**Methods in Ecology and Evolution**

# The simulation extrapolation technique meets ecology and evolution: A general and intuitive method to account for measurement error

Erica Ponzi[1,2], Lukas F. Keller[1,3] and Stefanie Muff[1,2]

Running title: **P-SIMEX to correct for pedigree error**

[1] Department of Evolutionary Biology and Environmental Studies, University of Zurich, Winterthurerstrasse 190, CH-8057 Zurich, Switzerland
[2] Department of Biostatistics, Epidemiology, Biostatistics and Prevention Institute, University of Zurich, Hirschengraben 84, CH-8001 Zurich, Switzerland
[3] Zoological Museum, University of Zurich, Karl Schmid-Strasse 4, CH-8006 Zurich, Switzerland

⋆ **Corresponding author:**
Stefanie Muff,
Department of Evolutionary Biology and Environmental Studies, University of Zürich, Winterthurerstrasse 190, 8057 Zürich, Switzerland
Department of Biostatistics, Epidemiology, Biostatistics and Prevention Institute, University of Zürich, Hirschengraben 84, 8001 Zürich, Switzerland
E-mail: stefanie.muff@uzh.ch

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1111/2041-210X.13255

# The simulation extrapolation technique meets ecology and evolution: A general and intuitive method to account for measurement error

Running title: **Simulation Extrapolation in Ecology and Evolution**

## Abstract

**1.** Measurement error and other forms of uncertainty are commonplace in ecology and evolution and may bias estimates of parameters of interest. Although a variety of approaches to obtain unbiased estimators are available, these usually require the formulation of an explicit (parametric) model for the error-prone variable, and a latent model for the unobserved (latent) error-free variable. In practice this is often difficult.

**2.** We propose to generalize the simulation extrapolation (SIMEX) technique, a heuristic approach to correct for measurement error, to situations where it is difficult to explicitly formulate an error model or latent model for a variable of interest. We illustrate the idea with the example of error in pedigrees. Pedigree errors cause error in estimates of inbreeding coefficients and the relatedness matrix, thus biasing estimates of inbreeding depression or heritability. Instead of formulating error models for inbreeding coefficients or the relatedness matrix, we directly apply the SIMEX idea to the pedigree. The initially known error proportion in the pedigree is progressively increased, all models are refitted, and the observed trend in the quantities of interest is extrapolated back to a hypothetical error-free pedigree to obtain bias-corrected estimates. We tested this pedigree-SIMEX (PSIMEX) method with simulated pedigrees and with data from a free-living population of song sparrows.

**3.** The simulation study indicates that the PSIMEX estimator is almost unbiased for inbreeding depression and heritability, and that it has a much lower mean squared error (MSE) than the naive estimator. In the application to the song sparrows, the error-corrected results could be validated against the actual values thanks to the availability of both an error-prone and an error-free pedigree. The results indicate that bias and MSE are reduced by PSIMEX. For easy accessibility of the method, we provide the R-package `PSIMEX`.

**4.** By transferring the SIMEX philosophy to error in pedigrees, we have illustrated how this heuristic approach can be generalized to situations where explicit error models are difficult to formulate. Thanks to the simplicity of the idea, many other error problems in ecology and evolution might be amenable to SIMEX-like error correction methods.

**Keywords:** Heritability, inbreeding coefficient, inbreeding depression, misassigned paternities, pedigree reconstruction, relatedness, SIMEX, uncertainty.

# Introduction

Measurement error and other forms of uncertainty in variables of interest are commonplace in ecology and evolution, and there is thus a need for methods and practical tools to account for such errors in statistical models (see *e. g.*, Solow, 1998; Macgregor et al., 2006; Reid et al., 2014; Steinsland et al., 2014; Wright et al., 2017; Mason et al., 2018). Measurement error can arise from countless sources in a wide range of studies, for example in the form of location error in telemetry and animal movement research (Montgomery et al., 2011; McClintock et al., 2014), error during the collection of phenotypic data (Hoffmann, 2000; Dohm, 2002; Macgregor et al., 2006; van der Sluis et al., 2010; Ge et al., 2017), misclassification in detection models and capture-recapture studies (Guillera-Arroita et al., 2014; Guélat and Kéry, 2018), or error caused by spatial variability or uncertainty in the observation of climate variables (Bishop and Beier, 2013; Stoklosa et al., 2014) or biodiversity metrics (Haila et al., 2014; Mason et al., 2018).

When variables measured with error or estimated with uncertainty are used as explanatory variables in statistical analyses, parameter estimates may be biased (*e. g.*, Fuller, 1987). To obtain unbiased parameter estimates, statistical models need to account for measurement error (see for example Gustafson, 2004; Carroll et al., 2006; Buonaccorsi, 2010, for an extensive treatment of frequentist and Bayesian measurement error correction techniques). Error correction methods always require that the error mechanism is known, which typically means that the error distribution and its parameters (*e. g.*, the error variance or the misclassification rate) must be specified prior to correcting for the error. In addition, some techniques require that latent (so-called "exposure") models specify the distributions of the unobserved (latent), true variables, in particular when errors are modeled in a Bayesian framework (Muff et al., 2015; Ponzi et al., 2018). However, the error-generating mechanisms that blur true variables can be rather complex, and specifying a model for the unobserved variables may also not be very straightforward. Consequently, it can be difficult or even practically impossible to formulate and fit a fully parametric model to obtain error-corrected estimates.

A general, heuristic method to correct for measurement error, which is also applicable in situations where a latent model for the unobserved variable is missing, is the simulation extrapolation (SIMEX) method. SIMEX was originally introduced by Cook and Stefanski (1994) to correct for measurement error in continuous covariates of regression models, and was later expanded to account for a broader range of regression models and error structures, such as non-additive error models (Eckert et al., 1997), misclassification error in discrete covariates or the response of regression models (Kuechenhoff et al., 2006), or to heteroschedastic error in covariates (Devanarayan and Stefanski, 2002). SIMEX is based on the rationale that more error leads to more bias in the estimated regression coefficients, and that progressively adding more error can reveal a pattern of the magnitude of the bias in dependence of the magnitude of the error. Based on this pattern, the algorithm extrapolates in the direction of less error, until the error-free estimate is reached. Thanks to its straightforward implementation without the need to formulate a model for the unobserved variable and its intuitive interpretation, SIMEX has been used extensively, with some applications also in ecology (*e. g.*, Solow, 1998; Gould et al., 1999; Hwang and Huang, 2003; Melbourne and Chesson, 2006). The possibility to cover a wide range of statistical models and error structures renders SIMEX a very general approach.

The main goal of this paper is to illustrate how the SIMEX approach can be further generalized to situations where it is not only difficult to specify the latent model for the unobserved variable, but also to find an explicit (*e. g.*, parametric) model for the error in a specific covariate. This occurs, for example, when the error mechanism does not directly act on a certain variable, but on a lower level of the data. The example that provided the motivation for this work is error in individual-specific values of inbreeding, given by the inbreeding coefficient $f$. Inbreeding coefficients are often used as a covariate in regression models to estimate the magnitude of the (usually negative) effects of inbreeding on fitness-related traits, denoted as *inbreeding depression* (Keller and Waller, 2002; Charlesworth and Willis, 2009). Inbreeding coefficients can be derived from pedigrees, but pedigrees are often erroneous. In free-living organisms, one of the major sources of pedigree error are incorrect paternities, when observed (social) behavior is used as a basis to assess parentage, but extra-pair paternities obscure the actual (genetic) relationships, leading to *misassigned paternities* (Keller et al., 2001; Griffith et al., 2002; Senneke et al., 2004; Jensen et al., 2007). These misassignments do not only affect the relatedness estimates of parents with their offspring, but all relatedness estimates among their descendants and their relatives. Consequently, the pedigree-based relatedness matrix $\mathbf{A}$ and the inbreeding coefficients of the individuals contain errors. We thus expect biased estimators for inbreeding depression (Keller et al., 2002; Visscher et al., 2002; Reid et al., 2014), but also for quantitative genetic measures that rely on the correctness of $\mathbf{A}$, most prominently estimators of heritability (Keller et al., 2001; Senneke et al., 2004; Charmantier and Reale, 2005).

While it is difficult to formulate explicit parametric error models for inbreeding coefficients or relatedness matrices when there are misassigned paternities, it is relatively straightforward to increase the error at the pedigree level and to repeatedly estimate the quantitative genetic measures with different levels of error. This is where the SIMEX idea enters: Instead of increasing the error variance of a continuous covariate as in the traditional SIMEX, we start from a known proportion of misassigned paternities and then successively increase this proportion in the pedigree to obtain information about the bias in quantitative genetic estimates (such as inbreeding depression or heritability) as the pedigree error is aggravated. In a second step, the observed trend upon increasing the error proportion is extrapolated back to that of a hypothetical error-free pedigree. This algorithm, which we will refer to as pedigree-SIMEX (PSIMEX) in the following, circumvents the formulation of an error model at the level of the inbreeding coefficients or the relatedness matrix. The only prerequisite to apply PSIMEX is that the proportion of misassigned paternities, as well as their distribution in the actual pedigree (*e. g.*, proportions varying over time), are known.

Here, we test the validity of the PSIMEX approach with different simulated pedigree topologies, and show that the method can substantially reduce or eliminate the bias in estimates of heritability and inbreeding depression. We then apply the PSIMEX algorithm to an empirical data set from a population of song sparrows, where apparent paternities (observed from social behavior) and actual (genetic) paternities are not always corresponding. Since paternities were determined both socially and genetically in this population, we were able to compare the PSIMEX estimates of heritability and inbreeding depression derived from the apparent pedigree to the estimates derived from the actual pedigree. Our application to the song sparrow data suggests that the PSIMEX method performs well not only in simulations but also with real field data. To facilitate the use of PSIMEX, we provide the novel R-package `PSIMEX` (Ponzi, 2017).

# Theory

## The original SIMEX algorithm

We start by outlining how SIMEX works in its simplest form, as originally proposed by Cook and Stefanski (1994). Assume that a continuous variable of interest $x$ is blurred by *classical additive measurement error*, such that only $w = x + u$ can be observed, where the measurement error $u$ is assumed to be independent of the latent variable $x$ and distributed as $u \sim \mathsf{N}(0, \sigma_u^2)$ with error variance $\sigma_u^2$. Further assume that $w$ instead of the unobservable $x$ is used as a covariate in a simple linear regression model, $y = \alpha + \beta_w w + \epsilon$. This is a typical measurement error, or errors-in-variables problem, known to lead to a biased regression parameter estimate, whenever $\sigma_u^2 > 0$ (Fuller, 1987; Carroll et al., 2006). Using $\sigma_w^2 = \sigma_x^2 + \sigma_u^2$ and the assumption that the error $u$ is independent of $x$, it is quite straightforward to see that the error-prone regression parameter $\beta_w$ is an estimator of

$$\beta_w = \frac{\sigma(w,y)}{\sigma_w^2} = \frac{\sigma(x,y)}{\sigma_x^2 + \sigma_u^2} \ ,$$

which is less than the true slope $\beta_x$ (from the regression of $y$ on $x$) by an *attenuation* factor $\lambda = \sigma_x^2/(\sigma_x^2 + \sigma_u^2)$, so that $\beta_w = \lambda\beta_x$. Although measurement error leads to underestimated effect sizes in this simple linear regression case, attenuation is not a general pattern. Even relatively standard regression models may yield upwardly biased parameter estimates, for example when an error-prone covariate is correlated with another covariate, in the presence of interactions, or in probit regression (Carroll et al., 2006; Freckleton, 2011; Muff and Keller, 2015).

