Published December 4, 2014

Technical note: An R package for fitting generalized linear mixed models in animal breeding[1]

A. I. Vazquez,*[2] D. M. Bates,† G. J. M. Rosa,* D. Gianola,*‡ and K. A. Weigel*

*Department of Dairy Science, †Department of Statistics, and ‡Department of Animal Sciences, University of Wisconsin, Madison 53706

ABSTRACT: Mixed models have been used extensively in quantitative genetics to study continuous and discrete traits. A standard quantitative genetic model proposes that the effects of levels of some random factor (e.g., sire) are correlated accordingly with their relationships. For this reason, routines for mixed models available in standard packages cannot be used for genetic analysis. The *pedigreemm* package of R was developed as an extension of the *lme4* package, and allows mixed models with correlated random effects to be fitted for Gaussian, binary, and count responses. Following the method of Harville and Callanan (1989), a correlation between levels of the grouping factor (e.g., sire) is induced by post-multiplying the incidence matrix of the levels of this random factor by the Cholesky factor of the corresponding (co)variance matrix (e.g., the numerator relationship matrix between sires). Estimation methods available in *pedigreemm* include approximations to maximum likelihood and REML. This note describes the classes of models that can be fitted using *pedigreemm* and presents examples that illustrate its use.

Key words: animal model, correlated random effect, generalized linear mixed model, quantitative genetics, R package, sire model

J. Anim. Sci. 2010. 88:497–504
doi:10.2527/jas.2009-1952

INTRODUCTION

Linear mixed models have been used extensively to estimate genetic parameters and predict breeding values associated with Gaussian traits (Henderson et al., 1959; Henderson, 1963, 1973). A more general class of mixed models is represented by the generalized linear mixed models (**GLMM**), which are appropriate for analysis of data from the exponential family of distributions (Tempelman, 1998). In linear mixed models under Gaussian assumptions, the marginal likelihood has a closed form, and maximum likelihood or REML estimation can be performed conveniently. In nonlinear models, however, the marginal likelihood does not have a closed form and must be approximated using, for example, a Laplacian approximation.

The R environment (R Development Core Team, 2008) is powerful, free software for statistical computing and graphics in which statistical theories can be implemented. R is an open source system written by volunteers and is extended via packages (www.r-project.org). The use of R and contributions to it have been growing in the scientific community over time. The *lme4* package (Bates and Maechler, 2008) fits linear models and GLMM to data. The program handles an arbitrary number of grouping factors, nested or cross-classified, and uses a combination of sparse and dense matrix representations to process large data sets at high speed. The use of *lme4* for genetic analysis has been limited because it does not allow correlations between clusters. If animals are related, the marginal likelihood must allow for covariance between individuals or groups. We developed a package called *pedigreemm* that uses the capabilities of *lme4* while allowing for correlations between levels of random effects, such as those attributable to genetic relationships. This package is available at the Comprehensive R Archive Network (**CRAN**; http://www.r-project.org/), and a developing version is located at http://r-forge.r-project.org/projects/pedigreemm/

In the present note, we discuss the approach used by *pedigreemm* for GLMM to include covariances between levels of random effects, and we show the use of the

[1]Partial financial support of K. Weigel by the National Association of Animal Breeders (Columbia, MO) is gratefully acknowledged. A. I. Vazquez was supported by USDA-ARS research contract 58-1265-3-151. Support by the Wisconsin Agriculture Experiment Station (Madison) and by grant DMS-NSF DMS-044371 to D. Gianola is acknowledged. We thank Sunduz Keles (University of Wisconsin, Madison) for providing access to the powerful server used during some of the development and testing of this software.

[2]Corresponding author: anainesvs@gmail.com

Received March 10, 2009.
Accepted October 2, 2009.

package with 2 examples that use data contained in the package. Last, we illustrate how related results can be extended.

