ROYAL STATISTICAL SOCIETY
DATA | EVIDENCE | DECISIONS

Journal of the Royal Statistical Society
Applied Statistics
Series C

Appl. Statist. (2015)
64, *Part* 2, *pp.* 231–252

Bayesian analysis of measurement error models using integrated nested Laplace approximations

Stefanie Muff,

University of Zurich, Switzerland

Andrea Riebler,

Norwegian University of Science and Technology, Trondheim, Norway

Leonhard Held,

University of Zurich, Switzerland

Håvard Rue

Norwegian University of Science and Technology, Trondheim, Norway

and Philippe Saner

University of Zurich, Switzerland

[Received July 2013. Final revision March 2014]

Summary. To account for measurement error (ME) in explanatory variables, Bayesian approaches provide a flexible framework, as expert knowledge can be incorporated in the prior distributions. Recently, integrated nested Laplace approximations have been proven to be a computationally convenient alternative to sampling approaches for Bayesian inference in latent Gaussian models. We show how the most common approaches to adjust for ME, the classical and the Berkson ME, fit into this framework. This is achieved through a reformulation with augmented pseudo-observations and a suitable extension of the latent Gaussian field. Two specific classes are described, which allow for a particularly simple implementation using integrated nested Laplace approximations. We present three applications within the framework of generalized linear (mixed) models with ME. To illustrate the practical feasibility, R code is provided in on-line supplementary material.

Keywords: Bayesian analysis; Berkson error; Classical error; Integrated nested Laplace approximation; Measurement error

1. Introduction

The existence and the effects of measurement error (ME) in statistical analyses have been recognized and discussed for more than a century; see for example Pearson (1902), Wald (1940), Berkson (1950), Fuller (1987) and Carroll *et al.* (2006). The sources of ME are manifold and imply much more than just instrumental imprecision in the measurement of physical variables, such as length and weight, but may include for instance biases due to preferential sampling, incomplete observations or misclassification.

Address for correspondence: Leonhard Held, Division of Biostatistics, Institute of Social and Preventive Medicine, University of Zurich, Hirschengraben 84, 8001 Zurich, Switzerland.
E-mail: leonhard.held@ifspm.uzh.ch

 0035–9254/15/64231

If ME is ignored, parameter estimates and confidence intervals may suffer from serious biases. If a regression model is multivariate and some covariates can be measured with and some without error, even the effects of the error-free measured covariates can be biased, where the direction of the bias depends on the correlation between covariates (Carroll *et al*., 1985; Gleser *et al*., 1987). Moreover, ME may cause a loss of power for detecting signals and connections among variables and may mask important features of the data. Given these facts, it is surprising that ME is often completely ignored or not treated properly. One reason might be that standard statistical textbooks on regression often pay very little attention to this aspect, although the problems have been recognized for a long time.

For successful error correction both the amount of error (i.e. the error variance) and the error model need to be specified correctly. Hence, information about the underlying measurement process is essential. Possible errors must be identified early in a study and the entire data collection process should be driven by such considerations. In recent decades, several approaches to model and correct for ME have been proposed, such as method-of-moments corrections (Fuller, 1987), simulation extrapolation (Cook and Stefanski, 1994), regression calibration (Carroll and Stefanski, 1990; Gleser, 1990) or Bayesian analyses (Clayton, 1992; Stephens and Dellaportas, 1992; Richardson and Gilks, 1993; Dellaportas and Stephens, 1995; Gustafson, 2004). A thorough overview of current state of the art methods is given in Carroll *et al*. (2006) and Buonaccorsi (2010).

In this paper, we focus on Bayesian approaches where prior knowledge, and in particular prior uncertainty, e.g. in variance estimates, can be incorporated in the model. Up to now, posterior marginal distributions in such ME models have been estimated by employing a Markov chain Monte Carlo (MCMC) sampler; see for example Stephens and Dellaportas (1992) or Richardson and Gilks (1993). The MCMC approach is very general, flexible and widely used. However, case-specific implementation may be challenging and sampling can become rather time consuming, although the use of generic software such as WinBugs (Lunn *et al*., 2000), OpenBugs (Lunn *et al*., 2009), JAGS (Plummer, 2003) or MCMC samplers in R, such as MCMCpack (Martin *et al*., 2011), facilitates the application of MCMC approaches.

Rue *et al*. (2009) proposed the integrated nested Laplace approximation (INLA) algorithm INLA, which is an alternative to MCMC sampling for the class of latent Gaussian models. INLA avoids sampling by accurately approximating posterior marginal distributions. Because of its flexibility in the choice of likelihood functions and latent models, the INLA approach is also an alternative to likelihood-based inference in particular for generalized linear mixed models (GLMMs) (Fong *et al*., 2010). INLA is implemented in C and runs under Linux, Windows and Macintosh systems via a freely available R interface (R Core Team, 2012). Models can be fitted in short time making prior sensitivity analysis (Roos and Held, 2011) and even bootstrap studies (see Section 5.2) feasible.

Here, we show how the most common approaches to adjust for ME, namely classical and Berkson ME, fit into the framework of latent Gaussian models. This is achieved through a reformulation with augmented pseudo-observations and a suitable extension of the latent field. Two specific classes, one for a simplified form of classical ME and one for Berkson ME, which allow for a particularly easy implementation in INLA, are also considered. We hope that our work facilitates the access to ME problems for a broad audience and that the solution that is presented here will stimulate the greater use of Bayesian methods for the analysis of data subject to ME. For this, R code is provided in the on-line supplementary material.

An alternative approach towards approximate inference in ME models is the variational Bayes method (Bishop, 2006; Ormerod and Wand, 2010), which has been recently applied by Pham *et al*. (2013) to the special case of simple (parametric and non-parametric) linear

regression with classical ME. Variational Bayes methods provide accurate point estimates; however, certain variational approximations tend to give too narrow posterior distributions, i.e. too optimistic uncertainty estimates. See Bishop (2006), page 467, and Rue *et al*. (2009), appendix A, for illustration of this problem in the case of latent Gaussian models. Novel variational approaches to improve on this are being actively explored; see for example Ormerod and Wand (2012).

This paper is organized as follows. Section 2 introduces three applications from the biological or medical field containing a linear regression combined with heteroscedastic classical error, a logistic model with a binary error-free covariate and a model suffering from classical error, and an overdispersed Poisson regression model with Berkson error. In Section 3 we review classical and Berkson ME and the effects on the estimates of regression coefficients. Bayesian analysis with INLA is introduced in Section 4, where we describe how to use this framework for model inference in the presence of classical and Berkson ME. Section 5 presents modelling details and the results of the three applications analysed with both the INLA and the MCMC approach. Finally, we provide a discussion and outlook in Section 6.

The data that are analysed in the paper and the programs that were used to analyse them can be obtained from

`http://wileyonlinelibrary.com/journal/rss-datasets`

2. Examples of measurement error problems

In this section we introduce three applications which will be analysed in detail in Section 5. Here, we mainly describe the problems at hand and the differences in the results depending on whether or not ME has been incorporated in the analysis. All parameter estimates in ME models are obtained by using INLA, as outlined in detail in subsequent sections.

2.1. Inbreeding in Swiss ibex populations

We analysed data described by Bozzuto *et al*. (2014) on 26 Alpine ibex populations in Switzerland, some of them monitored over the past 100 years. The study aimed to quantify the effect of inbreeding on populations' intrinsic growth rates. The intrinsic growth rate y_i of a population i is the theoretical maximal rate of increase, if there are no density-dependent effects. For each population the value y_i has been estimated as a parameter of a non-linear state space model based on time series data of abundance counts, harvest proportions and numbers of animals released. The inbreeding coefficient x_i of population i (which is often denoted as f_i) is a quantity between 0 and 1, with larger values indicating stronger inbreeding. Unfortunately, x_i cannot be measured exactly. A previous Bayesian analysis based on genotype experiments at 37 neutral microsatellite loci provided estimates for x_i, denoted by w_i, and error variances for each population i. Additional covariates that may influence the intrinsic growth rate include the number of years that a population was observed, the average precipitation in summer, an interaction between the two, and the average precipitation in winter. These covariates are treated as error free and subsumed in a row vector $\mathbf{z}_i$.

Fitting a linear regression model $E(y_i) = \beta_0 + \beta_x x_i + \mathbf{z}_i\boldsymbol{\beta}_z$ with highly dispersed independent priors (normal with zero mean and inverse variance equal to 0.0001) on β_0, β_x and the components of $\boldsymbol{\beta}_z$ in INLA, but using the proxy w_i instead of the true but unobserved x_i, the absolute value of the estimated slope parameter $|\hat{\beta}_x|$ is attenuated ($\hat{\beta}_x = -0.91$; 95% credible interval $[-2.18, 0.36]$). Indeed, after accounting for ME the effect of inbreeding on population growth dynamics is more pronounced ($\hat{\beta}_x = -1.82$; 95% credible interval $[-3.87, 0.13]$).

2.2. Influence of systolic blood pressure on coronary heart disease

The Framingham heart study is a large cohort study that aimed to understand the factors leading to coronary heart disease and, in particular, to characterize the relationship to systolic blood pressure SBP (Kannel *et al.*, 1986). The outcome $y_i \in \{0, 1\}$ is a binary indicator for presence or absence of the disease and is modelled via a logistic regression. We analysed data from $n = 641$ males that were originally presented in MacMahon *et al.* (1990). As in Carroll *et al.* (2006), section 9.10, we use $x_i = \log(\text{SBP}_i - 50)$ and a binary smoking status indicator $z_i \in \{0, 1\}$ as predictors. The transformation of SBP, which was originally proposed by Cornfield (1962), has also been used in Carroll *et al.* (1984, 1996, 2006). Since it is impossible to measure the long-term SBP, measurements at single clinical visits had to be used as a proxy. Note that, owing to daily variations or deviations in the measurement instrument, the single-visit measures might considerably differ from the long-term blood pressure (Carroll *et al.*, 2006). Hence, the ME in SBP has been a concern for many years in this study. Importantly, the magnitude of the error could be estimated, as SBP had been measured twice at different examinations. These proxy measures for x_i are denoted w_{1i} and w_{2i}. A naive approach ignoring ME would fit a logistic regression against the indicator of coronary heart disease:

$$\text{logit}\{\Pr(y_i = 1)\} = \beta_0 + \beta_x x_i + \beta_z z_i,$$

where the true covariate x_i is replaced by the centred mean of the two (suitably transformed) SBP-measurements. We estimated this naive regression with INLA assigning independent normal prior distributions with zero mean and precision (inverse variance) 0.01 to both the intercept β_0 and the fixed effects β_x and β_z. The slope β_x is attenuated in this naive regression ($\hat{\beta}_x = 1.66$; 95% credible interval [0.70, 2.63]) compared with the estimate that is obtained with error modelling ($\hat{\beta}_x = 1.89$; 95% credible interval [0.79, 3.01]).