To obtain estimates of the true slope $\beta_x$ instead of the biased $\beta_w$, the SIMEX algorithm is based on the heuristic that more error will generally lead to more bias. By systematically increasing the error in a simulation (SIM) step and then extrapolating (EX) the pattern of change in parameter estimates with increasing error backward, one approximates the parameter that one would obtain if there was no error in the data. Fig. 1 depicts the SIMEX idea. In the case of classical additive measurement error in a continuous covariate introduced above, the error variance $\sigma_u^2$ is artificially increased by adding more random error to the covariate of interest. For each error level (*i. e.*, each predefined increase of the error variance), the procedure is iterated $B$ times and regression parameters and standard errors are estimated and stored for each iteration. In the extrapolation phase, the observed trend upon increasing error is extrapolated in the direction of less error, and an error-corrected SIMEX estimate is obtained for zero error. The choice of the functional form for the extrapolation function is important, but it has been shown that, under rather general conditions, a function exists that leads to consistent estimators of the true parameters (Stefanski and Cook, 1995). Unfortunately, this function is rarely known, which is why SIMEX remains partially a heuristic method. In many applied problems, however, linear and quadratic functions have been shown to give approximately consistent and numerically stable results (*e. g.*, Kuechenhoff and Carroll, 1997; Carroll et al., 2006, pp. 108-110). Standard errors for error-corrected estimates are also obtained by an extrapolation step, using the same functional form for the extrapolation as for the point estimates, thereby accounting for the sampling error of the $B$ simulations plus the standard errors that are obtained from each of the $B$ regressions. Details about the computation of the standard error in the SIMEX algorithm are given in Appendix 1. In addition, the reader is referred to Stefanski and Cook (1995) and Apanasovich et al. (2009). Alternatively, the bootstrap could be used to obtain standard errors, but this is computationally extremely costly, because the application of SIMEX itself already requires that the model is iteratively refitted many times (Apanasovich et al., 2009).

To apply the SIMEX algorithm and to know when zero error is reached, the initial error *model* (*e. g.*, $u \sim \mathsf{N}(0, \sigma_u^2)$) and error model *parameter(s)* (*e. g.*, the value of $\sigma_u^2$) must be known, while we do *not* need a latent model for the unobserved covariate $x$ itself (*e. g.*, $x \sim \mathsf{N}(0, \sigma_x^2)$). The error model defines the mechanism according to which more error must be generated in the simulation phase, and the parameter value defines the fixed point from where back-extrapolation begins. Incorrect assumptions about the error model or error parameter(s) may result in a biased simulation mechanism and an incorrect extrapolation function and, therefore, may yield parameter estimates that are over or under-corrected. In the worst case, correction may be in the wrong direction. Thus, to use SIMEX – or, indeed, any measurement error correction technique – it is crucial to have a good knowledge of the error model and the error model parameters, obtained for example through repeated measurements or validation data.

SIMEX is a rather universal method that results in a substantial bias reduction in the estimators in the majority of applications. Successful bias reduction has been demonstrated for various error mechanisms in linear, logistic or log-linear mean models, also in the presence of random effects (*e. g.*, Wang et al., 1998; Apanasovich et al., 2009). However, there are some particular cases, such as regression in the presence of collinearity among predictors or correlation among errors, where the bias reduction is not considered sufficient and other methods might be preferable (*e. g.*, Fung and Krewski, 1999; Hwang and Huang, 2003).

## Extensions of SIMEX

Since the original contribution by Cook and Stefanski (1994), SIMEX has been extended to account for different types of errors and models. Examples include non-additive error models (Eckert et al., 1997), heteroschedastic error variances (Devanarayan and Stefanski, 2002), the mixture of classical and Berkson error models (Carroll et al., 2006), clustered data (Lin and Carroll, 2000), or frailty models (Li and Lin, 2003). Another extension is the so-called misclassification-SIMEX (Kuechenhoff et al., 2006), which accounts for error in discrete variables by increasing the misclassification probabilities in the simulation phase, and extrapolating back to zero misclassification. These extensions highlight the flexibility of SIMEX in situations where standard error modelling procedures are challenging to implement. The extension we suggest here covers the case when the error mechanism does not act on the variable itself, but on an underlying structure as in the case of error in the pedigree of a study population.

## SIMEX for pedigrees: PSIMEX

We now describe the extension of the SIMEX idea to the case of pedigree error, denoted as PSIMEX. The procedure starts from an initial, known proportion of misassigned parentages that occur in the pedigree, denoted as $\zeta_I$ (*e. g.*, $\zeta_I = 0.1$ for 10% error), and then randomly generates additional misassigned parentages to obtain error levels $\zeta > \zeta_I$. To this end, randomly selected parents are replaced by connecting the offspring to another individual according to a known error-generating mechanism (*e. g.*, some parents might be more likely to be chosen than others, see below). Importantly, to achieve a desired error proportion $\zeta > \zeta_I$ of misassigned parents, we cannot simply pick a proportion of $\zeta - \zeta_I$ parents and replace them, because this may include parents that were already incorrectly assigned, so that the effective proportion of error would be too low. To account for this circumstance, the actual proportion of additional misassignments needed at each step is calculated from the equation

$$1 - \zeta_X = \frac{1-\zeta}{1-\zeta_I} \ , \tag{1}$$

where $\zeta_X$ is the *effective* error proportion that has to be added to obtain a nominal error level of $\zeta$. As an example, assume that the initial parentage error rate is $\zeta_I = 0.17$ and that the aim is to increase this proportion to $\zeta = 0.30$ (0.17+0.13). If only a proportion of 0.13 of the parental relations is randomly picked and reassigned, some parents that were already misassigned will be randomly misassigned again. The effective error proportion is then less than 0.30. Equation (1) shows that we need to pick and re-assign $\zeta_X = 1 - \frac{1-0.30}{1-0.17} = 0.16$ of the parents to obtain an expected error proportion of 0.30.

For each error proportion $\zeta$ the procedure is repeated a fixed number of $B$ (*e. g.*, 100) times, and for each $b = 1, 2, \ldots, B$, the pedigree is recalculated including the newly misassigned connections. All the relevant quantities, such as inbreeding coefficients and the relatedness matrix, are newly derived each time from the pedigree and are then used to fit a model that estimates the quantity of interest $\hat{\Theta}_b(\zeta)$, such as inbreeding depression or heritability. The algorithm then averages over the $B$ values of $\hat{\Theta}_b(\zeta)$ to obtain the estimate $\hat{\Theta}(\zeta) = B^{-1}\sum_{b=1}^{B}\hat{\Theta}_b(\zeta)$ for error proportion $\zeta$. The procedure is repeated for a sequence of values $\zeta > \zeta_I$, which thus allows to estimate a functional dependence between $\hat{\Theta}(\zeta)$ and $\zeta$. The estimate at zero error $\hat{\Theta}(0)$, obtained by extrapolating in the direction of decreasing error, corresponds to the error-corrected estimate denoted as $\hat{\Theta}_{\text{simex}}$. Algorithm 1 reports a schematic explanation of the method.

Knowledge of the initial error proportion $\zeta_I$ and the mechanisms leading to this value are necessary when applying PSIMEX. An intrinsic assumption of the simulation phase, when additional error is generated, is that the same error-generating mechanism is used as in the observed data. As an example, the percentage of misassigned parents may fluctuate over time with fluctuating population size or sex ratio. Or misassignments might affect solely fathers and occur within the same generation, so it might be logical to assume replacement of only fathers with random individuals chosen from the same generation. Such structural aspects of the error-generating mechanisms must be taken into account in the PSIMEX procedure, because the relationship between error and bias may otherwise not reflect the true trend. Examples of the effects of incorrect assumptions about the initial error proportion or the error structure are given in Appendix 1 (Sections 6 and 7).

In PSIMEX, it is even less straightforward than in the standard SIMEX to derive a theoretical justification for a functional form of the extrapolation function. We therefore rely here on polynomial extrapolation functions that have been found to give stable and approximately consistent results in many setups (*e. g.*, Kuechenhoff and Carroll, 1997; Kuechenhoff et al., 2006). We describe how we select among them in later sections of the manuscript.

```
Simulation step:
for each paternity error rate $\zeta$ do
    for B number of simulations do
        select a proportion of $\zeta_X$ individuals in the pedigree according to (1);
        replace their parents with random individuals according to the assumed error mechanism;
        recalculate the new pedigree;
        derive quantities of interest from the new pedigree;
        estimate $\hat{\Theta}_b(\zeta)$ using the quantities from the new pedigree;
    end
    average across simulations $\hat{\Theta}(\zeta) = B^{-1} \sum_{b=1}^{B} \hat{\Theta}_b(\zeta)$;
end

Extrapolation step:
Estimate the functional dependency between $\hat{\Theta}(\zeta)$ and $\zeta$;
$\hat{\Theta}_{\text{simex}} = \hat{\Theta}(0)$
```

**Algorithm 1:** PSIMEX procedure

## Quantitative genetic measures

To illustrate how the PSIMEX idea works, we apply it to the estimation of two important quantitative genetic parameters, heritability and inbreeding depression. Heritability quantifies the proportion of phenotypic variance in a trait that is due to additive genetic factors, which is for example relevant to predict the response to selection (Lynch and Walsh, 1998). Inbreeding depression quantifies the reduction in fitness of offspring resulting from matings among relatives, which is key to understanding mating system evolution and dispersal (Keller and Waller, 2002; Charlesworth and Willis, 2009).