METHOD

Animal Care and Use committee approval was not obtained for this study because milk yield and clinical mastitis data were obtained from existing databases at the USDA-ARS Animal Improvement Programs Laboratory (Beltsville, MD) and Alta Genetics (Balzac, Alberta, Canada), respectively.

The general formulation of the GLMM is as follows:

$$g(\boldsymbol{\mu}_{\mathbf{Y}|\mathbf{U}}) = \mathbf{Zu} + \mathbf{X}\boldsymbol{\beta},$$

where $\mathbf{Y}$ is the random variable representing the response; $g(.)$ is a function that links the response with a model that is linear in the explanatory variables; $\boldsymbol{\mu}_{\mathbf{Y}|\mathbf{U}} = E[\mathbf{Y}|\mathbf{U} = \mathbf{u}]$ is the expectation of the response conditional to the random effects; $\boldsymbol{\beta}$ is a vector of fixed effects; $\mathbf{X}$ is the model matrix relating fixed effects to $g(\boldsymbol{\mu}_{\mathbf{Y}|\mathbf{U}})$; $\mathbf{u}$ is a vector of random effects, which can also accommodate multiple grouping factors, either nested or cross-classified; each grouping factor of random effect is distributed as $\mathbf{u} \sim N\left(\mathbf{0}, \mathbf{I}\sigma_u^2\right)$, where $N(.,.)$ represents the normal distribution with mean and variance indicated in the parentheses; σ_u^2 is a variance component; $\mathbf{I}$ is an identity matrix; and $\mathbf{Z}$ is the model matrix relating $g(\boldsymbol{\mu}_{\mathbf{Y}|\mathbf{U}})$ to $\mathbf{u}$. Some possible families for the conditional distribution of the data, given the random effects in GLMM, are the binomial, Gaussian, gamma, inverse Gaussian, and Poisson distributions, among others, all of them deriving from the exponential family of distributions. Parameter estimates are obtained by minimizing the Laplacian approximation to the deviance function (Bates, 2009). In *pedigreemm*, as in *lmer*, canonical links are the default, but other links are available (e.g., a probit link for the binomial family). Type `help(family)` for a complete list in the R environment.

In a standard Gaussian linear model, the variance-covariance matrix of the marginal distribution of the data is

$$V\left(\mathbf{y}\right) = \mathbf{Z}\left(\mathbf{I}\sigma_u^2\right)\mathbf{Z}' + \mathbf{R},$$

where where $V(\mathbf{y})$ is the variance of the response $\mathbf{y}$; σ_u^2 is the variance of the random factor (typically clustering observation, e.g., sires), whose levels are assumed to be independent and identically distributed; and $\mathbf{R}$ is the covariance matrix of the random residuals, often assumed to follow a normal distribution, independently and identically distributed. Different options for modeling $\mathbf{R}$ are available and are independently and identically distributed (i.e., $\mathbf{R} = \mathbf{I}\sigma_e^2$ is the default). In genetic analysis, the variance-covariance matrix of sire or animal effects, $\mathbf{u}$ (for a sire or animal model, respectively), typically results in an expression for the marginal distribution of the data of

$$V\left(\mathbf{y}\right) = \mathbf{Z}\left(\mathbf{A}\sigma_u^2\right)\mathbf{Z}' + \mathbf{R},$$

where $\mathbf{A}\sigma_u^2$ is the covariance matrix of the multivariate vector of random effects $\mathbf{u}$, and $\mathbf{A}$ is the additive relationship matrix. Animals are genetically related to each other, so their performance is expected to be correlated, unless σ_u^2 is 0.

