This material has been published in CAB Reviews, issue 7, No. 016, 2012, pp. 1–10, doi: 10.1079/PAVSNNR20127016, the only accredited archive of the content that has been certified and accepted after peer review.

The electronic version of this article is the definitive one. It is located here: http://www.cabi.org/cabreviews

Potential of GLMM in modelling invasive spread

J. Thiele[1*] and B. Markussen[2]

Address: [1] Institute of Landscape Ecology, University of Muenster, Robert-Koch-Str. 28, 48149 Muenster, Germany. [2] Department of Basic Sciences and Environment, University of Copenhagen, Thorvaldsensvej 40, 1871 Frederiksberg C, Denmark.

*Correspondence: J. Thiele. Email: jan.thiele@uni-muenster.de

Abstract

Utilization of Generalized Linear Mixed Models (GLMM) in invasion biology has increased exponentially during the last 5-10 years. GLMM are useful tools that can handle data with various distributions as well as spatial or temporal dependence which are involved in many study designs. We review the current state-of-the-art of GLMM with special focus on applications in invasion biology. This review covers all steps of data analysis with GLMM. We address frequently encountered practical problems, such as failure of convergence, and put some emphasis on validation of model assumptions. Further, we point towards possibilities of analysing zero-heavy data using combined GLMM. More detailed insight into practical applications of GLMM is provided in three worked examples in the supplementary material. Regarding applications of GLMM in invasion biology, a literature analysis showed that random effects are mostly used to account for non-independence of observations due to study design, but rarely for estimation of random variation. There may be some potential in using random-effect estimation more consciously, like in some recent studies of genetic variation of invasive species. Often, invasion biologists have to deal with count data or proportions. In such cases, several methods of parameter estimation are available, but their suitability depends on characteristics of the data at hand and, hence, they should be chosen carefully. Also repeated measures are common in invasion biology. In GLMM frameworks, the auto-correlation of such data can be modelled by structured co-variance matrices. This opportunity, however, has seldom been used.

Keywords: invasive species, convergence, mixed effects models, model validation, review, zero-heavy data

Introduction and Scope

Invasion biologists often study data that are not suitable for analysis with classical statistical procedures such as Ordinary Least Squares (OLS) regression and Analysis of Variance (ANOVA) which require normal distribution, homogeneity of variance and independence of residuals. In many cases, the variables used to assess biological phenomena are not normally distributed. Count data as well as binomial data are daily fare for invasion biologists (cf. Supplement A). Further, study designs that implicate

non-independence of the observations due to nested sampling, spatial dependence, (phylo-)genetic relatedness or repeated measures are common. Generalized Linear Mixed Models (GLMM) are flexible tools for analysing such data.

The fundamentals of GLMM have been established in the 1980s and the first software implementations occurred in the early 1990s (O'Hara 2009), but only after 2000 they have become widely available in statistical packages such as R (lme4 package first uploaded in 2003), SAS (PROC GLIMMIX became standard procedure in V9.2, 2008; first production version released in 2005), and ADMB (example of Poisson GLMM dated December 2006). Consequently, applications of GLMM in invasion biology have increased exponentially during the last few years (Fig. 1). The development of philosophies, sophisticated statistical procedures, and software implementations of GLMM is still continuing, but the basics are sufficiently mature for wide application in invasion biology. We will, however, not conceal that applying GLMM can be somewhat more complicated than classical linear modelling. There are different methods for estimating parameters and for testing significance and their suitability depends on the properties of the dataset. Further, it happens quite often that models do not converge, particularly when there are many fixed and random effects.