2.3. Seedling growth across different light conditions

The impact of shading (dark, middle and light) and defoliation (0%, 25%, 50% and 75% reduction of leaf surface) on plant seedling growth in the Malaysian rainforest has been investigated in a planned experiment that was described in Paine *et al.* (2012). The number of new leaves per plant after a 4-months growth phase was counted and used as the response variable for plant growth. Here, we analysed 60 seedlings from the species *Shorea fallax*, from which 20 plants were grown each under dark, middle and light shading conditions. There were five shade houses for each of the three shading conditions, and each shade house contained four seedlings. Each seedling in a shade house was exposed to a different degree of defoliation treatment: see Fig. 1. In experimental studies in ecology, it is common practice that the value for the target light intensity $\mathbf{w}$ (given in per cent and transformed to the log-scale) is assigned to all replicates within a treatment class (i.e. dark, middle and light). However, owing to external conditions the actual observed light availability $\mathbf{x}$ might considerably vary from the target value within replicates. Therefore, the target light intensity takes only three different values (one for dark, middle and light), whereas the actual light availability would be more variable and would take 15 different values (one for each shade house). This error structure is fundamentally different from that of the preceding examples in Sections 2.1 and 2.2, where mismeasured covariates were more variable than their error-free counterparts.

The selected regression model is Poisson with (log-) target light intensity as a proxy for the actual observed light availability, and additional unstructured random effects to account for potential overdispersion. In contrast with the preceding examples, where the inclusion of $\mathbf{w}$ instead of $\mathbf{x}$ in the regression attenuates the parameter estimates, theory for log-linear models

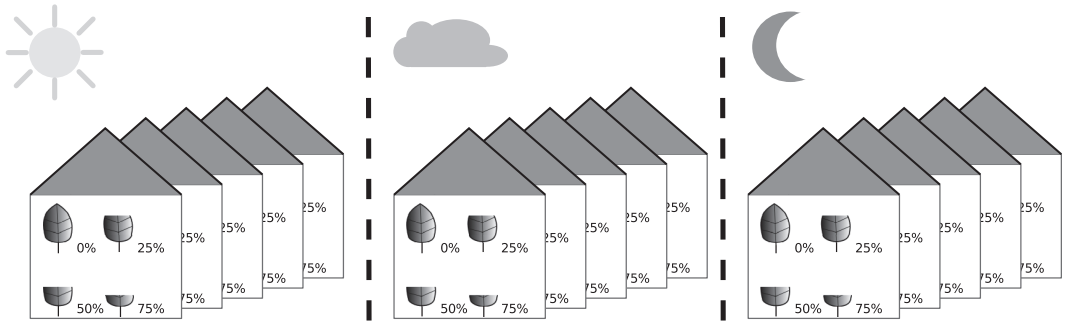

Fig. 1. Illustration of the shade house experiment: there were five shade houses per light condition and each shade house contained four seedlings; the seedlings in a shade house were each exposed to a different defoliation treatment, 0% indicating that the leaves had not been cut, 25% that a quarter of each leaf had been cut, etc.

with such a Berkson error structure (see Section 3.3) suggests that there is no bias in the regression coefficients (Carroll, 1989). However, it is not clear whether this result extends to models with random effects.

3. Measurement error models in regression

3.1. The generalized linear model

Assume that we have n observations in a generalized linear model. The data are given as $(\mathbf{y}, \mathbf{z}, \mathbf{x})$, with $\mathbf{y} = (y_1, \ldots, y_n)^{\mathrm{T}}$ denoting the response, $\mathbf{z} = (\mathbf{z}_1, \ldots, \mathbf{z}_p)$ a covariate matrix of dimension $n \times p$ for p error-free covariates and $\mathbf{x} = (x_1, \ldots, x_n)^{\mathrm{T}}$ a single error prone covariate whose true values are unobservable. The generalization to multiple independent error prone covariates is straightforward. Suppose that $\mathbf{y}$ is of exponential family form with mean $\mu_i = E(y_i|x_i)$, linked to the linear predictor η_i via

$$\begin{aligned} \mu_i &= h(\eta_i), \\ \eta_i &= \beta_0 + \beta_x x_i + \mathbf{z}_{[i,]}\boldsymbol{\beta}_z. \end{aligned} \tag{1}$$

Here, $h(\cdot)$ is a known monotonic inverse link (or response) function, β_0 denotes the intercept, β_x the fixed effect for the error prone covariate $\mathbf{x}$ and $\mathbf{z}_{[i,]}$ is $1 \times p$ with a corresponding vector $\boldsymbol{\beta}_z$ of fixed effects. This generalized linear model is extended to a GLMM by adding normally distributed random effects on the linear predictor scale (1).

Let $\mathbf{w} = (w_1, \ldots, w_n)^{\mathrm{T}}$ denote the observed version of the true, but unobserved, covariate $\mathbf{x}$. We distinguish two different ME processes: the classical error model and the Berkson error model (Berkson, 1950). The graphical structure of these models is very similar (Fig. 2), but their effects are fundamentally different.

3.2. Classical measurement error model

In the classical error model it is assumed that the covariate $\mathbf{x}$ can be observed only via a proxy $\mathbf{w}$, such that, in vector notation,

$$\mathbf{w} = \mathbf{x} + \mathbf{u},$$

with $\mathbf{u} = (u_1, \ldots, u_n)^{\mathrm{T}}$. Throughout the paper the components of the error vector $\mathbf{u}$ are assumed to be independent and normally distributed with mean 0 and variance τ_u^{-1}, i.e. $\mathrm{cov}(u_i, u_j) = 0$ for $i \neq j$. Note that in what follows we parameterize the normal distribution with mean and

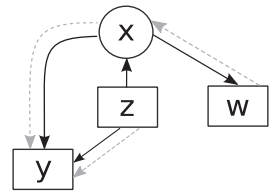

Fig. 2. Graphical structure of the error models (note the change in direction of the arrow between **x** and **w**): □, observed variables; ○, unknown variables; →, classical error model; - - ≫, Berkson error model

precision (or precision matrix in the multivariate context), rather than using the variance (or covariance matrix). We assume that the error term **u** is independent of the true covariate **x**, but also independent of any other covariates **z** and the response **y**. This implies a non-differential ME model, meaning that **y** and **w** are conditionally independent given **z** and **x**. In most applications this assumption is plausible as it implies that, given the true covariate **x** and covariates **z**, no additional information about the response variable **y** is gained through **w** (Carroll *et al.*, 2006).

More generally, the error structure can be heteroscedastic with $\mathbf{w}_j \sim \mathcal{N}(\mathbf{x}, \tau_u \mathbf{D})$, where the entries in the diagonal matrix **D** represent known weights $d_i > 0$. Heteroscedasticity is required when x_i can be measured with varying accuracy for different i. In fact, both the homoscedastic and the heteroscedastic cases are relevant in practice (see, for example, Subar *et al.* (2001) or the example in Section 5.1 presented here). Ideally, repeated measurements w_{ij}, $j = 1, \ldots, J_i$, of the true value x_i are available, so

$$w_{ij}|x_i \sim \mathcal{N}(x_i, \tau_u). \tag{2}$$

The repeated measurements w_{ij} are usually assumed to be conditionally independent and may be unbalanced. For notational convenience we assume in what follows that no repeated measurements are available, i.e. $J_1 = \ldots = J_n = 1$.

Estimates of β_x are usually attenuated in the classical ME setting if **w** is taken as a proxy for **x**. Consider for instance the least squares estimates in a simple linear regression with homoscedastic ME. Fitting the *naive* model $\mathbf{y} = \beta_0^* \mathbf{1} + \beta_x^* \mathbf{w} + \boldsymbol{\varepsilon}^*$ instead of the *true* model $\mathbf{y} = \beta_0 \mathbf{1} + \beta_x \mathbf{x} + \boldsymbol{\varepsilon}$ will result in $|\hat{\beta}_x^*| < |\hat{\beta}_x|$, if the ME variance $1/\tau_u$ is larger than 0. Another important effect is the significant increase in the variability around the regression line.

3.3. Berkson measurement error model

Berkson-type error can be observed in experimental settings where the value of a covariate may correspond to, for example, a predefined fixed dose, temperature or time interval, but the true values **x** may deviate from these planned values **w** because of imprecision in the realization. The second setting where Berkson-type error occurs is in epidemiological or biological studies, where, for example, averages of exposures in areas are assigned to individuals living or working close by. Examples are the application of fixed doses of herbicides in bioassay experiments (Rudemo *et al.*, 1989) or the radiation epidemiology study that was described in Kerber *et al.* (1993) and Simon *et al.* (1995). Such circumstances lead to the Berkson error model (Berkson, 1950)

$$\mathbf{x} = \mathbf{w} + \mathbf{u},$$

where **u** and **w** are independent, and

$$\mathbf{x}|\mathbf{w} \sim \mathcal{N}(\mathbf{w}, \tau_u \mathbf{D}), \tag{3}$$

with **D** denoting a diagonal matrix as in Section 3.2. As for classical ME, Berkson error is assumed to be non-differential. The effect of Berkson error is fundamentally different from that of classical error. In the linear regression model there is no attenuation effect, although the residual precision suffers from the same qualitative bias as in the classical ME model. Both the effects of classical and Berkson error in linear regression are illustrated in on-line supplementary Fig. 1. Issues become more involved for generalized linear (mixed) models. For instance, parameter estimates for logistic regression are only approximately consistent in the Berkson case (Burr, 1988; Bateson and Wright, 2010), which makes error modellings essential.

The difference between classical and Berkson error is reflected in the relationships between the error variances. Denote with τ_x^{-1} and τ_w^{-1} the variances of **x** and **w** respectively. Owing to the independence assumption of **x** and **u** in the classical and between **w** and **u** in the Berkson error case, the variances in the classical and Berkson ME models can be written respectively as

$$\tau_w^{-1} = \tau_x^{-1} + \tau_u^{-1},$$
$$\tau_x^{-1} = \tau_w^{-1} + \tau_u^{-1}.$$

Thus, the surrogate **w** is more variable than the true covariate **x** in the classical model, whereas the opposite is true in the Berkson case.

4. Analysis of measurement error models by using the integrated nested Laplace approximation approach

A Bayesian analysis of ME models dates back at least to Lindley and El-Sayyad (1968) and has been further developed and popularized by Clayton (1992). This approach is based on a three-level hierarchical model.

