and this can be easily applied to R^2 , too. Parametric bootstrapping quantifies the design-specific sampling variance by simulating response values from the fitted model followed by a re-estimation of the repeatability or R^2 . Under the assumption that the model is correctly specified, the variance among replicated simulations thus represents the sampling uncertainty of the estimate.

There are a number of other R packages that can estimate certain aspects of ICC and repeatabilities for more specific problems. The packages `IRR` (Gamer *et al.* 2012), `PSY` (Falissard 2012) and `PSYCH` (Revelle 2016) with their functions `icc` and `ICC` allow the estimation of rater agreement and consistency for simple (one- or two-way) designs. From the field of ecology and evolution, the `icc` package (Wolak, Fairbairn & Paulsen 2012) allows the estimation of ANOVA-based repeatability with a single response vector and a single grouping vector. Notably, the `icc` package allows the exploration of optimized sampling designs. However, none of these packages can estimate repeatabilities and their uncertainties for non-Gaussian models or control for confounding effects. The `MUMIN` package (Barton 2016) allows the quantification of R^2 from fitted generalized linear models, but does not allow the quantification of intra-class correlations.

Here, we introduce the `rptR` package for the free software environment R (R Core Team 2016). `rptR` provides general utilities for estimating adjusted and agreement repeatabilities for Gaussian, binomial (binary and proportion) and Poisson models. The package relies on mixed-effects models fitted by

the `lmer` and `glmer` functions from the `LME4` package (Bates *et al.* 2015). Confidence intervals for repeatabilities are estimated by parametric bootstrapping and statistical significance against $H_0: R = 0$ is tested by likelihood ratio and permutation tests. Moreover, it is also possible to estimate marginal R^2 and raw variance components along with their uncertainties and statistical significance.

Features

The package `rptR` as it is introduced here represents a complete rewrite of the `rptR` developmental package that we had written as part of our original repeatability review (Nakagawa & Schielzeth 2010). Compared to the developmental version of the package, the current version now provides several new features: (i) The package allows for the estimation of adjusted repeatabilities for Gaussian and well as non-Gaussian data. Adjusted repeatabilities are repeatabilities that control (adjust) for fixed effects (table 6 in Nakagawa & Schielzeth 2010 provides an overview of different types of repeatabilities). The variance explained by fixed effect is excluded from the denominator. (ii) The package allows estimating what we here call enhanced agreement repeatabilities. Enhanced agreement repeatabilities fit fixed effects, but include their variance in the denominator. (iii) The package allows estimating marginal R^2 along with repeatabilities. (iv) The package allows repeatability and marginal R^2 estimation from random-slope models, using the approach introduced by Johnson (2014). (v) The package allows the estimation of raw (unstandardized) variances. (vi) The package provides new plotting and summary functions as well as a detailed documentation in the form of a vignette (which is accessible via `vignette("rptR")`). (vii) We added the option to conduct bootstrapping and permutations in parallel for reduced waiting time, a progress bar for monitoring and an update option for stepwise increases of bootstrapping and/or permutations.

The package now features the four core functions `rptGaussian` for Gaussian error distributions with identity link, `rptPoisson` for Poisson error distributions with log or square root link, `rptBinary` for binary data with logit or probit link functions and `rptProportion` for proportion data (in the form of counts of success and failure events) following binomial distributions with logit or probit link. All functions can be called via the general `rpt` function by specifying the `datatype` argument. Results are returned as S3 class `rpt` objects and can be conveniently displayed via the generic `print`, `summary`, and `plot` functions. The package now has a simplified and unified formula interface and is based on the popular `LME4` package (Bates *et al.* 2015) as the central model-fitting engine. Grouping factors of interest are fitted as random effects and potentially confounding variables can be fitted as fixed effects. For non-Gaussian models, the package internally adds an observational level random effect (with the reserved term 'Overdispersion') for estimating overdispersion.