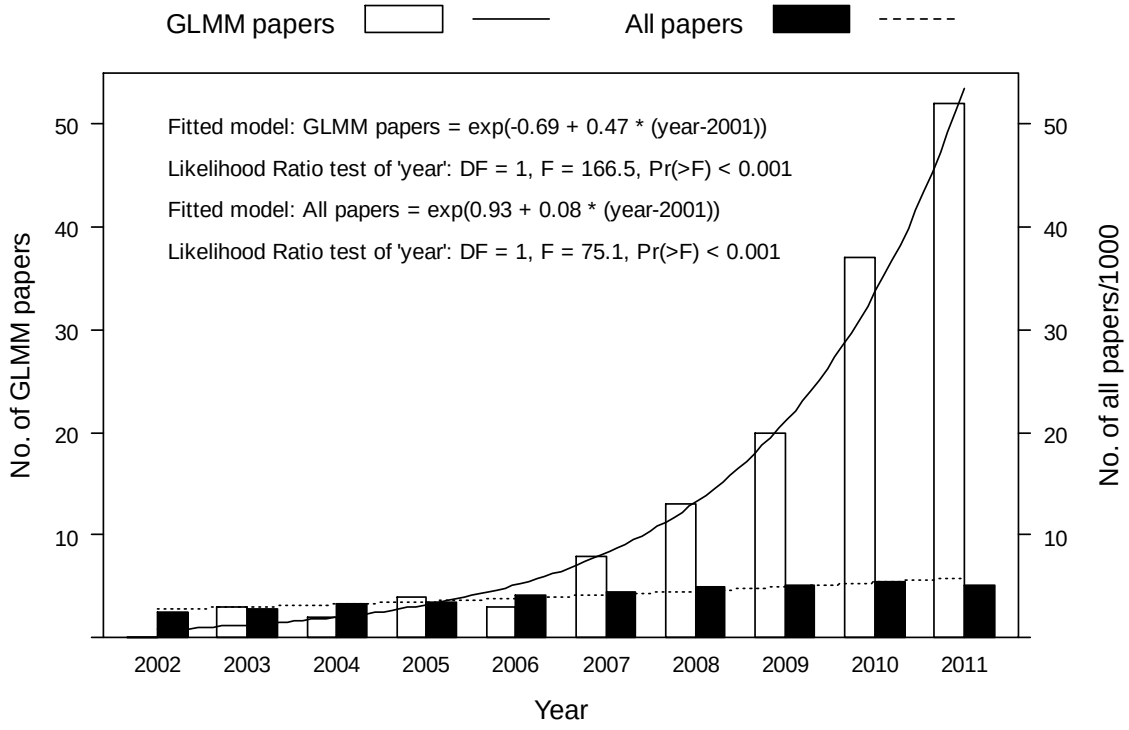

Figure 1 White bars show the number of peer-reviewed papers reporting applications of GLMM in invasion biology (GLMM papers) published from 2002 to 2011 according to the literature search documented in Supplement A. The figure for 2011 was linearly extrapolated from data covering 1 January to 12 July. In total, 116 GLMM papers were found in the literature searches. Black bars show the number of all papers found in Web of Science using the query "invasi* OR "non-native$" OR "non-indigenous" OR exotic$ OR alien$ OR weed*" and refining the results by subject areas as specified in Supplement A, ch. 1.1. The figure for 2011 was linearly extrapolated from data covering 1 January to 12 December 2011. Note that the number of all papers is given in thousands. The increase of the number of papers per year was modeled with GLM (quasi-Poisson, log-link).

In this review, we will try to elucidate applications of GLMM in the wide field of invasion biology, including not only spread, but also ecology of non-native species and impacts on native ecosystems. We will focus on mixed models, i.e. those including random effects, although some aspects, such as choosing suitable distributions or model

building are largely the same for Generalized Linear Models (GLM) without random effects.

The scope of this review is, first, to detect the most common issues that invasion biologist deal with when applying GLMM, and to identify common problems and possibly unused potentials, through analysing recent research articles. The results of this literature analysis are reported in Supplement A, if not in the main text. Second, we review the methodology of GLMM. We intend this review to serve as a guide to GLMM for invasion biologists. Hence, we will address all common issues of GLMM, if only briefly at times, by revisiting the steps of data analysis. Some emphasis is put on modelling zero-heavy data and validation of model assumptions. Third, we give worked examples of data analyses with GLMM in Supplement B–D that may help to clarify some details of model setups and interpretation of results.

Review Methodology

For the review of studies that applied GLMM in invasion biology, we searched Web of Science, Scopus and Google Scholar for articles published between 2002 and 2011 (effective July 12). We found 116 articles of which 50 were selected for analysis by assigning random numbers and sorting in ascending order. Details of the literature analysis are given in Supplement A.

What are GLMM?

GLMM are regression models that allow choosing among various distributions and link functions, just like GLM, in order to model a wide range of types of dependent variables through linear combinations of one or multiple predictor variables (fixed effects). Additionally, GLMM include random effects.

Random effects quantify the variation of regression intercept or slopes among the levels of a grouping variable by a probability distribution instead of estimating a fixed regression coefficient for each level. In GLMM this distribution is assumed to be Gaussian with zero mean and some variance to be estimated. Even though a random effect is described by a distribution, the values of its levels still may be estimated by maximum a posteriori estimation. In Linear Mixed Effects Models (LMM), i.e. GLMM with Gaussian response and identity link, this gives the best linear unbiased prediction (BLUP).

A grouping variable should be used as a random effect, if its levels may be conceived to be a random sample from a larger group (O'Hara 2009), e.g. some individuals drawn from a population, and if the variation among levels is more interesting than the effect of a single level. Often the levels of grouping variables are biologically meaningless, e.g. in the case of experimental blocks, but need to be taken into account in order to obtain valid p-values and estimates. Such ‘nuisance variables’ should be used as random effects when their variation is of interest. In all other cases variables should generally be used as fixed effects (Robinson 1991). Thus, it would be all right to use grouping or nuisance variables as fixed effects, but particularly with many levels, random effects have the practical advantage of using less degrees of freedom.

For auto-correlated data, such as repeated measures of the same individuals or spatially auto-correlated observations, it is sometimes sufficient to model the dependence by suitable random effects. However, the strength of correlation may depend on co-variates or co-factors, e.g. temporal or spatial distance. For instance, observations made at time point 1 (t1) and t2 may be more similar than at t1 and t3. GLMM can model complex temporal and spatial correlation structures (Dormann et al.

2007, Fieberg et al. 2010). Finally, also data where the random effects have heterogeneous variance can be analyzed with GLMM.

Choosing distribution and link function

The first step of the modelling process is to find a suitable distribution and link function for the data at hand. The natural distributions of count data are Poisson or, in case the variance is larger than the mean (overdispersion), the negative binomial distribution. Proportions and binary outcomes are naturally binomial variables. Depending on the software package, also other distributions may be available (Table 1).

Table 1 Some distributions used in GLM and GLMM, link functions and corresponding types of variables.