(a) The first level represents the *observation model* $\mathbf{y}|\mathbf{v}, \boldsymbol{\theta}_1$ defining distributional assumptions about the response variable **y** in dependence on some unknown (latent) parameters **v** and certain hyperparameters $\boldsymbol{\theta}_1$, e.g. variance or correlation parameters. Depending on the error model, the surrogate covariate **w** may also be interpreted as part of the response (see Sections 4.1 and 4.3) and the observation model must then be augmented accordingly.
(b) The second level describes the *latent model* or unobserved process $\mathbf{v}|\boldsymbol{\theta}_2$ depending on hyperparameters $\boldsymbol{\theta}_2$. Depending on the model, **v** is composed of different parameters; see details outlined in Sections 4.1–4.3, where various ME models are discussed.
(c) In the third level, *hyperpriors* are defined for the hyperparameters $\boldsymbol{\theta} = (\boldsymbol{\theta}_1^{\mathrm{T}}, \boldsymbol{\theta}_2^{\mathrm{T}})^{\mathrm{T}}$.

The posterior distribution of the unknowns **v** and $\boldsymbol{\theta}$ is then given by

$$p(\mathbf{v}, \boldsymbol{\theta}|\mathbf{y}) \propto p(\mathbf{y}|\mathbf{v}, \boldsymbol{\theta})\, p(\mathbf{v}|\boldsymbol{\theta})\, p(\boldsymbol{\theta}). \tag{4}$$

Of primary interest are often the posterior marginal distributions for components v_i of **v**, as well as posterior marginals of the hyperparameters θ_j. The former can be derived from expression (4) via

$$p(v_i|\mathbf{y}) = \int_{\boldsymbol{\theta}} \int_{\mathbf{v}_{-i}} p(\mathbf{v}, \boldsymbol{\theta}|\mathbf{y})\, \mathrm{d}\mathbf{v}_{-i} \mathrm{d}\boldsymbol{\theta}, \tag{5}$$

with $\mathbf{v}_{-i}$ denoting the latent field without the ith component. The computation of massively high integrals is, however, very difficult. Except for cases where everything can be computed analytically, exact inference is challenging. Hence, sampling-based approaches have been the standard tool (Gelfand and Smith, 1990).

Rue *et al*. (2009) proposed the INLA, an efficient computing methodology based on sufficiently accurate numerical approximations to perform Bayesian inference in a subclass of hierarchical models, namely latent Gaussian models. In this class the second level, the latent model, is assumed to be Gaussian.

INLA uses the fact that equation (5) can also be written as

$$p(v_i|\mathbf{y}) = \int_{\boldsymbol{\theta}} p(v_i|\boldsymbol{\theta}, \mathbf{y})\, p(\boldsymbol{\theta}|\mathbf{y})\, \mathrm{d}\boldsymbol{\theta},$$

and it approximates this term by a finite sum

$$\tilde{p}(v_i|\mathbf{y}) = \sum_k \tilde{p}(v_i|\boldsymbol{\theta}_k, \mathbf{y})\, \tilde{p}(\boldsymbol{\theta}_k|\mathbf{y}) \Delta_k.$$

Here, $\tilde{p}(v_i|\boldsymbol{\theta}, \mathbf{y})$ and $\tilde{p}(\boldsymbol{\theta}|\mathbf{y})$ denote approximations of $p(v_i|\boldsymbol{\theta}, \mathbf{y})$ and $p(\boldsymbol{\theta}|\mathbf{y})$ respectively. For $p(\boldsymbol{\theta}|\mathbf{y})$ a Laplace approximation is used, whereas for $p(v_i|\boldsymbol{\theta}, \mathbf{y})$ three different strategies are available; see Rue *et al*. (2009). The default is a simplified Laplace approximation. Finally, the sum is computed over suitable support points θ_k with appropriate area weights Δ_k. The value of Δ_k depends on the selection strategy of θ_k. If the points θ_k are laid out on a regular grid, say, all points are equally weighted. The default integration scheme in INLA is the central composite design strategy. Instead of laying out a dense grid of support points (grid strategy), only a limited number of well-chosen points are used, which are laid out in a q-dimensional space. Here, q denotes the number of hyperparameters, i.e. the dimension of $\boldsymbol{\theta}$. Using this strategy, a star-based design is used where centre points are augmented with a group of axial or star points, which allow the estimation of the curvature of $p(\boldsymbol{\theta}|\mathbf{y})$. For a graphical illustration see Fig. 1 in Martins *et al*. (2013). For more details, we refer to Rue *et al*. (2009), section 6.5. Posterior marginals for $p(\theta_j|\mathbf{y})$ can be obtained similarly from $\tilde{p}(\boldsymbol{\theta}|\mathbf{y})$.

The R package `r-inla` can be downloaded from `www.r-inla.org`. Models are specified in a modular way, and thus different types of regression and error models can be freely combined. Various types of random effects, such as independent or conditional auto-regressive models to account for spatial structure, can be incorporated. As discussed in Rue *et al*. (2009) and illustrated in a variety of applications, the approximation error of INLA is small compared with the Monte Carlo error and is negligible in practice; see for example Paul *et al*. (2010), Schrödle *et al*. (2011) and Riebler *et al*. (2012). In the following subsections we describe how various ME models fit into the hierarchical structure that is required by INLA.

4.1. Classical measurement error—general case

Consider a generalized linear (mixed) model regressing a response variable $\mathbf{y}$ on covariates $\mathbf{x}$ and $\mathbf{z}$. The p covariates in $\mathbf{z}$ can be observed directly, whereas instead of $\mathbf{x}$ only a surrogate $\mathbf{w}|\mathbf{x}, \boldsymbol{\theta} \sim \mathcal{N}(\mathbf{x}, \tau_u \mathbf{D})$, following the classical error model (2), is available. The distribution of $\mathbf{x}$, possibly depending on $\mathbf{z}$, is specified in the *exposure model* (Gustafson, 2004). In the most general case considered here, the covariate $\mathbf{x}$ is Gaussian with mean depending on $\mathbf{z}$, i.e.

$$\mathbf{x}|\mathbf{z} \sim \mathcal{N}(\alpha_0 \mathbf{1} + \mathbf{z}\boldsymbol{\alpha}_z, \tau_x \mathbf{I}). \tag{6}$$

Here, α_0 is the intercept, $\boldsymbol{\alpha}_z$ the $p \times 1$ vector of fixed effects and τ_x^{-1} the residual variance in the linear regression of $\mathbf{x}$ on $\mathbf{z}$. If $\mathbf{x}$ depends only on certain components of $\mathbf{z}$, the corresponding entries in $\boldsymbol{\alpha}_z$ are set to 0. The extreme case $\boldsymbol{\alpha}_z = \mathbf{0}$, where $\mathbf{x}$ is independent of $\mathbf{z}$, is discussed separately in Section 4.2.

The assumption that the distribution of the unobserved $\mathbf{x}$ given the observed covariates $\mathbf{z}$ follows a normal distribution is a prerequisite to apply INLA and often is justified. Because of recent extensions of the INLA approach (see Martins and Rue (2012)), $\mathbf{x}|\mathbf{z}$ could even follow a non-Gaussian distribution, so the normal assumption might be relaxed in the future. We can now formulate the latent Gaussian hierarchical model for classical ME.

(a) The observation model encompasses two components, namely the regression model and the error model:

$$E(\mathbf{y}|\mathbf{x}) = h(\beta_0\mathbf{1} + \beta_x\mathbf{x} + \mathbf{z}\boldsymbol{\beta}_z), \tag{7}$$

$$\mathbf{w} = \mathbf{x} + \mathbf{u}, \qquad \mathbf{u} \sim \mathcal{N}(0, \tau_u\mathbf{D}). \tag{8}$$

$\mathbf{w}$ is now part of the observation model, which is thus $\mathbf{y}, \mathbf{w}|\mathbf{v}, \boldsymbol{\theta}_1$ instead of $\mathbf{y}|\mathbf{v}, \boldsymbol{\theta}_1$.

(b) The latent part contains the exposure model for $\mathbf{x}$

$$\mathbf{x} = \alpha_0\mathbf{1} + \mathbf{z}\boldsymbol{\alpha}_z + \boldsymbol{\varepsilon}_x, \qquad \boldsymbol{\varepsilon}_x \sim \mathcal{N}(0, \tau_x\mathbf{I}), \tag{9}$$

as well as the specification of independent Gaussian priors for the regression coefficients. Thus the latent field is

$$\mathbf{v} = (\mathbf{x}^{\mathrm{T}}, \beta_0, \boldsymbol{\beta}_z^{\mathrm{T}}, \alpha_0, \boldsymbol{\alpha}_z^{\mathrm{T}})^{\mathrm{T}}.$$

We use independent normal prior distributions with zero mean and small precision for β_0 and the components of $\boldsymbol{\beta}_z$. Further, we try to elicit the mean and precision of α_0 and $\boldsymbol{\alpha}_z$ by incorporating prior or expert knowledge about the distribution of $\mathbf{x}|\mathbf{z}$. Note also that the exposure model (6) can be easily extended to include structured or unstructured random effects.

(c) The third level describes the prior distributions for all hyperparameters

$$\boldsymbol{\theta} = (\beta_x, \tau_u, \tau_x, \boldsymbol{\theta}_1^{\mathrm{T}})^{\mathrm{T}},$$

with $\boldsymbol{\theta}_1$ representing (possible) hyperparameters of the likelihood. Noteworthy, the regression coefficient β_x is also considered as an unknown hyperparameter, and not as part of the latent field (see the end of this section for an explanation). In our applications we assume a normal distribution with mean 0 and low precision for β_x. For τ_x and τ_u we assume gamma distributions where the corresponding shape and scale parameters are chosen on the basis of expert knowledge, but other prior distributions for τ_x and τ_u can be used in INLA; see Roos and Held (2011) for an example.

To fit this model in INLA, the exposure model (9) is reformulated as

$$\mathbf{0} = -\mathbf{x} + \alpha_0\mathbf{1} + \mathbf{z}\boldsymbol{\alpha}_z + \boldsymbol{\varepsilon}_x, \qquad \boldsymbol{\varepsilon}_x \sim \mathcal{N}(0, \tau_x\mathbf{I}), \tag{10}$$

so that it can also be interpreted as part of the observation model with pseudo-observations $\mathbf{0}$. Equations (7), (8) and (10) then encode for the regression, error and exposure models respectively. To analyse these three models jointly in INLA, the response variable $\mathbf{y}$ is augmented with the observed values $\mathbf{w}$ of the ME model (8) and the pseudo $\mathbf{0}$s from equation (10).

The three components of the observation model may follow different likelihood functions, or at least require a different specification of the hyperparameters. This is addressed by specifying the following response matrix in `r-inla`, which contains one separate column per equation, namely

$$\begin{pmatrix} y_1 & \text{NA} & \text{NA} \\ \vdots & \vdots & \vdots \\ y_n & \text{NA} & \text{NA} \\ \text{NA} & 0 & \text{NA} \\ \vdots & \vdots & \vdots \\ \text{NA} & 0 & \text{NA} \\ \text{NA} & \text{NA} & w_1 \\ \vdots & \vdots & \vdots \\ \text{NA} & \text{NA} & w_n \end{pmatrix}, \tag{11}$$

where 'NA' entries are ignored by INLA. Here, the first column follows the selected exponential family distribution for the response $\mathbf{y}$ with mean (7). The second is assumed to be Gaussian with hyperparameter τ_x; see expression (10). The third component is also Gaussian with hyperparameter τ_u, as specified in equation (8). If repeated measurements are available, the final block of matrix (11) must be extended to include a row for each individual measurement.