The methodology described by Harville and Callanan (1989) consists of post-multiplying the model matrix $\mathbf{Z}$ by the Cholesky decomposition of the numerator relationship matrix ($\mathbf{A}$). Note that $\mathbf{A}$ is a positive-definite matrix (unless identical twins or clones are in the pedigree, in which case it would be positive semi-definite). Let $\mathbf{A} = \mathbf{LL}'$, where $\mathbf{L}$ is the Cholesky factor. The matrix $\mathbf{L}$ can be written directly from the pedigree information (Henderson, 1976). Subsequently, let $\mathbf{Z}^* = \mathbf{ZL}$, then

$$\begin{aligned} V\left(\mathbf{Zu}\right) &= \mathbf{ZAZ}'\sigma_u^2 \\ &= \mathbf{ZLL}'\mathbf{Z}'\sigma_u^2 \\ &= \left(\mathbf{Z}^*\right)\left(\mathbf{Z}^*\right)'\sigma_u^2. \end{aligned}$$

Define $\mathbf{u}^* = \mathbf{L}^{-1}\mathbf{u}$, and rewrite $g(\boldsymbol{\mu}_{\mathbf{Y}|\mathbf{U}})$ as

$$\mathbf{ZL}(\mathbf{L}^{-1}\mathbf{u}) + \mathbf{X}\boldsymbol{\beta} = \mathbf{Z}^*\mathbf{u}^* + \mathbf{X}\boldsymbol{\beta}.$$

If one assumes $\mathbf{u} \sim N\left(\mathbf{0}, \mathbf{A}\sigma_u^2\right)$, the distribution of $\mathbf{u}^*$ is $N\left[\mathbf{0}, \mathbf{I}\sigma_u^2\right]$, because

$$\begin{aligned} \mathbf{L}^{-1}\mathrm{Var}\left(\mathbf{u}\right)\left(\mathbf{L}'^{-1}\right) &= \mathbf{L}^{-1}\mathbf{A}\left(\mathbf{L}'^{-1}\right)\sigma_u^2 \\ &= \mathbf{I}\sigma_u^2, \end{aligned}$$

so that the elements of $\mathbf{u}^*$ are mutually independent. The *lmer* procedure can then be applied to $g(\boldsymbol{\mu}_{\mathbf{Y}|\mathbf{U}}) = \mathbf{Z}^*\mathbf{u}^* + \mathbf{X}\boldsymbol{\beta}$ because the random effects are now independent. All levels of $\mathbf{u}$ should have records; for example, all sires present in the data set should have daughters on a sire model application, and all animals present in the data set should have observations on an animal model. The relationship matrix $\mathbf{A}$ can be expressed as $\mathbf{A} = \mathbf{TDT}'$, and $\mathbf{A}^{-1} = (\mathbf{T}^{-1})'\mathbf{D}^{-1}\mathbf{T}^{-1}$, where $\mathbf{T}^{-1}$ is a lower triangular matrix with ones in the diagonal and the only nonzero elements are -0.5 in the columns corresponding to the known parents; and $\mathbf{D}$ is a diagonal matrix produced from the inbreeding coefficients (Mrode, 2005). The inbreeding coefficients are calculated using the algorithm described in the appendix of

the paper by Sargolzaei and Iwaisaki (2005). Let **S** be the model matrix relating correlated random effects with the pedigree. The premultiplier of the model matrix $\mathbf{Z}'$ (the transpose of the model matrix relating random effects to the data) would be $\mathbf{L}' = \mathbf{D}^{1/2}\mathbf{T}'\mathbf{S}'$, such that the reparameterized $\mathbf{Z}'$ (i.e., $\mathbf{Z}^{*\prime}$) would be $\mathbf{Z}^{*\prime} = \mathbf{L}'\mathbf{Z}' = \mathbf{D}^{1/2}\mathbf{T}'\mathbf{S}'\mathbf{Z}'$. Calling $\mathbf{B} = \mathbf{D}^{1/2}\mathbf{T}'$, the Cholesky $\mathbf{L}'$ is obtained by solving for **B** in the system $(\mathbf{L}')^{-1}\mathbf{B} = \mathbf{S}$.

EXAMPLES

Two examples are presented to illustrate the use of *pedigreemm*. Example 1 fits an animal model to milk production records in Holstein cattle. Example 2 models mastitis counts during the first lactation of Holstein cows by using nonlinear sire models. All programs and data sets used are available with the package at the R CRAN archive (http://cran.r-project.org/). It is assumed that the user is familiar with the R language. If not, the user can refer to the manual *An Introduction to R* found online (http://cran.r-project.org/).

