

# Package ‘spaMM’

May 19, 2018

**Type** Package

**Title** Mixed-Effect Models, Particularly Spatial Models

**Encoding** UTF-8

**Version** 2.4.35

**Date** 2018-05-19

**Maintainer** François Rousset <francois.rousset@umontpellier.fr>

**Imports** methods, stats, graphics, Matrix, MASS, proxy, Rcpp (>= 0.12.10), nlme, nloptr, pbapply

**LinkingTo** Rcpp, RcppEigen

**Suggests** maps, testthat, lme4, rsae, rcdd, pedigreemm, minqa, lpSolveAPI (>= 5.5.0.14), foreach, multilevel, Infusion, IsoriX, blackbox

**Depends** R (>= 3.2.0)

**NeedsCompilation** yes

## Description

Inference in mixed-effect models, including generalized linear mixed models with spatial correlations and models with non-Gaussian random effects (e.g., Beta). Variation in residual variance (heteroscedasticity) can itself be represented by a generalized linear mixed model. Various approximations of likelihood or restricted likelihood are implemented, in particular h-likelihood (Lee and Nelder 2001 <doi:10.1093/biomet/88.4.987>) and Laplace approximation.

**License** CeCILL-2

**URL** <https://www.r-project.org>,  
<http://kimura.univ-montp2.fr/~rousset/spaMM.htm>

**RoxygenNote** 6.0.1

**ByteCompile** true

**Author** François Rousset [aut, cre, cph],  
Jean-Baptiste Ferdy [aut, cph],  
Alexandre Courtiol [aut],  
GSL authors [ctb] (src/gsl\_bessel.\*)

**Repository** CRAN

**Date/Publication** 2018-05-19 21:02:41 UTC

**R topics documented:**

adjlg . . . . .	3
arabidopsis . . . . .	4
autoregressive . . . . .	6
blackcap . . . . .	7
CauchyCorr . . . . .	9
COMPoisson . . . . .	10
confint.HLfit . . . . .	12
corMatern . . . . .	13
corrHLfit . . . . .	15
corrMatrix . . . . .	17
covStruct . . . . .	18
designL.from.Corr . . . . .	19
extractors . . . . .	21
fitme . . . . .	24
fixed . . . . .	26
fixedLRT . . . . .	27
freight . . . . .	30
get_ranPars . . . . .	31
good-practice . . . . .	31
HLCor . . . . .	32
HLfit . . . . .	34
inverse.Gamma . . . . .	40
Loaloa . . . . .	41
LRT . . . . .	43
make_scaled_dist . . . . .	45
mapMM . . . . .	46
MaternCorr . . . . .	50
multinomial . . . . .	52
negbin . . . . .	54
options . . . . .	55
pedigree . . . . .	57
phiHGLM . . . . .	58
plot.HLfit . . . . .	59
Poisson . . . . .	61
predict . . . . .	62
Predictor . . . . .	66
rankinfo . . . . .	68
salamander . . . . .	69
scotlip . . . . .	70
seaMask . . . . .	71
seeds . . . . .	73
separation . . . . .	74
simulate.HLfit . . . . .	74
spaMM . . . . .	75
spaMM-conventions . . . . .	78
spaMM.colors . . . . .	78

<i>adjlg</i>	3
spaMM.filled.contour . . . . .	79
spaMM_boot . . . . .	82
spaMM_glm.fit . . . . .	84
sparse_precision . . . . .	86
stripHLfit . . . . .	87
summary.HLfit . . . . .	88
sym_eigen . . . . .	89
update.HLfit . . . . .	90
wafers . . . . .	91
welding . . . . .	92
<b>Index</b>	<b>94</b>

---

<i>adjlg</i>	<i>Simulated data set for testing sparse-precision code</i>
--------------	---

---

### Description

This is used in tests/test-adjacency-long.R

### Usage

```
data("adjlg")
```

### Format

Includes an adjacency matrix `adjlgMat`. and a data frame `adjlg` with 5474 observations on the following 8 variables.

ID a factor with levels 1 to 1000

months a numeric vector

GENDER a character vector

AGE a numeric vector

X1 a numeric vector

X2 a numeric vector

month a numeric vector

BUY a numeric vector

### Source

The simulation code show in Example was suggested by Jeroen van den Ochtend.

## Examples

```

data(adjlg)
## See further usage in tests/test-adjacency-long.R
## Not run:
# as produced by:
library(data.table) ## Included data produced using version 1.10.4.3
library(igraph) ## Included data produced using version 1.2.1

rsample <- function(N=100, ## size of implied adjacency matrix
                    month_max=10,seed) {
  if (is.integer(seed)) set.seed(seed)
  dt <- data.table(ID=factor(1:N))
  dt$months <- sample(1:month_max,N,replace=T) ## # of liens for each level of ID
  dt$GENDER <- sample(c("MALE","FEMALE"),N,replace=TRUE)
  dt$AGE <- sample(18:99,N,replace=T)
  dt$X1 <- sample(1000:9900,N,replace=T)
  dt$X2 <- runif(N)

  dt <- dt[, c(.SD, month=data.table(seq(from=1, to=months, by = 1))), by = ID]
  dt[,BUY := 0]
  dt[month.V1==months,BUY := sample(c(0,1),1),by=ID]
  setnames(dt,"month.V1","month")

  ##### create adjacency matrix
  Network <- data.table(OUT=sample(dt$ID,N*month_max*4/10))
  Network$IN <- sample(dt$ID,N*month_max*4/10)
  Network <- Network[IN != OUT]
  Network <- unique(Network)
  g <- graph.data.frame(Network,directed=F)
  g <- add_vertices(g,sum(!unique(dt$ID) %in% V(g)),name=unique(dt[!dt$ID %in% V(g),list(ID)]))
  Network <- as_adjacency_matrix(g,sparse = TRUE,type="both")
  return(list(data=dt,adjMatrix=Network))
}

set.seed(123)
adjlg_sam <- rsample(N=1000,seed=NULL)
adjlg <- as.data.frame(adjlg_sam$data)
adjlgMat <- adjlg_sam$adjMatrix

## End(Not run)

```

## Description

For 948 “accessions” from European *Arabidopsis thaliana* populations, this data set merges the genotypic information at four single nucleotide polymorphisms (SNP) putatively involved in adaptation to climate (Fournier-Level et al, 2011, Table 1), with 13 climatic variables from Hancock et al. (2011).