Distribution	Range of variable	Common link functions	Scale parameter	Biological variables
Gaussian	Real axis	Identity	Yes	Metric
Transformed to Gaussian	Positive real axis	Log[1]	Yes	Metric on logarithmic scale
Transformed to Gaussian	Positive real axis	Box-Cox[1]	Yes	Metric
Beta	Reals strictly between 0 and 1	Logit, Probit, complementary log-log[2]	Yes	Proportion
Gamma	Positive real axis	Log, Power	Yes	Rates
Poisson[3]	0,1,2,…	Log, Identity	No	Counts; e.g. abundances, species numbers
Negative binomial	0,1,2,…	Log, Identity	Yes	Counts with overdispersion
Binomial[3]	0,1,…,N where N is an a priori given number of trials	Logit, Probit, complementary log-log	No	Binary: e.g. presence-absence; proportions: e.g. germination percentage
Polytomous[3]	1,2,..,K	Cumulative logit, Ordered probit	No	Ordinal scale

[1]Transformation is applied on the data points prior to a statistical analysis using the identity link. Although the transformation is not used as a link function, the interpretation will be the same as for the link function.
[2]Inverse CDF for the log-Weibull distribution.
[3]A scale parameter may be introduced via the quasi-likelihood approach.

The purpose of the link function is to transform values of the dependent variable so that they match the scale of the linear predictor, i.e. [-∞, ∞], and to linearise the relationship with the predictor variables (Table 2). For each distribution, there is a canonical ('natural') link function, but there are also less commonly used alternatives that may suit the data better in some cases. For instance, binomial data may be modelled with probit link, or count data with large means may in some instances be modelled with identity link (Table 1). It is advisable to fit models with different links to the dataset and to use the link that yields the best model fit and parameter interpretation.

Table 2 Some link functions used in GLM and GLMM and their interpretations.

Range of variable	**Link function name**	**Link function formula**	**Common interpretation of predictions**	**Common interpretation of contrasts[1]**
Real axis	Identity	μ	Position	Difference
Positive real axis	Log	$\log(\mu)$	Position Log rate	Log ratio
Positive real axis	Box-Cox	$(\mu^{\lambda}-1)/(\lambda*\gamma^{\lambda})$[2,3]	Position	Not available
Positive real axis	Power	μ^{λ}	Position	Not available
Reals strictly between 0 and 1	Logit	$\log(\mu/(1-\mu))$	Log odds	Log odds ratio
Reals strictly between 0 and 1	Probit	$\Phi^{-1}(\mu)$	Gaussian distribution of change point	Difference in susceptibility
Probability vector	Cumulative logit	$\log(\pi/(1-\pi))$[4]	Log odds	Log odds ratio, independent of cut-off point
Probability vector	Ordered probit (cumulative probit)	$\Phi^{-1}(\pi)$[4]	Vector of change points.	Difference in susceptibility

[1]Regression coefficients of the levels (dummy variables) of categorical predictor variables. Also applicable to one-unit changes in continuous predictor variables.
[2]The exponent λ is a constant that is estimated from the data so that the data are as close as possible to normal distribution and homogeneity of variance on the transformed scale.
[3]Here, γ is the geometric mean of the data.
[4]Here π is the cumulative probability.

Models for zero-heavy data

Non-negative observations with exceedingly many zeros are an often occurring data situation that cannot be modeled by the probability distributions listed in Table 1. Lambert (1992) introduced a zero-inflated Poisson model (ZIP) for zero-heavy count data by increasing the probability of zeros in a Poisson distribution. If there is additional overdispersion beside the inflated probability for zero counts, then a possibility is to use the zero-inflated Negative Binomial model (ZINB; Welsh et al. 1996). An alternative

description of the ZIP and the ZINB is to combine a separate model for the zeros with the conditional model for the non-zero counts, i.e. the truncated Poisson or Negative Binomial distribution. These parameterizations are known as hurdle models (Cameron and Trivedi 1998). The interpretation is that there is a hurdle to be surpassed to have non-zero response. Zero-inflated and hurdle models describe the same probability distributions, so the choice between these two model classes depends on the interpretation of the model parameters. This choice may be made considering the origin for the additional zeros (Martin et al. 2005). Algorithmically, hurdle models may be analyzed using a bivariate model with a binary response quantifying whether the observation is zero or positive, and an additional counting response with the actual count for the positive observations. For continuous responses we may also interpret the hurdle as censoring, i.e. negative responses are censored to zero and positive responses are observed as they are. This model introduced by Tobin (1958) is known as Tobit regression and consists of a binary component with a probit link quantifying whether the observations are negative or positive, and a conditional normal distribution with the identity link for the positive responses. A special feature of Tobit regression in comparison with the ZIP and the ZINB is that the parameters in the binary and the conditional components are shared since the definition of the probit function fits together with the censoring interpretation. In Table 3, we have collected examples of bivariate models that may be used to model zero-heavy data. The last model is used in the worked example in Supplement C.

Table 3 Examples of bivariate models for zero-heavy data.

Two-component model	**Range of variable**	**Bivariate recoding of response y**	**Model components**
Zero-inflated	0,1,2,..	a) 0 if y=0	a) Binary.
		b) (1,y) if y>0	b) Binary, Poisson (or Negative Binomial) truncated in 0.
Tobit	Zero and positive real axis	a) 0 if y=0	a)Binary with probit link.
		b) $(1,y^{\lambda})$ if y>0	b)Binary with probit link, Gaussian with identity link.
Conditional log Gaussian	Zero and positive real axis	a) 0 if y=0	a) Binary with log link.
		b) (1,log(y)) if y > 0	b) Binary with log link, Gaussian with identity link.

Estimation methods

Before actually calculating the model we need to consider which estimation method we can use. The choice depends on the dependent variable and on the random effects that are to be included in the model (cf. Bolker et al. 2009).

If the dependent variable can be modelled with a normal distribution, we will conduct a LMM using Restricted Maximum Likelihood (REML) for parameter estimation. For non-normal GLMM, exact integration over the random effects is only possible in special cases, and the practitioner is faced with the choice among a wealth of

approximate methods that may give different results. Here we only discuss the most popular methods that are implemented in SAS and R.