Data Description

Data Set 1. Milk production records of 3,397 lactations from first- through fifth-parity Holsteins were available. These records were from 1,359 cows, daughters of 38 sires in 57 herds. Records are in the *milk* data set in the *pedigreemm* package. The data were downloaded from the USDA site (http://www.aipl.arsusda.gov/). All lactation records represent cows with at least 100 d in milk, with an average of 347 d. Milk yield ranged from 4,065 to 19,345 kg estimated for 305 d, averaging 11,636 kg. There were 1,314, 1,006, 640, 334, and 103 records for first-, second-, third-, fourth-, and fifth-lactation animals, respectively. A 5-generation pedigree of the cows with a total of 6,547 animals was used in the analysis (http://www.aipl.arsusda.gov/). The pedigree information is available in the *pedCows* and *pedCowsR* pedigree objects also included in the package; the second one is a lighter pedigree (with 70% of the information on *pedCows*). The milk production data used in the first 2 examples are described below.

Data Set 2. The *pedigreemm* package can be used for discrete data that would be modeled with a GLMM. The number of cases of clinical mastitis (**NCM**) during the first lactation of each of 1,675 cows was used as the response variable. This data set is a subset of data used by Vazquez (2007), and is available in the *mastitis* data set in the *pedigreemm* package. Cows belonged to 41 herds and were daughters of 38 sires. There were 1,491 healthy cows, 134 cows had only 1 case of mastitis, 36 had 2 cases, and 14 had between 4 and 6 cases; overall, mastitis incidence was 0.11. Calving years for these records were from 2000 through 2005. A 3-generation pedigree of the sires was built (http://www.aipl.arsusda.gov/), with a total of 352 animals in the pedigree. The pedigree for the 38 sires is available in the *pedSires* object in the *pedigreemm* package.

Fitting the Models

Below, some information is provided to guide users on how to perform a genetic analysis using the *pedigreemm* package.

Example 1: Linear Animal Model. Standardized milk production (data set 1) was analyzed with the animal model

$$y_{ijk} = \beta_0 + \beta_1 L_i + \beta_2 \log(\mathrm{DIM})_{ij} + c_j + h_k + e_{ijk},$$

where y_{ijk} is the standardized milk production on the parity i for cow j; β_0 is an effect common to all records; L_i is the lactation number ($i = 1, 2, \ldots, 5$); $(\mathrm{DIM})_{ij}$ is the number of days in milk of cow j in her ith lactation; β_1 and β_2 are fixed regression coefficients of lactation and DIM, respectively; c_j is the random additive effect for cow j ($j = 1, 2, \ldots, 1{,}359$); h_k is a random effect for herd k ($k = 1, 2, \ldots, 57$); and e_{ijk} is a random residual. The following distribution was assumed for the vector of random effects:

$$\begin{pmatrix} \mathbf{c} \\ \mathbf{h} \\ \mathbf{e} \end{pmatrix} \sim N\left[\mathbf{0}, \begin{pmatrix} \mathbf{A}_{1,675}\sigma_c^2 & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_{57}\sigma_h^2 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{I}_{3,397}\sigma_e^2 \end{pmatrix}\right],$$

where, $\mathbf{c} = \{c_j\}$ is a vector of additive cow effects; $\mathbf{h} = \{h_k\}$ is a vector of herd effects; $\mathbf{e} = \{e_{ijk}\}$ is a vector of residuals; σ_c^2, σ_h^2, and σ_e^2 are the additive genetic, between-herd, and residual variances, respectively; **A** represents

the matrix of additive relationships between cows, with the dimension indicated in the subscript; and **I** is an identity matrix with the dimension also indicated in the subscript.

