Methods in Ecology and Evolution

British Ecological Society

Methods in Ecology and Evolution 2015, **6,** 27–37 doi: 10.1111/2041-210X.12281

Quantifying the predictability of behaviour: statistical approaches for the study of between-individual variation in the within-individual variance

Ian R. Cleasby[1,2]*, Shinichi Nakagawa[3] and Holger Schielzeth[4]

[1]*School of Biology, University of Leeds, Leeds LS2 9JT, UK;* [2]*Centre of Ecology and Conservation, University of Exeter, Cornwall Campus, Penryn TR10 9EZ, UK;* [3]*Department of Zoology, University of Otago, Dunedin, New Zealand; and* [4]*Department of Evolutionary Biology, Bielefeld University, Bielefeld, Germany*

Summary

1. Many aspects of animal behaviour differ consistently between individuals, giving rise to the growing field of animal personality research. While between-individual variation has long been of interest to biologists, the role of within-individual variation has received less attention. Indeed, many models assume that the extent of within-individual variation is the same across individuals despite the fact that individuals may often differ in their variability.

2. Recently, the importance of within-individual variability or predictability has been recognized within the field of animal behaviour. However, there is a lack of a consensus on how best to quantify it. This situation, in turn, has led to the development of a variety of different methods aimed at assessing how variable or predictable different individuals are.

3. Here, we review the indices that have been proposed as proxies of individual predictability. We then introduce existing techniques called hierarchical generalized linear models (HGLMs) and double-hierarchical generalized linear models (DHGLMs) as general tools for quantifying predictability. HGLMs and DHGLMs are extensions of random intercept mixed models that exploit the fact that variation in variances as well as variation in means can be modelled within a single overarching framework.

4. Explicit modelling of the within-individual residual variation by (D)HGLMs makes more efficient use of the data, performs better on unbalanced data sets and captures more of the uncertainty involved in modelling within-individual variation than other proposed indices. In addition, (D)HGLMs yield an estimator of population-wide variation in predictability, which can serve as a standardized effect size for comparisons across traits and studies. We call this estimator CV_P, the coefficient of variation in predictability.

5. The different methods described here and the standardized effect size CV_P should open new avenues for studying individuality in animal behaviour. Since sound understanding of individual variation is central to many studies in ecology and evolution, these methods have wide application both in the field of animal personality research and beyond.

Key-words: animal personality, behavioural consistency, coefficient of variation, dispersion models, intra-individual variability, mixed models, residual variance

General Introduction

The study of animal personalities has emerged as a major field of ecological research since the turn of the century (Sih, Bell & Johnson 2004; Dingemanse & Réale 2005; Réale *et al.* 2007). Animal personality research is mainly concerned with the two properties of behaviour in a population: (i) variation in individual means of behavioural traits or between-individual variance and (ii) covariance among behaviours (*sensu* Bell 2007; Réale *et al.* 2007; Dingemanse, Dochtermann & Nakagawa 2012). The former is often characterized by repeatability (Bell, Hankison & Laskowski 2009; Nakagawa & Schielzeth 2010), and the latter is often referred to as behavioural syndromes (Sih, Bell & Johnson 2004; Dingemanse, Dochtermann & Nakagawa 2012; Dingemanse & Dochtermann 2013). Similar concepts, such as individual niche specialization and the division of labour, are applied in slightly different contexts (Dall *et al.* 2012), but are largely concerned with the same behavioural properties (Lessells & Boag 1987; Bolnick *et al.* 2003).

Generalized linear mixed models (GLMMs) have provided an overarching framework for animal personality research. GLMMs decompose the observed (co)variance into within- and between-individual (co)variances allowing the calculation of both repeatability and between-individual correlations (Dingemanse & Dochtermann 2013). In addition, the use of random slope mixed models allows the estimation of interindi-

*Correspondence author. E-mail: i.r.cleasby@gmail.com

vidual variation in response to environmental gradients (i.e. phenotypic reaction norms, Dingemanse *et al*. 2010). A reaction norm perspective is relevant in animal personality research, because variation in reaction norms indicates variation in individual plasticity in response to the environment and might provide a mechanism that helps to maintain variation in animal personalities (Réale *et al*. 2007).

Although GLMMs are used frequently in animal personality research, the residual variation in behavioural traits is typically assumed to be homoscedastic, that is, the same for all individuals (Cleasby & Nakagawa 2011). However, this assumption may often be violated. Moreover, accounting for individual differences in residual variation allows us to estimate variation in within-individual variation, a behavioural property that is likely to be of biological importance (Westneat, Wright & Dingemanse 2014). The importance of modelling differences in within-individual variation has already been recognized in fields such as human psychology (Hoffman 2007) and quantitative genetics (Hill 1984; Hill & Zhang 2004; Rönnegård *et al*. 2010a). In contrast, the topic of within-individual variation in behavioural traits (termed 'behavioural predictability'; see our definition below) has only recently begun to be addressed by ecologists (Stamps, Briffa & Biro 2012; Briffa, Bridger & Biro 2013; Westneat, Schofield & Wright 2013) but will play an increasingly important role as it becomes better integrated into ecological theory (Dall *et al*. 2012; Westneat, Wright & Dingemanse 2014).

Currently, ecologists use various different terms when addressing within-individual variation (Dall *et al*. 2012), making it difficult to see connections between studies. For example, within-individual variation has been referred to as *predictability*, *consistency*, *intraindividual variation* or *specialization* among others (Matich, Heithaus & Layman 2011; Stamps, Briffa & Biro 2012; Briffa, Bridger & Biro 2013), and different indices have been proposed as proxies for these biological quantities (see below). In this paper, we reserve the term *plasticity* for variation across different environments (i.e. nonzero reaction norms and/or interindividual variation in reaction norms), while we treat within-individual behavioural variation in the same environment as *predictability*. The term *consistency* can be used as an umbrella term, subsuming within-individual variation both within and among environments (Fig. 1).

In practice, it may be difficult to disentangle plasticity from predictability as we cannot fully control the environment that individuals experience and what we might consider the same environment varies due to unmeasured microenvironmental factors (Rönnegård, Shen & Alam 2010b). The separation of terms is nevertheless useful, because we can, at least in principle, manipulate the environment that determines plasticity, while we cannot (by definition) manipulate the environmental gradients that produce variation in predictability. There is also a difference in focus between different fields. While the analysis of predictability in animal breeding is mostly concerned with maximizing robustness to microenvironmental perturbation (Rönnegård, Shen & Alam 2010b), behavioural ecologists are primarily interested in the potential for an adaptive unpredictable component to animal behaviour. If there is an adaptive unpredictable component, then microenvironmental gradients become less important.

Our aim in this paper is to explain how ecologists can use existing statistical techniques to appropriately estimate behavioural predictability (i.e. individual-specific residual variances) and provide a population-level measure of variation in behavioural predictability (i.e. the population-level variance in within-individual variances). We first review the proposed indices of individual-level predictability and identify and illustrate the disadvantages associated with these indices. Secondly, we provide a brief introduction to standard mixed models that allow an estimation of the between-individual variation in mean trait values but assume that the residual variance is homoscedastic (i.e. the same across all individuals). We then introduce hierarchical GLMs (HGLM) and double-hierarchical GLMs (DHGLM) that allow us to explicitly model the residual variance as a function of both fixed and random effects (Lee & Nelder 1996, 2006). The theory behind the hierarchical models (HGLM and DHGLM) that we discuss has been largely developed in the field of quantitative genetics (SanCristobal-Gaudy *et al*. 1998; Kizilkaya & Tempelman 2005; Lee & Nelder 2006; Rönnegård *et al*. 2010a) with application in other areas such as medicine (Lee & Nelder 1996, 2002). Similar models are known as dispersion models in the social sciences (Raudenbush & Bryk 1987; Hoffman 2007). To our knowledge, the first and only application to date of a DHGLM in the context of animal personality was the analysis of provisioning rates in red-winged blackbirds (*Agelaius phoeniceus*) (Westneat, Schofield & Wright 2013) with a few more ecological examples in Lee & Nelder (1996).