**Usage**

```
data("arabidopsis")
```

**Format**

The data frame includes 948 observations on the following variables:

**pos1046738, pos5510910, pos6235221, pos8132698** Genotypes at four SNP loci

**LAT** latitude

**LONG** longitude

**seasonal, tempWarmest, tempColdest, preciWettest, preciDriest, preciCV, PAR\_SPRING,**

**growingL, conseqCold, conseqFrFree, RelHumidSp, dayLSp, aridity** Thirteen climatic variables.

See Hancock et al. (2011) for details about these variables.

**Details**

The response is binary so `HLmethod="PQL/L"` seems warranted (see Rousset and Ferdy, 2014).

**Source**

The data were retrieved from <http://bergelson.uchicago.edu/regmap-data/climate-genome-scan> on 22 February 2013 (they may no longer be available from there).

**References**

Fournier-Level A, Korte A., Cooper M. D., Nordborg M., Schmitt J., Wilczek AM (2011). A map of local adaptation in *Arabidopsis thaliana*. *Science* 334: 86-89.

Hancock, A. M., Brachi, B., Faure, N., Horton, M. W., Jarymowycz, L. B., Sperone, F. G., Toomajian, C., Roux, F., and Bergelson, J. 2011. Adaptation to climate across the *Arabidopsis thaliana* genome, *Science* 334: 83-86.

Rousset F., Ferdy, J.-B. (2014) Testing environmental and genetic effects in the presence of spatial autocorrelation. *Ecography*, 37: 781-790. <http://dx.doi.org/10.1111/ecog.00566>

**Examples**

```
data("arabidopsis")
if (spaMM.getOption("example_maxtime")>8) {
  HLCor(cbind(pos1046738,1-pos1046738)~seasonal+Matern(1|LAT+LONG),
        ranPars=list(rho=0.1192779,nu=0.2369892,lambda=8.599),
        family=binomial(),HLmethod="PQL/L",data=arabidopsis)
}
## The above ranPars are deduced from the following fit:
if (spaMM.getOption("example_maxtime")>46) {
  SNPfit <- fitme(cbind(pos1046738,1-pos1046738)~seasonal+Matern(1|LAT+LONG),
                verbose=c(TRACE=TRUE),
                family=binomial(),method="PQL/L",data=arabidopsis)
  summary(SNPfit) # p_v=-125.0392
}
```

## Description

Two autoregressive(AR) models are currently implemented: the adjacency model (a conditional AR, i.e., CAR), and the AR1 model for time series. Implementation of further models (in particular, of simultaneous AR, i.e., SAR) is to be expected in the future. Efficient algorithms for CAR models have been widely discussed in particular in the econometric literature (e.g., LeSage and Pace 2009), but these models are not necessarily recommended for irregular lattices (see Wall, 2004 and Martellosio, 2012 for some insights on the implications of autoregressive models). The fastest method in spaMM for large data sets is implemented in the `fitme` function. For small data sets (as in the example below), `HLCor` may be fastest. It is suggested to use `fitme` generally unless one has a large number of small data sets to analyze. For non-LMMs, `corrHLfit(*,HLmethod="PQL/L")` can be quite fast (but does not return an ML or REML fit).

An AR1 random effect is specified as `AR1(1|<grouping factor>)`. It describes correlations between realizations of the random effect for (typically) successive time-steps by a correlation  $\phi$ , denoted `ARphi` in function calls. Nested AR1 effects can be specified by a nested grouping factor, as in `AR1(1|<time index> %in% <nesting factor>)`.

A CAR random effect is specified as `adjacency(1|<grouping factor>)`. The correlations among levels of the random effect form a matrix  $(\mathbf{I} - \rho \text{adjMatrix})^{-1}$ , in terms of an `adjMatrix` matrix which must be provided, and of the scalar  $\rho$ , denoted `rho` in function calls. The rows and columns of `adjMatrix` must be ordered as increasing values of the levels of the geographic location index specifying the spatial random effect. For example, if the model formula is `y ~ adjacency(1|geo.loc)` and `<data>$geo.loc` is 2,4,3,1,... the first row/column of the matrix refers to `geo.loc=1`, i.e. to the fourth row of the data.

## Details

For **AR1** models, and large data sets, the fitting functions by default select methods that exploits the sparsity of the precision matrix of the random effects. the dimension of the implied precision matrix is determined by the extreme values of grouping factor (typically interpreted as a time index), as all intermediate values must be considered. Thus, the precision matrix may be quite large even if few levels are represented in the data.

For **CAR** models, different fitting strategies may be used:

A call to `HLCor` uses the spectral decomposition of the adjacency matrix as further detailed below. This is fast for small datasets but `fitme` may be preferable otherwise.

A call to `corrHLfit` with the additional argument `init.HLfit=list(rho=0)` should be equivalent in speed and result to the `HLCor` call.

A call to `corrHLfit` without this argument does not use the spectral decomposition. It performs a generic numerical maximization of the likelihood (or restricted likelihood) as function of the correlation parameter  $\rho$ . The ML fits by `corrHLfit` and `HLCor` should be practically equivalent. The REML fits should slightly differ from each other, due to the fact that the REML approximation for GLMMs does not maximize a single likelihood function.

In the adjacency model, the covariance matrix of random effects  $\mathbf{u}$  can be described as  $\lambda(\mathbf{I}-\rho\mathbf{W})^{-1}$  where  $\mathbf{W}$  is the (symmetric) adjacency matrix. HLCor uses the spectral decomposition of the adjacency matrix, written as  $\mathbf{W}=\mathbf{V}\mathbf{D}\mathbf{V}'$  where  $\mathbf{D}$  is a diagonal matrix of eigenvalues  $d_i$ . The covariance of  $\mathbf{V}'\mathbf{u}$  is  $\lambda(\mathbf{I}-\rho\mathbf{D})^{-1}$ , which is a diagonal matrix with elements  $\lambda_i=\lambda/(1-\rho d_i)$ . Hence  $1/\lambda_i$  is in the linear predictor form  $\alpha+\beta d_i$ . This can be used to fit  $\lambda$  and  $\rho$  efficiently. If HLCor is used, the results are reported as the coefficients  $\alpha$  (Intercept) and  $\beta$  (adjd) of the predictor for  $1/\lambda_i$ , in addition to the resulting values of  $\rho$  and of the common  $\lambda$  factor.