Penalized Quasi-Likelihood (PQL) (Breslow and Clayton 1993) is widely used since it is computationally fast. However, PQL estimates are known to be biased and should be avoided for Poisson variables when the mean counts within groups are less than 5 and for binomial variables when the mean number of either successes or failures are less than 5. Further, GLMM can be approximated by LMM using pseudo-data (Wolfinger and O'Connell 1993). This method is known as Pseudo-Likelihood and is the default method in PROC GLIMMIX. Neither PQL nor pseudo-likelihood provides an approximation to the actual likelihood of the data, and hence these methods cannot be used to compare models by either likelihood ratio tests or by information criteria. The standard methods to attain such an approximation listed in increasing order of accuracy, but also computational costs, are the Laplace approximation, Gauss-Hermite Quadrature (GHC), and Monte Carlo integration (Gilks et al. 1996). In practice, GHC is too slow when the number of random effects is larger than three. Monte Carlo integration is closely related to the Markov Chain Monte Carlo (MCMC) techniques extensively used in Bayesian statistics, and many variants exist allowing for the analysis of very complicated models. This, however, is outside the scope of this review.

Convergence problems

With many fixed predictor variables compared to sample size, and with more than one random grouping variable, GLMM computation algorithms may fail to converge. This is a common problem in biological applications (see also worked examples in Supplement B–D).

Among several recommendations, Cheng et al. (2010) advice that centering, standardising and full-rank coding of the predictor variables and reduction of collinearity, if present, is done in order to alleviate convergence problems. In non-complete designs, full-rank coding means that treatment combinations not used should be removed from the design matrix (cf. Supplement C). Further, with caution, one may try out, if manipulations of the data table, such as aggregation of levels of categorical predictor variables, facilitate convergence. One might also try to change the settings for the algorithm, e.g. to increase the number of scoring steps and/or the maximum number of iterations, and to loosen the convergence criteria. If these measures are not successful, the reason for failing convergence is most likely a (too) complicated random-effects structure (Cheng et al. 2010), and a solution could be to simplify the random and/or fixed effects model.

Analysis strategies

In the literature analysis (Supplement A), a gross classification gives that the primary modelling aim was inference in 66 % of the analyses and prediction/forecasting in 29 %, while estimation of random variation accounted for the remaining 5 %. The purpose of inference is to provide significance tests of relationships between dependent and predictor variables, i.e. to state p-values answering the question: *Is there an effect?* Thereafter, parameter estimates and confidence intervals are stated as answers to the follow-up question: *What is the effect?* Further, one often calculates local group means and their confidence intervals to assess the impact of the predictor variables on the dependent variable. In prediction, the present dataset is used for calculating the expected ('predicted') mean values, whereas in forecasting a new data set of the same predictor variables is used. We speak of projection, if the new dataset represents a hypothetical scenario rather than measured data. For prediction, forecasting and projection, we may

ask: *Which model is best*? In all cases, model validation should be done answering the question: *Can the conclusions be trusted?*

Is there an effect?

Computation of p-values for the significance of fixed effects in a GLMM is often done by either Wald or Likelihood Ratio (LR) tests.

Wald tests compare parameter estimates against their standard error like t-tests in classical regression analysis for the null hypothesis that a regression coefficient equals zero. In the case of distributions with fixed dispersion, e.g. Poisson and binomial, we can use Wald Chi-square tests, while in the case of distributions where the dispersion or variance is estimated, e.g. normal, quasi-Poisson or quasi-binomial, we need to use Wald F tests which require the denominator degrees of freedom (df). Several methods for estimating df in GLMM have been proposed (e.g. Satterthwaite 1941, Kenward-Roger 1997; cf. also Bolker et al. 2009), but these do not always give reliable p-values. In particular, caution should be exercised when the standard errors are large. Large standard errors also may result in inflated p-values in cases with less identifiable parameters. The phenomenon e.g. occurs in quasi-separated binary data, where a regression parameter simply should be large. However, the parameter will be estimated at some value which can be outweighed by the standard error.

Likelihood Ratio tests compare nested models, with and without the effect to be tested, and test the null hypothesis of no difference in residual deviance. Like with Wald tests, there are different variants of the LR tests for distributions where the variation of the data needs to be estimated (→ LR *F* test) and distributions with a priori defined variation (→ LR *Chi²* test). When conducting LR tests of fixed effects in LMM, it is recommended to use Maximum Likelihood (ML) for estimating parameters (Bolker et al. 2009, Cheng et al. 2010, Pinheiro and Bates 2000, Zuur et al. 2009). It is also possible to use REML, but then the restricted likelihood should be defined using the design under the reduced model (Welham and Thompson 1997). This method, however, is not implemented in standard software.

Regarding random effects, significance can be tested with LR tests comparing nested models which differ by one random intercept or slope (Morrell 1998). This tests the hypothesis that a variance component equals zero against the alternative that it is positive. Hence, the hypothesis lies on the boundary for the possible values of the variance component, and the LR test should be evaluated in a mixture of a point distribution in zero and the chi-square distribution (Bolker et al. 2009, Self and Liang 1987, Molenberghs and Verbeke 2007). For random effects in LMM, it is recommended that REML is used to define the LR statistic (Morrell 1998).