In many situations, we believe that behavioural and evolutionary ecologists will be most interested in quantifying the variation in individual-level predictabilities in order to compare across different populations. Therefore, we propose a statistic, termed CV_P (the coefficient of variation for predictability), which quantifies the population-level variation in predictability and facilitates cross-study comparisons in a similar way to more familiar measures such as repeatability (Bell, Hankison & Laskowski 2009; Sonesson, Ødegård & Rönnegård 2013). We finish by discussing the advantages and limitations of HGLMs and DHGLMs and outline how such models could be used to tackle specific biological hypotheses.

Individual-level predictability indices and their problems

There currently exist several different approaches for quantifying within-individual variation within the ecological literature (Table 1). For instance, calculating the residual individual standard deviation (riSD) involves fitting statistical models and using the difference between observed and fitted values (model residuals) as the basis for estimating individual-level variation. The standard deviation of these residuals for each individual (riSD) then constitutes an estimate of the individual-specific within-individual variation (Briffa, Bridger & Biro 2013). Similarly, estimating relative specialization involves

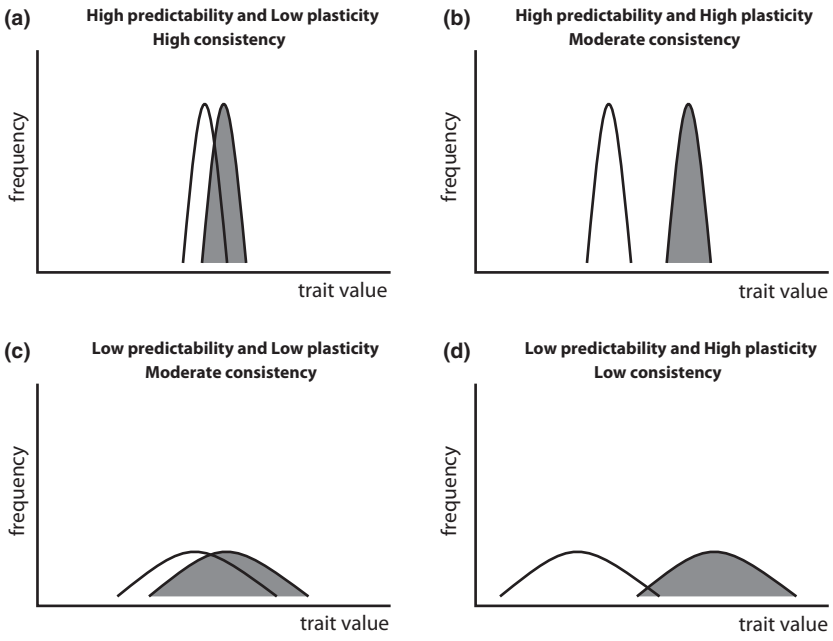

Fig. 1. Schematic illustration of the terminology used in this paper. Each plot shows the distribution of observations of a single individual in two different environments (shown in white and grey shades). Predictability refers to the variation within environments, while plasticity refers to the variation between environments. Consistency serves as an umbrella term that covers within- and between-environment variation.

Table 1. Overview of the different methods for calculating point estimates of individual predictability

Method	Equation	Range of values	Comparability across studies	Notes	Reference
Coefficient of relative plasticity (CRP)	$CRP_i = \frac{V_i}{V_P}$	0 to 1*	No, because populations can differ in average within-individual consistency and between-individual variance in means	The CRP for each individual i is calculated as the ratio of the variance of the individual V_i over the overall phenotypic variance V_P	Réale & Dingemanse (2008)
Individual stability statistic (ISS)	$ISS_{t_1,t_2} = 1 - \left(\frac{Z_{t1} - Z_{t2}}{2}\right)^2$	0 to 1	No, because populations can differ in average within-individual consistency and between-individual variance in means	Z_{t1} and Z_{t2} are normalized phenotypic values measured on the same individual at times t_1 and t_2	Asendorpf (1990)
Residual individual standard deviation (riSD)	$riSD = \sqrt{\frac{(Y_{ij} - E_{ij})^2}{N_i - 1}}$	0 to ∞	Yes, but comparisons among traits or systems are difficult because of lack of standardization	Y_{ij} represents observation j for the ith individual, and E_{ij} is the expected value (mean) generated from a regression model. N_i is the number of observations for individual i. The riSD may also be referred to as intraindividual variability (IIV)	Stamps, Briffa & Biro (2012)
Pruitt's choosiness (PC)	$PC = \frac{\bar{V}_{Pi}}{V_i}$	0 to ∞	Yes, but comparisons among traits or systems are difficult because of lack of standardization	Where $\bar{V}_{Pi}$ is the average phenotypic variance across all individuals in a population and V_i is the variance observed for individual i	Pruitt *et al.* (2011)
Relative specialization (RS)	$RS = \frac{E(SS)_B}{(Y_{ij} - \mu_i)^2}$	0 to ∞	Yes, but comparisons among traits or systems are difficult because of lack of standardization	Where $E(SS_B)$ is the sum of squares between groups for all individuals in a population; Y_{ij} represents observation j for the ith individual, and μ_i is the average value for individual i	Matich, Heithaus & Layman (2011)

*Values larger than 1 are mathematically possible, but are biologically unlikely.

calculating the difference between an individual's observed values and that individual's predicted mean, and dividing by the variability between individuals in a population (Matich, Heithaus & Layman 2011). Alternative approaches such as the coefficient of relative plasticity (Réale & Dingemanse 2008) and Pruitt's choosiness (Pruitt *et al*. 2011) are based on dividing raw estimates of individual-level variance with some measure of population-wide variance to create a ratio between the two. A further approach is the individual stability statistic (ISS), which directly compares the difference between measurements taken from the same individual at different time points (Asendorpf 1992).

While all these techniques generate point estimates of within-individual variability at the individual level, they are not ideal for inferences about population-level variation in predictability. In particular, population-level inferences are often based on two-step analyses, in which within-individual variability is estimated and then fitted as the response variable in a follow-up analysis. There are at least three disadvantages with such two-step analyses. First, while point estimates of within-individual variation are carried forward in the modelling process, the uncertainty around such estimates is not. By ignoring this uncertainty, we run the risk of inflating type 1 errors. Secondly, two-step models are not the most parsimonious and powerful way of modelling within-individual variance (Hoffman 2007). Thirdly, if between-individual variation in predictability exists, the model assumptions of homoscedastic residuals are violated. Hence, estimates such as riSD are based on models that make assumptions that are violated if the phenomenon they seek to describe is real. Each of these factors will influence our ability to make accurate population-level inferences and the variance in the estimated values of the residual variance will be an upwardly biased estimate of the true residual variance.

Moreover, all indices listed in Table 1 are strongly influenced by unbalanced sampling (Fig. 2). Consequently, within-individual predictabilities will be estimated with unequal precision and may simply reflect sampling effort per individual, and hence, population-level predictability is difficult to estimate without bias. A simple simulation illustrates how the