## References

- LeSage, J., Pace, R.K. (2009) Introduction to Spatial Econometrics. Chapman & Hall/CRC.
- Martellosio, F. (2012) The correlation structure of spatial autoregressions, *Econometric Theory* 28, 1373-1391.
- Wall M.M. (2004) A close look at the spatial structure implied by the CAR and SAR models: *Journal of Statistical Planning and Inference* 121: 311-324.

## Examples

```
##### AR1 random effect:
ts <- data.frame(lh=lh,time=seq(48)) ## using 'lh' data from stats package
HLCor(lh~1 +AR1(1|time), data=ts, ranPars=list(ARphi=0.5,lambda=0.25,phi=0.001))

##### CAR random effect:
data("scotlip")
# CAR by Laplace with 'outer' estimation of rho
if (spaMM.getOption("example_maxtime")>0.7) {
  corrHLfit(cases~I(prop.ag/10) +adjacency(1|gridcode)+offset(log(expec)),
            adjMatrix=Nmatrix,family=poisson(),data=scotlip,HLmethod="ML")
}
if (spaMM.getOption("example_maxtime")>0.8) {
  fitme(cases~I(prop.ag/10) +adjacency(1|gridcode)+offset(log(expec)),
        adjMatrix=Nmatrix,family=poisson(),data=scotlip)
}

# CAR by Laplace with 'inner' estimation of rho
HLCor(cases~I(prop.ag/10) +adjacency(1|gridcode)+offset(log(expec)),
      adjMatrix=Nmatrix,family=poisson(),data=scotlip,HLmethod="ML")
```

## Description

This data set is extracted from a study of genetic polymorphisms potentially associated to migration behaviour in the blackcap (*Sylvia atricapilla*). Across different populations in Europe and Africa, the average migration behaviour was found to correlate with average allele size (dependent on the number of repeats of a small DNA motif) at the locus ADCYAP1, encoding a neuropeptide. This

data set is quite small and ill-suited for separating random-effect variance from residual variance. The likelihood surface for the Matérn model actually has local maxima.

### Usage

```
data("blackcap")
```

### Format

The data frame includes 14 observations on the following variables:

**latitude** latitude, indeed.

**longitude** longitude, indeed.

**migStatus** migration status as determined by Mueller et al, from 0 (resident populations) to 2.5 (long-distance migratory populations)

**means** Mean allele sizes in each population

**pos** Numerical index for the populations

### Details

Migration status was coded as : pure resident populations as '0', resident populations with some migratory restlessness as '0.5', partial migratory populations as '1', completely migratory populations migrating short-distances as '1.5', intermediate-distance migratory populations as '2' and distinct long-distance migratory populations as '2.5'.

### Source

Data from Mueller et al. (2011), including supplementary material retrieved from <http://rspb.royalsocietypublishing.org/content/suppl/2011/02/11/rspb.2010.2567.DC1>.

### References

Mueller, J. C., Pulido, F., and Kempenaers, B. 2011. Identification of a gene associated with avian migratory behaviour, Proc. Roy. Soc. (Lond.) B 278, 2848-2856.

### Examples

```
## see 'corrHLfit' and 'fixedLRT' for examples involving these data
```



CauchyCorr

*Cauchy correlation function and Cauchy formula term***Description**

The Cauchy family of correlation functions is useful to describe spatial processes with power-law decrease of correlation at long distance. It is valid for Euclidean distances in spaces of any dimension, and for great-circle distances on spheres of any dimension. It has a scale parameter ( $\rho$ , as in the Matérn correlation function), a shape (or “smoothness”, Gneiting 2013) parameter, and a long-memory dependence (or, more abstractly, “shape”; Gneiting 2013) parameter (Gneiting and Schlater 2004). The present implementation also accepts a Nugget parameter. The family can be invoked in two ways. First, the CauchyCorr function evaluates correlations, using distances as input. Second, a term of the form `Cauchy(1|<...>)` in a formula specifies a random effect with Cauchy correlation function, using coordinates found in a data frame as input. In the latter case, the correlations between realizations of the random effect for any two observations in the data will be the value of the Cauchy function at the scaled distance between coordinates specified in `<...>`, using “+” as separator (e.g., `Cauchy(1|latitude + longitude)`).

**Usage**

```
CauchyCorr(d, rho=1, shape, longdep, Nugget=NULL)
# Cauchy(1|...)
```

**Arguments**

<code>d</code>	Euclidean or great-circle distance
<code>rho</code>	The scaling factor for distance, a real $>0$ .
<code>shape</code>	The shape parameter, a real $0 < \leq 2$ for Euclidean distances and $0 < \leq 1$ for great-circle distances.
<code>longdep</code>	The long-memory dependence parameter, a real $>0$ . It gives the exponent of the asymptotic decrease of correlation with distance: the <b>smaller</b> longdep is, the longer the dependence.
<code>Nugget</code>	(Following the jargon of Kriging) a parameter describing a discontinuous decrease in correlation at zero distance. Correlation will always be 1 at $d = 0$ , and from which it immediately drops to $(1 - \text{Nugget})$ . Defaults to zero.
<code>...</code>	Names of coordinates, using “+” as separator (e.g., <code>Matern(1 latitude + longitude)</code> )

**Details**

The correlation at distance  $d > 0$  is

$$(1 - \text{Nugget}) \left( 1 + (\rho d)^{\text{shape}} \right)^{-\text{longdep}/\text{shape}}$$

**Value**

Scalar/vector/matrix depending on input.

## References

Gneiting, T. and Schlater M. (2004) Stochastic models that separate fractal dimension and the Hurst effect. *SIAM Rev.* 46: 269–282.