Both quantities can be estimated by fitting (generalized) linear mixed models to phenotypic or fitness data, using the so called *animal model* (Henderson, 1976; Lynch and Walsh, 1998; Kruuk, 2004). For a continuous trait $y_{ij}$ with (repeated) measurements $j$ for individuals $i$, the animal model can be written as

$$y_{ij} = \mu + \boldsymbol{x}_{ij}^{\top}\boldsymbol{\beta} + \beta_f f_i + a_i + id_i + R_{ij} \ , \tag{2}$$

where $\mu$ is the population mean, $\boldsymbol{x}_{ij}$ is the vector of covariates for individual $i$ at the $j^{\text{th}}$ measurement occasion, $\boldsymbol{\beta}$ is a vector of fixed effects, $f_i$ is the inbreeding coefficient of individual $i$, which reflects how related an animal's parents are, and $\beta_f$ is the respective fixed effect, with a negative slope indicating the presence of inbreeding depression (Lynch and Walsh, 1998). The last three components of Equation (2) are random effects, namely the additive genetic effects (or breeding values) $a_i$ with dependency structure $(a_1, \ldots, a_n)^T \sim \mathsf{N}(0, \sigma_A^2 \mathbf{A})$, the independent effects for the animal identity $id_i \sim \mathsf{N}(0, \sigma_{PE}^2)$ accounting for the permanent environmentally-induced differences among individuals, and an independent Gaussian residual term $R_{ij} \sim \mathsf{N}(0, \sigma_R^2)$ that captures the remaining (unexplained) variability. The dependency structure of the breeding values $a_i$ is given by the additive genetic relatedness matrix $\mathbf{A}$ (Lynch and Walsh, 1998), which, four our purposes here, is derived from the pedigree. The entry $(i, j)$ of $\mathbf{A}$ is $2 \cdot \theta_{ij}$, where $\theta_{ij}$ is the probability that an allele drawn at random from individual $i$ will be identical by descent to an allele drawn at random from individual $j$, also known as the coefficient of co-ancestry. In the same framework, the narrow-sense heritability $h^2$ is defined as

$$h^2 = \frac{\sigma_A^2}{\sigma_A^2 + \sigma_{PE}^2 + \sigma_R^2} \ . \tag{3}$$

Not all components of the animal model (2) are always necessary. For example, the breeding values $a_i$ are often omitted when estimating inbreeding depression, although this can result in biased estimates of inbreeding depression and is thus not recommended (Reid et al., 2008; Becker et al., 2016). Also, inbreeding $f_i$ is not always included as a covariate when estimating heritability, although it is generally recommended to account for potentially higher phenotypic similarity between individuals with similar levels of inbreeding (Reid et al., 2006; Reid and Keller, 2010), but also because ignoring the effects of inbreeding may lead to biases in estimates of the additive genetic variance $\sigma_A^2$ (Wolak and Keller, 2014). On the other hand, model (2) can be further expanded to include additional variance components, such as the maternal variance $\sigma_M^2$, or nest or time effects (Kruuk and Hadfield, 2007; Wilson et al., 2010). Moreover, it can be formulated as a generalized linear mixed model (GLMM), for example for binary or count traits. We will use a binary GLMM in our application to the song sparrow data below.

In the presence of misassigned parentages in the pedigree, the relatedness matrix $\mathbf{A}$ will necessarily suffer from error, which, in turn, may bias the estimates of the variance components and thus the estimates of heritability $h^2$. Similarly, the accurate quantification of the individual-specific inbreeding coefficients $f_i$ will be hampered by misassigned parentages in the pedigree, which will result in biased estimates of inbreeding depression $\beta_f$. Note that, for the scope of this paper, we assume the error to be present solely in the form of pedigree error. Therefore, the only entities in model (2) that suffer from error are the covariate $f_i$ and the matrix $\mathbf{A}$, while all covariates that are not depending on knowledge of the pedigree are assumed to be observed correctly. Without the PSIMEX approach we would need to formulate error models for $f_i$ and $\mathbf{A}$, which may be very challenging. In fact, the error in the inbreeding coefficients does not follow any standard distribution, and the direction and magnitude of the error seems to depend on the true value of $f_i$ (see Section 2.1 in Appendix 1), which is in contrast to the standard assumptions of most error modeling techniques. It is even less clear how an error model for the entries in $\mathbf{A}$ would look like. Increasing the error in the pedigree via the PSIMEX algorithm is therefore a very convenient and practical alternative.

# Simulation study

To illustrate and test the PSIMEX procedure we carried out a simulation study with a set of simulated pedigrees, generated using the function `generatePedigree()` from the R package `GeneticsPed` (Gorjanic and Henderson, 2007). We then introduced increasing proportions of misassignments of fathers, which we replaced with a randomly chosen male individual from the same generation, using a constant misassignment probability across all generations in the pedigree. After quantifying the effects of the errors on estimates of heritability and inbreeding depression, we applied the PSIMEX procedure to obtain error-corrected estimates of the two measures, starting from different initial error proportions.

## Effects on heritability

To understand the effect of pedigree error on heritability, we generated a total of 50 different pedigrees, each with 50 mothers and 50 fathers in a total of 30 generations. We simulated phenotypic traits $y_i$ for each individual $i$ in each pedigree using the animal model

$$y_i = \mu + \beta_{\text{sex}} sex_i + a_i + R_i \ , \tag{4}$$

with $\mu = 10$, $\beta_{\text{sex}} = 2$, $(a_1, \ldots, a_n)^\top \sim \mathsf{N}(0, \sigma_A^2 \mathbf{A})$ and independent $R_i \sim \mathsf{N}(0, \sigma_R^2)$, using $\sigma_A^2 = 0.3$ and $\sigma_R^2 = 0.1$, which corresponds to a heritability of $h^2 = \sigma_A^2/(\sigma_A^2 + \sigma_R^2) = 0.75$. Breeding values $a_i$ were generated with the `rbv()` function from the R package `MCMCglmm` (Hadfield, 2010) that accounts for the dependency structure given by the relatedness matrix $\mathbf{A}$ from the pedigree.

For each pedigree, we then disrupted paternal assignments to obtain error proportions $\zeta$ ranging from 0 to 1.0 in steps of 0.1, and the procedure was repeated 100 times for each error level and each pedigree. Estimates (posterior modes) of $h^2$ were obtained by fitting model (4) with a Bayesian approach. For efficiency reasons, we used integrated nested Laplace approximations (INLA, Rue et al., 2009; Holand et al., 2013), but posterior distributions could alternatively be obtained from Markov chain Monte Carlo (MCMC) samples, for example by using the popular `MCMCglmm` package. For each pedigree and each error level, the naive $\hat{h}^2$ was extracted together with its standard error. Following Wilson et al. (2010, Supplementary File 5), inverse gamma priors were used for all variance components, namely $\mathsf{IG}(1/2, \sigma_P^2/2)$, where $\sigma_P^2$ is the total phenotypic variance of the trait, but the results were not sensitive to the prior choice. For example, we obtained the same results when assigning $\sigma_A^2 \sim \mathsf{IG}(1/2, \sigma_P^2/3)$ and $\sigma_R^2 \sim \mathsf{IG}(1/2, \sigma_P^2/6)$, which reflects a prior belief that a larger proportion of variance is captured by the additive genetic component rather than by the environment. We also repeated the whole procedure with different heritabilities, namely $h^2 = 0.50$ and $h^2 = 0.25$, by increasing the values of the environmental variance to $\sigma_E^2 = 0.3$ and $\sigma_E^2 = 0.9$, respectively, while holding the remaining parameters constant.

## Effects on inbreeding depression

Given that the relatedness structure and inbreeding in a population depend – among other things – on the effective population size and the variance in reproductive success (Reid and Keller, 2010), it seemed likely that the topology of the pedigree would influence the effects of pedigree error on inbreeding depression. We therefore considered three different pedigree topologies that differed in reproductive skew and, hence, induced different levels of inbreeding in a population by simulating pedigrees with varying $N_e/N_c$ ratios, where $N_e$ is the effective population size and $N_c$ is the census size (Frankham, 1995; Palstra and Fraser, 2012). The $N_e/N_c$ ratios was modified by changing the number of reproductive males and females per generation, while keeping the population size in each generation constant at 100. In the first case, both numbers of fathers and mothers per generation were set to 50, which led to $N_e/N_c = 1$, so that all individuals had offspring. As a consequence, the average degree of relatedness among individuals was comparatively low, implying a low average inbreeding coefficient throughout generations. The second case corresponded to pedigrees with only 15 mothers and fathers per generation ($N_e/N_c = 0.3$) and the third case to pedigrees with only 5 mothers and fathers per generation ($N_e/N_c = 0.1$). As the 100 individuals of a generation thus originated from only 30 or 10 parents, respectively, the mean relatedness among individuals increased, and their matings thus led to more highly inbred offspring. Note that this change in pedigree structure does not affect heritability estimates, which only depend on the initial variance components and not on the topology of the pedigree.

Inbreeding coefficients $f_i$ were derived from the pedigree using the `calcInbreeding()` function from the R package `pedigree` (Coster, 2013). Fitness traits $y_i$ for individual $i$ were then simulated according to

$$y_i = \mu + \beta_{\text{sex}} sex_i + \beta_f f_i + R_i \ , \tag{5}$$

with a population mean of $\mu = 10$, sex effect of $\beta_{\text{sex}} = 2$, coefficient for inbreeding depression $\beta_f = -7$, and a residual term $R_i \sim \mathsf{N}(0, 0.1)$. We generated 100 distinct error-free pedigrees for each of the three pedigree topologies, and each pedigree was then disrupted with error proportions ranging from $\zeta = 0$ to 1.0 in steps of 0.1. This procedure was repeated 100 times for each error level and each originally generated pedigree, and in each iteration, the linear regression model (5) was fitted in a standard likelihood framework, using as covariates the inbreeding coefficients $f_i$ that were derived from the erroneous pedigrees. The estimated inbreeding depression $\hat{\beta}_f$ was stored in each iteration.

## Application of PSIMEX to simulated pedigrees

To assess the performance of the PSIMEX algorithm in recovering error-free estimates of heritability and inbreeding depression, we selected one erroneous pedigree for each of the 100 that were initially generated for the analyses of the previous two subsections, and applied PSIMEX to the selected pedigree starting from initial error proportions of $\zeta_I = 0.1, 0.2, 0.3$ and 0.4.

In each case, we provided the algorithm with the known initial error proportion $\zeta_I$ and correct assumptions about the error mechanism, that is, random replacement of fathers with male individuals from the same generations and with constant misassignment probability across all generations in the pedigree. During the simulation phase, error was increased according to error proportions ranging from $\zeta_I$ to 1 in steps of 0.1, and inbreeding coefficients and the relatedness matrix were derived each time from the respective (erroneous) pedigree and then used to fit the model to estimate the quantity of interest with $B = 100$ iterations per error level. Error-corrected estimates $\hat{\Theta}_{\text{simex}}$ for $h^2$ and $\beta_f$ were obtained by fitting a linear, a quadratic and a cubic extrapolation function to the trend upon increasing error proportions, and the function with the lowest AICc was retained. The results were compared to the corresponding naive estimates $\hat{\Theta}_{\text{naive}}$ that were obtained using the error-prone pedigrees to derive $f_i$ and $\mathbf{A}$. The code for all simulations is given in Appendices 2 and 3.

# Simulation results

## The effects of pedigree error on heritability

Misassigned paternity error caused a clear decreasing trend in estimates of heritability, irrespective of the actual value of the heritability ($h^2 = 0.75, 0.50$, or $0.25$), with a continuous decline in the estimates as the error proportions increased (see Fig. 2 for $h^2 = 0.75$ and Figs S4 and S5 in Appendix 1 for the other cases). This is a consequence of the decrease in the estimates of $\sigma_A^2$ due to the increasing pedigree error, which causes information loss in the relatedness matrix $\mathbf{A}$.