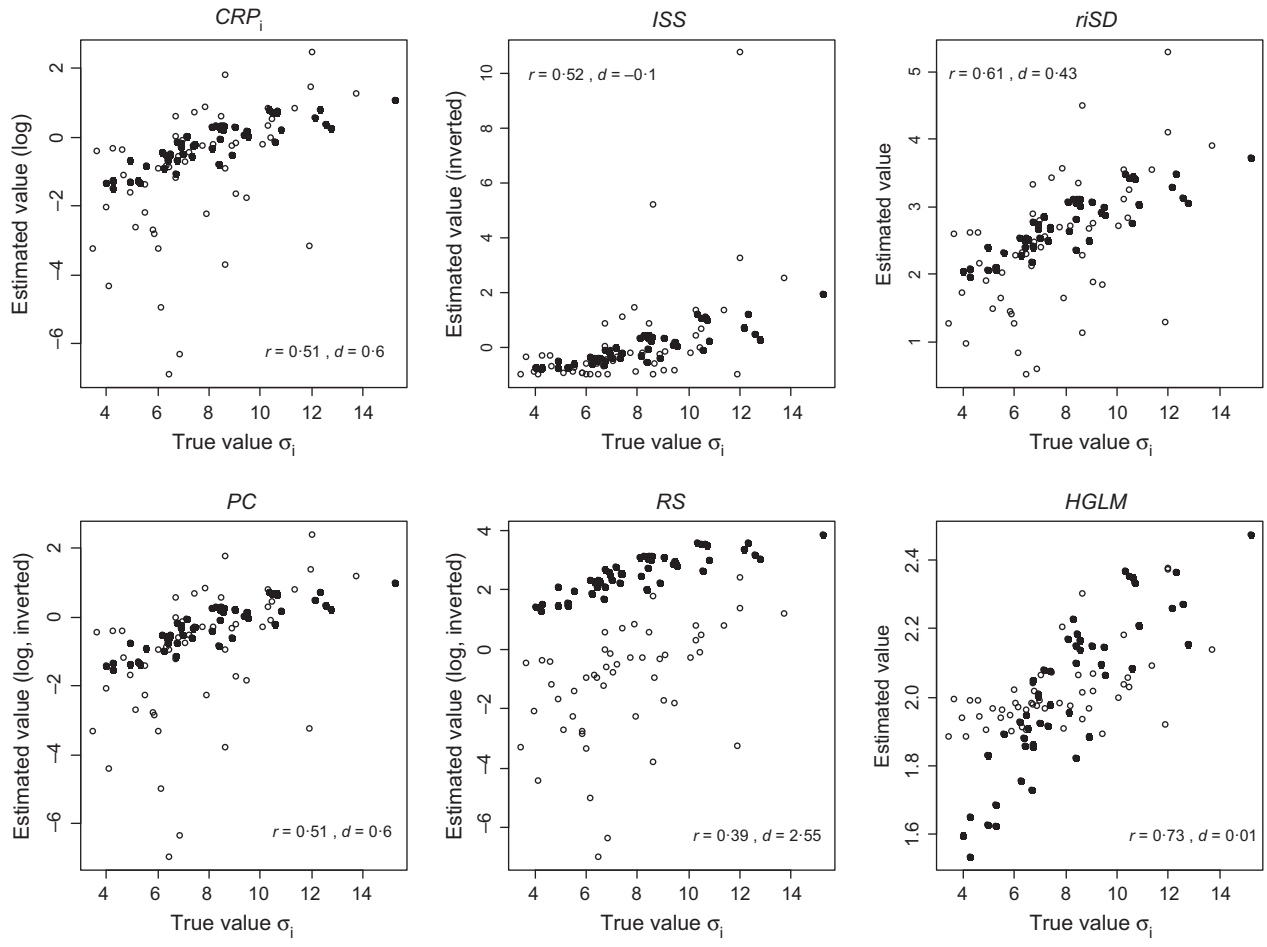

Fig. 2. Performance of different proxies of predictability in a simulated case with unbalanced sampling design. For educational purposes, we chose a scenario with an extreme difference in the number of observations, 50 individuals with 18 observations each (filled dots) and 50 individuals with 2 observations each. The estimated index values are plotted against the simulated within-individual variance (variances drawn from a log-normal distribution with $\mu_{ln} = 2$ and $\sigma_{ln} = 0{\cdot}3$). For simplicity, the simulated data set did not include additional confounding effects and no between-individual variation in intercepts. Details on index calculated are listed in Table 1. The statistics *r* and *d* refer to the correlation of estimates with simulated values and the standardized difference between the two classes of individual (18 vs. 2 observations).

individual-level estimators scatter depending on sample size (Fig. 2): with 18 observations per individual, the different indices correlate well with the simulated individual-specific variance, but two observations per individual yield large scatter. Average indices differ between the two classes of individuals, illustrating that the calculated values partly reflect sample size and hence an aspect of the sampling design rather than biology. Problematically, the estimates for groups with few observations tend to be more extreme and will therefore have large undue influence on follow-up analyses. In contrast, HGLM estimates of predictability (from models presented below) tend to shrink towards the population mean when the sample size is low. This occurs because the partial pooling in hierarchical models represents a balance between individual-level and population-level means, and therefore, shrinkage tends to be greater for groups with fewer observations (Gelman & Hill 2007).

A GLMM for estimating variation in mean trait values

A Gaussian random intercept GLMM, which would typically be used when calculating repeatability, can be written in matrix notation as follows (model 1):

$$\mathbf{y} = \boldsymbol{\eta} + \boldsymbol{\varepsilon}, \quad \text{eqn 1}$$

$$\boldsymbol{\eta} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\boldsymbol{\alpha}, \quad \text{eqn 2}$$

$$\boldsymbol{\alpha} \sim N(\mathbf{0}, \mathbf{I}_m\sigma^2_\alpha), \quad \text{eqn 3}$$

$$\boldsymbol{\varepsilon} \sim N(\mathbf{0}, \mathbf{I}_n\sigma^2_\varepsilon), \quad \text{eqn 4}$$

where $\mathbf{y}$ is a response vector of length n (n is the total number of observation, single observations will be indexed by j), $\boldsymbol{\beta}$ is a vector of length k that represents the fixed effects (k is the number of fixed effects in the model including the population intercept), $\boldsymbol{\alpha}$ is a vector of length m that represents the individual-specific random effect deviations (m is the total number of individuals, single individuals will be indexed by i), $\mathbf{X}$ of dimensions $n \times k$ is a design matrix relating observations to fixed effects, $\mathbf{Z}$ of dimensions $n \times m$ is a design matrix relating observations to random effects, and $\mathbf{I}_m$ and $\mathbf{I}_n$ are identity matrices of dimensions $m \times m$ and $n \times n$, respectively. The term $\boldsymbol{\eta} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\boldsymbol{\alpha}$ represents the linear predictor for the mean, and the residual deviations from the prediction are represented by a vector $\boldsymbol{\varepsilon}$ of length n. The between-individual random effect variance in means is assumed to be normally distributed with the mean of 0 and variance σ^2_α, and $\mathbf{I}_m\boldsymbol{\sigma}^{\mathbf{2}}_{\boldsymbol{\alpha}}$ signifies that the random effects are independently and identically distributed (i.i.d.). The variance σ^2_α constitutes a population-level hyperparameter for the variance in expected values and provides an estimate of between-individual variation (note the slightly different meaning of 'hyperparameter' specific to Bayesian analyses, Schielzeth & Nakagawa 2013). Similarly, σ^2_ε is the residual variance and $\mathbf{I}_n\sigma^2_\varepsilon$ signifies that the residuals are independently and identically distributed.

Using this type of model, the repeatability R of phenotypic traits can be calculated as:

$$R = \frac{\sigma^2_\alpha}{\sigma^2_\alpha + \sigma^2_\varepsilon}. \quad \text{eqn 5}$$

Quantifying repeatability for behavioural traits has become a routine in ecology and evolution (Lessells & Boag 1987; Nakagawa & Schielzeth 2010). However, in other fields, Equation 5 is more generally referred to as the intraclass correlation (ICC) (McGraw & Wong 1996). Equation 5 is also similar to the formula used for defining heritability, which demonstrates how variance component models developed for quantitative genetics may transfer to ecological studies.. The ICC conveys information on between-individual variance in means, but within-individual variation is assumed to be homoscedastic (the same expected value for the whole population). However, from a biological point-of-view, differences in residual variation between individuals are likely to be common in ecological data sets (Cleasby & Nakagawa 2011), necessitating the use of models in which heteroscedasticity is incorporated as we describe below.

The linear predictor $\boldsymbol{\eta}$ with its components the fixed effect predictors $\mathbf{X}\boldsymbol{\beta}$ and the (random) individual deviations $\mathbf{Z}\boldsymbol{\alpha}$, given in Equations 2–3, represents the *mean part of the model* since they are concerned with estimating means and variation in means, respectively. Equation 4 represents the residual variance and is termed the *dispersion part of the model*. While the dispersion part of the model is usually limited to the estimation of a single residual variance common to all observations (Equation 4), mixed models can be extended to explicitly model systematic and random patterns in the residual variance.