Product structures of two unknown parameters, such as $\beta_x\mathbf{x}$ in equation (7), are generally not supported in INLA, because of the inherent requirement that the latent field $\mathbf{v}$ must be Gaussian. Even if β_x and $\mathbf{x}$ are assumed to be normally distributed, their product is not. As a consequence the latent field $\mathbf{v}$ would not be Gaussian if $\beta_x\mathbf{x}$ is included directly as part of $\mathbf{v}$, and the approximations that are used throughout the INLA methodology would not be accurate; see Martins and Rue (2012). To incorporate product structures, one of the factors, say β_x, must be treated as a hyperparameter, i.e. an element of $\boldsymbol{\theta}$, so that conditionally on β_x the latent field is still Gaussian. In the context of ME models as specified here, $\mathbf{x}$ appears in all three components (7), (8) and (10) of the observation model. This allows us to use the so-called copy option that is available in INLA, which was implemented to use the same random field (here $\mathbf{x}$) multiple times, possibly with different scalings (here $\beta_x\mathbf{x}$). The unknown regression coefficient β_x is no longer a component of the latent field $\mathbf{v}$, but a scaling parameter, i.e. hyperparameter, of $\mathbf{x}$.

In practice, an identical copy $\mathbf{x}^*$ of $\beta_x\mathbf{x}$ is created, which is then used in equation (7) as a replacement for $\beta_x\mathbf{x}$. This is achieved by extending the latent vector $\mathbf{x}$ to $\mathbf{x}_c = (\mathbf{x}^{\mathrm{T}}, \mathbf{x}^{*\mathrm{T}})^{\mathrm{T}}$ with $\pi(\mathbf{x}_c|\beta_x) = p(\mathbf{x})\,p(\mathbf{x}^*|\mathbf{x}, \beta_x)$ and

$$p(\mathbf{x}^*|\mathbf{x}, \beta_x) \propto \exp\left\{ -\frac{\tau}{2}(\mathbf{x}^* - \beta_x\mathbf{x})^{\mathrm{T}}(\mathbf{x}^* - \beta_x\mathbf{x}) \right\}.$$

The precision τ, fixed to some large value, controls the similarity between $\mathbf{x}^*$ and $\beta_x\mathbf{x}$ (default value 10^9). The regression coefficient β_x is treated as unknown, and thus as a hyperparameter of the model. This is in contrast with other applications where the respective coefficient is often fixed to 1 (Martins *et al.*, 2013). With the extension of $\mathbf{x}$ to $\mathbf{x}_c$, the latent field $\mathbf{v}$ is necessarily extended. However, since we condition on β_x instead of treating it as a component of the latent field, the latent Gaussian condition is not violated. Posterior marginal distributions for $\mathbf{x}$, a component of $\mathbf{v}$, and β_x, a component of $\boldsymbol{\theta}$, are derived as outlined in Section 4. For exact specification within `r-inla`, see the application in Section 5.2 and the corresponding section in the on-line supplementary material.

4.2. Classical measurement error—independent exposure model

To facilitate the use of simple ME structures in INLA, we provide a specific ME model which does not require the specification of a joint model. The new class covers the case where the

exposure model (6) for $\mathbf{x}$ does not depend on the other (error-free) covariates $\mathbf{z}$, i.e.

$$\mathbf{x} \sim \mathcal{N}(\alpha_0 \mathbf{1}, \tau_x \mathbf{I}).$$

The derivation of the model is sketched in what follows and its use is shown in Section 5.1 and in the corresponding section of the on-line supplementary material.

Without loss of generality, we omit the parameters β_0 and $\boldsymbol{\beta}_z$ in equation (7) and consider the simplified model

$$\begin{aligned} E(\mathbf{y}|\mathbf{x}) &= h(\beta_x \mathbf{x}), \\ \mathbf{x} &= \alpha_0 \mathbf{1} + \boldsymbol{\varepsilon}_x, \qquad \boldsymbol{\varepsilon}_x \sim \mathcal{N}(0, \tau_x \mathbf{I}), \\ \mathbf{w} &= \mathbf{x} + \mathbf{u}, \qquad \mathbf{u} \sim \mathcal{N}(0, \tau_u \mathbf{D}). \end{aligned} \tag{12}$$

Here, α_0 is also considered a hyperparameter (see the explanation below); thus the latent field $\mathbf{v}$ now consists of only $\mathbf{x}$, and $\boldsymbol{\theta} = (\beta_x, \tau_x, \tau_u, \alpha_0, \boldsymbol{\theta}_1^{\mathrm{T}})^{\mathrm{T}}$, where $\boldsymbol{\theta}_1$ may again contain additional hyperparameters of the likelihood. The posterior distribution of $\mathbf{x}$ and $\boldsymbol{\theta}$ is then

$$\begin{aligned} p(\mathbf{x}, \boldsymbol{\theta}|\mathbf{y}, \mathbf{w}) &\propto p(\boldsymbol{\theta})\, p(\mathbf{x}|\boldsymbol{\theta})\, p(\mathbf{w}|\mathbf{x}, \boldsymbol{\theta})\, p(\mathbf{y}|\mathbf{x}, \boldsymbol{\theta}) \\ &\propto p(\boldsymbol{\theta})\, p(\mathbf{x}|\mathbf{w}, \boldsymbol{\theta})\, p(\mathbf{w}|\boldsymbol{\theta})\, p(\mathbf{y}|\mathbf{x}, \boldsymbol{\theta}), \end{aligned}$$

using that $p(\mathbf{x}|\boldsymbol{\theta})\, p(\mathbf{w}|\mathbf{x}, \boldsymbol{\theta}) = p(\mathbf{x}|\mathbf{w}, \boldsymbol{\theta})\, p(\mathbf{w}|\boldsymbol{\theta})$. Now

$$\mathbf{w}|\boldsymbol{\theta} \sim \mathcal{N}[\alpha_0 \mathbf{1}, \{(\tau_u \mathbf{D})^{-1} + (\tau_x \mathbf{I})^{-1}\}^{-1}]$$

and

$$\begin{aligned} p(\mathbf{x}|\mathbf{w}, \boldsymbol{\theta}) &\propto p(\mathbf{x}|\boldsymbol{\theta}) p(\mathbf{w}|\mathbf{x}, \boldsymbol{\theta}) \\ &\propto \exp\left\{ -\frac{\tau_x}{2}(\mathbf{x} - \alpha_0 \mathbf{1})^{\mathrm{T}}(\mathbf{x} - \alpha_0 \mathbf{1}) - \frac{\tau_u}{2}(\mathbf{x} - \mathbf{w})^{\mathrm{T}} \mathbf{D} (\mathbf{x} - \mathbf{w}) \right\}. \end{aligned}$$

Combining these quadratic forms gives

$$\mathbf{x}|\mathbf{w}, \boldsymbol{\theta} \sim \mathcal{N}\{(\tau_x \alpha_0 \mathbf{1} + \tau_u \mathbf{D}\mathbf{w})(\tau_x \mathbf{I} + \tau_u \mathbf{D})^{-1}, \tau_x \mathbf{I} + \tau_u \mathbf{D}\},$$

so the posterior distribution $p(\mathbf{x}, \boldsymbol{\theta}|\mathbf{y}, \mathbf{w})$ can be evaluated explicitly. An alternative formulation is obtained by considering $\boldsymbol{\nu} = \beta_x \mathbf{x}$ instead of $\mathbf{x}$, where

$$\boldsymbol{\nu}|\mathbf{w}, \boldsymbol{\theta} \sim \mathcal{N}\left\{ \beta_x (\tau_x \alpha_0 \mathbf{1} + \tau_u \mathbf{D}\mathbf{w})(\tau_x \mathbf{I} + \tau_u \mathbf{D})^{-1}, \frac{\tau_x \mathbf{I} + \tau_u \mathbf{D}}{\beta_x^2} \right\}. \tag{13}$$

This model is termed '`mec`' within `r-inla` and has four hyperparameters: β_x, τ_x, τ_u and α_0. Note that α_0, which was originally part of the latent field in Section 4.1, now directly enters distribution (13) and is now a hyperparameter of the `mec` model. This formulation leads to a considerable simplification of the `r-inla` call; see the on-line supplementary material for code examples.

4.3. Berkson measurement error

We again consider a generalized linear (mixed) model (1) but replace the classical error (2) by a Berkson error (3):

$$\mathbf{x}|\mathbf{w}, \boldsymbol{\theta} \sim \mathcal{N}(\mathbf{w}, \tau_u \mathbf{D}).$$

Since $\mathbf{x}$ is defined conditionally on the observations $\mathbf{w}$, the exposure model (6) is obsolete. The latent Gaussian hierarchical model for Berkson ME is thus given by

(a) the observation model, which contains only the regression model

$$E(\mathbf{y}|\mathbf{x}) = h(\beta_0\mathbf{1} + \beta_x\mathbf{x} + \mathbf{z}\boldsymbol{\beta}_z),$$

(b) the latent field, given by $\mathbf{v} = (\mathbf{x}^{\mathrm{T}}, \beta_0, \boldsymbol{\beta}_z^{\mathrm{T}})^{\mathrm{T}}$ with Gaussian priors and the error model

$$\mathbf{x} = \mathbf{w} + \mathbf{u}, \qquad \mathbf{u} \sim \mathcal{N}(0, \tau_u\mathbf{D}),$$

(c) the hyperparameters $\boldsymbol{\theta} = (\beta_x, \tau_u, \boldsymbol{\theta}_1^{\mathrm{T}})^{\mathrm{T}}$, where the error precision τ_u is given a suitable gamma and the coefficient β_x a Gaussian prior, and $\boldsymbol{\theta}_1$ representing (possible) hyperparameters of the likelihood.

Analogously to Section 4.2, where $\mathbf{x}$ did not depend on the other covariates $\mathbf{z}$, we can define a latent Gaussian model for the Berkson ME model. Indeed, the same simplifications as in model (12) lead to the hierarchical model

$$\begin{aligned} E(\mathbf{y}|\mathbf{x}) &= h(\beta_x\mathbf{x}), \\ \mathbf{x} = \mathbf{w} + \mathbf{u}, &\qquad \mathbf{u} \sim \mathcal{N}(0, \tau_u\mathbf{D}). \end{aligned}$$

Again, $\mathbf{v} = \mathbf{x}$ is the latent field, and the hyperparameters are $\boldsymbol{\theta} = (\beta_x, \tau_u, \boldsymbol{\theta}_1^{\mathrm{T}})^{\mathrm{T}}$. Importantly, the latent model $\mathbf{x}|\mathbf{w}, \boldsymbol{\theta}$ does now exactly correspond to the error model. It is thus straightforward to calculate the posterior distribution

$$p(\mathbf{x}, \boldsymbol{\theta}|\mathbf{y}, \mathbf{w}) \propto p(\boldsymbol{\theta})\, p(\mathbf{x}|\mathbf{w}, \boldsymbol{\theta})\, p(\mathbf{y}|\mathbf{x}, \boldsymbol{\theta}).$$

The reparameterization $\boldsymbol{\nu} = \beta_x\mathbf{x}$ is again useful and leads to

$$\boldsymbol{\nu}|\mathbf{w}, \boldsymbol{\theta} \sim \mathcal{N}\left(\beta_x\mathbf{w}, \frac{\tau_u}{\beta_x^2}\mathbf{D}\right).$$

This model is termed '`meb`' within the R package `r-inla` and has as hyperparameters β_x and τ_u.