In our simulations, all mothers were always correctly assigned to their offspring. Thus, even when all paternities were assigned randomly ($\zeta = 1$), we do not expect heritability estimates of $h^2 = 0$, since at least half of the parent-offspring pairs in the pedigree were still correct. With all mothers assigned correctly and all fathers assigned incorrectly, one would expect heritability estimates to equal half the true heritability (the heritability estimate from a mother-offspring regression, Lynch and Walsh, 1998). This expectation was confirmed in the case of $h^2 = 0.50$ and $h^2 = 0.25$, where replacing all fathers led to average estimates of heritability equal to $h^2 = 0.26$ and $h^2 = 0.11$, respectively. However, for $h^2 = 0.75$, we obtained an average heritability of $h^2 = 0.44$ (95% quantile interval from 0.40 to 0.49) for $\zeta = 1$. This is higher than the expected $h^2 = 0.375$, presumably because the misassigned fathers were, on average, still related to the true fathers due to the small population size (100 individuals per generation). We confirmed this interpretation by reducing the relatedness between true and randomly assigned fathers through increasing the number of individuals in each generation to 1000 in each pedigree. Doing so resulted in an average heritability estimate close to expectations ($h^2 = 0.37$, 95% quantile interval from 0.34 to 0.40).

## The effects of pedigree error on inbreeding depression

Interestingly, the trend in estimates of inbreeding depression depended on the structure of the simulated pedigree (Fig. 3). Adding misassigned paternities to pedigrees with $N_e/N_c = 1$ (low average inbreeding) led to underestimation of inbreeding depression, with a trend towards zero as the error proportion increased (Fig. 3a). An opposite pattern was observed in pedigrees with $N_e/N_c = 0.1$ (high average inbreeding), where the estimated inbreeding depression became stronger with increasing paternity error (Fig. 3c). Intermediate pedigrees with $N_e/N_c = 0.3$ showed an initial trend of increased inbreeding depression, followed by a decrease and a trend towards zero, resulting in a U-shaped pattern (Fig. 3b). The differences among these patterns may be due to the fact that the pedigree structure affects the consequences of randomly replacing fathers in a pedigree. In a pedigree that is characterized by few breeding individuals in the population (*i. e.*, $N_e/N_c = 0.1$) and, hence, high reproductive skew, randomly replacing fathers causes, on average, a decrease in estimated inbreeding coefficients. On the other hand, when reproductive skew is low (*i. e.*, $N_e/N_c = 1$), random replacement of fathers leads to both higher and lower estimates of inbreeding. Consequently, the way in which pedigree error affects estimates of inbreeding depression may depend in complex ways on the pedigree structure.

### PSIMEX estimates of heritability and inbreeding depression

The PSIMEX procedure yielded heritability estimates that were essentially unbiased for initial error proportions of $\zeta_I \leq 0.3$, irrespective of the true value of $h^2$ (Fig. 4a and Tables 1a and S1). Even when starting from an initial error proportion as high as $\zeta_I = 0.4$, the PSIMEX estimators were much less biased than the naive estimators that did not account for the error. Moreover, the PSIMEX estimators had a considerably lower mean squared error (MSE) compared to the naive estimators for all three cases and irrespective of the starting error (Tables 1a and S1). The quadratic extrapolation function had the smallest AICc in 96% of the cases for $h^2 = 0.75$, 47% for $h^2 = 0.50$ and 32% for $h^2 = 0.25$, respectively, and was thus chosen to obtain the PSIMEX estimate. The linear function was selected in in no case for $h^2 = 0.75$, 47% of the cases for $h^2 = 0.50$ and 62% for $h^2 = 0.25$. The cubic function was selected in the remaining cases.

PSIMEX also yielded estimates of inbreeding depression that were much closer to the simulated true values for all three pedigree topologies and initial error proportions (Fig. 4b and Tables 1b and S2). Even though the PSIMEX estimators were less consistent for larger $\zeta_I$, they always considerably reduced the bias with respect to the corresponding naive estimators that ignored the error. As with estimates of heritability, the MSE was always much lower for the PSIMEX than for the naive estimators (Tables 1b and S2). The cubic extrapolation function had the smallest AICc and was thus chosen in 58%, 19% and 100% of the cases for pedigrees with $N_e/N_c = 1$, $N_e/N_c = 0.3$, and $N_e/N_c = 0.1$, respectively, whereas a quadratic extrapolation function was selected in the remaining cases.

## Empirical example: Song sparrows

### Study population

We also applied the PSIMEX approach to data from empirical study of a population of free-living song sparrows (*Melospiza melodia*) on Mandarte Island, Canada. Over more than 40 years, 31 generations of song sparrows have been monitored, yielding a pedigree of 6095 individuals, together with data on morphological and life history traits. In our analyses we focused exclusively on adult birds and used only pedigree data from 1993 onwards, because individuals were not genotyped in the years before. For details on data collection and methods see Smith and Keller (2006).

Thanks to the small population size (roughly 30 breeding pairs per season), pedigree data are essentially complete. However, because extra-pair matings are common (*e. g.*, Reid et al., 2014), the so-called *apparent pedigree*, which was inferred from observations of parental care, is prone to misassigned paternities (wrong assignments of fathers), while maternities are correct, since all mothers are the genetic parents of the offspring they feed (Reid et al., 2014; Germain et al., 2016). Genetic information from 13 microsatellite loci has been used to assign each offspring its genetic father, yielding an almost exact reconstruction of the *actual pedigree* (Sardell et al., 2010). A comparison of the apparent and actual pedigrees used in our study here yielded an estimated proportion of 0.17 misassigned paternities. Note that this proportion differs from extra-pair paternity rates reported elsewhere, because we restricted the analysis to individuals with records of tarsus length and juvenile survival (2883 observations on 1056 individuals for tarsus length, and 3472 individuals with one measurement each for survival, respectively). The circumstance that we have both an error-prone and an essentially error-free (actual) version of the same pedigree renders the song sparrows an ideal study system to test the PSIMEX algorithm, because we can directly compare the naive (from the observed pedigree), actual (from the actual pedigree) and error-corrected estimates.

## PSIMEX for the song sparrow analyses

Our analyses focused on the heritability of tarsus length and on inbreeding depression in juvenile survival, which is known to exhibit inbreeding depression in this population (*e. g.*, Reid et al., 2014). We applied PSIMEX to obtain error-corrected estimates of the heritability of tarsus length and the magnitude of inbreeding depression in juvenile survival from the apparent (error-prone) pedigree with a paternity error rate of $\zeta_I = 0.17$. We assumed this proportion to be constant across generations, and that misassigned fathers were randomly selected from the same generation, although there is weak indication that extra-pair paternities over-proportionally involved paired territorial males (Sardell et al., 2010), and that the value of the focal trait itself may play a role (Firth et al., 2015). We assumed that matings were random, although relatively inbred animals might be slightly more likely to mate with closely related individuals (Reid et al., 2006). The corresponding estimates from the actual (error-free) pedigree served as a benchmark for the error-corrected estimates $\hat{\Theta}_{\text{SIMEX}}$.

To estimate heritability of tarsus length we fitted the model

$$y_{ij} = \mu + \beta_{\text{sex}} sex_i + \beta_f f_i + a_i + id_i + R_{ij} \ , \tag{6}$$

where $y_{ij}$ was the $j^{\text{th}}$ measurement of tarsus length for individual $i$, and $sex_i$ and inbreeding coefficient $f_i$ were fixed effects. Random effects were as defined in Equation (2). The model was fitted with INLA, and inverse gamma priors $\mathsf{IG}(1/2, 1/6\sigma_P^2)$ were used for the three variance components $\sigma_A^2$, $\sigma_{PE}^2$ and $\sigma_R^2$, and heritability was estimated as in Equation (3).

On the other hand, juvenile survival is a binary trait indicating survival of an individual from the age of six days to one year (1=yes, 0=no). We therefore used a GLMM with the binary survival variable $y_i$ as outcome to estimate inbreeding depression $\beta_f$. The binary survival variable can be interpreted as the realization of a slow count process, thus we used a complementary log-log (cloglog) link. The model was thus given as

$$g[\mathsf{E}(y_i)] = \mu + \beta_f f_i + \beta_{\text{sex}} sex_i + \beta_{\text{year}} year_i \ ,$$

where the expected value $\mathsf{E}(y_i) = p_i$ is the survival probability of individual $i$, and $g$ is the complementary log-log link function. Besides the inbreeding coefficient $f_i$ and the sex of the individual, we also included the year a bird was born as a categorical covariate. More complex models can be formulated, for example including an interaction term between sex and inbreeding (see Reid et al., 2014), but we ignored these extensions for the purpose of this paper. The model was fitted in a likelihood framework using `lme4` in `R`, as in Reid et al. (2014).

To obtain error-corrected PSIMEX estimates of $h^2$ and $\beta_f$, we set the number of iterations per error level to $B = 100$ and used error proportions $\zeta$ ranging from 0.2 to 1.0 in steps of 0.1. Linear, quadratic and cubic functions were used to extrapolate to error-corrected estimates $\hat{\Theta}_{\text{simex}}$ of $h^2$ and $\beta_f$, respectively, and the function with lowest AICc was chosen as the "best" extrapolation function. In order to assess the quality of the standard errors derived from the extrapolation via the PSIMEX (as described in Appendix 1), we also ran a bootstrap with 50 iterations for the best fitting extrapolation function. We restricted the bootstrap to the analysis of inbreeding depression, since bootstrapping the animal model given in Equation (6) was too computationally intensive. The code for this application is given in Appendices 6 and 7.

## Song sparrows results

For both, heritability and inbreeding depression, bias and MSE were reduced by the PSIMEX estimators obtained from the best fitting (*i. e.*, minimum AICc) extrapolating functions. In both cases, the AICc-criterion suggested that a linear extrapolation function fitted the error trend best (Tables 2 and 3). Fig. 5 shows the trend of the simulated values for increasing error proportions, as well as the extrapolation to zero error given by PSIMEX. Different extrapolation functions yielded quite different estimates, which underlines the importance of the choice of the extrapolation function. The bootstrap on the error-corrected estimator of $\beta_f$ indicates that PSIMEX yields standard errors that are not too dissimilar from the true values, with a standard error of 0.80 derived from the linear extrapolation function, and a standard error of 1.02 from the bootstrap, although it is of course difficult to draw definite conclusions from a single case.