A GLMM with structured dispersion

When heteroscedasticity due to interindividual variation in predictability occurs, we can allow the residual variance to differ between individuals within a mixed model in order to incorporate heterogeneous within-individual variances (Hoffman 2007). Using matrix notation, we can write such a mixed model as (model 2):

$$\mathbf{y} = \boldsymbol{\eta} + \boldsymbol{\varepsilon}, \quad \text{eqn 6}$$

$$\boldsymbol{\eta} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\boldsymbol{\alpha}, \quad \text{eqn 7}$$

$$\boldsymbol{\alpha} \sim N(\mathbf{0}, \mathbf{I}_m\sigma^2_\alpha), \quad \text{eqn 8}$$

$$\boldsymbol{\varepsilon} \sim N(\mathbf{0}, Diag\{\mathbf{Z}\boldsymbol{\sigma}^{\mathbf{2}}_{\boldsymbol{\varepsilon}}\}), \quad \text{eqn 9}$$

where notation is as above, except for $\boldsymbol{\sigma}^{\mathbf{2}}_{\boldsymbol{\varepsilon}}$ that is now a vector of length m and *Diag* {} is used to create an m × m diagonal matrix with elements $\sigma^2_{\varepsilon_i}$. Since we are here only concerned with between-individual variation, it is sensible to fit the same design matrix $\mathbf{Z}$ that clusters observation by individual identities in the mean part of the model (here in Equations 7–8) and in the dispersion part (Equation 9). In the general case, however, these matrices might differ. In fitting such a GLMM, we can interpret individuals with the smaller

residual variances, $\sigma^2_{\varepsilon_i}$ as being more predictable than those with larger residual variances. Extending mixed models to include structured dispersion is relatively straightforward and can be achieved using a variety of statistical software (Gelman & Hill 2007; Rönnegård, Shen & Alam 2010b).

Specifying the model in this way means every individual has its own unique residual variance. Values of $\sigma^2_{\varepsilon_i}$ are estimated for each individual independently rather than coming from a common population-level distribution (Lunn *et al*. 2013), which would be the case if adopting a hierarchical approach (HGLM or DHGLM, see below). Unfortunately, model 2 may require a large numbers of repeated measurements per individual in order to provide precise estimates of within-individual variances and will be susceptible to unbalanced sampling. Consequently, measures of predictability for individuals with fewer observations will tend to be underestimated and estimated with less confidence than for individuals with many observations. In addition, it does not generate a population-level estimate of the variance in within-individual variability.

An HGLM for modelling variation in predictability

Instead of modelling the residual variance for each individual independently, we can model the between-individual variation in predictability using hyperparameters. This is possible in a framework of flexible extensions of GLMMs that are called HGLMs (for comprehensive reviews of HGLMs, we recommend Lee & Nelder 1996, 2006; Lee, Nelder & Pawitan 2006). While GLMMs are characterized by modelling of normal random effect distributions and normal residual distributions (possibly at a link scale), HGLMs allow non-normal distributions of these components by modelling a set of interlinked GLMs. The parameters of the separate sets of interlinked GLMs can be fitted iteratively and traditional model checking techniques can be used to assess model assumptions (Lee, Nelder & Pawitan 2006). One of the key advantages of HGLMs is that they allow thick tails in the distributions of random effects as well as residuals, which makes the models more robust against outliers (Lee & Nelder 1996). While robustness is a genuine advantage of HGLM, we here present HGLM for their potential to explicitly model variance in predictability.

The concept of modelling with hyperparameters is based on the assumption that lower level processes follow parametric distributions and that the parameters of these distributions can be estimated (Gelman & Hill 2007). The gamma and log-normal distributions are convenient choices for modelling the residual variance because they only allow positive values (but other distributions are possible, Lunn *et al*. 2013). An example of such a model can be written as (model 3):

$$\mathbf{y} = \boldsymbol{\eta} + \boldsymbol{\varepsilon}, \qquad \text{eqn 10}$$

$$\boldsymbol{\eta} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\boldsymbol{\alpha}, \qquad \text{eqn 11}$$

$$\boldsymbol{\alpha} \sim N(\mathbf{0}, \mathbf{I}_m\sigma^2_\alpha), \qquad \text{eqn 12}$$

$$\boldsymbol{\varepsilon} \sim N(\mathbf{0}, Diag\{\mathbf{Z}\boldsymbol{\sigma}^2_{\boldsymbol{\varepsilon}}\}), \qquad \text{eqn 13}$$

$$\log(\boldsymbol{\sigma}_{\boldsymbol{\varepsilon}}) \sim N(\mathbf{d}_0, \mathbf{I}_m\omega^2_\sigma), \qquad \text{eqn 14}$$

where $\mathbf{d_0}$ represents a vector of length m denoting the mean at the log scale. $\boldsymbol{\sigma}^2_{\boldsymbol{\varepsilon}}$ is still a vector of length m containing individual-specific variances $\sigma^2_{\varepsilon_i}$, but individual-specific residual standard deviations are now assumed to be following a log-normal distribution with mean at the log scale of d_0 and variance at the log scale of ω^2_σ (see Fig. 3 for a schematic). Equations 11–12 represent the mean part of the model, while Equations 13–14 represent the dispersion part. The mean part of the model is focussed upon modelling the mean values of the response variable, whereas the dispersion part of the model is concerned with modelling the residual variance (Fig. 4).

The value of using hyperparameters to characterize the distribution of residual variance is that it allows us to model the distribution of individual residual variances jointly as originating from a common distribution, which is a useful

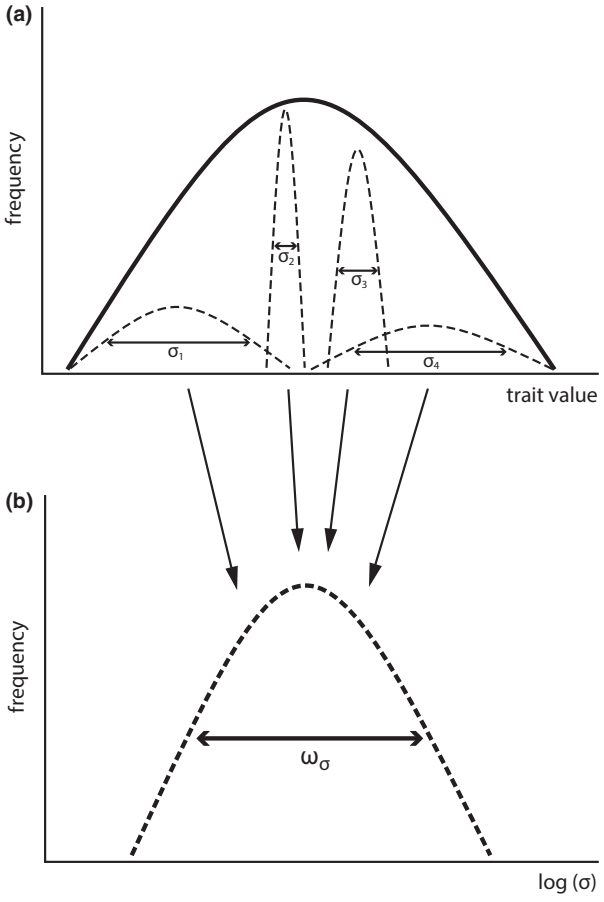

Fig. 3. Schematic illustration of the individual-level predictors of predictability (a) and the population-wide variation in predictability (b).