As in Section 4.1, a joint model formulation can also be used for Berkson ME models. However, here it does not add to the generality of the model specification as no exposure model is involved. Thus, we recommend the use of the `meb` model and just illustrate the analogous formulation for completeness. After rewriting the error model similarly to equation (10), the respective joint model contains only the two components

$$\begin{aligned} E(\mathbf{y}|\mathbf{x}) &= h(\beta_0\mathbf{1} + \beta_x\mathbf{x} + \mathbf{z}\boldsymbol{\beta}_z), \\ -\mathbf{w} = -\mathbf{x} + \mathbf{u}, &\qquad \mathbf{u} \sim \mathcal{N}(0, \tau_u\mathbf{D}), \end{aligned}$$

and the response matrix simplifies to

$$\begin{pmatrix} y_1 & \mathrm{NA} \\ \vdots & \vdots \\ y_n & \mathrm{NA} \\ \mathrm{NA} & -w_1 \\ \vdots & \vdots \\ \mathrm{NA} & -w_n \end{pmatrix}.$$

As in Section 4.1, the copy feature is employed to fit this model in INLA.

5. Applications

In what follows we demonstrate how to define the different ME applications that were introduced in Section 2. The results that are presented are based on the `r-inla` version updated on December 4th, 2013, and the `r-inla` code for all examples is given in the on-line supplementary material. A comparison of the results that were obtained by INLA with those obtained by an independent MCMC implementation is provided for each application to underline the accuracy of INLA. We centre all continuous covariates at zero in the following analyses, as otherwise the efficiency of the MCMC method might be reduced (Gelfand *et al.*, 1995, 1996), and additional adjustments of the default parameters in the numerical optimization routine of INLA might be needed.

5.1. Inbreeding in Swiss ibex populations

The ibex data that were introduced in Section 2.1 were analysed by using a linear model with classical heteroscedastic error variances. The observation model is

$$\mathbf{y}|\mathbf{x} \sim \mathcal{N}(\beta_0\mathbf{1} + \beta_x\mathbf{x} + \mathbf{z}\boldsymbol{\beta}_z, \tau_\varepsilon\mathbf{I})$$

with $\mathbf{y}$ denoting the intrinsic growth rates, $\mathbf{x}$ the inbreeding coefficients of the populations and $\mathbf{z}$ the matrix of additional covariates, as listed in Section 2.1. The level of inbreeding x_i in population $i = 1, \ldots, 26$ was estimated as w_i from a Bayesian analysis, which, as a by-product, also provided an estimated population-specific error precision $\hat{\tau}_{u_i}$. Since larger values of $\mathbf{w}$ have more uncertainty, i.e. smaller precision, as shown in Fig. 3, it is natural to formulate a heteroscedastic classical error model

$$\mathbf{w}|\mathbf{x} \sim \mathcal{N}(\mathbf{x}, \tau_u\mathbf{D})$$

with entries $\hat{\tau}_{u_i}$ in the diagonal matrix $\mathbf{D}$. Since $\mathbf{x}$ is assumed to be uncorrelated with the covariates $\mathbf{z}$, the exposure model (6) reduces to

$$\mathbf{x} \sim \mathcal{N}(\alpha_0\mathbf{1}, \tau_x\mathbf{I}).$$

Here, $\alpha_0 = 0$ was fixed because of the preceding covariate centring. Note that the range of $\mathbf{x}$ (before centring) is limited to the interval $[0, 1]$, which may be in conflict with the normality assumption for small precision τ_x. To make sure that this is not a critical point here, we ran

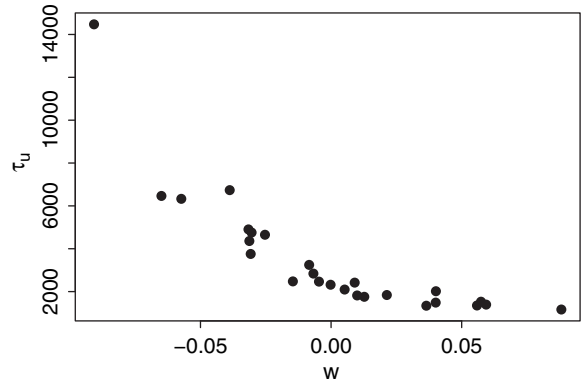

Fig. 3. Uncertainty in the covariate $\mathbf{x}$ in the ibex study, depending on the estimate $\mathbf{w}$: larger values can be estimated with less precision (i.e. with larger variance $1/\tau_u$)

an identical analysis with a normal distribution truncated to $[0, 1]$, using otherwise the same parameters, and obtained virtually identical results. The unknowns in this example are the latent field $\mathbf{v} = (\mathbf{x}^{\mathrm{T}}, \beta_0, \boldsymbol{\beta}_z^{\mathrm{T}})^{\mathrm{T}}$ and the hyperparameters $\boldsymbol{\theta} = (\beta_x, \tau_u, \tau_x, \tau_\varepsilon)^{\mathrm{T}}$. We assigned independent $\mathcal{N}(0, 10^{-4})$ priors to all β-coefficients. The assignment of the prior distributions to the precision parameters is more delicate. We used gamma distributions, where the corresponding shape and rate parameters were chosen on the basis of expert or prior knowledge. In practice, the inbreeding coefficient $\mathbf{x}$ of sexually breeding species is not observed over the whole theoretical range $[0, 1]$. For populations of similar age and size to those in the current study, $\mathbf{x}$-values are expected to lie within $[0, 0.45]$ (Biebach and Keller, 2010). Assuming a uniform distribution within this range, this corresponds to the precision $12/0.45^2 \approx 59$, which we take as a lower limit for τ_x. In the absence of prior knowledge from other studies, we assume that the difference between the largest and the smallest value of $\mathbf{x}$ is at least 0.05, which gives an upper limit of 4800, again assuming a uniform distribution. The parameters of the corresponding gamma distribution with 2.5% quantile at 59 and 97.5% quantile at 4800 are determined by numerical optimization and subsequent rounding, leading to a $G(1.194, 0.000\,85)$ prior for τ_x.

The precision τ_u represents a possible multiplicative bias in the estimates $\hat{\tau}_{u_i}$. Here, we assume that this bias is between 0.5 and 2 with probability 0.95, leading to $G(8.5, 7.5)$. To obtain a lower bound for the prior of τ_ε we assumed a uniform distribution of $\mathbf{y}$ in $[0, 1]$, because populations are usually growing in the absence of density-dependent effects ($\mathbf{y} > 0$) and their growth is bounded by the number of offspring per animal and year (here $\mathbf{y} < 1$). As upper limit we used 100 divided by the sample variance of $\mathbf{y}$, so the coefficient of determination is $R^2 = 0.99$. Using these values as 2.5% and 97.5% quantiles, we obtain a $G(0.903, 0.0014)$ prior distribution for τ_ε.

An MCMC simulation was run for 100 000 iterations with a burn-in of 10 000 iterations and a saving frequency of 10. Hereby, the estimates that were obtained from INLA were chosen as starting values. Convergence was visually checked. An excellent agreement between MCMC samples and the posterior marginals of INLA can be seen in on-line supplementary Fig. 2. The parameter estimates are graphically compared with the naive Bayesian analysis in Fig. 4, including $\mathbf{w}$ instead of $\mathbf{x}$ and using the same priors for the respective parameters. The absolute value of the slope $|\beta_x|$ and the residual precision τ_ε are underestimated in this naive regression, as predicted by the theory. The other parameters are much less affected by the error in $\mathbf{x}$.

5.2. Influence of systolic blood pressure on coronary heart disease

The outcome $y_i \in \{0, 1\}$ in this study is an indicator for coronary heart disease, assumed to be Bernoulli distributed. The observation model is logistic, using an indicator for smoking, $\mathbf{z}$, and the transformed (unobserved) long-term blood pressure $\mathbf{x} = \log(\mathrm{SBP} - 50)$ as binary and continuous covariates respectively. Hence, the linear predictor is

$$\mathrm{logit}\{E(\mathbf{y}|\mathbf{x})\} = \beta_0 \mathbf{1} + \beta_x \mathbf{x} + \beta_z \mathbf{z}.$$

Since SBP has been measured at two different examinations, the magnitude of the ME of these surrogate measures can be quantified. Here, we assume that the repeated measurements $\mathbf{w}_1 = (w_{11}, \ldots, w_{n1})^{\mathrm{T}}$ and $\mathbf{w}_2 = (w_{12}, \ldots, w_{n2})^{\mathrm{T}}$ at examinations 1 and 2 respectively are independent and normally distributed with mean $\mathbf{x}$ and precision τ_u, leading to the classical homoscedastic error model

$$\mathbf{w}_j|\mathbf{x} \sim \mathcal{N}(\mathbf{x}, \tau_u \mathbf{I}), \qquad j = 1, 2.$$