Likelihood ratio tests are returned by default. Randomization procedures are also implemented, but since randomization is time-consuming, this option is deactivated in the default

state (argument `npermut = 0`). Randomization in the `rptR` package is implemented as a permutation of residuals of the fitted null model (excluding the grouping factor of interest). This ensures that the remaining data structure and effects are represented in the simulated data while the dependence with the grouping factor of interest is broken.

Example analysis

We will illustrate the features of `rptR` by estimating adjusted repeatabilities for Poisson data with log link for a dataset that was generated for estimating R^2 in GLMMs (Nakagawa & Schielzeth 2013). The data represent counts of eggs for an imaginary species of beetle and we estimate repeatabilities at two hierarchical level of organization, among populations and among housing containers, while controlling for an experimental food manipulation. The data are distributed with `rptR` and can be loaded and analyzed by the following:

```
library(rptR)
data(BeetlesFemale)
```

The repeatability can be estimated through the general `rpt` function or directly with the more specialized `rptPoisson` function. The syntax for the formula argument is the same as for the `glmer` function from the `LME4` package. We here specify two random effects (population and container) and aim to estimate the agreement repeatabilities at both levels simultaneously (controlled by the `grname` argument).

```
rptPoisson(Egg ~ 1 + (1|Container) + (1|Population),
  grname=c("Container", "Population"),
  data=BeetlesFemale, link="log")
```

The `nboot` argument controls the number of parametric bootstrap iterations for confidence interval estimation and defaults to 500 bootstraps. It may be advisable to reduce this number initially, but to increase the number to 1000 or more for the final analysis. The `npermut` argument controls the number of randomizations for permutation-based null hypothesis testing and defaults to 0. We recommend starting with the default and to use a substantial number of permutations (e.g. 1000) for the final analysis. More iterations for bootstrapping and permutation can be added, using the `update` and `rptObj` arguments (see package documentation for details).

With the example dataset, it is advisable to control for important design effects, in particular the nutritional treatment that may have inflated the phenotypic variance experimentally. Therefore, we estimate adjusted repeatabilities by including a dummy coded binary predictor as a covariate in the fixed part on the right-hand side of the formula argument (while using the option `parallel = TRUE` for reduced waiting time).

```
rep1 <- rptPoisson(Egg ~ Treatment + (1|Container) +
  (1|Population), grname=c("Container", "Population"),
  data=BeetlesFemale, link="log", nboot=1000,
  parallel=TRUE)
```

The output can be viewed using the generic `print`, `summary` and `plot` functions (Table 1). Unlike the

agreement repeatability estimated above, the adjusted repeatability has removed the variance explained by `Treatment` from the repeatability estimation. Adjusted repeatabilities are also known as consistency repeatabilities, particularly in the psychological literature (Shrout & Fleiss 1979; McGraw & Wong 1996), or as narrow sense repeatabilities (Biro & Stamps 2015). Depending on the sampling designs, adjusted repeatabilities are often, but not always, larger than unadjusted (agreement) repeatabilities. The unadjusted repeatability without any control for systematic effects is also known as the broad sense or agreement repeatability (Shrout & Fleiss 1979; Nakagawa & Schielzeth 2010; Biro & Stamps 2015).

Plots can be customized by additional arguments handed over to `plot.default` (Fig. 1):

```
print(rep1)
summary(rep1)
plot(rep1, grname="Population", scale="link",
  cex.main=0.8, main="Population variance", las=1)
plot(rep1, grname="Container", scale="link",
  cex.main=0.8, main="Container variance", las=1)
```

Sometimes it is desired to fit models that include fixed effects, but to add the variance explained by fixed effects in the denominator of the repeatability estimation. In `rptR`, it is possible to estimate the variance explained by fixed effects as the variance in the linear predictor (i.e. the variance in fitted values) and to include this variance in the repeatability estimation. This functionality is controlled by the `adjusted` argument. By setting `adjusted = FALSE`, the model estimates repeatabilities with the variance explained by fixed effect in the denominator. We call this enhanced agreement repeatabilities.

```
rep2 <- rptPoisson(Egg ~ Treatment + (1|Container) +
  (1|Population), grname=c("Container",
  "Population"), data=BeetlesFemale, link="log",
  nboot=1000, parallel=TRUE, adjusted=FALSE)
```