This model was fitted in R as follows:

```
milk <- within(milk, sdMilk <- milk / sd(milk))
system.time(
 fm1 <- pedigreemm(sdMilk ~ lact + log(dim) + (1|id) + (1|herd),
 data = milk, pedigree = list(id = pedCowsR)))
 user     system   elapsed
 209.889  2.036    212.115
```

Here, the vector $\mathbf{y} = \{y_{ijk}\}$ is `sdMilk` and the milk production record is `milk$milk`. The `fm1` object is class *pedigreemm* and contains the model fitted in example 1. Object `fm1` has the properties defined in its class. Note that at the left of the symbol `~` is the response variable, and at the right is the linear model. Fixed effects are included, and an intercept is added by default (to exclude the intercept, the model should include a *−1*). Terms in parentheses, such as `(1|herd)`, represent random effects; the *1* before the symbol `|` indicates the fitting of a random intercept. A regression here would be a subject-specific random regression term. The parameter `pedigree` lists the random effects that should be linked with the nonlinear covariance matrix, in this case the additive relationship matrix. All levels of this random effect should be represented in the pedigree. The function `system.time()` is used to measure how long the procedure takes to run the analysis. As the result shows, it took 209.889 s (approximately 3.5 min) for this example, run on a CPU with a 2-GHz AMD Athlon 64 processor (dual core, but R is single threaded) with 4 GB of memory (but the process used approximately 7% of the memory). The object provided would be class `pedigree`. Storing this object allows the user to reuse it without fitting the model again. An alternative to save `fm1` for future use would be `save(fm1, file = "fm1.rda")`. To restore this object to the environment, type `load("fm1.rda")`. Some of the results for `fm1` can by seen just by typing `fm1`, or by typing `summary(fm1)`, which returns the following summary:

```
Linear mixed model fit by REML
 Linear mixed model fit by REML
 Formula: sdMilk ~lact + log(dim) + (1 | id) + (1 | herd)
   Data: milk
  AIC   BIC     logLik   deviance  REMLdev
  8420  8457    -4204    8393      8408
 Random effects:
  Groups  Name         Variance  Std.Dev.
  id      (Intercept)  0.28064   0.52976
  herd    (Intercept)  0.20412   0.45179
  Residual            0.48606   0.69718
 Number of obs: 3397, groups: id, 1359; herd, 57
 Fixed effects:
              Estimate Std.    Error t     value
 (Intercept)  1.73551          0.26865    6.460
 Lact         -0.10666          0.01236    -8.631
 log(dim)     0.72606          0.04389    16.543
 Correlation of Fixed Effects:
             (Intr)    lact
 lact        -0.297
 log(dim)    -0.961    0.221
```

The variance attributable to additive effect was 0.281, the herd variance was 0.204, and the residual variance was 0.486. The estimates of fixed effects were 1.74 for the intercept, 0.73 for the regression coefficient on `log(dim)` and −0.11 for the regression on the lactation effect. The structure of the object shows all the information listed on it. This information can be seen by typing `str(fm1)`. Some important information contained in the object includes the model matrices used, **Z**, or the transformed **Z**, the call, among other relevant information.

If there is enough information, a linear animal model with permanent environmental effects, could be fitted in R by using the code

```
milk <- within(milk, idPE <- id)
fm2 <- pedigreemm(
```

```
 sdMilk ~ lact + (1|id) + log(dim) + (1|idPE) + (1|herd),
 data = milk, pedigree = list(id = pedCows))
save(fm2, file = "fm2.rda")
```

The random cow effect is included twice in the code for the model: one time includes a correlation structure equivalent to that of the additive relationship matrix, and the second time includes independence between the animals. The first term will capture the additive genetic variance, whereas the second term will capture the variance between animals attributable to effects other than additive genetics.