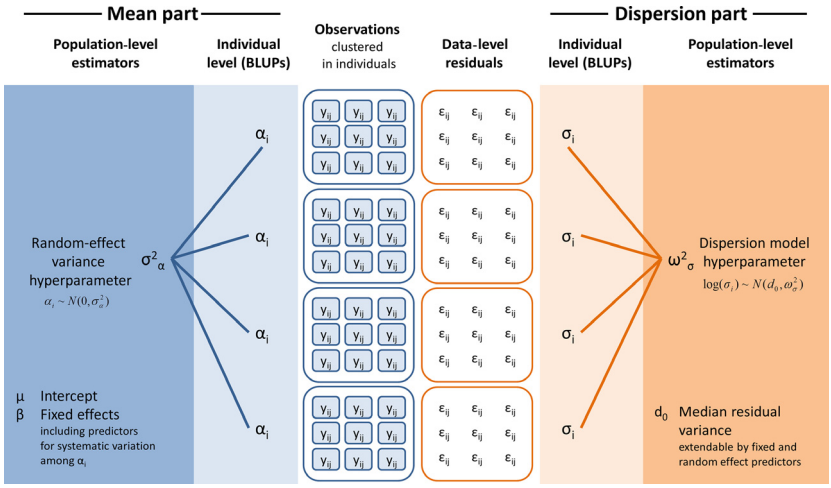

Fig. 4. Schematic overview of the structure of normal–log-normal hierarchical GLM (HGLM). The middle part shows the observations and the data-level residuals. The left part (in blue) shows the more familiar mean part of the model that constitutes the backbone of a random intercept mixed model. The right part (in orange) shows the dispersion part of the model that is expanded in a double-hierarchical model in order to model the structure in residual variance. Best linear unbiased predictions (BLUPs) provide point estimates of a random effect within a mixed model.

assumption if subjects were sampled randomly from a homogeneous population. Such an approach makes better use of a data by partially pooling data across individuals (Gelman & Hill 2007) as opposed to assuming every individual has the same residual variance (homoscedasticity) or estimating independent residual variances for each individual (previous section). Thus, it provides a better way of allowing for unequal within-individual variances when we only have a few measures per individual and/or our sampling is unbalanced (Gelman & Hill 2007).

The log-normal distribution is useful because it facilitates modelling the residual variance as a function of covariates, and the parameters of the log-normal distribution are readily interpretable (Limpert, Stahel & Abbt 2001; Lunn *et al.* 2013). In addition, we can exploit the relationship between the log-normal and normal distributions in order to calculate parameters on the log-normal scale such as the coefficient of variation (CV). Indeed, the use of coefficients of variation to quantify variability is already well-established in quantitative genetics (Houle 1992; Hill & Mulder 2010). The mean and standard deviation of the log-normal distribution are as follows:

$$\mu_{\ln} = \exp(d_0 + \omega_\sigma/2), \qquad \text{eqn 15}$$

$$\sigma_{\ln} = \exp(d_0 + \omega_\sigma/2)\sqrt{\exp(\omega_\sigma^2) - 1}. \qquad \text{eqn 16}$$

We can also calculate the coefficient of variation (CV) on the log-normal scale using ω_σ^2, which we term CV_P (the coefficient of variation for predictability).

$$CV_P = \sqrt{\exp(\omega_\sigma^2) - 1}. \qquad \text{eqn 17}$$

CV_P constitutes an appropriate population-level estimator of the degree of variation in within-individual predictabilities, allowing us to compare and contrast differences in within-individual variation across different traits and different studies. Conveniently, CV_P is not affected by the mean of the trait we are measuring and also not by the average predictability in the population. Since d_0 represents the median residual standard deviation at the logarithm scale, the repeatability of the HGLM can be calculated as:

$$R = \frac{\sigma_\alpha^2}{\sigma_\alpha^2 + \exp(d_0)^2}. \qquad \text{eqn 18}$$

Adding predictors to the dispersion part of HGLMs

Specifying separate models for the mean and dispersion allows us to incorporate fixed effects predictors of dispersion parameters just as we do for the mean; such models are termed HGLMs with structured dispersion (Lee, Nelder & Pawitan 2006). An HGLM with structured dispersion includes fixed effects within the dispersion part of the model as follows (model 4):

$$\mathbf{y} = \boldsymbol{\eta} + \boldsymbol{\varepsilon}, \qquad \text{eqn 19}$$

$$\boldsymbol{\eta} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\boldsymbol{\alpha}, \qquad \text{eqn 20}$$

$$\boldsymbol{\alpha} \sim N(\mathbf{0}, \mathbf{I}_m\sigma_\alpha^2), \qquad \text{eqn 21}$$

$$\boldsymbol{\varepsilon} \sim N(\mathbf{0}, Diag\{\mathbf{Z}\boldsymbol{\sigma}_{\boldsymbol{\varepsilon}}^{\mathbf{2}}\}), \qquad \text{eqn 22}$$

$$\log(\boldsymbol{\sigma}_{\boldsymbol{\varepsilon}}) \sim N(\boldsymbol{\eta}_{\mathbf{d}}, \mathbf{I}_m\omega_\sigma^2), \qquad \text{eqn 23}$$

$$\boldsymbol{\eta}_{\mathbf{d}} = \mathbf{X}_{\mathbf{d}}\boldsymbol{\beta}_{\mathbf{d}}, \qquad \text{eqn 24}$$

where $\boldsymbol{\sigma}_{\boldsymbol{\varepsilon}}^{\mathbf{2}}$ is a vector of length m containing individual-specific variances $\sigma_{\varepsilon_j}^2$, $\boldsymbol{\eta}_{\mathbf{d}}$ is the linear predictor in the dispersion part of the model and $\boldsymbol{\beta}_{\mathbf{d}}$ represents a vector of fixed effects (including a population intercept of d_0) with the corresponding design matrix $\mathbf{X}_{\mathbf{d}}$ within the dispersion part of the model. Here, each residual is assumed to be independently normally distributed but with a unique variance that differs among groups or gradients defined by $\mathbf{X}_{\mathbf{d}}$. Equations 20–21 represent the mean part and Equations 22–24 the dispersion part of model 4. The structured HGLM presented here allows the estimation of individual-specific residual variances, because this creates an

estimator of the between-individual variance in predictability. However, structured HGLM can be modified to model observation-specific residual variances if required (e.g. see model 5 below).

Lee & Nelder (2006) extended HGLMs further by including random effects in both the mean and dispersion parts of the model to create DHGLMs. Thus, additional structure in the residual variance caused by random sources of variation can be modelled by incorporating a random effect within the dispersion part of a model. Such a model can be written as (model 5):

$$\mathbf{y} = \boldsymbol{\eta} + \boldsymbol{\varepsilon}, \quad \text{eqn 25}$$

$$\boldsymbol{\eta} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\boldsymbol{\alpha}, \quad \text{eqn 26}$$

$$\boldsymbol{\alpha} \sim N(\mathbf{0}, \mathbf{I}_m \sigma^2_\alpha), \quad \text{eqn 27}$$

$$\boldsymbol{\varepsilon} \sim N(\mathbf{0}, Diag\{\boldsymbol{\sigma}^2_\varepsilon\}), \quad \text{eqn 28}$$

$$\log(\boldsymbol{\sigma}_\varepsilon) = \boldsymbol{\eta}_\mathbf{d}, \quad \text{eqn 29}$$

$$\boldsymbol{\eta}_\mathbf{d} = \mathbf{X}_\mathbf{d}\boldsymbol{\beta}_\mathbf{d} + \mathbf{Z}_\mathbf{d}\boldsymbol{\alpha}_\mathbf{d}, \quad \text{eqn 30}$$

$$\boldsymbol{\alpha}_\mathbf{d} \sim N(\mathbf{0}, \mathbf{I}_m \omega^2_{\alpha_d}), \quad \text{eqn 31}$$

where $\boldsymbol{\sigma}^2_\varepsilon$ is a vector of length n containing observation-specific variances $\sigma^2_{\varepsilon i}$, and the linear predictor in the dispersion part of the model $\boldsymbol{\eta}_\mathbf{d}$ now includes a random effect component $\mathbf{Z}_\mathbf{d}\boldsymbol{\alpha}_\mathbf{d}$ with $\boldsymbol{\alpha}_\mathbf{d}$ being a vector of appropriate length in the dispersion part and $\mathbf{Z}_\mathbf{d}$ relating observations to dispersion part random effects. For convenience, $\boldsymbol{\alpha}_\mathbf{d}$ is assumed to be normally distributed with mean of zero and variance $\sigma^2_{\alpha_d}$. Random effects in the dispersion part might include environmental sources of variation (e.g. year identities or group identities) or genetic sources of variation (such as an additive genetic relatedness matrix). The dispersion part of model 5 is represented in Equations 28–31. Note that in this formulation, the mean part (Equation 26) and the dispersion part (Equation 30) of the same model could have the same fixed and random terms (e.g. Westneat, Schofield & Wright 2013).