Furthermore, it is sometimes of interest to estimate the variances directly rather than as ratios of variances as represented by the repeatability. By setting the argument `ratio = FALSE`, `rptR` will estimate raw variances rather than repeatabilities. There are three reserved terms to the `grname` argument, 'Overdispersion', 'Residual', and 'Fixed' that allow estimating the overdispersion variance (on the latent scale modelled as an observation specific random effect), residual variance (the sum of overdispersion and distribution-specific variance) and the variance explained by fixed effects.

```
rep3 <- rptPoisson(Egg ~ Treatment + (1|Container) +
  (1|Population), grname=c("Container", "Population",
  "Fixed", "Overdispersion", "Residual"),
  data=BeetlesFemale, link="log", nboot=1000,
  parallel=TRUE, ratio=FALSE, adjusted=FALSE)
```

Limitations

While point estimation is convenient and fast using `rptR`, the two Monte Carlo simulation steps can be slow. For a

Table 1. Examples of print and summary displays of an rpt object

Print	Summary
Repeatability estimation using the glmm method and log link	Repeatability estimation using glmer method
Repeatability for Container	Call = rptPoisson(formula = Egg ~ Treatment + (1 Container) + (1 Population), grname = c("Container", "Population"), data = BeetlesFemale, link = "log", nboot = 1000, npermut = 1000, parallel = TRUE)
Link-scale approximation:	Data: 480 observations
R = 0.038	
SE = 0.026	
CI = [0.003, 0.107]	
P = 0.00874 [LRT]	
0.086 [Permutation]	Container (60 groups)
Original-scale approximation:	Repeatability estimation overview:
R = 0.032	R SE 2.5% 97.5% P_permut
SE = 0.024	Org 0.0319 0.0237 0.00276 0.0966 0.041
CI = [0.003, 0.097]	Link 0.0380 0.0261 0.00333 0.1071 0.086
P = 0.00874 [LRT]	
0.041 [Permutation]	Bootstrapping:
Repeatability for Population	N Mean Median 2.5% 97.5%
Link-scale approximation:	Org 1000 0.0369 0.0334 0.00276 0.0966
R = 0.52	Link 1000 0.0428 0.0392 0.00333 0.1071
SE = 0.124	Permutation test:
CI = [0.234, 0.684]	N Mean Median 2.5% 97.5% P_permut
P = 1.19e-16 [LRT]	Org 1000 0.00742 0.00132 0 0.0347 0.041
0.001 [Permutation]	Link 1000 0.01126 0.00203 0 0.0518 0.086
Original-scale approximation:	Likelihood ratio test:
R = 0.504	logLik full model = -1178.331
SE = 0.124	logLik red. model = -1181.155
CI = [0.221, 0.672]	D = 5.65, d.f. = 1, P = 0.00874
P = 1.19e-16 [LRT]	
0.001 [Permutation]	Population (12 groups)
	Repeatability estimation overview:
	R SE 2.5% 97.5% P_permut
	Org 0.504 0.124 0.221 0.672 0.001
	Link 0.520 0.124 0.234 0.684 0.001
	Bootstrapping:
	N Mean Median 2.5% 97.5%
	Org 1000 0.459 0.463 0.221 0.672
	Link 1000 0.473 0.478 0.234 0.684
	Permutation test:
	N Mean Median 2.5% 97.5% P_permut
	Org 1000 0.184 0.184 0.129 0.240 0.001
	Link 1000 0.244 0.246 0.176 0.313 0.001
	Likelihood ratio test:
	logLik full model = -1178.331
	logLik red. model = -1211.963
	D = 67.3, d.f. = 1, P = 1.19e-16

The data show the analysis of a toy dataset of Poisson distributed count data analyzed with log link, 1000 parameteric bootstrap iterations for estimating confidence intervals and 1000 permutations for significance testing. R = repeatability, SE = standard error, CI = confidence interval, P = error probability (P -value), LRT = likelihood ratio test.