Also a single random effect can be tested with LR comparing the model including the random effect with a model that does not include it (which is an LM or GLM), but many GLMM software packages do not offer this option. It is possible to fit the model without the random effect in another program and, then, to conduct the LR test, but it is important to make sure that the log-likelihoods are commensurate in both programs (http://glmm.wikidot.com/random-effects-testing). Alternatively, the random effect can be tested with parametric bootstrap (see Supplement B). Although the parametric bootstrap does not account for the variability of the parameter estimates, it is often more trustworthy than the LR test since it does not rely on the asymptotical distribution of the test statistic.

What is the effect?

Once we have tested the significance of effects, we might want to know how strong the effects are. How much does the dependent variable change given a unit change in a predictor variable? We need to take into account that we did not model the dependent variable itself, but its transformation by the link function (Table 2). Hence, if we want to know the effects on the original scale, we need to backtransform the predictions, e.g. if we conducted a Poisson-GLMM with log link we have to exponentiate the estimates of fixed predictor variables (for an example of logistic modelling see Supplement B).

The interpretation of the estimates may depend on whether random effects are included or not. Suppose, for instance, that we have two observations from each subject, where the subjects are representatives from some population, and that we estimate some fixed effect. In the model with random intercepts the fixed effect will then be subject-specific. In the model without random intercepts the fixed effect will be averaged over the population.

Like all statistics calculated from a sample, model predictions are not the true value of the population, but estimates that include uncertainty. The range in which the true value of the effect is likely to be found is given by confidence intervals that are routinely provided by most GLMM standard software.

Only few studies in invasion biology aim at estimating the effect of random variables on the dependent variable (see Supplement A). But the random variation among subjects or groups may be interesting as well (O'Hara 2009). A random intercept effect measures how much the group-specific intercepts vary around the global intercept, and the strength of the random effect may be assessed comparing its estimated standard deviation to the size of the fixed effect, i.e. the global intercept, in this case. For instance, if the global intercept was 10 and the estimated standard deviation of the random intercept was 2.5, then approximately 95 % of group-specific intercepts would be in the range of 10± 2*2.5, i.e. between 5 and 15. If there was a fixed slope estimate of 2 and the standard deviation of the random slope effect was 0.5, then we would have to expect group-specific slopes within the range of approximately 1 and 3. A worked example of how to interpret random effects is given in Douglas Bates new book on "mixed effects modeling in R" (Bates 2010, lme4.r-forge.r-project.org/book/front.pdf).

Which model is best?

According to our literature analysis, 34 % of GLMM applications in invasion biology conducted model building before final parameter estimation and inference. In our view, this is particularly appropriate whenever the purpose of the study is prediction, forecasting or projection. Full models give unbiased estimates, but may not be good for prediction because they may contain insignificant predictors (Whittingham et al. 2006) or, more generally, be over-fit (Crawley 2002). For inference, model selection is only advisable, if the number of predictor variables is large. Otherwise, the full model containing all available predictor variables should be used. Generally, we would tend to keep biologically meaningful variables in the model, even if they are not significant (cf. Cheng et al. 2010).

Strategies for model building are forward selection, backward elimination and best subset (Bolker et al. 2009, Cheng et al. 2010). The stepwise procedures have been repeatedly criticized because the order of parameter entry or deletion can influence the selection result, multiple tests involved in the procedures inflate type I errors, and parameter estimates may be biased (Burnham & Anderson 2002, Whittingham et al. 2006). For inference, we recommend that model building is done by backward model selection. Forward model selection should only be used, if there are too many

predictors. For prediction and forecasting, we recommend best subset modelling and possibly model averaging (cf. e.g. Johnson and Omland 2004). In any case, the candidate models must have the same random-effects structure when selecting fixed effects, and vice versa if selecting random effects (e.g. Cheng et al. 2010).

With GLMM, best subset modelling may easily become computationally expensive, when there are several fixed or random effects. Hence, it will often be necessary to decide on a sensible maximum model, i.e. a subset of all possible fixed and random effects and interactions that can be calculated in reasonable time (cf. Bolker et al. 2009, Cheng et al. 2010).

In best subset modelling, Information Criteria (IC) are used for evaluating candidate models. IC consider both model fit (deviance) and complexity (df of the model parameters), and they also can compare nonnested models (contrary to stepwise procedures). Aikaike's Information Criterion (AIC) is the most widely used IC in invasion biology and ecology. For small sample sizes, it is recommended to use corrected AIC (AICc) which penalizes more strongly for model complexity (Burnham & Anderson 2002). For overdispersed data, quasi-AIC (QAIC) can be used, although this has been criticized (cf. Bolker et al. 2009). Bayes Information Criterion (BIC) is very similar to AIC, but used less commonly. BIC tends to favor less complex models compared to AIC (Keselman et al. 1998). Both AIC and BIC require estimating the degrees of freedom of the parameters in the model which is problematic with random effects (Vaida & Blanchard 2005). The choice of IC is largely subjective as no variant is consistently superior to the others (Cheng et al. 2010). An alternative is the Deviance Information Criterion (DIC) which is calculated using MCMC sampling and takes the effective number of model parameters into account (Spiegelhalter et al. 2002, Miaou & Song 2005). DIC has recently gained popularity in ecology (Bolker et al. 2009).

The aim of model building is often to find one 'best model' that is used for parameter estimation and inference. However, several different models may fit the data similarly well, so that model selection may be uncertain. Stepwise procedures and selection of a single best model do not account for such uncertainty (Whittingham et al. 2006). With IC and best subset modelling it is possible to identify similarly good models that are within a certain range of IC values, e.g. Δ AIC < 4, and then, to average parameter estimates among them using Akaike's weights. Multimodel averaging has increasingly been advocated and applied in ecological studies (Johnson & Omland 2004, Dormann et al. 2008, Bolker et al. 2009) and is recommendable particularly for prediction and forecasting (Whittingham et al. 2006).