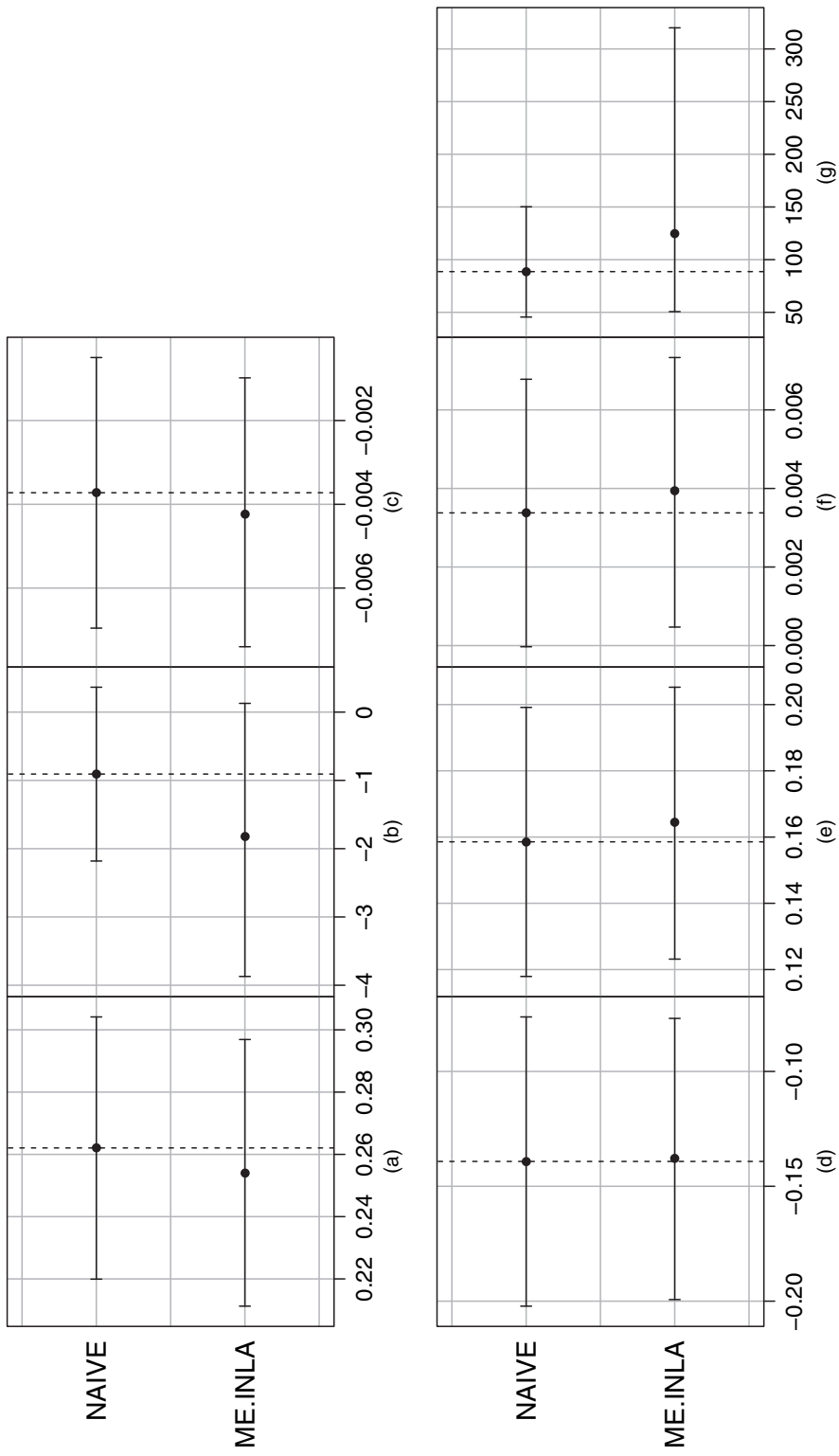

Fig. 4. Posterior means and 95% credible intervals in the ibex data analysis (the broken vertical lines indicate the naive Bayesian estimates without error modelling; only parameters for which a naive estimate is available are included): (a) β_0; (b) β_x; (c) β_{z1}; (d) β_{z2}; (e) β_{z3}; (f) β_{z4}; (g) τ_ε

Finally, the exposure model (6) comes in its most general form

$$\mathbf{x}|\mathbf{z} \sim \mathcal{N}(\alpha_0 \mathbf{1} + \alpha_z \mathbf{z}, \tau_x \mathbf{I}).$$

The latent field in this model is $\mathbf{v} = (\mathbf{x}^{\mathrm{T}}, \beta_0, \beta_z, \alpha_0, \alpha_z)^{\mathrm{T}}$, and the hyperparameters are $\boldsymbol{\theta} = (\beta_x, \tau_u, \tau_x)^{\mathrm{T}}$.

For β_0, β_x and β_z we assigned independent $\mathcal{N}(0, 10^{-2})$ priors. The remaining prior distributions are specified on the basis of subject matter considerations. We assume that 90 mm Hg and 180 mm Hg can be regarded as the respective 2.5% and 97.5% quantile of SBP, and that $\mathrm{SBP} - 50 \sim \mathrm{LogNormal}(\mu, \sigma^2)$. Through optimization we determined $\mu \approx 4.3$ and $\sigma^2 \approx 0.1$, so that the log-normal distribution has the desired quantiles. Consequently, we used $1/\sigma^2$ as expected value for τ_x. Assuming equal mean and variance for τ_x we specified $\tau_x \sim G(10, 1)$, and further $\alpha_0 \sim \mathcal{N}(0, 1)$, whereas $\mu = 0$ is used instead of $\mu = 4.3$ because of the centring of $\mathbf{w}_1$ and $\mathbf{w}_2$. Rothe and Kim (1980) found the ME of SBP to be as much as 20 mm Hg, meaning that our assumed mean SBP of 135 mm Hg varies between 115 and 155. This corresponds to an error factor of 1.15, from which we derive an expected value of approximately 100 for τ_u. Assuming again equal mean and variance of the prior for the precision, we set $\tau_u \sim G(100, 1)$. For α_z we assume a mean of 0, and set $\alpha_z \sim \mathcal{N}(0, 1)$. Note that these prior specifications might deviate from the reference example in Carroll *et al.* (2006), where the exact parameters were not given in the text. Furthermore, Carroll *et al.* (2006) used the quantity $\Delta := \tau_x/\tau_u$ instead of τ_u and gave it a uniform prior in the interval $(0, 0.5)$. Since this is not straightforward with INLA, the model was modified as described.

To obtain posterior marginals with MCMC sampling, regression coefficients of generalized linear models cannot directly be sampled from a standard full conditional distribution. Here, samples for the regression coefficients $\boldsymbol{\beta} = (\beta_0, \beta_x, \beta_z)$ were obtained on the basis of the block sampling algorithm by Gamerman (1997). This approach uses transition densities that combine the weighted least squares method with a prior on $\boldsymbol{\beta}$ (McCullagh and Nelder, 1989; West, 1985). The full conditionals for all unknowns are given in section 3.1 of the on-line supplementary material. The simulation was run for 100 000 iterations with a burn-in of 10 000, and every fifth value was saved. Starting values for $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ were chosen from the INLA output. For τ_u and τ_x, the means of their respective prior distribution were used as initial estimates.

The agreement between the MCMC and INLA output is almost perfect; see on-line supplementary Fig. 3. Fig. 5 shows parameter estimates for β_x and β_z obtained by the naive regression model including $\mathbf{w}_1$ and $\mathbf{w}_2$ instead of $\mathbf{x}$, and four error correction approaches. Carroll *et al.*

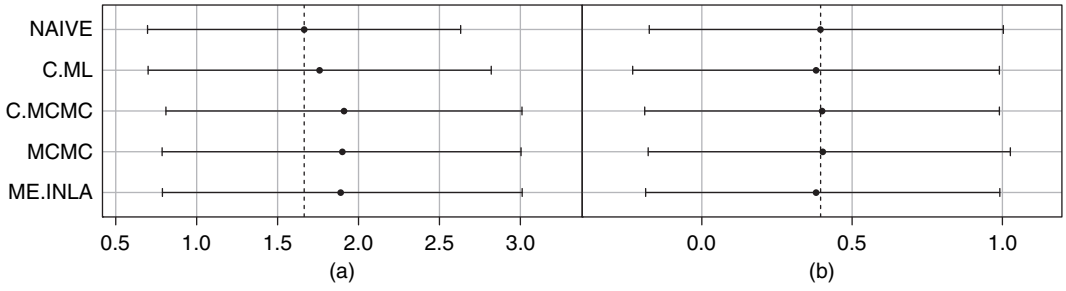

Fig. 5. Posterior means and 95% credible intervals for the Framingham data analysis (for NAIVE, MCMC and the error-corrected analysis ME.INLA, posterior means are used as point estimates; C.MCMC and C.ML denote the Bayesian and the maximum likelihood analyses that were conducted in Carroll *et al.* (2006); the broken vertical lines indicate the naive Bayesian estimates without error modelling): (a) β_x; (b) β_z

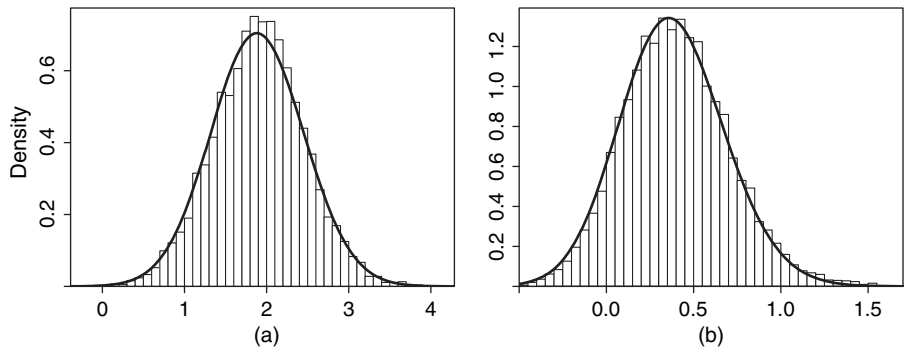

Fig. 6. Histograms of the 10000 posterior means from the bootstrap simulation with INLA, compared with the posterior marginal densities from the INLA analysis using the original data (———): (a) β_x; (b) β_z

(2006) used an ME model fitted via a maximum likelihood method and a Bayesian approach using MCMC sampling, denoted here as methods C.ML and C.MCMC. The fourth and fifth rows show the results that were obtained by our MCMC implementation and INLA. All error-corrected estimates and the credible intervals are similar. Although the coefficient β_z of the error-free measured smoking status seems unbiased, the effect of systolic blood pressure is clearly attenuated in the naive analysis. Adjusting for ME leads to a more pronounced effect, as expected, however, with a larger assigned uncertainty.

In practice, MCMC techniques are commonly applied via easy-to-use software such as BUGS or JAGS. Code length to implement the current model in JAGS is similar to the INLA call. To give a fairer comparison of the methods, we thus also ran JAGS via the R interface `rjags`, similarly to the analysis that was presented in Carroll *et al.* (2006). The simulation of the 110000 iterations including the burn-in finished in 6 min 45 s, whereas INLA required 2.6 s on the same IntelCore i7-2640M 2.80 GHz processor.

It is often of interest to study frequentist properties of Bayesian estimates (Bayarri and Berger, 2004). To do so, we performed a non-parametric bootstrap simulation, iteratively sampling $n = 641$ observations with replacement from the $n = 641$ distinct data records of the original data set. ME model fitting was carried out iteratively with INLA, and posterior means of β_x and β_z were stored after each run. Priors were kept as in the original analysis. We did 10000 bootstrap iterations, which took less than 8 h in total. Histograms of the posterior mean estimates of β_x and β_z are shown in Fig. 6, together with the posterior marginals from the INLA analysis using the original data (the curves). Importantly, the distributions are in very good agreement, not only for the coefficient β_z of the error-free covariate, but also for β_x, the regression coefficient of the mismeasured covariate **x**. This small simulation study thus illustrates excellent frequentist properties of the Bayesian estimates, as well as computational robustness and efficiency of INLA.