Gneiting T. (2013) Strictly and non-strictly positive definite functions on spheres. *Bernoulli* 19: 1327-1349.

## Examples

```
data("blackcap")
HLCor(migStatus ~ means+ Cauchy(1|latitude+longitude), data=blackcap,
      HLmethod="ML", ranPars=list(longdep=0.5, shape=0.5, rho=0.05))
## The Cauchy family can be used in Euclidean spaces of any dimension:
set.seed(123)
randpts <- matrix(rnorm(20), nrow=5)
distMatrix <- as.matrix(proxy::dist(randpts))
CauchyCorr(distMatrix, rho=0.1, shape=1, longdep=10)
```

---

COMPoisson

*Conway-Maxwell-Poisson (COM-Poisson) GLM family*

---

## Description

The COM-Poisson family is a generalization of the Poisson family which can describe over-dispersed as well as under-dispersed count data. It is indexed by a parameter  $\nu$  that quantifies such dispersion. It includes the Poisson, geometric and Bernoulli as special (or limit) cases (see Details). The COM-Poisson family is here implemented as a `family` object, so that it can be fitted by `glm`, and further used to model conditional responses in mixed models fitted by this package's functions (see Examples).  $\nu$  is distinct from the dispersion parameter  $\nu = 1/\phi$  considered elsewhere in this package and in the GLM literature, as  $\nu$  affects in a more specific way the log-likelihood. The "canonical link"  $\theta(\mu)$  between the canonical GLM parameter  $\theta$  and the expectation  $\mu$  of the response does not have a known expression in terms of elementary functions. The link inverse is  $\mu(\theta) = \sum_{i=0}^{\infty} \lambda^i / (i!)^\nu$  for  $\lambda = e^\theta$  (hence the link is here nicknamed "loglambda").

## Usage

```
COMPoisson(nu = stop("COMPoisson's 'nu' must be specified"),
           link = "loglambda")
```

## Arguments

<code>link</code>	GLM link function. Cannot be modified.
<code>nu</code>	Under-dispersion parameter. The <code>fitme</code> and <code>corrHLfit</code> functions called with <code>family=COMPoisson()</code> (no given <code>nu</code> value) will estimate this parameter. In other usage of this family, <code>nu</code> must be specified. <code>COMPoisson(nu=1)</code> is the Poisson family.

**Details**

For  $\nu > 1$ , the distribution is under-dispersed. The limit as  $\nu \rightarrow \infty$  is the Bernoulli distribution with expectation  $\lambda/(1 + \lambda)$ .

The link inverse function, as shown in Description, involves an infinite summation. In this summation and related computations for the COM $Poisson$  model, the sum can be easily approximated by a finite sum for large  $\nu$  but not when  $\nu$  approaches zero. For this reason, the code may fail to fit distributions with  $\nu$  approaching 0 (strong residual over-dispersion). The case  $\nu=0$  itself is the geometric distribution with parameter  $\lambda$  and is fitted by an ad hoc algorithm devoid of such problems. Otherwise, spaMM truncates the sum, and uses numerical integrals to approximate missing terms (which slows down the fitting operation). In addition, it applies an ad hoc continuity correction to ensure continuity of the result in  $\nu=1$  (Poisson case). These corrections affect numerical results for the case of residual overdispersion but are negligible for the case of residual underdispersion. Alternatively, spaMM uses Gaunt et al.'s approximations when the condition defined in `spaMM.getOption("CMP_asympto_cond")` is satisfied.

The name `COMP_nu` should be used to set values of  $\nu$  in control arguments of the fitting functions (e.g., `fitme(., init=list(COMP_nu=1))`).

**Value**

A family object.

**References**

Gaunt, Robert E. and Iyengar, Satish and Olde Daalhuis, Adri B. and Simsek, Burcin. An asymptotic expansion for the normalizing constant of the Conway–Maxwell–Poisson distribution. To appear in *Annals of the Institute of Statistical Mathematics*.

G. Shmueli, T. P. Minka, J. B. Kadane, S. Borle and P. Boatwright (2005) A useful distribution for fitting discrete data: revival of the Conway-Maxwell-Poisson distribution. *Appl. Statist.* 54: 127-142.

Sellers KF, Shmueli G (2010) A Flexible Regression Model for Count Data. *Ann. Appl. Stat.* 4: 943–961

**Examples**

```
# Fitting COM $Poisson$  model with estimated nu parameter:
data("freight") ## example from Sellers & Shmueli, Ann. Appl. Stat. 4: 943961 (2010)
fitme(broken ~ transfers, data=freight, family = COM $Poisson$ ())
# GLMM with under-dispersed conditional response
HLfit(broken ~ transfers+(1|id), data=freight, family = COM $Poisson$ (nu=10),HLmethod="ML")

## Not run:
data("freight")
# Equivalence of poisson() and COM $Poisson$ (nu=1):
COMPglm <- glm(broken ~ transfers, data=freight, family = poisson())
coef(COMPglm)
logLik(COMPglm)
COMPglm <- glm(broken ~ transfers, data=freight, family = COM $Poisson$ (nu=1))
coef(COMPglm)
logLik(COMPglm)
```

```
HLfit(broken ~ transfers, data=freight, family = COMpoisson(nu=1))
## End(Not run)
```

---

confint.HLfit                      *Confidence intervals for fixed-effect parameters*

---

### Description

This computes confidence intervals for a given fixed effect parameter, based on the  $p_v$ -based approximation of the profile likelihood ratio for this parameter. The profiling is other all other fitted parameters: other fixed effects, as well as variances of random effects and spatial correlations if these were fitted.

### Usage

```
## S3 method for class 'HLfit'
confint(object, parm, level=0.95, verbose=TRUE,...)
```

### Arguments

object	An object of class HLfit, such as return object of HLfit, HLCor or corrHLfit calls;
parm	The name of a parameter to be fitted, or its position in the the object's \$fixef vector. Valid names are those of the object's \$fixef;
level	The coverage of the interval;
verbose	whether to print the interval or not. As the function returns its more extensive results invisibly, this printing is the only visible output;
...	Additional arguments (maybe not used, but conforming to the generic definition of confint).