Model building of random effects appears to be of less importance in invasion biology, because most studies use single random intercepts or random-effects structures that are predetermined by study design (cf. Supplement A). In principle, however, model building is as sensible for random effects as for fixed effects and can be conducted in a similar way. If model building of random effects is desired, this should be done before selection of fixed effects, i.e. using the full or maximum model, because the results of fixed-effects model building may depend on the random-effects structure (Zuur et al. 2009, Cheng et al. 2010).

Can the conclusions be trusted?

GLMM rely on assumptions that need to be met in order to get valid estimates and p-values. In case of backward model selection these should be validated for the initial and the final model, and in case of best-subset selection validation should be done for the selected model. The assumptions of GLMM are:

a) A response distribution.

b) A link function.
c) Linearity against the predictors on the scale of the link function.
d) Gaussian distribution of the random effects.

Before describing possibilities to validate GLMM, we first discuss the special methods that are available for the validation of LMM. In Gaussian models, the specification of the link function is replaced by transformations of the actual observations of the response variable, if necessary, and hence the link is the identity function. The standard validation methods for Gaussian models investigate the statistical properties of the residuals and of the predicted random effects. Basically, there exist two sets of residuals for LMM. The (unconditional) residuals are the differences between the observations and the estimated fixed effects, while for the conditional residuals the predicted random effects, e.g. the random intercepts and effects of slopes of the group levels, also are subtracted. The residuals, the conditional residuals, and the predicted random effects are all assumed to be Gaussian, and the conditional residuals are approximately independent. As a consequence of this, the model assumptions of LMM may be assessed by the following graphical diagnostics:

I. The Gaussian distribution is validated by a normal quantile plot of the conditional residuals. Variance homogeneity is validated by a scatter plot of the conditional residuals against the predicted values. Independence of the error terms may be validated by an autocorrelation plot of the conditional residuals.
II. The appropriateness of the identity link is assessed with scatter plot of the residuals and of the conditional residuals against the predicted values.
III. Linearity against the predictors is assessed with scatter plots of the residuals and of the conditional residuals against the individual covariates.
IV. The Gaussian distribution of random effects is validated by a normal quantile plot of the estimated random effects (BLUPs).

The normal quantile plots may be accompanied by goodness-of-fit tests based on an adequate statistic, e.g. the Shapiro-Wilks, Kolmogorov-Smirnov, Cramer-von-Mises, or the Anderson-Darling statistic (see D'Agostino and Stephens 1986). These tests, however, may have too large power in the sense that they may reject the normal distribution for non-important deviations (see Supplement C). Furthermore, not even the conditional residuals are strictly independent, and hence the type I error of the tests may not be at the significance level. Ritz (2004) devised a goodness-of-fit test for the distribution of the random effect taking the dependence between the predictions of the random effects into account, but to our knowledge this test is not readily available in the standard software packages.

We are not aware of any standard methods for a detailed assessment of the response distribution in GLM(M). Instead, the choice of the distribution is often based on qualitative properties of the experimental design, e.g. the Poisson and the negative binomial distributions are the natural choices for count data (Table 1). However, Pearson or deviance goodness-of-fit tests for overdispersion in distributions with fixed dispersion are often performed. There have been several attempts to define useful residuals in framework of GLM (see Pierce and Schafer 1986). But the distributional properties of such residuals are not explicitly known. So the interpretation of classical residual plots for GLM is difficult, and for GLMM things just get worse. For graphical assessment of link function and linearity, cumulative residuals and associated goodness-of-fit tests have been proposed for GLM (Lin et al 2002). It is possible to describe the asymptotic distribution of the cumulative residuals using simulations, and the method

extends to GLMM invoking the Generalized-Estimating-Equations (GEE) approach (Liang and Zeger 1986). Concerning the distribution of the random effects, the goodness-of-fit test proposed by Ritz (2004) was extended to GLMM, but it has low power for the logistic regression (Waagepetersen 2006). It is still possible to make a normal quantile plot of the predicted random effects, but due to unknown distributional properties of the predicted random effects, there is no justification for alarm even for less nice looking plots. In summary, GLMM may be validated as follows:

a) Make a histogram of the raw observations to see if the chosen response distribution is completely off. For distributions with fixed dispersion perform a Pearson or deviance goodness-of-fit test.
b) To assess the appropriateness of the link function, plot the cumulative residuals against the linear predictor, possibly accompanied by a goodness-of-fit test.
c) To assess linearity against the predictor variables, plot the cumulative residuals against the individual continuous predictors, possibly accompanied by a goodness-of-fit test.
d) Make a normal quantile plot of the predicted random effects. This plot, however, may only be used to find comfort and cannot be used to invalidate the distributional assumption.

The cumulative residuals may be done in PROC GENMOD in SAS, which also provides the Kolmogorov-Smirnov goodness-of-fit test and allows for correlation via the GEE-approach (see Supplement C). In R, cumulative residuals and the associated Kolmogorov-Smirnov and Cramer-von-Mises tests may be done via the gof-package (Holst 2011). This package, however, does not include the GEE-approach and hence only works for GLM. To use the R-package on GLMM, the random effects should either be removed or reused as fixed effects in the validation step (see Supplement B).

How to report the model?

In the literature analysis, we found that many papers did not report crucial aspects of GLMM. For instance, 62 % did not report the method of parameter estimation (PQL, Laplace etc.). Not a single paper reported all of the information necessary for evaluation of the methods.