In line with the simulations, the naive $h^2$ estimator was found to be an attenuated version of the actual value. A closer inspection of the estimates of the additive ($\sigma^2_A$), residual ($\sigma^2_R$) and permanent environmental ($\sigma^2_{PE}$) variance components indicates that this is a consequence of a higher estimate of $\sigma^2_{PE}$ and a lower estimate of $\sigma^2_A$ when the apparent pedigree is used (Fig. S3 in Appendix 1). For inbreeding depression, the error in the apparent pedigree induced an attenuation bias in $\beta_f$ that was comparable to the case with $N_e/N_c = 1$ in the simulations (Fig. 5b). The song sparrow pedigree has an architecture with $N_e/N_c \approx 0.6$, which corresponds to a value between the simulated cases with $N_e/N_c = 1$ and $N_e/N_c = 0.3$. It is, however, unclear if the patterns observed in the simulations are universal, in particular when traits are not continuous, like the binary juvenile survival trait studied here.

Interestingly, pedigree error seemed to bias estimators of inbreeding depression more severely (in relative terms) than estimates of heritability. A potential reason for this difference is that information about inbreeding requires both parents to be known correctly, while at least some information about additive genetic effects can be obtained from a single correctly assigned parent.

## Discussion

We have employed the case of pedigree error to promote a very general strategy to account for measurement error in ecology and evolution. By adapting the philosophy of the SIMEX approach, originally proposed to account for measurement error in continuous regression covariates (Cook and Stefanski, 1994), we illustrate how increasing error in the assignment of parents to their offspring in the pedigree can yield information about the resulting bias in parameters that are estimated in downstream statistical analyses, such as inbreeding depression or heritability. The observed trend with increasing error is then back-extrapolated to the hypothetical situation of zero error, yielding error-corrected, approximately consistent estimators for the parameters of interest. SIMEX is an intuitive way to assess and correct for the effects of any type of error in very general applications, especially when an explicit model for the unobserved component is difficult to formulate, and when it is thus not possible to embed an explicit error model directly in the statistical analysis. To facilitate the accessibility of the PSIMEX method, the code to perform all the analyses is provided via the novel `R` package `PSIMEX`, which is available from the CRAN repository (Ponzi, 2017).

We used simulation studies and a dataset from wild-living song sparrows to illustrate that the approach is successful in recovering error-corrected estimates of inbreeding depression and heritability in the presence of pedigree error. Interestingly, the simulations also revealed that pedigree error does not necessarily lead to attenuated versions of quantitative genetic measures. In fact, inbreeding depression was over- *or* underestimated with increasing misassigned paternity error, and the direction of the bias depended on the pedigree topology. This result indicates that, without observing or simulating the actual effect of the error, the direction of the bias in the naive estimator cannot be known a priori, even in the simple case of completely random error. Encouragingly, the $\hat{\Theta}_{\text{simex}}$ estimates obtained through the PSIMEX procedure captured these effects correctly and had smaller bias and MSE than the naive estimates that ignore pedigree error. In the simulation study, the MSE was even considerably lower for PSIMEX than for the naive estimators, in particular for heritability, although it should be noted that the true standard errors are smaller than in most realistic situations. The application of the method to a wild population of song sparrows, where PSIMEX estimators could be compared to actual values derived from error-free pedigrees, confirmed that the method is performing well, with a reduction in bias and MSE, and approximately correct standard errors (as verified with a bootstrap procedure for inbreeding depression), although with a bit more uncertainty in the estimators than the naive versions. This tradeoff between bias and variance is well known in measurement error theory (*e. g.*, Carroll et al., 2006, p. 62–63). Our view on this is that the lower uncertainty in the naive estimators is actually too optimistic, because the information loss due to the error is not accounted for. Error correction yields less biased point estimates and more correct (*i. e.*, wider) estimators of the actual uncertainty.

An important prerequisite for error correction, not only in the context discussed here, is knowledge of the error-generating mechanism and the error model parameter(s), as otherwise error models are nonidentifiable (Fuller, 1987; Gustafson, 2005; Carroll et al., 2006). Unlike in the case of other error-correction techniques, however, the SIMEX approach does not require that an explicit model for the unobserved, true covariate without measurement error is formulated. Moreover, we have illustrated in this paper that the specification of an error model for the variable of interest can be circumvented by knowing the error generating mechanism at the lowest level of the data-generating process. In the case of PSIMEX, both the relatedness matrix $\mathbf{A}$ and the inbreeding coefficients $f_i$ of individuals are deduced from the pedigree, thus we did not have to formulate error models for $\mathbf{A}$ or $f_i$, which would have been challenging, as illustrated in Appendix 1 (Section 2.1). Instead, we could directly work with the error structure in the pedigree. Of course, this required that information about the underlying error generating mechanism and the error proportion in the pedigree was available. In our application to the song sparrows, we could approximately estimate the proportion of erroneously assigned fathers in the pedigree from a comparison of the error-prone and the error-free pedigree, and we assumed a random error mechanism where fathers are replaced with random individuals from the same generation (Reid et al., 2015).

To obtain information about the error mechanism and error model parameters, it is crucial to collect data that allow the quantification of the error. False assumptions may lead to biased SIMEX estimators, as illustrated in Appendix 1 (Sections 6 and 7). Ideally, error estimation should be part of the study design, because it is much harder to obtain error estimates retrospectively. Quite often it will be sufficient to take error-prone and error-free measurements on a small subset of all study subjects. In the case of pedigrees, for example, it can be useful to genetically verify a subset of all parents in order to estimate the error proportion. In the absence of precise information about the error, similar studies or comparable populations might provide useful information, which can be used as *prior knowledge*. Of course, transportation of such information across study systems bears the risk that potentially inappropriate but untestable assumptions enter the modelling process, and it is therefore advisable to obtain some error estimates in the actual study system for comparison.

An important aspect of SIMEX procedures is the choice of the extrapolation function. Since it is not straightforward to derive a theoretical justification for a functional form of the extrapolation function for PSIMEX, we used polynomial extrapolation functions, because these are known as stable and approximately consistent alternatives to more complicated functions (*e. g.*, Kuechenhoff et al., 2006, p. 109). We suggest to select the "best" extrapolant via the AICc criterion, which should balance between model complexity and model fit to obtain good predictions.

The PSIMEX methodology can be adapted to correct for different error mechanisms in the pedigree, for example when misassignments do not only affect fathers but also mothers, when the proportion of misassigned paternities varies across the study period, or when the replacement of fathers is more likely to occur with phenotypically or genotypically similar individuals. These error generating mechanisms can easily be handled by the PSIMEX algorithm; see Appendix 1 (Section 7) for some examples. Moreover, the PSIMEX idea can be applied to virtually any quantity that is derived from pedigrees, for example to error-correct the estimates of variance parameters, but also to estimates of sexual selection, linkage, penetrance, the response to selection, genetic correlations, etc. Note, however, that we assumed that only pedigree error is present and that all variables that are not related to information derived from the pedigree are error-free. If measurement error in other variables is present at the same time, a possible strategy would be to apply the PSIMEX approach to a model that accounts for the (parametric) error in such covariates, for example to a Bayesian hierarchical error model (see *e. g.*, Ponzi et al., 2018).

Although we have employed and illustrated the SIMEX algorithm only for the particular application to pedigree error, the same generic principle can be adapted to many other situations. As an example, the SIMEX procedure could be used to account for location error in habitat selection studies, where parameters of interest, such as measures of distance and velocity, classifications of an animal's activities, or an animal's presence or absence at a given location may be erroneous (Ganskopp and Johnson, 2007; McKenzie et al., 2009). Instead of formulating an error model for the biased covariates themselves, it might often be easier to focus directly on the location error, using information on the accuracy of the measurements (*e. g.*, GPS error) and the mechanisms that might obscure it, which can be used to obtain error model parameters for a SIMEX correction.

# Conclusions

The conceptual simplicity of the SIMEX philosophy allows its implementation even in situations when it is difficult or impossible to formulate or incorporate an explicit error model for an erroneous variable. The only prerequisites to apply the SIMEX algorithm are that the error-generating mechanism is known, and that it is possible to make the error "worse" in a controllable, quantitative way. We believe that many other applications in ecology and evolution will benefit from this simple and practical approach to obtain error-corrected parameter estimates in the presence of measurement error.

**Supporting information:**

**Appendix 1**: Supplementary text and figures (pdf)

**Appendix 2**: R script for PSIMEX on inbreeding in simulated data

**Appendix 3**: R script for PSIMEX on heritability in simulated data

**Appendix 4**: R script for PSIMEX on inbreeding

**Appendix 5**: R script for PSIMEX on heritability

**Appendix 6**: R script for PSIMEX on inbreeding in Song Sparrows

**Appendix 7**: R script for PSIMEX on heritability in Song Sparrows

**Author contributions:**
S.M. and L.F.K. conceived the research idea. E.P. designed and conducted the simulations and analyses. E.P. and S.M. wrote the manuscript. All authors provided feedback during the writing process and gave final approval.

**Conflict of interest statement:**
The authors declare they have no competing interests.

**Acknowledgements:**
We thank Peter Arcese and Jane Reid for the use of the song sparrow data, Pirmin Nietlisbach for the help and explanation of the data set, and Amy B. Marr who helped motivating this work. Our validation was greatly facilitated by access to data collected by dozens of fieldworkers and funded by NSERC Canada from 1975 to present.

**Data accessibility:**
All code to replicate the analyses is available in the supplementary materials. The code and data are deposited on Zenodo: https://doi.org/10.5281/zenodo.3264813.

## References

Apanasovich, T. V., R. J. Carroll, and A. Maity (2009). SIMEX and standard error estimation in semiparametric measurement error models. *Electronic Journal of Statistics 3*, 318–348.

Becker, P. J., J. Hegelbach, L. F. Keller, and E. Postma (2016). Phenotype-associated inbreeding biases estimates of inbreeding depression in a wild bird population. *Journal of Evolutionary Biology 29*, 35–46.

Bishop, D. A. and C. M. Beier (2013). Assessing uncertainty in high-resolution spatial climate data across the US Northeast. *PLOS One 8*, e70260.

Buonaccorsi, J. (2010). *Measurement error: models, methods and applications*. Boca Raton: CRC press.

Carroll, R. J., D. Ruppert, L. A. Stefanski, and C. M. Crainiceanu (2006). *Measurement error in nonlinear models, a modern perspective*. Boca Raton: Chapman and Hall.

Charlesworth, D. and J. H. Willis (2009). The genetics of inbreeding depression. *Nature Reviews Genetics 10*, 783–796.

Charmantier, A. and D. Reale (2005). How do misassigned paternities affect the estimation of heritability in the wild? *Molecular Ecology 14*, 2839–2850.

Cook, J. R. and L. A. Stefanski (1994). Simulation-extrapolation estimation in parametric measurement error models. *Journal of the American Statistical Association 89*, 1314–1328.

Coster, A. (2013). *pedigree: Pedigree functions*. R package version 1.4.

Devanarayan, V. and L. A. Stefanski (2002). Empirical simulation extrapolation for measurement error models with replicate measurements. *Statistics and Probability Letters 59*, 219–225.

Dohm, M. R. (2002). Repeatability estimates do not always set an upper limit to heritability. *Functional Ecology 16*, 273–280.

Eckert, R. S., R. J. Carroll, and N. Wang (1997). Transformations to additivity in measurement error models. *Biometrics 53*, 262–272.