5.3. Seedling growth across different light conditions

Let **y** denote the number of new leaves per plant after a 4-months growth phase. The covariate **z** denotes the degree of defoliation and $\mathbf{x} = \log(\%\text{light})$ the (transformed) light intensity, where **w** is the target value. Using **w** instead of **x** in the analysis leads to the homoscedastic Berkson error with

$$\mathbf{x}|\mathbf{w} \sim \mathcal{N}(\mathbf{w}, \tau_u \mathbf{I}).$$

In what follows we centred both covariates **w** and **z**. The nested design of the study, as described in Section 2.3, naturally leads to a Poisson regression model. To account for overdispersion, independent normal random effects $\gamma_{ijk} \sim \mathcal{N}(0, \tau_\gamma)$ were added, extending the generalized linear model to a GLMM:

$$\log\{E(y_{ijk}|x_{ij}, \gamma_{ijk})\} = \beta_0 + \beta_x x_{ij} + \beta_z z_k + \gamma_{ijk},$$

with $i = 1, 2, 3$ denoting the light condition, $j = 1, \ldots, 5$ the shade house per light condition and $k = 1, \ldots, 4$ the degrees of defoliation. The unknowns are $\mathbf{v} = (\mathbf{x}^{\mathrm{T}}, \beta_0, \beta_z)^{\mathrm{T}}$ and $\boldsymbol{\theta} = (\beta_x, \tau_u, \tau_\gamma)^{\mathrm{T}}$.

The β-parameters were assigned independent $\mathcal{N}(0, 10^{-2})$ priors, and the overdispersion precision τ_γ a highly dispersed but proper $G(1, 0.005)$ prior with mean 200. For the error precision τ_u it was assumed that the actual light values **x** do not interfere with the target values **w** from other light levels. The (centred and log-transformed) target light values are 1.22, 0.10 and -1.32 for dark, middle and light conditions; thus the interval between middle and light measurements is 1.42. Interpreting this as one branch of a 95% confidence interval of a Gaussian variable, we obtain $\sigma_u = 1.42/1.96 = 0.72$, yielding a lower bound for τ_u of $1/0.72^2 = 1.93$. For the upper bound a tenth of the variation is assumed, leading to an upper limit of $1/0.072^2 = 193$. The gamma prior distribution with the 2.5% and 97.5% quantiles is $\tau_u \sim G(1.12, 0.0203)$.

The results from the regression in INLA were compared with an MCMC run with 100 000 iterations, a burn-in of 10 000 and a saving frequency of 10. Sampling was based on a reparameterization as proposed by Besag *et al.* (1995), where all except one full conditional distributions are standard and can be Gibbs sampled. The MCMC samples and posterior marginals fit very well; see on-line supplementary Fig. 4. The parameter estimates from the naive analysis including **w** and the error-corrected estimates of INLA are shown in Fig. 7. Our results did not reveal a difference in the regression coefficients after accounting for ME, yet there is a small bias in the precision of the random effect τ_γ. Moreover, the lengths of the credible intervals for β_0, β_x and τ_γ are slightly increased. Note that the same framework as presented here can be used for logistic regression models, where Berkson error is known to cause bias in the parameter estimates (Burr, 1988; Bateson and Wright, 2010).

6. Discussion

ME in covariates may lead to serious biases in parameter estimates and confidence intervals of statistical models. A variety of approaches to model such error have been proposed in recent decades, among which Bayesian methods probably provide the most flexible framework. Bayesian treatments, employing MCMC samplers, have been successfully applied for more than 20 years, but their application has never become part of standard regression analyses.

The aim of this work was to illustrate how the most common ME models (classical and Berkson error) can be included in GLMMs by using the recently proposed INLA framework, which gives fast and accurate approximations instead of doing any sampling. The R code provided should help to make such models accessible to a broader audience. Note that INLA provides a much larger variety of likelihood functions and latent models than we could illustrate here, and the modular structure adds to the flexibility. It is, for instance, straightforward to handle multiple independent mismeasured covariates, to introduce a systematic bias into the error model or to include any structured random term. Gaussian classical and Berkson error naturally fit into the INLA framework of latent Gaussian models, and thus the error prone covariates that were used here are always continuous.

The treatment of more general error models is also possible. For instance, the linear regression error ε might be correlated with the error **u** in a covariate. Another interesting application,

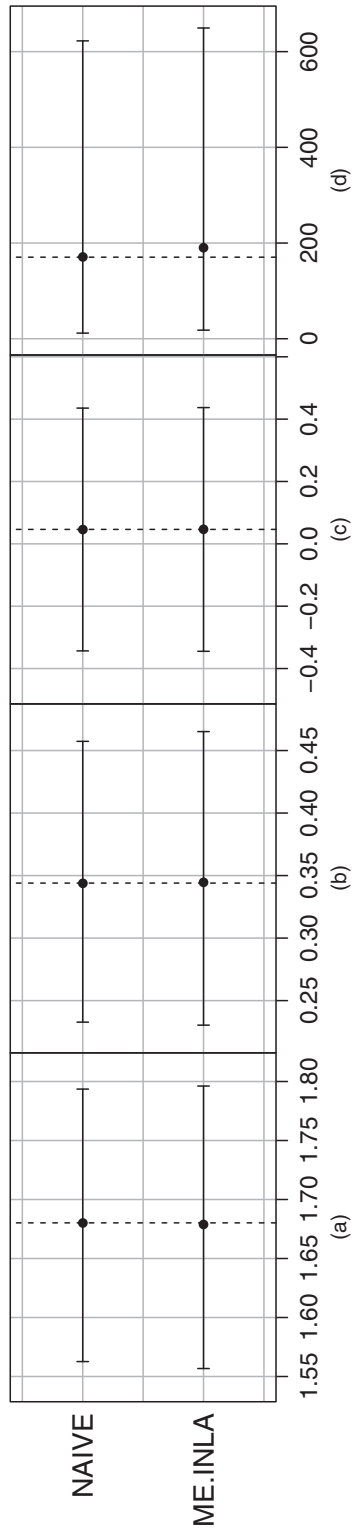

Fig. 7. Posterior means and 95% credible intervals for the seedling growth example (the broken vertical lines indicate the naive Bayesian estimates without error modelling; only parameters where a naive estimate was available are included here): (a) β_0; (b) β_x; (c) β_z; (d) τ_γ

which is relevant for example in ecology, is the use of non-Gaussian error models, e.g. a Poisson or negative binomial model where, instead of the true and positive (but unobserved) continuous covariate $\mathbf{x}$, a discrete proxy $\mathbf{w}$ with mean $\mathbf{x}$ is observed. More general models are also useful, e.g. a log-linear model with mean $E(\mathbf{w}) = a\mathbf{x}^{\beta_x}$ or a logistic model for binomial proxies; see for example Bagchi *et al.* (2014). Furthermore, it might not always be appropriate to assume that the components of $\mathbf{x}$ are independent and identically distributed. Hence, $\mathbf{x}$ could follow a more complex Gaussian Markov random-field structure (Rue and Held, 2005) to account for temporal and/or spatial dependences; see Bernardinelli *et al.* (1997) for such a formulation in an epidemiological context. Both of these extensions can be handled with INLA. Another direct extension concerns additive (possibly heteroscedastic) error in the response, which is also known as equation error. This has been discussed in linear regression (Fuller, 1987; Buonaccorsi, 2010), and Bayesian inference has been proposed (Kelly, 2007; Andreson and Hurn, 2013; deCastro *et al.*, 2013). The treatment in INLA is possible by adding random effects to the regression equation.

One of the biggest challenges with mismeasured variables is the estimation of the error variance, either from repeated measurements, instrumental variables or from previous studies. The advantage of a Bayesian approach, as taken here, is that uncertainty of such estimates can be automatically incorporated in prior distributions, which has been shown to be more beneficial than fixing the error variance at a best guess value (Gustafson, 2005). Even if nothing is known and a flat prior is used, Gustafson (2005) has shown that there can be a surprising amount of indirect learning. However, it is questionable whether ME modelling makes sense in the absence of prior knowledge about the error variance, the more so because the model might not be identifiable.

7. Supplementary material

For brevity, the R code for all the examples that are presented here is described in detail in the on-line supplementary document. Furthermore, this document contains full conditionals and posterior marginals for Section 5.2. On `www.r-inla.org/examples/case-studies/muff-etal-2014` selected data and R code are provided to download.

Acknowledgements

The authors acknowledge constructive comments by three reviewers, the Associate Editor and the Joint Editor, which helped to improve the manuscript. We especially thank Lukas Keller for stimulating the ME project and many interesting discussions. We are grateful to him as well as to Iris Biebach and Claudio Bozzuto, for providing us with the ibex data. We also thank Robert Bagchi, Helmut Küchenhoff and Malgorzata Roos for useful discussions and suggestions for this work.

Stefanie Muff and Andrea Riebler contributed equally to this work.

References

Andreon, S. and Hurn, M. (2013) Measurement errors and scaling relations in Astrophysics: a review. *Statist. Anal. Data Minng*, **6**, 15–33.

Bagchi, R., Gallery, R. E., Gripenberg, S., Gurr, S. J., Narayan, L., Addis, C. E., Freckleton, R. P. and Lewis, O. T. (2014) Pathogens and insect herbivores drive rainforest plant diversity and composition. *Nature*, to be published, doi 10.1038/nature12911.

Bateson, T. F. and Wright, J. M. (2010) Regression calibration for classical exposure measurement error in environmental epidemiology studies using multiple local surrogate exposures. *Am. J. Epidem.*, **172**, 344–352.

Bayarri, M. J. and Berger, J. O. (2004) The interplay of Bayesian and frequentist analysis. *Statist. Sci.*, **19**, 58–80.

Berkson, J. (1950) Are there two regressions? *J. Am. Statist. Ass.*, **45**, 164–180.

Bernardinelli, L., Pascutto, C., Best, N. G. and Gilks, W. R. (1997) Disease mapping with errors in covariates. *Statist. Med.*, **16**, 741–752.