Example 2: Poisson Sire Model. This illustration uses a GLMM fitting a sire model. The NCM (data set 2) has been modeled previously with Poisson models (Vazquez et al., 2009). The GLMM uses the log as a link function between NCM and a predictor that is linear, as follows:

$$\log(\lambda_{ijkm}) = \beta_0 + B_i + CY_j + s_k + h_m,$$

where λ_{ijkm} is a Poisson parameter specific to observation *ijkm*; β_0 is an intercept; B_i is the effect of birth year i of a cow; CY_j is the fixed effect of the calving year ($j = 2000, 2001, \ldots, 2005$); s_k is a random effect of sire k ($k = 1, 2, \ldots, 38$); and h_m is a random effect of herd ($m = 1, 2, \ldots, 41$). Because the log is used as a link function, then $E(NCM_{ijkm}|s_m,h_k) = \lambda_{ijkm}$, where NCM_{ijkm} is the number of clinical mastitis cases associated with observation *ijkm*.

The following distribution was assumed for the vectors of random herd (h) and sire (s) effects:

$$\begin{pmatrix}\mathbf{s}\\ \mathbf{h}\end{pmatrix} \sim N\left[\mathbf{0}, \begin{pmatrix}\mathbf{A}_{38}\sigma_s^2 & \mathbf{0}\\ \mathbf{0} & \mathbf{I}_{41}\sigma_h^2\end{pmatrix}\right],$$

where σ_s^2 is the between-sire variance, $\mathbf{A}$ represents the matrix of additive relationships between sires, and $\mathbf{I}_{41}$ is an identity matrix of order 41.

The model above was fitted using the following code:

```
fm3 <- pedigreemm(
 NCM ~ birth + calvingYear + (1|sire) + (1|herd),
 data = mastitis, pedigree = list(sire = pedSires),
 family = "poisson")
save(fm3, file= "fm3.rda")
```

Note that for fitting a sire model, the specified random effect would be the sire rather than the animal, and the pedigree would be the sire pedigree. The NCM is a count response, and could be modeled with a Poisson distribution. The family parameter should be specified; otherwise, a linear model with a count response would be applied by default. The herd variance was 0.905 and the sire variance was 0.058. Some larger applications have been used for counts of mastitis cases and for a binary outcome for mastitis (Vazquez, 2007; Vazquez et al., 2009).

To fit a model with a noncanonical link, the desired link has to be specified (see `help(family)`). An example of a call to the function using the noncanonical link *probit* for the binomial family, rather than using the default *logit*, is

```
fm4 <- pedigreemm(
 mastitis ~ birth + calvingYear + (1|sire) + (1|herd),
 data = mastitis, pedigree = list(sire = pedSires),
 family = "binomial"(link = "probit"))
```

Model Output

This section illustrates a few properties and uses of the *pedigreemm* objects with the fitted models to obtain the information of interest.

Random and Fixed Effects. The objects containing the fitted models have many properties of potential interest to the user. If an object is not in the environment, it should be loaded using `load('fm1.rda')`. From this object, the predicted random effects can be extracted as follows:

```
Random.fm1 <- ranef(fm1)
```

The object called `Random.fm1` will be a list with as many elements as random effects. To see its structure, type `str(Random.fm1)`. The predicted herd effects can be obtained by typing `Random.fm1$herd`.

The sire effects, by default, are expressed in their original scale (u), defined as $u^* = (\mathbf{L}^{-1})u$ (BLUP in example 1 with a Gaussian distribution). To calculate u, the transformed predicted sire effects (u^*) are premultiplied by $\mathbf{L}$, the Cholesky factor of the relationship matrix. If, for some reason, the transformed predicted effects u are desired, it should be specified in the function by typing `ranef(fm1, pedigree=FALSE)`. The (u^*) can be seen as follows:

```
cow.effects <- ranef(fm1)$id
```

The 6 animals with the larger predictions can be returned by ordering the vector of predicted sires as follows:

```
cowId <-labels(ranef(fm1)$id)[[1]]
effect <- data.frame(effects=as.numeric(cow.effects[[1]]), cow = cowId)
best <- tail(effect[order(effect$effects),])
```