Firth, J. A., J. D. Hadfield, A. W. Santure, J. Slate, and B. C. Sheldon (2015). The influence of nonrandom extra-pair paternity on heritability estimates derived from wild pedigrees. *Evolution 69*, 1336–1344.

Frankham, R. (1995). Effective population size/adult population size ratios in wildlife: a rewiev. *Genetic Research 66*, 95–107.

Freckleton, R. P. (2011). Dealing with collinearity in behavioural and ecological data: model averaging and the problems of measurement error. *Behavioral Ecology and Sociobiology 65*, 91–101.

Fuller, W. A. (1987). *Measurement Error Models.* New York: John Wiley & Sons.

Fung, K. Y. and D. Krewski (1999). On measurement error adjustment methods in Poisson regression. *Environmetrics 10*, 213–224.

Ganskopp, D. C. and D. D. Johnson (2007). GPS error in studies addressing animal movements and activities. *Rangeland Ecology and Management 60*, 350–358.

Ge, T., A. J. Holmes, R. L. Buckner, J. W. Smoller, and M. R. Sabuncu (2017). Heritability analysis with repeat measurements and its application to resting-state functional connectivity. *PNAS 114*, 5521–5526.

Germain, R. R., M. E. Wolak, P. Arcese, S. Losdat, and J. M. Reid (2016). Direct and indirect genetic and fine-scale location effects on breeding date in song sparrows. *Journal of Animal Ecology 85*, 1613–1624.

Gorjanic, G. and D. A. Henderson (2007). *GeneticsPed: Pedigree and genetic relationship functions.* R package version 1.34.0.

Gould, W. R., L. A. Stefanski, and K. H. Pollock (1999). Use of simulation-extrapolation estimation on catch-effort analysis. *Canadian Journal of Fisheries and Aquatic Sciences 56*, 1234–1240.

Griffith, S. C., I. P. Owens, and K. A. Thuman (2002). Extrapair paternity in birds: a review of interspecific variation and adaptive function. *Molecular Ecology 11*, 2195–2212.

Guélat, J. and M. Kéry (2018). Effects of spatial autocorrelation and imperfect detection on species distribution models. *Methods in Ecology and Evolution 9*, 1614–1625.

Guillera-Arroita, G., J. J. Lahoz-Monfort, D. I. MacKenzie, B. A. Wintle, and M. A. McCarthy (2014). Ignoring imperfect detection in biological surveys is dangerous: a response to 'Fitting and interpreting occupancy models'. *PLOS One 9*, 1–14.

Gustafson, P. (2004). *Measurement Error and Misclassification in Statistics and Epidemiology: Impacts and Bayesian Adjustments.* Boca Raton: Chapman & Hall/CRC.

Gustafson, P. (2005). On model expansion, model contraction, identifiability and prior information: Two illustrative scenarios involving mismeasured variables. *Statistical Science 20*, 111–140.

Hadfield, J. D. (2010). MCMC methods for multi-response generalized linear mixed models: The MCMCglmm R package. *Journal of Statistical Software 33*, 1–22.

Haila, Y., K. Henle, E. Apostolopoulou, J. Cent, E. Framstad, C. Goerg, K. Jax, R. Klenke, W. Magnuson, Y. G. Matsinos, B. Mueller, R. Paloniemi, J. Pantis, F. Rauschmayer, I. Ring, J. Settele, J. Simila, K. Touloumis, J. Tzanopoulos, and G. Pe'er (2014). Confronting and coping with uncertainty in biodiversity research and praxis. *Nature Conservation 8*, 45–75.

Henderson, C. R. (1976). Simple method for computing inverse of a numerator relationship matrix used in prediction of breeding values. *Biometrics 32*, 69–83.

Hoffmann, A. A. (2000). Laboratory and field heritabilities: Lessons from *Drosophila*. In T. Mousseau, S. B., and J. Endler (Eds.), *Adaptive Genetic Variation in the Wild*. New York, Oxford: Oxford Univ Press.

Holand, A., I. Steinsland, S. Martino, and H. Jensen (2013). Animal models and integrated nested Laplace approximations. *G3 3*, 1241–1251.

Hwang, W. H. and S. Y. H. Huang (2003). Estimation in capture-recapture models when covariates are subject to measurement errors. *Biometrics 59*, 1113–1122.

Jensen, H., E. M. Bremset, T. H. Ringsby, and B. E. Sæther (2007). Multilocus heterozygosity and inbreeding depression in an insular house sparrow metapopulation. *Molecular Ecology 16*, 4066–4078.

Keller, L. F., P. R. Grant, B. R. Grant, and K. Petren (2001). Heritability of morphological traits in Darwin's Finches: misidentified paternity and maternal effects. *Heredity 87*, 325–336.

Keller, L. F., P. R. Grant, B. R. Grant, and K. Petren (2002). Environmental conditions affect the magnitude of inbreeding depression in survival of Darwin's Finches. *Evolution 56*, 1229–1239.

Keller, L. F. and D. M. Waller (2002). Inbreeding effects in wild populations. *Trends in Ecology and Evolution 17*, 230–241.

Kruuk, L. E. B. (2004). Estimating genetic parameters in natural populations using the 'animal model'. *Philosophical Transactions of the Royal Society of London 359*, 873–890.

Kruuk, L. E. B. and J. D. Hadfield (2007). How to separate genetic and environmental causes of similarity between relatives. *Journal of Evolutionary Biology 20*, 1890–1903.

Kuechenhoff, H. and R. J. Carroll (1997). Segmented regression with errors in predictors: Semi-parametric and parametric methods. *Statistics in Medicine 16*, 169–188.

Kuechenhoff, H., S. M. Mwalili, and E. Lesaffre (2006). A general method for dealing with misclassification in regression: The misclassification SIMEX. *Biometrics 62*, 85–96.

Li, Y. and X. Lin (2003). Functional inference in frailty measurement error models for clustered survival data using the SIMEX approach. *Journal of the American Statistical Association 98*, 191–203.

Lin, X. and R. J. Carroll (2000). Nonparametric function estimation for clustered data when the predictor is measured with/without error. *Journal of the American Statistical Association 95*, 520–534.

Lynch, M. and B. Walsh (1998). *Genetics and Analysis of Quantitative Traits.* Sunderland, MA: Sinauer Associates.

Macgregor, S., B. K. Cornes, N. G. Martin, and P. M. Visscher (2006). Bias, precision and heritability of self-reported and clinically measured height in Australian twins. *Human Genetics 120*, 571–580.

Mason, N. W. H., R. J. Holdaway, and S. J. Richardson (2018). Incorporating measurement error in testing for changes in biodiversity. *Methods in Ecology and Evolution 9*, 1296–1307.

McClintock, B. T., J. M. London, M. F. Cameron, and P. L. Boveng (2014). Modelling animal movement using the Argos satellite telemetry location error ellipse. *Methods in Ecology and Evolution 6*, 266–277.

McKenzie, H. W., C. L. Jerde, D. R. Visscher, E. H. Merrill, and M. A. Lewis (2009). Inferring linear feature use in the presence of GPS measurement error. *Environmental and Ecological Statistics 16*, 531–546.

Melbourne, B. A. and P. Chesson (2006). The scale transition: scaling up population dynamics with field data. *Ecology 87*, 1478–1488.

Montgomery, R. A., G. J. Roloff, and J. M. Ver Hoef (2011). Implications of ignoring telemetry error on inference in wildlife resource use models. *Journal of Wildlife Management 75*, 702–708.

Muff, S. and L. F. Keller (2015). Reverse attenuation in interaction terms due to covariate error. *Biometrical Journal 57*, 1068–1083.

Muff, S., A. Riebler, L. Held, H. Rue, and P. Saner (2015). Bayesian analysis of measurement error models using integrated nested Laplace approximations. *Journal of the Royal Statistical Society, Applied Statistics Series C 64*, 231–252.

Palstra, F. P. and D. J. Fraser (2012). Effective/census population size ratio estimation: a compendium and appraisal. *Ecology and Evolution 2*, 2357–2365.

Ponzi, E. (2017). *PSIMEX: SIMEX Algorithm on Pedigree Structures*. University of Zurich. R package version 1.1.

Ponzi, E., L. F. Keller, T. Bonnet, and S. Muff (2018). Heritability, selection, and the response to selection in the presence of phenotypic measurement error: Effects, cures, and the role of repeated measurements. *Evolution 0*(0).

Reid, J. M., P. Arcese, and L. F. Keller (2006). Intrinsic parent-offspring correlation in inbreeding level in a song sparrow (*Melospiza melodia*) population open to immigration. *The American Naturalist 168*, 1–13.

Reid, J. M., P. Arcese, and L. F. Keller (2008). Individual phenotype, kinship, and the occurrence of inbreeding in song sparrows. *Evolution 62*, 887–899.

Reid, J. M., P. Arcese, L. F. Keller, R. R. Germain, A. B. Duthie, S. Losdat, M. E. Wolak, and P. Nietlisbach (2015). Quantifying inbreeding avoidance through extra-pair reproduction. *Evolution 69*, 59–74.

Reid, J. M. and L. F. Keller (2010). Correlated inbreeding among relatives: occurrence, magnitude and implications. *Evolution 64*, 973–985.

Reid, J. M., L. F. Keller, A. B. Marr, P. Nietlisbach, R. J. Sardell, and P. Arcese (2014). Pedigree error due to extra-pair reproduction substantially biases estimates of inbreeding depression. *Evolution 3*, 802–815.

Rue, H., S. Martino, and N. Chopin (2009). Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations (with discussion). *Journal of the Royal Statistical Society. Series B (Statistical Methodology) 71*, 319–392.

Sardell, R. J., L. F. Keller, P. Arcese, T. Bucher, and J. M. Reid (2010). Comprehensive paternity assignment: genotype, spatial location and social status in song sparrows *Melospiza melodia*. *Molecular Ecology 19*, 4352 – 4364.

Senneke, S. L., M. D. MacNeil, and L. D. Van Vleck (2004). Effects of sire misidentification on estimates of genetic parameters for birth and weaning weights in Hereford cattle. *American Society of Animal Science 82*, 2307–2312.

Smith, J. N. M. and L. F. Keller (2006). Genetics and demography of small populations. In J. Smith, L. Keller, A. Marr, and P. Arcese (Eds.), *Conservation and biology of small populations: the song sparrows of Mandarte Island*. Oxford University Press.

Solow, A. R. (1998). On fitting a population model in the presence of observation error. *Ecology 79*, 1463–1466.

Stefanski, L. A. and J. R. Cook (1995). Simulation-extrapolation: The measurement error jackknife. *Journal of the American Statistical Association 90*, 1247–1256.

Steinsland, I., C. Larsen, A. Roulin, and H. Jensen (2014). Quantitative genetic modelling and inference in the presence of nonignorable missing data. *Evolution 68*, 1735–1747.