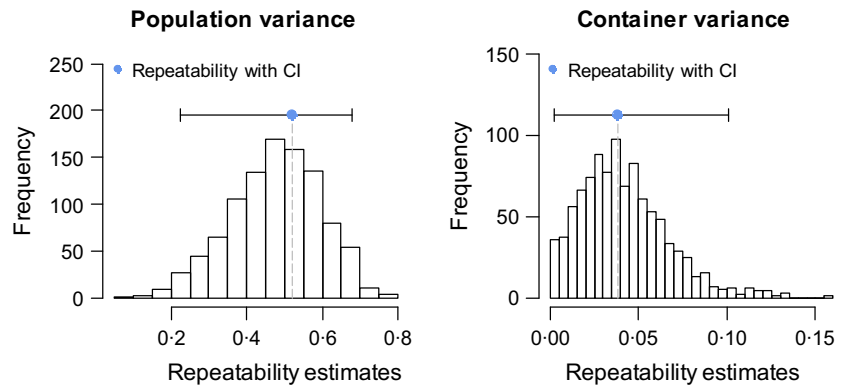


Fig. 1. Examples for plots of rpt objects. The data show the analysis of a toy dataset of Poisson distributed count data analyzed with log link. Link scale repeatabilities are shown. CI = confidence interval.

large dataset with multiple random effects such as the beetle toy dataset, a decent number of bootstraps or permutations can take several minutes. This time can be substantially shortened by conducting the simulations on multiple cores in parallel. For many applications, we suggest to initially fit the model with bootstrapping and permutations switched off (`nboot = 0`, `npermut = 0`). The final analysis should be done with larger numbers of bootstraps and permutations (possibly in separated steps, since computation times add up). It is also possible to increase the number of bootstraps and permutations by calling the function again with the arguments `update = TRUE` and the rpt object from the previous function call being handed over to the `rptObj` argument.

With the focus on `glmer` from the `LME4` package (Bates *et al.* 2015) as the central mixed-model fitting engine, we are limited to modelling additive overdispersion models. We have previously described repeatability estimation also for multiplicative overdispersion models (Nakagawa & Schielzeth 2010), but since we find the model fits of `glmer` accurate, fast and efficient, we currently do not see the need for implementing estimation based on multiplicative overdispersion models.

The package does not work properly for data exhibiting underdispersion relative to the chosen GLMM family. Since the additive overdispersion term cannot drop below zero, the total residual variance cannot become smaller than the distribution-specific variance. What seems like a limitation is in fact an unavoidable feature. While overdispersion can easily arise from unmodelled confounding effects, underdispersion signifies that the GLMM family is simply not appropriate. There is no formal warning build in in rptR for potential underdispersion, hence users are advised to check their data themselves. The lack of any link-scale residual variance might, however, suggest potential underdispersion.

Availability

A stable version of the package (rptR 0.9.2 at the time of writing) can be downloaded from Comprehensive R Archive Network (CRAN, <https://cran.r-project.org/package=rptR>). This version will run with the current version of R (R 3.3.3 at the time of writing) and will provide the features

introduced in this publication. At the same time, we aim to further develop the package, potentially improving plot and display features as well as supporting additional GLMM families. A developmental version will be available on Github (<https://github.com/mastoffel/rptR>), where it will be tested before being released on CRAN. The current stable Github version can be downloaded manually or via `install_github` from the devtools suite of functions (Wickham & Chang 2016): `install_github("mastoffel/rptR", build_vignettes = TRUE)`.

Authors' contributions

S.N. and H.S. conceived the idea for the package and drafted a first version. M.A.S. has implemented the new features with help of H.S. All authors have contributed to testing the packages, drafting and revising the manuscript.