### Value

A list including the confidence interval for the target parameter, and the fits lowerfit and upperfit giving the profile fits at the confidence bounds. This is returned invisibly.

### Examples

```
## Not run:
data("wafers")
wfit <- HLfit(y ~X1+(1|batch), family=Gamma(log), data=wafers, HLmethod="ML")
confint(wfit,"X1")

## End(Not run)
```

corMatern

*Matern Correlation Structure as a corSpatial object***Description**

This implements the Matérn correlation structure (see [Matern](#)) for use with lme or glmmPQL. Usage is as for others corSpatial objects such as corGaus or corExp, except that the Matérn family has an additional parameter. This function was defined for comparing results obtained with corrHLfit to those produced by lme and glmmPQL. There are problems in fitting (G)LMMs in the latter way, so it is not a recommended practice.

**Usage**

```
corMatern(value = c(1, 0.5), form = ~1, nugget = FALSE, nuScaled = FALSE,
          metric = c("euclidean", "maximum", "manhattan"), fixed = FALSE)
```

**Arguments**

- |          |  |
|----------|--|
| value    | <p>An optional vector of parameter values, which serves as initial values or as fixed values depending on the fixed argument. It has either two or three elements, depending on the nugget argument.</p> <p>If nugget is FALSE, value should have two elements, corresponding to the "range" and the "smoothness" <math>\nu</math> of the Matérn correlation structure. If value has zero length, the default is a range of 90% of the minimum distance and a smoothness of 0.5 (exponential correlation). <b>Warning:</b> the range parameter used in corSpatial objects is the inverse of the scale parameter used in <a href="#">MaternCorr</a> and thus they have opposite meaning despite both being denoted <math>\rho</math> elsewhere in this package or in nlme literature.</p> <p>If nugget is TRUE, meaning that a nugget effect is present, value can contain two or three elements, the first two as above, the third being the "nugget effect" (one minus the correlation between two observations taken arbitrarily close together). If value has length zero or two, the nugget defaults to 0.1. The range and smoothness must be greater than zero and the nugget must be between zero and one.</p> |
| form     | <p>(Pasted from corSpatial) a one sided formula of the form <math>\sim S1 + \dots + Sp</math>, or <math>\sim S1 + \dots + Sp \mid g</math>, specifying spatial covariates S1 through Sp and, optionally, a grouping factor g. When a grouping factor is present in form, the correlation structure is assumed to apply only to observations within the same grouping level; observations with different grouping levels are assumed to be uncorrelated. Defaults to <math>\sim 1</math>, which corresponds to using the order of the observations in the data as a covariate, and no groups.</p>   |
| nugget   | <p>an optional logical value indicating whether a nugget effect is present. Defaults to FALSE.</p>   |
| nuScaled | <p>If nuScaled is set to TRUE the "range" parameter <math>\rho</math> is divided by <math>2\sqrt{\nu}</math>. With this option and for large values of <math>\nu</math>, corMatern reproduces the calculation of corGaus. Defaults to FALSE, in which case the function compares to corGaus with range parameter <math>2(\sqrt{\nu})\rho</math> when <math>\nu</math> is large.</p>  |

metric	(Pasted from corSpatial) an optional character string specifying the distance metric to be used. The currently available options are "euclidean" for the root sum-of-squares of distances; "maximum" for the maximum difference; and "manhattan" for the sum of the absolute differences. Partial matching of arguments is used, so only the first three characters need to be provided. Defaults to "euclidean".
fixed	an optional logical value indicating whether the coefficients should be allowed to vary in the optimization, or kept fixed at their initial value. Defaults to FALSE, in which case the coefficients are allowed to vary.

### Details

This function is a constructor for the corMatern class, representing a Matérn spatial correlation structure. See [MaternCorr](#) for details on the Matérn family.

### Value

an object of class corMatern, also inheriting from class corSpatial, representing a Matérn spatial correlation structure.

### Note

The R and C code for the methods for corMatern objects builds on code for corSpatial objects, by D.M. Bates, J.C. Pinheiro and S. DebRoy, in a circa-2012 version of nlme.

### References

Mixed-Effects Models in S and S-PLUS, José C. Pinheiro and Douglas M. Bates, Statistics and Computing Series, Springer-Verlag, New York, NY, 2000.

### See Also

[glmmPQL](#), [lme](#)

### Examples

```
## LMM
data("blackcap")
blackcapD <- cbind(blackcap,dummy=1) ## obscure, isn't it?
## With method= 'ML' in lme, The correlated random effect is described
## as a correlated residual error and no extra residual variance is fitted:
nlme::lme(fixed = migStatus ~ means, data = blackcapD, random = ~ 1 | dummy,
          correlation = corMatern(form = ~ longitude+latitude | dummy),
          method = "ML")

## Binomial GLMM
if (spaMM.getOption("example_maxtime")>32) {
  data("Loaloe")
  LoaloeD <- cbind(Loaloe,dummy=1)
  MASS::glmmPQL(fixed = cbind(npos,ntot-npos)~elev1+elev2+elev3+elev4+maxNDVI1+seNDVI,
                data = LoaloeD, random = ~ 1 | dummy,family=binomial,
```

```

    correlation = corMatern(form = ~ longitude+latitude | dummy))
  }

```

corrHLfit

*Fits a mixed model, typically a spatial GLMM.***Description**

corrHLfit performs the joint estimation of correlation parameters, fixed effect and dispersion parameters.

**Usage**

```

corrHLfit(formula, data, init.corrHLfit = list(),
           init.HLfit = list(), ranFix = list(), lower = list(),
           upper = list(),
           objective = NULL, resid.model = ~1, resid.formula,
           control.dist = list(), control.corrHLfit = list(),
           processed = NULL, family = gaussian(),
           nb_cores = NULL, ...)