Note that the function `tail` will extract the last 6 elements of the list. Without this function, the object `cow.effects[order(cow.effects$effects),]` would be the entire list of predicted random effects, ordered from smallest to largest. For instance, the largest predicted values `best` are

```
        effects      cow
960     0.9246586    960
132     0.9481742    132
152     0.9486676    152
244     1.0016387    244
1031    1.0090903    1031
1083    1.0197981    1083
```

The units of the predicted random effects are those of standardized milk production. The worst animals can be found as follows:

```
worst <- head(effect[order(effect$effects),])
```

Other functions can take these objects (`fm1, herd.effects`, etc.) to produce new objects (e.g., to draw a density plot):

```
plot(density(Random.fm1$herd[[1]]), xlab = 'Herd Effect',
 ylab = 'Density', main = '')
```

To get estimates of fixed effects, one can use

```
fixef(fm1)
(Intercept)    lact          log(dim)
1.7355129      -0.1066584    0.7260614
```

As for random effects, fixed effects are an attribute of the fitted model objects and can be extracted directly from the object.

Residual Values. The residuals are an attribute of the `fm1` object and could be obtained either directly or with the `resid()` function. The `head` function displays the residuals of the first 6 records:

```
 head(resid(fm1))
[1] -0.3849845 0.3842404 -0.5010199 -0.1163802 -0.1408525 0.3466662
```

The residuals could be evaluated vs. the fitted values by plotting

```
hist(resid(fm1), xlab = 'Residuals from fm1')
```

Inbreeding Coefficients and Additive Relationship Matrix.

The inbreeding coefficient can be obtained by providing a pedigree object as a parameter as follows:

```
Inbreeding <- inbreeding(pedSires)
```

In this case, the object assigned to `Inbreeding` will be a vector with the inbreeding coefficients of the animals in the pedigree object, `pedSire`, ordered after the number of animals in the pedigree object.

The relationship matrix of any pedigree can be built by using the function `relfactor()`, which returns the right component of the Cholesky decomposition of the relationship matrix, an upper triangular matrix, so

```
U <- relfactor(pedCows)
A <- t(U) %*% U
```

Coefficients, Bayesian Information Criterion, Akaike Information Criterion, Deviance. Other statistics can be obtained from any of the fitted model objects. For example, the coefficients

```
fixef(summary(fm1))
(Intercept)    lact          log(dim)
1.7355129      -0.1066584    0.7260614
```

In R, the results can always be indexed to obtain a single coefficient:

```
fixef(summary(fm1))[[3]]
[1] 0.7260614
```

In the case of model-fitting information, such as the Akaike information criterion (**AIC**) or the Bayesian information criterion (**BIC**), such statistics can be seen by typing `summary(fm1)`:

```
…
AIC     BIC     logLik    deviance    REMLdev
8420    8457    -4204      8393        8408
…
```

The covariance matrix of fixed effect estimates can be obtained as follows:

```
 vcov(fm1)
3 x 3 Matrix of class "dpoMatrix"
                 [,1]            [,2]            [,3]
[1,]    0.0721745513   -0.0009844112   -0.0113318251
[2,]   -0.0009844112    0.0001527018    0.0001200142
[3,]   -0.0113318251    0.0001200142    0.0019263519
```

Again, for a specific element of the covariance matrix, one can index the `vcov` function, for example, `vcov(fm1)[1,2]`.

ANOVA and Likelihood Ratio Test. Another analysis that uses the fitted model object is the ANOVA function `anova(fm1)`:

```
Analysis of Variance Table
           Df    Sum Sq     Mean Sq    F value
lact        1     77.265     77.265     158.96
log(dim)    1     132.825    132.825    273.27
```

Additionally, let the model `fm1_nested` be equal to `fm1` but without the `log(dim)`,

```
fm1_nested <-pedigreemm(
  formula = sdMilk ~lact + (1 | id) + (1 | herd),
  data = milk, pedigree = list(id= pedCowsR),
 verbose = TRUE)
```