The ability to include both fixed and random effects for the dispersion gives us considerable flexibility when seeking to model the data. In most cases, the use of HGLMs with structured dispersion and DHGLMs allows us to include all the relevant information as fixed and random effects within the dispersion part of the model so that two-step analyses are not necessary. However, estimates of within-individual residual variance from (D)HGLMs will be conditioned upon both the dispersion and mean part of the model. Thus, any measures of within-individual variation will depend on the other predictors included in the model (as for estimating repeatability, Nakagawa & Schielzeth 2010).

Table 2. A summary of the current software packages in which (D)HGLMs can be fitted

Software/ Package	Features	Remarks	Freely available	Reference
BUGS	Bayesian MCMC-based approach. Example code provided in appendix and in Gelman & Hill (2007) and Lunn *et al.* (2013)	Allows easy calculation of 95% credible intervals. Permits flexible modelling, but researchers must write code for models and specify appropriate prior distributions	Yes	Lunn *et al.* (2000)
JAGS	Bayesian MCMC-based approach. Example code available in Westneat, Schofield & Wright (2013)	Allows easy calculation of 95% credible intervals. Permits flexible modelling, but researchers must write code for models and specify appropriate prior distributions	Yes	Plummer (2003)
hglm in R	Likelihood approach based on the h-likelihood of Lee & Nelder (2006). Examples given in Rönnegård *et al.* (2010a)	Relatively easy to use for current R practitioners. Can handle non-Gaussian response variables. Uses gamma rather than log-normal distribution for modelling the standard deviations	Yes	Rönnegård, Shen & Alam (2010b)
ASreml	Use h-likelihood approach. For example see Rönnegård *et al.* (2010a)	Powerful and flexible program but knowledge about coding models in ASReml required	No	Gilmour *et al.* (2009)
Genstat	Use h-likelihood approach of Lee & Nelder (2006)	Allows fixed effects in dispersion part of the model. Uses a gamma distribution rather than log-normal	No	VSN International (2011)
SAS®	Uses h-likelihood based on algorithm from Rönnegård *et al.* (2010a)	dhglm algorithms may take a long time to run	No	

Table 3. A comparison of the precision and accuracy of estimates of ω (the standard deviation of the logarithm) and CV_P for different sample sizes

	Simulated values			
Sample size	ω = 0·1; CV_P = 0·1	ω = 0·3; CV_P = 0·31	ω = 0·5; CV_P = 0·53	ω = 0·8; CV_P = 0·95
10 obs/10 individuals	ω = 0·22 (0·03–0·56) CV_P = 0·22 (0·03–0·61)	ω = 0·41 (0·19–0·84) CV_P = 0·43 (0·20–1·01)	ω = 0·52 (0·28–1·01) CV_P = 0·56 (0·29–1·37)	ω = 0·87 (0·53–1·64) CV_P = 1·07 (0·58–3·75)
20 obs/10 individuals	ω = 0·10 (0·01–0·30) CV_P = 0·10 (0·01–0·31)	ω = 0·34 (0·18–0·68) CV_P = 0·35 (0·18–0·77)	ω = 0·45 (0·26–0·86) CV_P = 0·46 (0·27–1·04)	ω = 0·84 (0·54–1·54) CV_P = 1·01 (0·56–3·13)
50 obs/10 individuals	ω = 0·071 (0·04–0·16) CV_P = 0·070 (0·04–0·16)	ω = 0·30 (0·18–0·57) CV_P = 0·31 (0·18–0·62)	ω = 0·58 (0·37–1·08) CV_P = 0·64 (0·38–1·47)	ω = 0·79 (0·54–1·38) CV_P = 0·88 (0·56–2·12)
10 obs/20 individuals	ω = 0·077 (0·01–0·24) CV_P = 0·072 (0·01–0·24)	ω = 0·28 (0·08–0·49) CV_P = 0·28 (0·08–0·53)	ω = 0·49 (0·32–0·76) CV_P = 0·52 (0·32–0·89)	ω = 0·72 (0·52–1·07) CV_P = 0·83 (0·55–1·48)
20 obs/20 individuals	ω = 0·08 (0·02–0·20) CV_P = 0·08 (0·02–0·20)	ω = 0·28 (0·17–0·43) CV_P = 0·28 (0·17–0·46)	ω = 0·51 (0·36–0·76) CV_P = 0·55 (0·37–0·76)	ω = 0·83 (0·61–1·23) CV_P = 0·99 (0·67–1·87)
50 obs/20 individuals	ω = 0·12 (0·06–0·21) CV_P = 0·12 (0·06–0·21)	ω = 0·30 (0·21–0·45) CV_P = 0·31 (0·21–0·45)	ω = 0·51 (0·38–0·76) CV_P = 0·55 (0·39–0·87)	ω = 0·85 (0·63–1·24) CV_P = 1·03 (0·63–1·92)
10 obs/50 individuals	ω = 0·11 (0·02–0·24) CV_P = 0·11 (0·02–0·24)	ω = 0·28 (0·18–0·40) CV_P = 0·29 (0·18–0·42)	ω = 0·47 (0·36–0·61) CV_P = 0·49 (0·37–0·67)	ω = 0·77 (0·61–0·96) CV_P = 0·88 (0·68–1·23)
20 obs/50 individuals	ω = 0·11 (0·02–0·19) CV_P = 0·11 (0·02–0·19)	ω = 0·31 (0·24–0·40) CV_P = 0·32 (0·24–0·42)	ω = 0·52 (0·42–0·66) CV_P = 0·56 (0·44–0·74)	ω = 0·82 (0·68–0·98) CV_P = 0·99 (0·77–1·38)
50 obs/50 individuals	ω = 0·11 (0·08–0·17) CV_P = 0·11 (0·08–0·17)	ω = 0·31 (0·25–0·39) CV_P = 0·32 (0·25–0·41)	ω = 0·52 (0·42–0·64) CV_P = 0·55 (0·44–0·72)	ω = 0·85 (0·66–1·00) CV_P = 0·96 (0·75–1·30)

Different sample sizes include both the number of individuals sampled and the number of replications per individual to generate a balanced data set. For each scenario, we replicated the simulation 100 times, and means and 95% credible intervals were calculated across replicates. Values in the table represent coefficient estimates with the 95% credible intervals beneath in brackets. Data were generated under the model shown in Equations 10–14 (model 3), and parameters were estimated in WinBUGS based on the same model. For simplicity, no fixed effects were included within the model. Note that CV_P is approximately ω when ω is small (<0·5).

Benefits, limitations and extensions

A key benefit of both HGLMs and DHGLMs is that they provide us with a way to explicitly model the residual variance in the data. There are also a variety of different software packages that can fit such models (Table 2). The ability to model the residual variance allows us to address a whole range of novel research questions relating to the variance in our data (Geiler-Samerotte *et al.* 2013). (D)HGLMs also provide a means of assessing interindividual variation in behavioural predictability within one modelling framework rather than using two-stage approaches (Hoffman 2007). Moreover, our proposed measure of CV_P may provide a useful means for quantifying the extent of between-individual variation in predictability allowing for comparison across different studies.