```

**Arguments**

formula	Either a linear model <a href="#">formula</a> (as handled by various fitting functions) or a predictor, i.e. a formula with attributes (see <a href="#">Predictor</a> and examples below). See Details in <a href="#">spaMM</a> for allowed terms in the formula.
data	A data frame containing the variables in the response and the model formula.
init.corrHLfit	An optional list of initial values for correlation and/or dispersion parameters, e.g. <code>list(rho=1, nu=1, lambda=1, phi=1)</code> where rho and nu are parameters of the Matérn family (see <a href="#">Matern</a> ), and lambda and phi are dispersion parameters (see Details in <a href="#">spaMM</a> for the meaning of these parameters). All are optional, but giving values for a dispersion parameter changes the ways it is estimated (see Details). rho may be a vector (see <a href="#">make_scaled_dist</a> ) and, in that case, it is possible that some or all of its elements are NA, for which corrHLfit substitute automatically determined values.
init.HLfit	See identically named <a href="#">HLfit</a> argument.
ranFix	A list similar to <code>init.corrHLfit</code> , but specifying fixed values of the parameters not estimated. See <a href="#">ranFix</a> for further information.
lower	An optional list of values of parameters specified through <code>init.corrHLfit</code> , used as lower values in calls to <code>optim</code> . See Details for default values.
upper	Same as lower, but upper values.
objective	For development purpose, not documented (this had a distinct use in the first version of <a href="#">spaMM</a> , but has been deprecated as such).
resid.model, resid.formula	See identically named <a href="#">HLfit</a> arguments.

control.dist	See control.dist in <a href="#">HLCor</a>
control.corrHLfit	This may be used control the optimizer. See <a href="#">spaMM.options</a> for default values.
processed	For programming purposes, not documented.
family	Either a <a href="#">family</a> or a <a href="#">multi</a> value.
nb_cores	<b>Not yet operative</b> , only for development purposes. Number of cores to use for parallel computations.
...	Optional arguments passed to <a href="#">HLCor</a> , <a href="#">HLfit</a> or <a href="#">designL.from.Corr</a> , for example the <code>distMatrix</code> argument of <a href="#">HLCor</a> . Arguments that do not fit within these functions are detected and a warning is issued.

## Details

Under the Matérn correlation model, `corrHLfit` typically performs a optimization over the  $\rho$  and  $\nu$  parameters, with maximum possible values as set by [spaMM.options](#).

By default `corrHLfit` will estimate correlation parameters by maximizing the objective value returned by `HLCor` calls wherein the dispersion parameters are estimated jointly with fixed effects for given correlation parameters. If dispersion parameters are specified in `init.corrHLfit`, they will also be estimated by maximizing the objective value, and `HLCor` calls will not estimate them jointly with fixed effects. This means that in general the fixed effect estimates may vary depending on `init.corrHLfit` when any form of REML correction is applied.

Correctly using `corrHLfit` for likelihood ratio tests of fixed effects may then be tricky. It is safe to perform full ML fits of all parameters (using `HLmethod="ML"`) for such tests (see Examples). The higher level function [fixedLRT](#) is a safe interface for likelihood ratio tests using some form of REML estimation in `corrHLfit`.

`attr(<fitted object>,"optimInfo")$lower` and `...$upper` gives the lower and upper bounds for optimization of correlation parameters. These are the default values if the user did not provide explicit values. For the adjacency model, the default values are the inverse of the maximum and minimum eigenvalues of the `adjMatrix`. For the Matérn model, the default values are not so easily summarized: they are intended to cover the range of values for which there is statistical information to distinguish among them.

## Value

The return value of an `HLCor` call, with additional attributes. The `HLCor` call is evaluated at the estimated correlation parameter values. These values are included in the return object as its `$corrPars` member. The attributes added by `corrHLfit` include the original call of the function (which can be retrieved by `getCall(<fitted object>)`), and information about the optimization call within `corrHLfit`.

## See Also

See more examples on data set [Loaloo](#). See [fixedLRT](#) for likelihood ratio tests.



## Examples

```
# Example with an adjacency matrix (autoregressive model):
# see 'adjacency' documentation page

#### Examples with Matérn correlations
## A likelihood ratio test based on the ML fits of a full and of a null model.
if (spaMM.getOption("example_maxtime")>1.4) {
  data("blackcap")
  fullfit <- corrHLfit(migStatus ~ means+ Matern(1|latitude+longitude),data=blackcap,
                      HLmethod="ML")
  summary(fullfit)
  nullfit <- corrHLfit(migStatus ~ 1 + Matern(1|latitude+longitude),data=blackcap,
                      HLmethod="ML",init.corrHLfit=list(phi=1e-6))
  summary(nullfit)
  ## p-value:
  1-pchisq(2*(logLik(fullfit)-logLik(nullfit)),df=1)
}

## see data set Loaloe for additional examples
```

---

 corrMatrix

*Using a corrMatrix argument*


---

## Description

corrMatrix is an argument of HLCor, of calls dist or matrix, with is used if the model formula contains a term of the form corrMatrix(1|<...>). It describes a correlation matrix, possibly as a dist object. A covariance matrix can actually be passed through this argument, but then it must be a full matrix, not a dist object. The way the rows and columns of the matrix are matched to the rows of the data depends on the nature of the grouping term <...>.

## Details

The simplest case is illustrated in the first two examples below: the grouping term is identical to a single variable which is present in the data, whose levels match the rownames of the corrMatrix. As illustrated by the second example, the order of the data does not matter in that case, because the factor levels are used to match the data rows to the appropriate row and columns of the corrMatrix. The corrMatrix may even contain rows (and columns) in excess of the levels of the grouping term, in which case these rows are ignored.

These convenient properties no longer hold when the grouping term is not a single variable from the data (third example below), or when its levels do not correspond to row names of the matrix. In these cases, (1) no attempt is made to match the data rows to the row and column names of the corrMatrix. Such attempt could succeed only if the user had given names to the matrix matching those that the the called function could create from the information in the data, in which case the user should find easier to specify a single variable that can be matched; (2) the order of data and corrMatrix matter; Internally, a single factor variable is constructed from all levels of the variables in the grouping term (i.e., from all levels of latitude and longitude, in the third example), with levels 1,2,3... that are matched to rows 1,2,3... of the corrMatrix. Thus the first row of the data is

always associated to the first row of the matrix; (3) further, the dimension of the matrix must match the number of levels implied by the grouping term. For example, one might consider the case of 14 response values but of correlations between only 7 levels of a random effect, with two responses for each level. Then the matrix must be of dimension  $7 \times 7$ .