The models are nested, and a likelihood ratio test can be performed as follows:

```
anova(fm1, fm1_nested)
Data: milk
Models:
fm1_nested: sdMilk ~ lact + (1 | id) + (1 | herd)
fm1: sdMilk ~ lact + log(dim) + (1 | id) + (1 | herd)
             Df    AIC         BIC       logLik    Chisq Chi   Df   Pr(>Chisq)
fm1_nested    5     8666.5    8697.1    -4328.2
fm1           6     8405.2    8442.0    -4196.6    263.27      1    < 2.2e-16 ***
```

The result indicates that the model stored in `fm1` is significantly better than the model stored in `fm1_nested`, the model without `log(dim)` as a covariate. This likelihood ratio test result is consistent with the AIC and BIC criteria, which were much smaller for the `fm1` model.

Concluding Remarks

The strategy used on *pedigreemm* uses the Cholesky decomposition of the (co)variance structure of the random effects. This approach allowed us to expand an existing package easily to accommodate covariance among random effects, and it also may have computational advantages because convergence may be faster when a model is parameterized in terms of independent random variables. However, this may not be a convenient representation when the (co)variance structure is not sparse, such as in the case of densely connected pedigrees. In such circumstances, large example memory capacities are necessary and the convergence process becomes slower. On the other hand, this tool is available in R, which is free and powerful statistical software. As developed, *pedigreemm* uses records as input data and a pedigree and fits generalized mixed models with correlated random effects. In this case, an additive relationship matrix is used in the (co)variance structure for the random effects. Additionally, a few public functions allow the relationship matrix or inbreeding coefficient to be obtained for a certain pedigree.

LITERATURE CITED

Bates, D. M. 2009. Index of http://lme4.r-forge.r-project.org/slides/2009-07-21-Seewiesen/ Section: 4Precision. Accessed Aug. 5, 2009.

Bates, D., and M. Maechler. 2008. The Comprehensive R Archive Network. http://cran.r-project.org/ Accessed Jan. 2009.

Harville, D. A., and T. P. Callanan. 1989. Computational aspects of likelihood-based inference for variance components. Pages 136–176 in Advances in Statistical Methods for Genetic Improvement of Livestock. D. Gianola and K. Hammond, ed. Springer-Verlag, Berlin, Germany.

Henderson, C. R. 1963. Selection index and expected genetic advance. Pages 141–163 in Statistical Genetics and Plant Breeding. W. D. Handson and H. F. Robinson, ed. Natl. Acad. Sci. and Natl. Res. Counc., Washington, DC.

Henderson, C. R. 1973. Sire evaluation and genetic trends. Pages 10–41 in Proc. Anim. Breed. Genet. Symp., Blacksburg, VA. Am. Soc. Anim. Sci., Champaign, IL.

Henderson, C. R. 1976. A simple method for computing the inverse of a numerator relationship matrix used in prediction of breeding values. Biometrics 32:69–83.

Henderson, C. R., O. Kempthorne, S. R. Searle, and C. N. VonKrosig. 1959. Estimation of environmental and genetic trends from records subject to culling. Biometrics 15:192–218.

Mrode, R. 2005. Linear Models for the Prediction of Animal Breeding Values. 2nd ed. CAB Int., New York, NY.

R Development Core Team. 2008. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.

Sargolzaei, M., and H. Iwaisaki. 2005. Comparison of four direct algorithms for computing inbreeding coefficients. Anim. Sci. J. 76:401–406.

Tempelman, R. J. 1998. Generalized linear mixed models in dairy cattle breeding. J. Dairy Sci. 81:1428–1444.

Vazquez, A. I. 2007. Analysis of number of episodes of clinical mastitis in Norwegian Red and Holstein cows with Poisson and categorical data mixed models. MS Thesis. Univ. Wisconsin, Madison.

Vazquez, A. I., D. Gianola, D. Bates, K. A. Weigel, and B. Heringstad. 2009. Assessment of Poisson, logit and linear models for genetic analysis of clinical mastitis in Norwegian Red Cows. J. Dairy Sci. 92:739–748.