Besag, J., Green, P., Higdon, D. and Mengersen, K. (1995) Bayesian computation and stochastic systems (with discussion). *Statist. Sci.*, **10**, 3–66.
Biebach, I. and Keller, L. (2010) Inbreeding in reintroduced populations: the effects of early reintroduction history and contemporary processes. *Conservn Genet.*, **11**, 527–538.
Bishop, C. M. (2006) *Pattern Recognition and Machine Learning*. New York: Springer.
Bozzuto, C., Biebach, I., Muff, S. and Keller, L. (2014) Inbreeding effects on population dynamics—the case of Swiss Ibex populations. *Technical Report*. Institute of Evolutionary Biology and Environmental Studies, University of Zurich, Zurich.
Buonaccorsi, J. P. (2010) *Measurement Error: Models, Methods, and Applications*. Boca Raton: CRC Press.
Burr, D. (1988) On errors-in-variables in binary regression—Berkson case. *J. Am. Statist. Ass.*, **83**, 739–743.
Carroll, R. (1989) Covariance analysis in generalized linear measurement error models. *Statist. Med.*, **8**, 1075–1093.
Carroll, R., Gallo, P. and Gleser, L. (1985) Comparison of least squares and errors-in-variables regression, with special reference to randomized analysis of covariance. *J. Am. Statist. Ass.*, **80**, 929–932.
Carroll, R. J., Küchenhoff, H., Lombard, F. and Stefanski, L. A. (1996) Asymptotics for the SIMEX estimator in nonlinear measurement error models. *J. Am. Statist. Ass.*, **91**, 242–250.
Carroll, R., Ruppert, D., Stefanski, L. and Crainiceanu, C. (2006) *Measurement Error in Nonlinear Models: a Modern Perspective*, 2nd edn. Boca Raton: Chapman and Hall.
Carroll, R. J., Spiegelman, C. H., Lan, K. K. G., Bailey, K. T. and Abbott, R. D. (1984) On errors-in-variables for binary regression models. *Biometrika*, **71**, 19–25.
Carroll, R. and Stefanski, L. (1990) Approximate quasilikelihood estimation in models with surrogate predictors. *J. Am. Statist. Ass.*, **85**, 652–663.
deCastro, M., Bolfarine, H. and Galea, M. (2013) Bayesian inference in measurement error models for replicate data. *Environmetrics*, **24**, 22–30.
Clayton, D. (1992) Models for the longitudinal analysis of cohort and case-control studies with inaccurately measured exposures. In *Statistical Models for Longitudinal Studies of Health* (eds J. H. Dwyer, M. Feinleib, P. Lippert and H. Hoffmeister), pp. 301–331. Oxford: Oxford University Press.
Cook, J. and Stefanski, L. (1994) Simulation-extrapolation estimation in parametric measurement error models. *J. Am. Statist. Ass.*, **89**, 1314–1328.
Cornfield, J. (1962) Joint dependence of risk of coronary heart disease on serum cholesterol and systolic blood pressure: a discriminant function analysis. *Fed. Proc.*, **21**, 58–61.
Dellaportas, P. and Stephens, D. (1995) Analysis of errors-in-variables regression models. *Biometrics*, **51**, 1085–1095.
Fong, Y., Rue, H. and Wakefield, J. (2010) Bayesian inference for generalized linear mixed models. *Biostatistics*, **11**, 397–412.
Fuller, W. A. (1987) *Measurement Error Models*. New York: Wiley.
Gamerman, D. (1997) Sampling from the posterior distribution in generalized linear mixed models. *Statist. Comput.*, **7**, 57–68.
Gelfand, A. E., Sahu, S. K. and Carlin, B. P. (1995) Efficient parametrisations for normal linear mixed models. *Biometrika*, **82**, 479–488.
Gelfand, A. E., Sahu, S. K. and Carlin, B. P. (1996) Efficient parametrizations for generalized linear mixed models. In *Bayesian Statistics 5* (eds J. M. Bernardo, J. O. Berger, A. P. Dawid and A. F. M. Smith), pp. 165–180. New York: Oxford University Press.
Gelfand, A. E. and Smith, A. F. M. (1990) Sampling-based approaches to calculating marginal densities. *J. Am. Statist. Ass.*, **85**, 398–409.
Gleser, L. (1990) Improvements of the naive approach to estimation in nonlinear errors-in-variables regression models. In *Statistical Analysis of Measurement Error Models and Applications* (eds P. J. Brown and W. A. Fuller). Providence: American Mathematical Society.
Gleser, L., Carroll, R. and Gallo, P. (1987) The limiting distribution of least squares in an errors-in-variables regression model. *Ann. Statist.*, **15**, 220–233.
Gustafson, P. (2004) *Measurement Error and Misclassification in Statistics and Epidemiology: Impacts and Bayesian Adjustments*. Boca Raton: Chapman and Hall–CRC.
Gustafson, P. (2005) On model expansion, model contraction, identifiability and prior information: two illustrative scenarios involving mismeasured variables. *Statist. Sci.*, **20**, 111–140.
Kannel, W., Neaton, J. and Wentworth, D. (1986) Overall and coronary heart-disease mortality-rates in relation to major risk-factors in 325,348 men screened for the MRFIT. *Am. Hrt J.*, **112**, 825–836.
Kelly, B. (2007) Some aspects of measurement error in linear regression of astronomical data. *Astrophys. J.*, **665**, 1489–1506.
Kerber, R. A., Till, J. E., Simon, S. L., Lyon, J. L., Thomas, D. C., Preston-Martin, S., Rallison, M. L., Lloyd, R. D. and Stevens, W. (1993) A cohort study of thyroid disease in relation to fallout from nuclear weapons testing. *J. Am. Med. Ass.*, **270**, 2076–2082.
Lindley, D. V. and El-Sayyad, G. M. (1968) The Bayesian estimation of a linear functional relationship. *J. R. Statist. Soc.* B, **30**, 190–202.

Lunn, D., Spiegelhalter, D., Thomas, A. and Best, N. (2009) The BUGS project: evolution, critique, and future directions. *Statist. Med.*, **28**, 3049–3067.
Lunn, D., Thomas, A., Best, N. and Spiegelhalter, D. (2000) WinBUGS—a Bayesian modelling framework: concepts, structure, and extensibility. *Statist. Comput.*, **10**, 325–337.
MacMahon, S., Peto, R., Cutler, J., Collins, R., Sorlie, P., Neaton, J., Abbott, R., Godwin, J., Dyer, A. and Stamler, J. (1990) Blood pressure, stroke and coronary heart disease: Part 1, prolonged differences in blood pressure: prospective observational studies corrected for the regression dilution bias. *Lancet*, **335**, 765–774.
Martin, A. D., Quinn, K. M. and Park, J. H. (2011) MCMCpack: Markov Chain Monte Carlo in R. *J. Statist. Softwr.*, **42**, 01–22.
Martins, T. G. and Rue, H. (2012) Extending INLA to a class of near-Gaussian latent models. *Preprint arXiv:1210.1434*.
Martins, T., Simpson, D., Lindgren, F. and Rue, H. (2013) Bayesian computing with INLA: new features. *Computnl Statist. Data Anal.*, **67**, 68–83.
McCullagh, P. and Nelder, J. A. (1989) *Generalized Linear Models*, 2nd edn. New York: Chapman and Hall.
Ormerod, J. T. and Wand, M. P. (2010) Explaining variational approximations. *Am. Statistn*, **64**, 140–153.
Ormerod, J. T. and Wand, M. P. (2012) Gaussian variational approximate inference for generalized linear mixed models. *J. Computnl Graph. Statist.*, **21**, 2–17.
Paine, T., Stenflo, M., Philipson, C., Saner, P., Bagchi, R., Ong, R. and Hector, A. (2012) Differential growth responses in seedlings of ten species of Dipterocarpaceae to experimental shading and defoliation. *J. Trop. Ecol.*, **28**, 377–384.
Paul, M., Riebler, A., Bachmann, L. M., Rue, H. and Held, L. (2010) Bayesian bivariate meta-analysis of diagnostic test studies using integrated nested Laplace approximations. *Statist. Med.*, **29**, 1325–1339.
Pearson, K. (1902) On the mathematical theory of errors of judgement. *Philos. Trans. R. Soc. Lond.* A, **198**, 235–299.
Pham, T. H., Ormerod, J. T. and Wand, M. P. (2013) Mean field variational Bayesian inference for nonparametric regression with measurement error. *Computnl Statist. Data Anal.*, **68**, 375–387.
Plummer, M. (2003) JAGS: a program for analysis of Bayesian graphical models using Gibbs sampling. In *Proc. 3rd Int. Wrkshp Distributed Statistical Computing, Vienna* (eds K. Hornik, F. Leisch and A. Zeileis). (Available from `http://www.r-project.org/conferences/DSC-2003/proceedings/`.)
R Core Team (2012) *R: a Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing.
Richardson, S. and Gilks, W. (1993) Conditional independence models for epidemiological studies with covariate measurement error. *Statist. Med.*, **12**, 1703–1722.
Riebler, A., Held, L. and Rue, H. (2012) Estimation and extrapolation of time trends in registry data—borrowing strength from related populations. *Ann. Appl. Statist.*, **6**, 304–333.
Roos, M. and Held, L. (2011) Sensitivity analysis in Bayesian generalized linear mixed models for binary data. *Baysn Anal.*, **6**, 259–278.
Rothe, C. F. and Kim, K. C. (1980) Measuring systolic arterial blood pressure: possible errors from extension tubes or disposable transducer domes. *Crit. Care Med.*, **8**, 683–689.
Rudemo, M., Ruppert, D. and Streibig, J. C. (1989) Random-effect models in nonlinear regression with applications to bioassay. *Biometrics*, **45**, 349–362.
Rue, H. and Held, L. (2005) *Gaussian Markov Random Fields: Theory and Applications*. Boca Raton: Chapman and Hall–CRC.
Rue, H., Martino, S. and Chopin, N. (2009) Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations (with discussion). *J. R. Statist. Soc.* B, **71**, 319–392.
Schrödle, B., Held, L., Riebler, A. and Danuser, J. (2011) Using integrated nested Laplace approximations for the evaluation of veterinary surveillance data from Switzerland: a case-study. *Appl. Statist.*, **60**, 261–279.
Simon, S. L., Till, J. E., Lloyd, R. D., Kerber, R. L., Thomas, D. C., Preston-Martin, S., Lyon, J. L. and Stevens, W. (1995) The Utah leukemia case-control study: dosimetry methodology and results. *Hlth Phys.*, **68**, 460–471.
Stephens, D. and Dellaportas, P. (1992) Bayesian analysis of generalised linear models with covariate measurement error. In *Bayesian Statistics 4* (eds J. M. Bernardo, J. O. Berger, A. P. Dawid and A. F. M. Smith). Oxford: Oxford University Press.
Subar, A. F., Thompson, F. E., Kipnis, V., Midthune, D., Hurwitz, P., McNutt, S., McIntosh, A. and Rosenfeld, S. (2001) Comparative validation of the Block, Willett, and National Cancer Institute Food Frequency Questionnaires: the Eating at America's Table Study. *Am. J. Epidem.*, **154**, 1089–1099.
Wald, A. (1940) The fitting of straight lines if both variables are subject to error. *Ann. Math. Statist.*, **11**, 284–300.
West, M. (1985) Generalized linear models: scale parameters, outlier accommodation and prior distributions. In *Bayesian Statistics 2* (eds J. M. Bernardo, M. H. DeGroot, D. V. Lindley and A. F. M. Smith), pp. 531–558. Amsterdam: North-Holland.

Supporting information

Additional 'supporting information' may be found in the on-line version of this article:

'Supplementary material for "Bayesian analysis of measurement error models using INLA"'.