### Examples

```
data("blackcap")
## Here we manually reconstruct the correlation matrix
## of the ML fit produced by corrHLfit:
MLcorMat <- MaternCorr(proxy::dist(blackcap[,c("latitude", "longitude")]),
  nu=0.6285603, rho=0.0544659)
blackcap$name <- as.factor(rownames(blackcap))
## (1) Single variable present in the data
HLCor(migStatus ~ means+ corrMatrix(1|name), data=blackcap,
  corrMatrix=MLcorMat, HLmethod="ML")
## (2) Same, permuted: still gives correct result
perm <- sample(14)
# Permuted matrix (with permuted names)
pmat <- as.matrix(MLcorMat)[perm, perm]
HLCor(migStatus ~ means+ corrMatrix(1|name), data=blackcap,
  corrMatrix=as.dist(pmat), HLmethod="ML")
## (3) Other grouping terms:
HLCor(migStatus ~ means+ corrMatrix(1|latitude+longitude), data=blackcap,
  corrMatrix=MLcorMat, HLmethod="ML")
```

---

covStruct

*Specifying correlation structures*

---

### Description

covStruct is a formal argument of HLCor, also handled by fitme and corrHLfit, that allows one to specify the correlation structure for different types of random effects, It is an alternative to other ad hoc formal arguments such as corrMatrix or adjMatrix.

It is a list of matrices with names specifying the type of matrix considered: covStruct=list(corrMatrix=<some matrix> or covStruct=list(adjMatrix=<some matrix>), where the “corrMatrix” or “adjMatrix” names are used to specify the type of information provided (accordingly, the names can be repeated: covStruct=list(corrMatrix=<. >, corrMatrix=<. >)).

The covariance structure of a corrMatrix(1|<grouping factor>) formula term can be specified in two ways (see Examples): either by a correlation matrix factor (covStruct=list(corrMatrix=<some matrix>)), or by a precision matrix factor  $\mathbf{Q}$  such that the covariance factor is  $\lambda \mathbf{Q}^{-1}$ , using the type name “precision”: covStruct=list(precision=<some matrix>). In this case, an algorithm efficient for **sparse** precision matrices is used to fit the model.

### Details

covStruct can also be specified as a list with an optional “types” attribute, e.g. structure(list(<some matrix>, types="corrMatrix"))

**See Also**

[pedigree](#) for a type of applications where declaring a precision matrix is useful.

**Examples**

```
## Not run:
data("blackcap")
# a 'dist' object can be used to specify a corrMatrix:
MLdistMat <- MaternCorr(proxy::dist(blackcap[,c("latitude","longitude")] ),
  nu=0.6285603,rho=0.0544659) # a 'dist' object!
blackcap$name <- as.factor(rownames(blackcap))
HLCor(migStatus ~ means + corrMatrix(1|name), data=blackcap,
  corrMatrix=MLdistMat, HLmethod="ML")
#### Same result by different input and algorithm:
m <- as.matrix(MLdistMat) ## leaves 0 on the diagonal!
diag(m) <- 1 ## so that m is true correlation matrix
prec_mat <- solve(m) ## precision factor matrix
HLCor(migStatus ~ means + corrMatrix(1|name), data=blackcap,
  covStruc=list(precision=prec_mat), HLmethod="ML")

## End(Not run)
```

---

designL.from.Corr

*Computation of “square root” of correlation matrix*


---

**Description**

This function is not usually directly called by users, but arguments may be passed to it through higher-level calls (see Examples). For given correlation matrix  $\mathbf{C}$ , it computes a “design matrix”  $\mathbf{L}$  such that  $\mathbf{C} = \mathbf{L} * \mathbf{t}(\mathbf{L})$ .  $\mathbf{t}(\text{chol}(\cdot))$  (Cholesky factorization) is a fast method for this computation, but it is not robust numerically and may even return an error, in which cases more robust methods (eigen or svd) are used. Matrix roots are not unique (for example, they are lower triangular for  $\mathbf{t}(\text{chol}(\cdot))$ , and symmetric for  $\text{svd}(\cdot)$ ). As matrix roots are used to simulate samples under the fitted model (in particular in the parametric bootstrap implemented in `fixedLRT`), this implies that for given seed of random numbers, these samples will differ with these different methods (although their distribution should be identical).

**Usage**

```
designL.from.Corr(m = NULL, symSVD = NULL, try.chol = TRUE, try.eigen = FALSE,
  threshold = 1e-06, SVDfix = 1/10)
```

**Arguments**

**m** The matrix which ‘root’ is to be computed. This argument is ignored if `symSVD` is provided.

symSVD	A list representing the symmetric singular value decomposition of the matrix which 'root' is to be computed. Must have elements \$u, a matrix of eigenvectors, and \$d, a vector of eigenvalues.
try.chol	If try.chol=TRUE, the Cholesky factorization will be tried.
try.eigen	The default behavior is to try chol, and use svd if chol fails. If try.eigen=TRUE, the eigen factorization will be tried before svd. eigen is a compromise between speed and accuracy, but in our experience it may *hang* so by default it is not tried.
threshold	A correction threshold for low eigenvalues is the case and eigensystem or singular-value decomposition are used.
SVDfix	A solution to failures of svd: see Details.

### Details

The function may call `svd`, for singular value decomposition (SVD) of a matrix  $\mathbf{M}$ . `svd` may return “error code 1 from Lapack routine 'dgesdd'” (cf. unhelpful discussions on R forums). This can be circumvented by computing the SVD of  $(1 - x)\mathbf{I} + x\mathbf{M}$  and deducing the singular values of  $\mathbf{M}$  in a trivial way. The  $x$  value to be used in this fix is provided by the `SVDfix` argument.

svd errors have occurred for correlation matrices that were close to the identity matrix except for a few large non-diagonal elements. Such matrices may occur in particular for the Matérn correlation model with low  $\nu$ , high  $\rho$ , and if some samples are spatially close. Then, an alternative fix to the svd problem may be to restrict the  $\nu$  and/or  $\rho$  ranges, using the lower and upper arguments of `corrHLfit`, although one should make sure that this has no bearing on the inferences.