However, accurate and precise estimates of within-individual variation require a large number of observations per individual. Such a problem is common when seeking to model variance where it is estimated that approximately twice as many observations are required to detect a proportional effect on variance as would be required to detect the same effect on the mean (Hill & Mulder 2010). Providing exact guidelines on sample size required is difficult as different studies will have different characteristics. However, an exploration of MCMC estimates with simulated data showed that even when varying the number of observations per individual or the number of individuals sampled, the estimated values of ω (the standard deviation of the logarithm) and CV_P were close to their true simulated values in most cases (Table 3). However, when samples sizes were small, precision was low, making inference difficult. In general, the optimal sample size required to estimate ω or CV_P with high precision increases as the σ decreases (Hill & Mulder 2010). For example, sample sizes of 20 observations from 20 individuals may be suitable when ω is 0·3 or above (under the optimistic scenario that all model assumptions are strictly fulfilled as in our simulation). When ω is 0·1, substantially larger sample sizes are required. Nevertheless, DHGLMs have already proved useful in the study of avian provisioning where sample sizes can be high enough to permit such approaches (Westneat, Schofield & Wright 2013), and we believe (D)HGLMs will become more common as ecologists learn how to use and interpret them.

Aside from the study of animal personalities, the ability to include fixed effects for the dispersion in structured HGLMs permits a variety of analyses. For instance, such an approach may prove useful to studies investigating sex-specific sensitivity to environmental conditions during early life (Jones, Nakagawa & Sheldon 2009) or studies investigating how variation differs between control and treatment groups (e.g. Cleasby *et al.* 2011). An advantage here is that high levels of replication per

individual are not required to address such questions, if sampling is sufficiently dense per class.

(D)HGLMs are also suitable for behavioural observations with non-normal distributions, and the hglm R package (Rönnegård, Shen & Alam 2010b) will allow fitting (D)HGLMs for both Poisson and binomial response variables. Further extensions could also include incorporating random slopes within the dispersion part of the model. However, while technically feasible, we are unaware of any published studies that have done this and would suggest caution if adopting this approach as the performance of such models is not clear. Finally, (D)HGLMs have recently been developed that estimate the correlation between the random effects for the mean and the dispersion parts of the model rather than treating them as independent (Felleki *et al*. 2012; Rast, Hofer & Sparks 2012). In the future, we anticipate that the use of (D)HGLMs will increase and as a consequence that further extensions to the basic HGLM framework will allow us to perform increasingly detailed analyses. As such, we recommend Lee & Nelder (2006) and Lee, Nelder & Pawitan (2006) as good descriptions of the inherent flexibility of (D)HGLMs.

We should always bear in mind that residual variation can be generated by a variety of different processes with biological and non-biological explanations (Westneat, Wright & Dingemanse 2014). In particular, inferences about residual variation are affected by the scale at which data are analysed and can be misleading if data are skewed (Yang, Christensen & Sorensen 2011). For example, mean–variance relationships or the failure to properly account for changes in the mean is capable of generating variation in residual variation (Sonesson, Ødegård & Rönnegård 2013), which might then be incorrectly taken as evidence for individual heterogeneity. Yang, Christensen & Sorensen (2011) recommended using Box–Cox transformations (Box & Cox 1964) to find the appropriate scale on which to analyse the data and ensure that the distributional assumptions of the model are met. Model estimates based on untransformed and transformed data can be compared to examine the sensitivity of model estimates to skewed data and scale effects (Rönnegård *et al*. 2013), and model selection techniques may be used to compare different transformations (Yang, Christensen & Sorensen 2011). Ultimately, researchers need to consider how processes such as measurement error and incomplete models may contribute to heterogeneity in residual variances in addition to considering which biological phenomena may be important.

Conclusions

The study of variance is a key component of ecological and evolutionary research. While between-individual variation has long been of interest to ecologists, the role of within-individual variation has received much less attention (Bell, Hankison & Laskowski 2009). However, the biological importance of between-individual variation in within-individual variance is increasingly being recognized (Westneat, Wright & Dingemanse 2014). To date, various approaches have been used to assess within-individual variation, but many do not fully utilize the fact that the residual variance can be modelled analogously to means within a single framework. Here, we describe how we can use existing techniques (HGLMs and DHGLMs) to model both mean effects and residual variances simultaneously. We also propose a metric CV_P, which can be used as a population-level estimator of the degree of variation in within-individual consistencies and should facilitate comparisons between studies. As within-individual variation is integrated within ecological theory, we predict that (D)HGLMs will play an increasingly important role in the study of animal behaviour and phenotypic variation, enabling researchers to test a variety of new and exciting hypotheses. Moreover, the ability to model the variance in our data has application far beyond biological studies and will be of interest to researchers addressing any questions concerned with variance as well as means.

Acknowledgements

We thank Yimen Araya-Ajoy, Stuart Bearhop, Tom Bodey, Alexa Cleasby, Niels Dingemanse, Dave Fisher, Keith Hamer, Barbara Morrissey, Steve Votier, Ewan Wakefield and Alastair Wilson for advice and discussion during the writing of the manuscript. We thank Jarrod Hadfield, Julien Martin and one anonymous reviewer for comments that improved the manuscript. We are grateful to David Westneat, Niels Dingemanse and Jonathan Wright for providing us with a copy of an unpublished manuscript. We are grateful to Losia Lagisz for help in making Figs 1 and 3, and to Nils Cordes for advice on Fig. 4. HS was supported by an Emmy Noether fellowship from the German Research Foundation (DFG; SCHI 1188/1-1). SN was supported by a Rutherford Discovery Fellowship, New Zealand.

Data accessibility

R scripts and WinBUGS code for running the models described are available in the electronic supplement together with a brief description of the different open source packages available for running HGLMs.

References

Asendorpf, J.B. (1990) The measurement of individual consistency. *Methodika*, **4**, 1–23.

Asendorpf, J.B. (1992) Beyond stability: predicting inter-individual differences in intra-individual change. *European Journal of Personality*, **6**, 103–117.

Bell, A.M. (2007) Future directions in behavioural syndromes research. *Proceedings of the Royal Society B: Biological Sciences*, **274**, 755–761.

Bell, A.M., Hankison, S.J. & Laskowski, K.L. (2009) The repeatability of behaviour: a meta-analysis. *Animal Behaviour*, **77**, 771–783.

Bolnick, D.I., Svanbäck, R., Fordyce, J.A., Yang, L.H., Davis, J.M., Hulsey, C.D. & Forister, M.L. (2003) The ecology of individuals: incidence and implications of individual specialization. *The American Naturalist*, **161**, 1–28.

Box, G.E. & Cox, D.R. (1964) An analysis of transformations. *Journal of the Royal Statistical Society, Series B*, **26**, 211–252.

Briffa, M., Bridger, D. & Biro, P.A. (2013) How does temperature affect behaviour? Multilevel analysis of plasticity, personality and predictability in hermit crabs. *Animal Behaviour*, **86**, 47–64.

Cleasby, I.R. & Nakagawa, S. (2011) Neglected biological patterns in the residuals. *Behavioral Ecology & Sociobiology*, **65**, 2361–2372.

Cleasby, I.R., Burke, T., Schroeder, J. & Nakagawa, S. (2011) Food supplements increase adult tarsus length, but not growth rate, in an island population of house sparrows (*Passer domesticus*). *BMC Research Notes*, **4**, 431.

Dall, S.R.X., Bell, A.M., Bolnick, D.I. & Ratnieks, F.L.W. (2012) An evolutionary ecology of individual differences. *Ecology Letters*, **15**, 1189–1198.

Dingemanse, N.J. & Dochtermann, N.A. (2013) Quantifying individual variation in behaviour: mixed-effect modelling approaches. *Journal of Animal Ecology*, **82**, 39–54.

Dingemanse, N.J., Dochtermann, N.A. & Nakagawa, S. (2012) Defining behavioural syndromes and the role of 'syndrome deviation' in understanding their evolution. *Behavioral Ecology & Sociobiology*, **66**, 1543–1548.