We suggest that the following list of information should be routinely provided in papers (modified after Bolker et al. 2009, their supplementary material): study design, sample size, number of levels of random grouping variables, software package, type of dependent variable, distribution, over-/underdisperion (for Poisson data and proportions modelled with binomial distribution), link function, method of parameter estimation, test methods of fixed and random effects, estimation of df of the residuals (when using Wald F tests) and of the random effects (when using AIC or BIC or their variants), model selection criteria and strategy; for Poisson data: mean and variance; for proportions: minimum number of successes/ failures, results of model validation, magnitude of random effects.

Conclusion/Summary

GLMM are important tools in invasion biology, because study designs often involve non-Gaussian dependent variables and independence of observations due to spatial or temporal grouping. Applications of GLMM have rapidly increased after standard software had become available, and they are likely to increase further in future. It is difficult to say, if invasion biologists manage GLMM all right or if flawed applications

are common, because most papers do not report sufficient details. GLMM are as flexible and powerful as they are complicated and challenging. Users should be aware of the different methods of significance testing and of estimating parameters. PQL is often not suitable for studies in invasion biology. Laplace approximation is a good compromise between precision and computational speed and will be suitable for most studies. Currently, MCMC techniques are becoming more commonly available. They may help to solve some of the difficulties in inference and estimation (Bolker et al. 2009).

Model validation is hardly ever reported, but of crucial importance for valid inference and estimation. Perhaps, users should pay more attention to validation of model assumptions. Generally, we encourage reporting on methods more rigorously, if not in the paper itself, then in online supplements.

Structured co-variance matrices are good tools for modelling temporally, spatially or phylogenetically correlated data (cf. Supplement D). Repeated measures are common in invasion biology (32 % of reviewed GLMM analyses), but so far most such studies have used unstructured co-variance matrices, although explicit modelling of temporal auto-correlation would give more precise p-values. For some recent studies that modelled auto-correlation of repeated measures see McEachern (2009), Chun et al. (2010), Tognetti et al. (2010). An application of GLMM to modelling spatial auto-correlation of grid-based distribution data can be found in Gassó et al. (2009). One difficulty is that correlation structures currently are only widely implemented for LMM, but rarely for GLMM. The exception appears to be PROC GLIMMIX in SAS, GEE (Carl and Kühn 2007) and many of the Bayesian approaches. In view of the further rapid software development, the potential for modelling correlated data may increase in future. Generally, we encourage using structured co-variance matrices for modelling correlated data whenever possible.

Usually invasion biologists are not interested in testing or interpreting the random effects. This is obviously due to the fact that random variables most often are nuisance variables. However, there may be some potential in using random-effects modelling more consciously. For instance, a species invasion potential may not only depend on mean traits of the population, but also on genetic variation at genotype or population level that can be measured in GLMM as random effects (see e.g. Buckley et al. 2003, Brodersen et al. 2008, Xu et al. 2010).

Acknowledgements

We thank Thure Hauser, Rikke Bagger Jørgensen, Anne-Marie Thonning-Skou, Lene Sigsgaard and Johannes Kollmann for providing datasets used in Supplements C and D. We thank Ingolf Kühn and two anonymous referees for many helpful comments and suggestions on an earlier version of this paper.

References

O'Hara RB. How to make models add up – a primer on GLMMs. Annales Zoologici Fennici 2009;46:124–137.

Robinson GK. That BLUP is a good thing: The estimation of random effects. Statistical Science 1991;6:15–32.

Dormann CF, McPherson JM, Araújo MB, Bivand R, Bolliger J, Carl G et al. Methods to account for spatial auto-correlation in the analysis of species distributional data: a review. Ecography 2007;30:609–628.

Fieberg J, Matthiopoulos J, Hebblewhite M, Boyce MS, Frair JL. Correlation and studies of habitat selection: problem, red herring or opportunity? Philosophical Transactions of the Royal Society B 2010;365:2233–2244.
Lambert D. Zero-inflated Poisson regression, with an application to defects in manufacturing. Technometrics 1992;34:1–14.
Welsh AH, Cunningham RB, Donelly CF, Lindenmayer DB. Modelling the abundance of rare species: statistical models for counts with extra zeros. Ecological Modelling 1996;88:297–308.
Cameron AC, rivedi PK. Regression analysis of count data. New York, USA: Cambridge University Press;1998.
Martin TG, Wintle BA, Rhodes JR, Kuhnert PM, Field SA, Low-Choy SJ et al. Zero tolerance ecology: improving ecological inference by modelling the source of zero observations. Ecology Letters 2005;8:1235–1246.
Tobin J. Estimation of relationships for limited dependent variables. Econometrica 1958;26:24–36.
Bolker BM, Brooks ME, Clark CJ, Geange SW, Poulsen JR, Stevens MHH, et al. Generalized linear mixed models: a practical guide for ecology and evolution. Trends In Ecology & Evolution 2009;24:127–135.
Breslow NE, Clayton DG. Approximate inference in Generalized Linear Mixed Models. Journal of the American Statistical Association 1993;88:9–25.
Wolfinger R, O'Connell M. Generalized Linear Mixed Models: A Pseudo-likelihood Approach. Journal of Statistical Computation and Simulation 1993;4:233–243.
Gilks WR, Richardson S, Spiegelhalter DJ. Introducing Markov chain Monte Carlo. In: Gilks WR, Richardson S, Spiegelhalter DJ, editors. Markov Chain Monte Carlo in practice. Boca Raton, USA: Chapman and Hall; 1996. p. 1–19.
Cheng J, Edwards LJ, Maldonado-Molina MM, Komro KA, Muller KE. Real longitudinal data analysis for real people: Building a good enough mixed model. Statistics In Medicine 2010;29:504–520.
Satterthwaite FF. Synthesis of variance. Psychometrika 1941;6:309–316.
Kenward MG, Roger JH. Small sample inference for fixed effects from restricted likelihood. Biometrics 1997;53:983–997.
Pinheiro JC, Bates DM. Mixed-Effects Models in S and SPLUS. New York, USA: Springer; 2000.
Zuur AF, Ieno EN, Walker NJ, Saveliev AA, Smith GM. Mixed effects models and extensions in ecology with R. New York, USA: Springer; 2009.
Welham S, Thompson R. Likelihood ratio tests for fixed model terms using residual maximum likelihood. Journal of the Royal Statistical Society B 1997;59:701–714.
Morrell CH. Likelihood ratio testing of variance components in the linear mixed-effects model using restricted maximum likelihood. Biometrics 1998;54:1560–568.
Self SG, Liang KY. Asymptotic Properties of Maximum Likelihood Estimators and Likelihood Ratio Tests under Nonstandard Conditions. Journal of the American Statistical Association 1987;398:605–610.
Molenberghs G, Verbeke G. Likelihood ratio, score, and Wald tests in a constrained parameter space. American Statistician 2007;61:22–27.
Bates D. lme4: Mixed-effects modeling with R. Springer / lme4.**r**-forge.**r**-project.org/book/front.pdf; 2010.
Whittingham MJ, Stephens PA, Bradbury RB, Freckleton RP. Why do we still use stepwise modelling in ecology and behaviour? Journal of Animal Ecology 2006;75: 1182–1189.