Stoklosa, J., C. Daly, S. D. Foster, M. B. Ashcroft, and D. I. Warton (2014). A climate of uncertainty: accounting for error in climate variables for species distribution models. *Methods in Ecology and Evolution 6*, 412–423.

van der Sluis, S., M. Verhage, D. Posthuma, and C. V. Dolan (2010). Phenotypic complexity, measurement bias, and poor phenotypic resolution contribute to the missing heritability problem in genetic association studies. *PLOS One 5*, e13929.

Visscher, P. M., J. A. Woolliams, J. Smith, and J. L. Williams (2002). Estimation of pedigree errors in the UK dairy population using microsatellite markers and the impact on selection. *Journal of Dairy Science 85*, 2368–2375.

Wang, N., X. Lin, R. G. Gutierrez, and R. J. Carroll (1998). Bias analysis and SIMEX approach in generalized linear mixed measurement error models. *Journal of the American Statistical Association 93*, 249–261.

Wilson, A. J., D. Reale, M. N. Clements, M. M. Morrissey, E. Postma, C. A. Walling, L. E. B. Kruuk, and D. H. Nussey (2010). An ecologist's guide to the animal model. *Journal of Animal Ecology 79*, 13–26.

Wolak, M. E. and L. F. Keller (2014). Dominance genetic variance and inbreeding in natural populations. In A. Charmantier, D. Garant, and L. E. B. Kruuk (Eds.), *Quantitative Genetics in the Wild*. New York: Oxford University Press.

Wright, W. J., K. M. Irvine, J. M. Warren, and J. K. Barnett (2017). Statistical design and analysis for plant cover studies with multiple sources of observation errors. *Methods in Ecology and Evolution 8*, 1832–1841.

**Figures:**

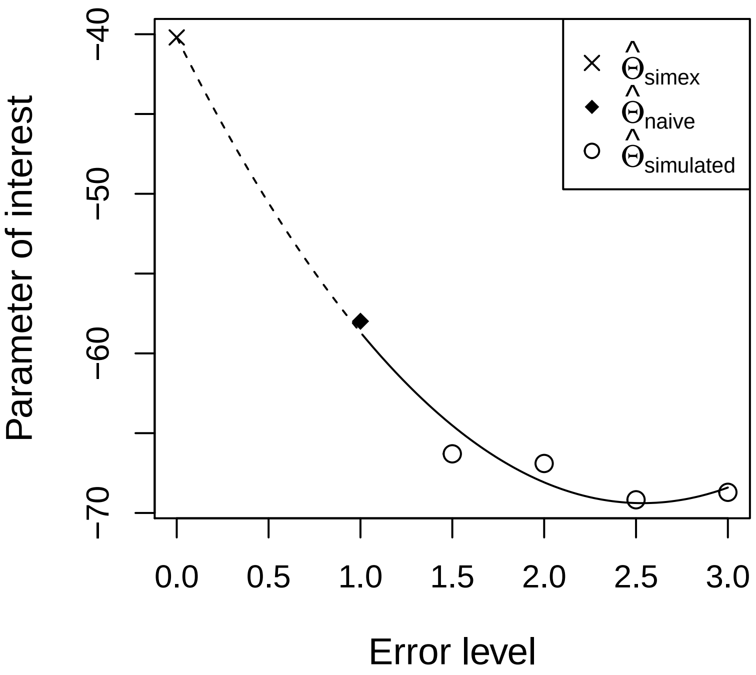

Figure 1: Illustration of the SIMEX procedure. The error level is increased in predefined interval steps, the parameter of interest is estimated at each error level, and a function is fitted on the observed trend upon increasing error. An error-corrected estimate is obtained by extrapolating the function back to an error level of zero.

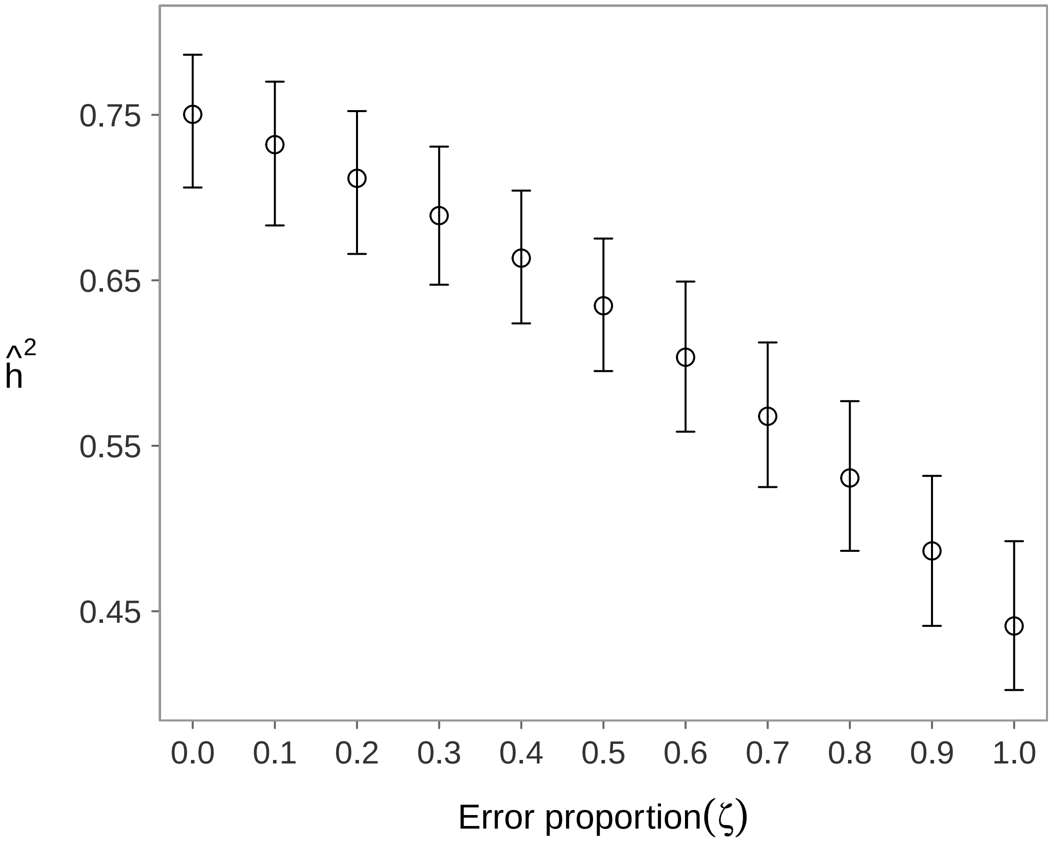

Figure 2: Effect of increasing the paternity error proportion on heritability in the simulated pedigrees with actual heritability $h^2 = 0.75$. Mean estimates of heritability $\hat{h}^2$ from 100 simulations for each pedigree with increasing error rates $\zeta = 0.1, ..., 1$ are shown with their 5% to 95% sample quantile intervals at each error proportion. A clear decreasing trend is observed in the estimate as the error proportion increases.

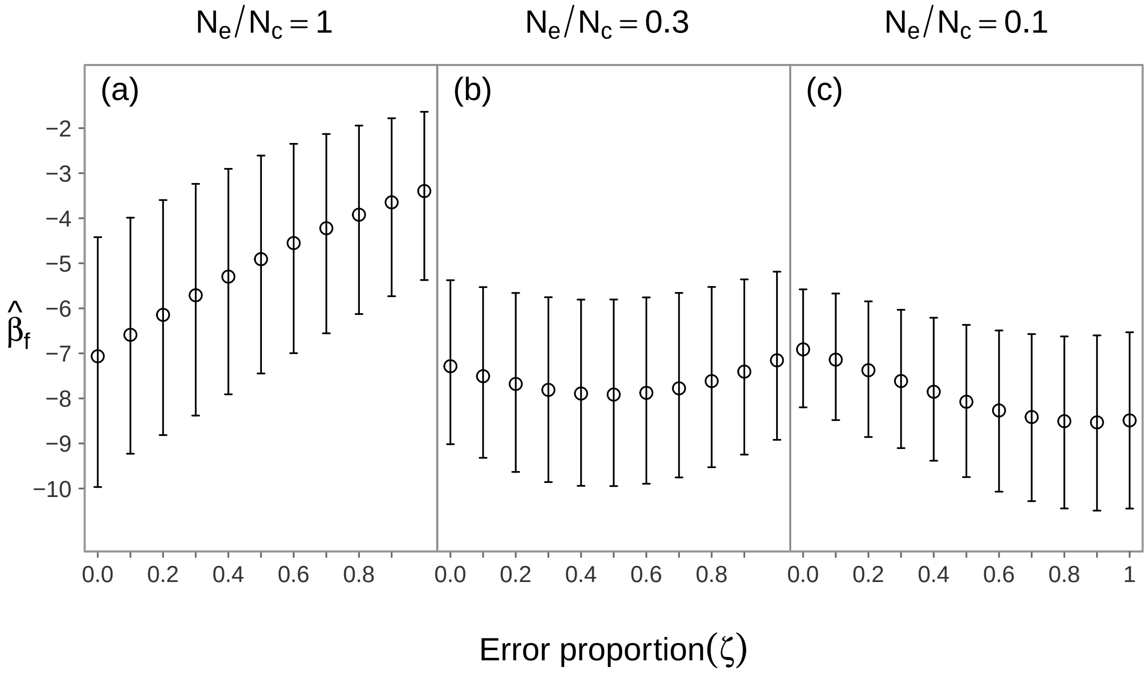

Figure 3: Effects of increasing paternity error proportion on inbreeding depression in simulated pedigrees with each of the three different pedigree topologies ($N_e/N_c = 1$, 0.3, and 0.1). Mean estimates of inbreeding depression $\hat{\beta}_f$ from 100 simulations with increasing error rates $\zeta = 0.1, ..., 1$ are shown with their 5% to 95% sample quantile intervals at each error proportion.

**a) Heritability**

| | $h^2 = 0.75$ | | $h^2 = 0.5$ | | $h^2 = 0.25$ | |
|---|---|---|---|---|---|---|
| Estimator | Bias | MSE | Bias | MSE | Bias | MSE |
| Naive | $-0.039$ | 0.0016 | $-0.038$ | 0.0016 | $-0.030$ | 0.0010 |
| PSIMEX | $-0.003$ | 0.0002 | 0.011 | 0.0003 | 0.0002 | 0.0002 |

**b) Inbreeding depression**

| | $N_e/N_c = 1$ | | $N_e/N_c = 0.3$ | | $N_e/N_c = 0.1$ | |
|---|---|---|---|---|---|---|
| Estimator | Bias | MSE | Bias | MSE | Bias | MSE |
| Naive | 0.919 | 0.864 | $-0.394$ | 0.183 | $-0.463$ | 0.243 |
| PSIMEX | $-0.031$ | 0.031 | 0.041 | 0.032 | $-0.003$ | 0.037 |

Table 1: Mean bias and MSE in the naive and PSIMEX estimates from all simulations for the three levels of heritability (a), and the three pedigree topologies for inbreeding depression (b). The initial error proportion was $\zeta_I = 0.2$, and the PSIMEX estimate was always extracted using the extrapolating function with minimal AICc.