Dingemanse, N.J. & Réale, D. (2005) Natural selection and animal personality. *Behaviour*, **142**, 9–10.
Dingemanse, N.J., Kazem, A.J., Réale, D. & Wright, J. (2010) Behavioural reaction norms: animal personality meets individual plasticity. *Trends in Ecology & Evolution*, **25**, 81–89.
Felleki, M., Lee, D., Lee, Y., Gilmour, A.R. & Rönnegård, L. (2012) Estimation of breeding values for mean and dispersion, their variance and correlation using double hierarchical generalized linear models. *Genetics Research*, **94**, 307–317.
Geiler-Samerotte, K.A., Bauer, C.R., Li, S., Ziv, N., Gresham, D. & Siegal, M.L. (2013) The details in the distributions: why and how to study phenotypic variability. *Current Opinion in Biotechnology*, **24**, 1–18.
Gelman, A. & Hill, J. (2007) *Data Analysis Using Regression and Multilevel/Hierarchical Models*. Cambridge University Press, New York, NY, USA.
Gilmour, A.R., Gogel, B.J., Cullis, B.R. & Thompson, R. (2009) *ASReml user guide release 3.0*. VSN International Ltd., Hemel Hempstead, UK.
Hill, W.G. (1984) On selection among groups with heterogeneous variance. *Animal Production*, **39**, 473–477.
Hill, W.G. & Mulder, H. (2010) Genetic analysis of environmental variation. *Genetics Research*, **92**, 381–395.
Hill, W.G. & Zhang, X.S. (2004) Effects on phenotypic variability of directional selection arising through genetic differences in residual variability. *Genetical Research*, **83**, 121–132.
Hoffman, L. (2007) Multilevel models for examining individual differences in within-person variation and covariation over time. *Multivariate Behavioral Research*, **42**, 609–629.
Houle, D. (1992) Comparing evolvability and variability of quantitative traits. *Genetics*, **130**, 195–204.
Jones, K.S., Nakagawa, S. & Sheldon, B.C. (2009) Environmental sensitivity in relation to size and sex in birds: meta-regression analysis. *The American Naturalist*, **174**, 122–133.
Kizilkaya, K. & Tempelman, R.J. (2005) A general approach to mixed effects modeling of residual variances in generalized linear mixed models. *Genetics, Selection, Evolution: GSE*, **37**, 1–26.
Lee, Y. & Nelder, J.A. (1996) Hierarchical generalized linear models (with discussion). *Journal of the Royal Statistical Society Series B*, **58**, 619–678.
Lee, Y.J. & Nelder, J.A. (2002) Analysis of ulcer data using hierarchical generalized linear models. *Statistics in Medicine*, **21**, 191–202.
Lee, Y. & Nelder, J.A. (2006) Double hierarchical generalized linear models (with discussion). *Journal of the Royal Statistical Society, Series C*, **55**, 139–185.
Lee, Y., Nelder, J.A. & Pawitan, Y. (2006) *Generalized Linear Models with Random Effects: Unified Analysis via H-likelihood*. Chapman & Hall/CRC, Boca Raton, Florida, USA.
Lessells, C.M. & Boag, P.T. (1987) Unrepeatable repeatabilities: a common mistake. *The Auk*, **104**, 116–121.
Limpert, E., Stahel, W.A. & Abbt, M. (2001) Log-normal distributions across the sciences: keys and clues. *BioScience*, **51**, 341–352.
Lunn, D.J., Thomas, A., Best, N. & Spiegelhalter, D. (2000) WinBUGS – a Bayesian modelling framework: concepts, structure, and extensibility. *Statistics and Computing*, **10**, 325–337.
Lunn, D., Jackson, C., Best, N., Thomas, A. & Spiegelhalter, D. (2013) *The BUGS Book – A Practical Introduction to Bayesian Analysis*. CRC Press/Chapman & Hall, Boca Raton, FL, USA.
Matich, P., Heithaus, M.R. & Layman, C.A. (2011) Contrasting patterns of individual specialization and trophic coupling in two marine apex predators. *Journal of Animal Ecology*, **80**, 294–305.
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.
Plummer, M. (2003). JAGS: A program for analysis of Bayesian graphical models using Gibbs sampling. *Proceedings of the 3rd International Workshop on Distributed Statistical Computing*, Vienna, Austria
Pruitt, J.N., DiRienzo, N., Kralj-Fišer, S., Johnson, J.C. & Sih, A. (2011) Individual-and condition-dependent effects on habitat choice and choosiness. *Behavioral Ecology & Sociobiology*, **65**, 1987–1995.
Rast, P., Hofer, S.M. & Sparks, C. (2012) Modelling individual differences in within-person variation of negative and positive affect in a mixed effects location scale model using BUGS/JAGS. *Multivariate Behavioral Research*, **47**, 177–200.
Raudenbush, S.W. & Bryk, A.S. (1987) Examining correlates of diversity. *Journal of Educational and Behavioral Statistics*, **12**, 241–269.
Réale, D. & Dingemanse, N.J. (2008) Personality and individual social specialisation. *Social Behaviour: Genes, Ecology and Evolution* (eds T. Székely, A.J. Moore & J. Komdeur), pp. 417–441. Cambridge University Press, New York, USA.
Réale, D., Reader, S.M., Sol, D., McDougall, P.T. & Dingemanse, N.J. (2007) Integrating animal temperament within ecology and evolution. *Biological Reviews*, **82**, 291–318.
Rönnegård, L., Shen, X. & Alam, M. (2010b) hglm: a package for fitting hierarchical generalized linear models. *The R Journal*, **2**, 20–28.
Rönnegård, L., Felleki, M., Fikse, F., Mulder, H.A. & Strandberg, E. (2010a) Genetic heterogeneity of residual variance-estimation of variance components using double hierarchical generalized linear models. *Genetics, Selection, Evolution: GSE*, **42**, 8.
Rönnegård, L., Felleki, M., Fikse, W.F., Mulder, H.A. & Strandberg, E. (2013) Variance component and breeding value estimation for genetic heterogeneity of residual variance in Swedish Holstein dairy cattle. *Journal of Dairy Science*, **96**, 2627–2636.
SanCristobal-Gaudy, M., Elsen, J.M., Bodin, L. & Chevalet, C. (1998) Prediction of the response to a selection for canalisation of a continuous trait in animal breeding. *Genetics, Selection, Evolution: GSE*, **30**, 423–451.
Schielzeth, H. & Nakagawa, S. (2013) Nested by design: model fitting and interpretation in a mixed model era. *Methods in Ecology & Evolution*, **4**, 14–24.
Sih, A., Bell, A. & Johnson, J.C. (2004) Behavioral syndromes: an ecological and evolutionary overview. *Trends in Ecology & Evolution*, **19**, 372–378.
Sonesson, A.K., Ødegård, J. & Rönnegård, L. (2013) Genetic heterogeneity of within-family variance of body weight in Atlantic salmon (*Salmo salar*). *Genetics, Selection, Evolution: GSE*, **45**, 41.
Stamps, J.A., Briffa, M. & Biro, P.A. (2012) Unpredictable animals: individual differences in intraindividual variability (IIV). *Animal Behaviour*, **83**, 1325–1334.
VSN International (2011) *GenStat for Windows*, 14th edn. VSN International, Hemel Hempstead, UK. Available at: www.GenStat.co.uk.
Westneat, D.F., Schofield, M. & Wright, J. (2013) Parental behavior exhibits among-individual variance, plasticity, and heterogeneous residual variance. *Behavioral Ecology*, **24**, 598–604.
Westneat, D.F., Wright, J. & Dingemanse, N.J. (2014) The biology hidden inside residual within-individual phenotypic variation. *Biological Reviews*. doi:10.1111/brv.12131
Yang, Y., Christensen, O.F. & Sorensen, D. (2011) Analysis of a genetically structured variance heterogeneity model using the Box–Cox transformation. *Genetics Research*, **93**, 33–46.

Received 15 March 2014; accepted 28 August 2014
Handling Editor: Jarrod Hadfield

Supporting Information

Additional Supporting Information may be found in the online version of this article.

Data S1. Full DHGLM.

Data S2. Model with half cauchy variances.

Data S3. Model with independent variances.

Data S4. Model with log normal variances.

Data S5. Model with log normal variances and including group-level predictor.

Data S6. R script for DHGLMs.

Data S7. Random intercept model.

Appendix S1. Quantifying the predictability of behaviour: statistical approaches for the study of between-individual variation in the within-individual variance.

Appendix S2. The link between normalmixed models and HGLM.