Acknowledgements

The package was motivated by the continued interest in and requests concerning an initial draft package on r-forge. We greatly appreciate the input and suggestions that we received from many colleagues. The manuscript and vignette have benefitted from valuable input by Paul Johnson and two anonymous reviewers. The programming of the package was financially supported by the Research Centre for Mathematical Modelling (RCM²). H.S. was supported by an Emmy Noether fellowship from the German Research Foundation (SCH1188/1-2). S.N. is supported by a Future Fellowship, Australia (FT130100268).

Data accessibility

The simulated data used for illustration is distributed with the package. Once the package has been installed as described above, the data can be viewed by a call to `data(package = "rptR")`. Data deposition: <https://cran.r-project.org/package=rptR>

References

- Barton, K. (2016) MuMIn: multi-model inference. R package version 1.15.6. Available at: <http://CRAN.R-project.org/package=MuMIn>.
- Bates, D., Mächler, M., Bolker, B. & Walker, S. (2015) Fitting linear mixed-effects models using LME4. *Journal of Statistical Software*, **67**, 1–48.
- Bell, A.M., Hankison, S.J. & Laskowski, K.L. (2009) The repeatability of behaviour: a meta-analysis. *Animal Behaviour*, **77**, 771–783.
- Biro, P.A. & Stamps, J.A. (2015) Using repeatability to study physiological and behavioural traits: ignore time-related change at your peril. *Animal Behaviour*, **105**, 223–230.
- Boake, C.R.B. (1989) Repeatability: its role in evolutionary studies of mating behavior. *Evolutionary Ecology*, **3**, 173–182.
- Falissard, B. (2012) psy: various procedures used in psychometry. R package version 1.1. Available at: <https://cran.r-project.org/package=psy>.

- Gamer, M., Lemon, J., Fellows, I. & Singh, P. (2012) irr: various coefficients of interrater reliability and agreement. R package version 0.84. Available at: <http://cran.r-project.org/package=irr>.
- Johnson, P.C.D. (2014) Extension of Nakagawa & Schielzeth's R^2 GLMM to random slopes models. *Methods in Ecology and Evolution*, **5**, 944–946.
- McGraw, K.O. & Wong, S.P. (1996) Forming inferences about some intraclass correlation coefficients. *Psychological Methods*, **1**, 30–46.
- Nakagawa, S. & Schielzeth, H. (2010) Repeatability for Gaussian and non-Gaussian data: a practical guide for biologists. *Biological Reviews*, **85**, 935–956.
- Nakagawa, S. & Schielzeth, H. (2013) A general and simple method for obtaining R^2 from generalized linear mixed-effects models. *Methods in Ecology and Evolution*, **4**, 133–142.
- R Core Team (2016) *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.
- Revelle, W. (2016) psych: procedures for personality and psychological research. R package version 1.6.6. Available at: <http://cran.r-project.org/package=psych>.
- Shrout, P.E. & Fleiss, J.L. (1979) Intraclass correlations: uses in assessing rater reliability. *Psychological Bulletin*, **86**, 420–428.
- de Villemereuil, P., Schielzeth, H., Nakagawa, S. & Morrissey, M. (2016) General methods for evolutionary quantitative genetic inference from generalised mixed models. *Genetics*, **204**, 1281–1294.
- Wickham, H. & Chang, W. (2016) devtools: tools to make developing R packages easier. R package version 1.12.0. Available at: <http://cran.r-project.org/package=devtools>.
- Wolak, M.E., Fairbairn, D.J. & Paulsen, Y.R. (2012) Guidelines for estimating repeatability. *Methods in Ecology and Evolution*, **3**, 129–137.

Received 1 March 2017; accepted 30 March 2017

Handling Editor: Sarah Goslee

Supporting Information

Details of electronic Supporting Information are provided below.

Appendix S1. Code contained within rptR: repeatability estimation and variance decomposition by generalized linear mixed-effects models.