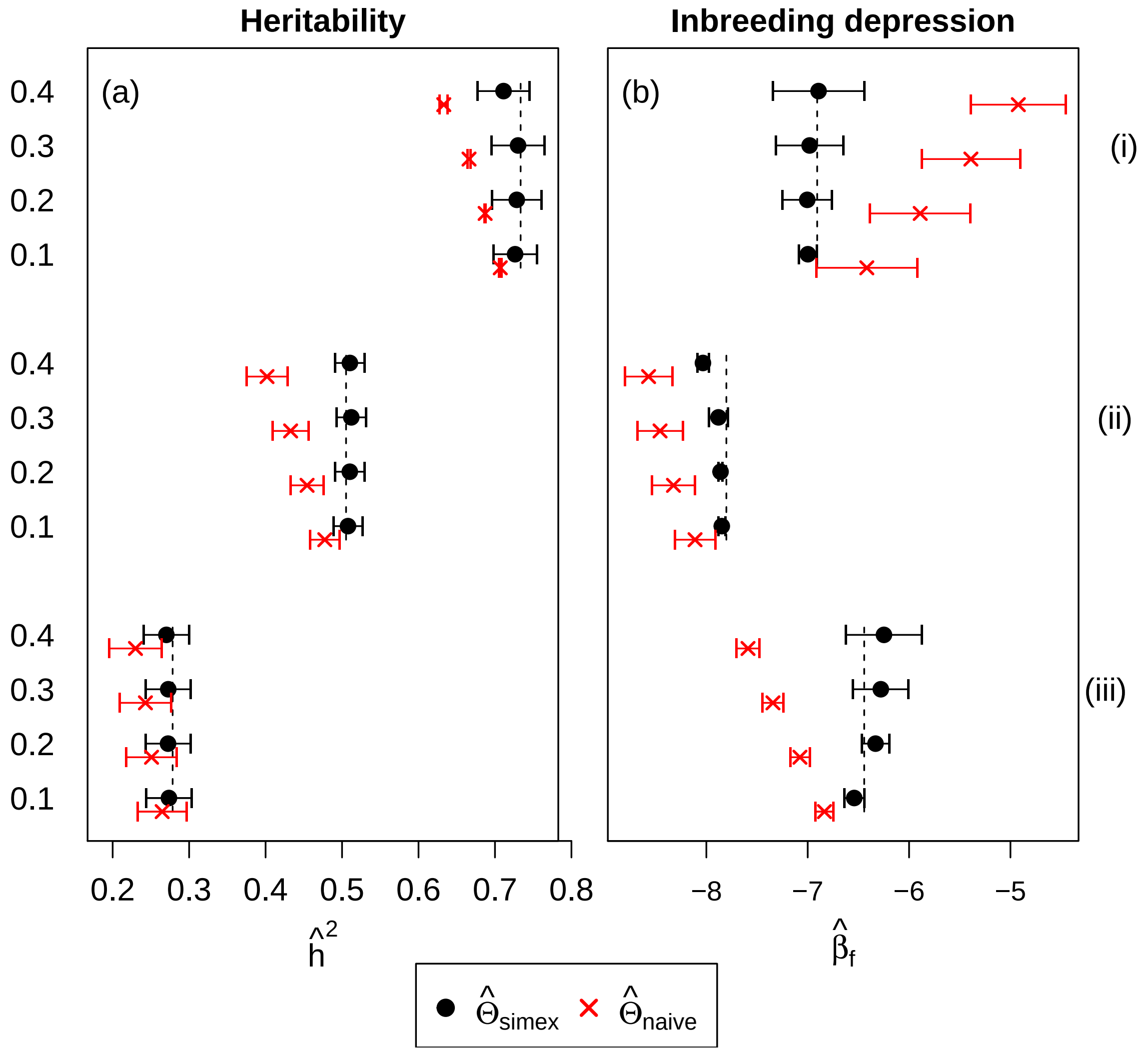

Figure 4: Naive ($\hat{\Theta}_{\text{naive}}$) and error-corrected ($\hat{\Theta}_{\text{simex}}$) estimates of heritability and inbreeding depression from one simulated pedigree per case. For the error-corrected PSIMEX estimates, means and 95% confidence intervals from the simulations for PSIMEX and naive estimates are reported. The dashed line represents the actual simulated value for the respective pedigree. Four different estimates are given, corresponding to four initial error proportions ($\zeta_I = 0.1$, 0.2, 0.3 and 0.4). For heritability (a), results are reported for values of heritability (i) $h^2 = 0.75$, (ii) $h^2 = 0.5$ and (iii) $h^2 = 0.25$. For inbreeding depression (b), results are reported for pedigree topologies (i) $N_e/N_c = 1$, (ii) $N_e/N_c = 0.3$ and (iii) $N_e/N_c = 0.1$.

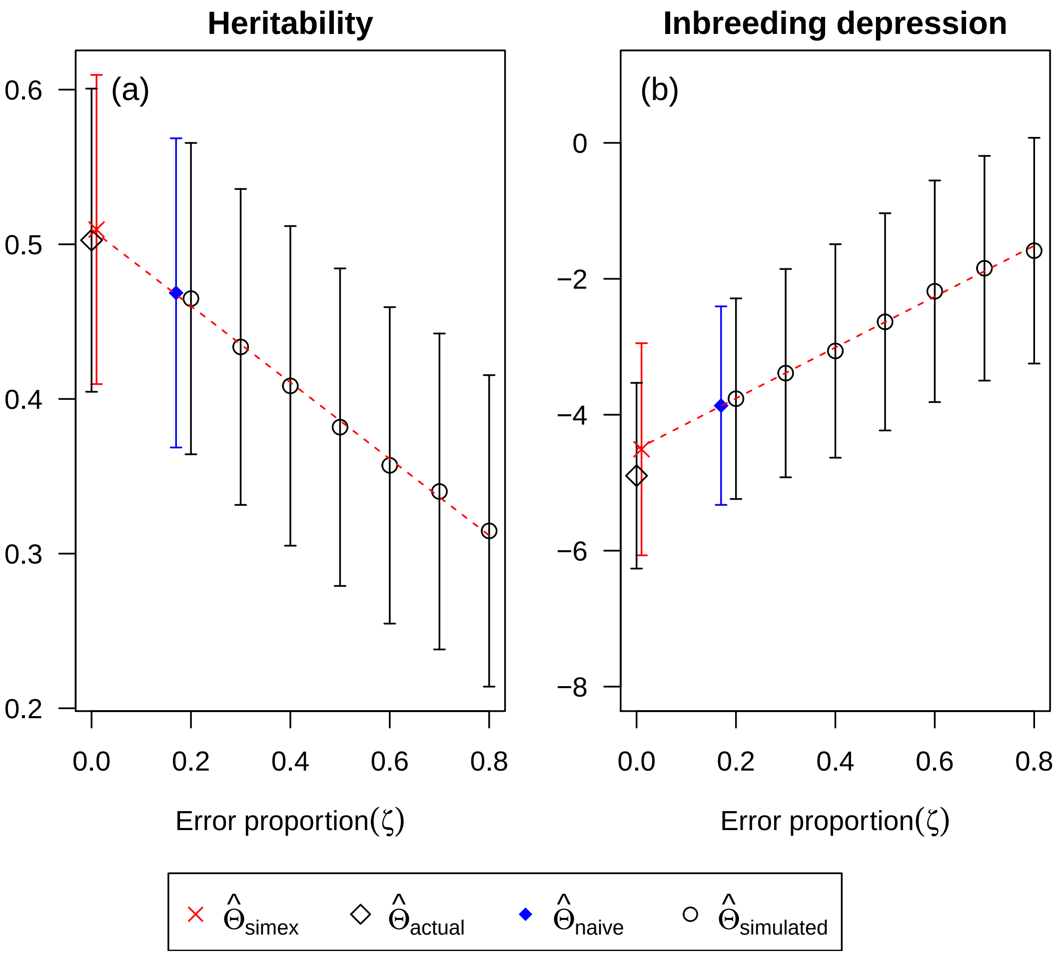

Figure 5: Results of the PSIMEX procedure when error-correcting heritability of tarsus length (a) and inbreeding depression of juvenile survival (b) in the song sparrow dataset. The initial error proportion was $\zeta_I = 0.17$. The trend upon increasing error proportions is shown together with the extrapolated values obtained from the best extrapolation function (linear). The naive and the actual estimates from the genetic pedigree are also shown. PSIMEX estimates are much closer to the actual values than the naive estimates both for heritability and inbreeding depression.

| Estimator | $\hat{h}^2$ | 95% CI | Bias | MSE | AICc |
|---|---|---|---|---|---|
| Actual | 0.503 | 0.401 to 0.596 | - | - | - |
| Naive | 0.469 | 0.371 to 0.571 | $-0.032$ | 0.0036 | - |
| PSIMEX (linear) | **0.510** | **0.409 to 0.611** | **0.007** | **0.0026** | $\mathbf{-56.6}$ |
| PSIMEX (quadratic) | 0.522 | 0.423 to 0.622 | 0.0195 | 0.0035 | $-55.6$ |
| PSIMEX (cubic) | 0.523 | 0.421 to 0.625 | 0.021 | 0.0041 | $-37.0$ |

Table 2: Estimates of heritability ($\hat{h}^2$) of tarsus length in the song sparrow dataset. Actual, naive and PSIMEX estimates (using linear, quadratic and cubic extrapolation functions) are reported together with their 95% credible intervals (CIs), as well as bias and MSE with respect to the actual value. The linear extrapolation function was the one with lowest AICc (in bold).

| Estimator | $\hat{\beta}_f$ | 95% CI | Bias | MSE | AICc |
|---|---|---|---|---|---|
| Actual | $-4.90$ | $-6.26$ to $-3.53$ | - | - | - |
| Naive | $-3.87$ | $-5.33$ to $-2.40$ | 1.03 | 1.62 | - |
| PSIMEX (linear) | $\mathbf{-4.51}$ | $\mathbf{-6.07}$ **to** $\mathbf{-2.95}$ | **0.39** | **0.79** | $\mathbf{-15.1}$ |
| PSIMEX (quadratic) | $-4.58$ | $-7.18$ to $-1.97$ | 0.32 | 1.86 | $-6.7$ |
| PSIMEX (cubic) | $-4.13$ | $-6.92$ to $-1.34$ | 0.77 | 2.61 | 1.7 |

Table 3: Estimates of inbreeding depression ($\hat{\beta}_f$) in juvenile survival in the song sparrow dataset. Actual, naive and PSIMEX estimates (using linear, quadratic and cubic extrapolation functions) are reported together with their 95% confidence intervals (CIs), as well as bias and MSE with respect to the actual value. The linear extrapolation function was the one with lowest AICc (in bold).