Crawley MJ. Statistical computing: an introduction to data analysis using S-Plus. Chichester: Wiley; 2002.
Burnham KP, Anderson DR. Model selection and multimodel inference: a practical information-theoretic approach. 2nd ed. New York, USA: Springer; 2002.
Johnson JB, Omland KS. Model selection in ecology and evolution. Trends in Ecology and Evolution 2004;19:101–108.
Keselman HJ, Algina J, Kowalchuk RK, Wolfinger RD. A comparison of two approaches for selecting covariance structures in the analysis of repeated measures. Communications in Statistics - Simulation and Computation 1998;27:591–604.
Vaida F, Blanchard S. Conditional Akaike information for mixed-effects models. Biometrika 2005;92:351–370.
Spiegelhalter DJ, Best N, Carlin BP, Linde A. Bayesian measure of model complexity and fit. Journal of the Royal Statistical Society B 2002;64:583–639.
Miaou SP, Song JJ. Bayesian ranking of sites for engineering safety improvements: Decision parameter, treatability concept, statistical criterion, and spatial dependence. Accident Analysis and Prevention 2005;37:699–720.
Dormann CF, Schweiger O, Arens P, Augenstein I, Aviron S, Bailey D et al. Prediction uncertainty of environmental change effects on temperate European biodiversity. Ecology Letters 2008;11:235–244.
D'Agostino RB, Stephens, MA. Goodness-of-fit techniques. New York, USA: Marcel Dekker; 1986.
Ritz C. Goodness-of-fit tests for mixed models. Scandinavian Journal of Statistics 2004;31:443–458.
Pierce DA, Schafer DW. Residuals in generalized linear models. JASA 1986;81:977–986.
Lin DY, Wei LJ, Ying Z. Model-Checking Techniques Based on Cumulative Residuals. Biometrics 2002;58:1–12.
Liang KY, Zeger SL. Longitudinal data analysis using generalized linear models. Biometrika 1986;72:13–22.
Waagepetersen R. A simulation-based goodness-of-fit test for random effects in generalized linear mixed models. Scandinavian Journal of Statistics 2006;33:721–731.
Holst KK. "gof", R-package for model-diagnostics based on cumulative residuals, version 0.7-6. 2011.
McEachern AK, Thomson DM, Chess KA. Climate alters response of an endemic island plant to removal of invasive herbivores. Ecological Applications 2009;19:1574–1584.
Chun YJ, Kim CG, Moloney KA. Comparison of life-history traits between invasive and native populations of purple loosestrife (*Lythrum salicaria*) using nonlinear mixed effects model. Aquatic Botany 2010;93:221–226.
Tognetti PM, Chaneton EJ, Omacini M, Trebino HJ, León RJC. Exotic vs. native plant dominance over 20 years of old-field succession on set-aside farmland in Argentina. Biological Conservation 2010;143;2494–2503.
Gassó N, Sol D, Pino J, Dana ED, Lloret F, Sanz-Elorza M, Sobrino E, Vilà M. Exploring species attributes and site characteristics to assess plant invasions in Spain. Diversity and Distributions 2009;15:50–58.
Carl G, Kühn I. Analyzing spatial autocorrelation in species distributions using Gaussian and logit models. Ecological Modelling 2007;207;159–170.

Buckley YM, Downey P, Fowler SV, Hill R, Memmot J, Norambuena H et al. Are invasives bigger? A global study of seed size variation in two invasive shrubs. Ecology 2003;84:1434–1440.

Brodersen C, Lavergne S, Molofsky J. Genetic variation in photosynthetic characteristics among invasive and native populations of reed canarygrass (*Phalaris arundinacea*). Biological Invasions 2008;10:1317–1325.

Xu CY, Julien MH, Fatemi M, Girod C, Van Klinken RD, Gross CL et al. Phenotypic divergence during the invasion of *Phyla canescens* in Australia and France: evidence for selection driven evolution. Ecology Letters 2010;13:32–44.