### Value

The “square root of the input matrix”. Its rows and columns are labelled according to the columns of the original matrix.

### Examples

```
## Not run:
## try.chol argument passed to designL.from.Corr
## through the '...' argument of higher-level functions
## such as HLCor, corrHLfit, fixedLRT:
data("scotlip")
HLCor(cases~I(prop.ag/10) +adjacency(1|gridcode)+offset(log(expec)),
      ranPars=list(rho=0.174),adjMatrix=Nmatrix,family=poisson(),
      data=scotlip,try.chol=FALSE)

## End(Not run)
```

**Description**

logLik extracts the log-likelihood (exact or approximated). get\_any\_IC computes model selection/information criteria such as AIC. See Details for more information about these criteria. An AIC method is also defined for HLfit objects, which is equivalent to get\_any\_IC. dev\_resids returns a vector of squared (unscaled) deviance residuals (the summands in McCullagh and Nelder 1989, p. 34). deviance returns the sum of squares of these (unscaled) deviance residuals, that is (consistently with stats::deviance) the unscaled deviance. fitted extracts fitted values (see [fitted.values](#)). residuals extracts residuals of the fit. fixef extracts the fixed effects coefficients,  $\beta$ . ranef extracts the predicted random effects,  $L\mathbf{v}$  (default since version 1.12.0), or  $\mathbf{u}$  (see Details in [HLfit](#) for definitions), print.ranef controls their printing. vcov returns the variance-covariance matrix of the fixed-effects coefficients. Corr returns a correlation matrix of random effects. getDistMat returns a distance matrix for a Matérn correlation model. get\_ZALMatrix returns the design matrix for the random effects  $\mathbf{v}$ . get\_RLRTSim\_args returns a list of arguments suitable for calls to RLRTSim::RLRTSim() get\_boot\_response returns the response vector of a given bootstrap replicate in a fixedLRT run (see Details).

**Usage**

```
## S3 method for class 'HLfit'
logLik(object,which,...)
## S3 method for class 'HLfit'
fitted(object,...)
## S3 method for class 'HLfit'
fixef(object,...)
## S3 method for class 'HLfit'
ranef(object, type="correlated", ...)
## S3 method for class 'ranef'
print(x, max.print=40L, ...)
## S3 method for class 'HLfit'
vcov(object,...)
## S3 method for class 'HLfit'
deviance(object,...)
## S3 method for class 'HLfit'
residuals(object, type = c("deviance", "pearson", "response"), ...)
getDistMat(object, scaled=FALSE, which=1L)
Corr(object,...)
dev_resids(object,...)
get_any_IC(object, ..., verbose=interactive())
get_RLRTSim_args(object,...)
get_ZALMatrix(object, as_matrix)
get_boot_response(object, replicate)
## S3 method for class 'HLfit'
AIC(object, ..., k, verbose=interactive())
```

## Arguments

object	A object of class <code>HLfit</code> , as returned by the fitting functions in <code>spaMM</code> .
type	For <code>ranef</code> , use <code>type="correlated"</code> (default) to display the correlated random effects ( <b>L</b> <i>v</i> ), whether in a spatial model, or a random-coefficient model. Use <code>type="uncorrelated"</code> to pretty-print the elements of the <code>&lt;object&gt;\$ranef</code> vector ( <b>u</b> ). For <code>residuals</code> , the type of residuals which should be returned. The alternatives are: "deviance" (default), "pearson", and "response".
which	For <code>logLik</code> , the name of the element of the APHLs list to return. The default depends on the fitting method. In particular, if it was REML or one of its variants, the function returns the log restricted likelihood (exact or approximated). For <code>getDistMat</code> , as integer, to select a random effect from several for which a distance matrix may be constructed.
scaled	If <code>FALSE</code> , the function ignores the scale parameter <i>rho</i> and returns unscaled distance.
as_matrix	Deprecated.
verbose	Whether to print the model selection criteria or not.
x	For <code>print.ranef</code> : the return value of <code>ranef.HLfit</code> .
max.print	Controls options("max.print") locally.
k	For AIC: unused by <code>HLfit</code> method, but included to conform to the generic.
replicate	integer; to get the response vector of the given bootstrap replicate.
...	Other arguments that may be needed by some method.

## Details

See [residuals.glm](#) for more information about the types of residuals.

## Value

Return values are numeric (for `logLik`), vectors (most cases), matrices (for `vcov`), matrices or dist objects (for `getDistMat`). `ranef` returns a list of vectors or matrices (the latter for random-coefficient terms).

`Corr` returns a list, for the different random effect terms, of unconditional correlation matrix of the random effects "v" (as defined in see Details of `HLfit`), if there is any such matrix. Otherwise the list elements

`get_any_IC` computes, optionally prints, and returns invisibly the following quantities. The **conditional AIC** (Vaida and Blanchard 2005) is a relative measure of quality of prediction of new realizations of a mixed model, conditional on the realized values of the random effects. It involves the conditional likelihood, and degrees of freedom for (i) estimated residual error parameters and (ii) the overall linear predictor characterized by the **Effective degrees of freedom** already discussed by previous authors including Lee and Nelder (2001), which gave a general formula for it in HGLMs. Both a plug-in "asymptotic" estimate of the conditional AIC and of this effective df are returned by `get_any_IC`. Note that these may be biased estimates of conditional AIC and effective df, and that more refined formulas are discussed in the literature (e.g. Overholser and Xu 2014), some of which may be implemented in future versions of `get_any_IC`. Lee et al. (2006) and Ha et al (2007) defined a corrected AIC [i.e.,  $AIC(D^*)$  in their eq. 7] which is here interpreted as the conditional

AIC. Also returned are the **marginal AIC** (Akaike's classical AIC), and a focussed AIC for dispersion parameters (**dispersion AIC**) discussed by Ha et al (2007; eq.10). This diversity of criteria should encourage users to think twice before applying model selection automatically, which is no better although more fashionable than misuses of simple null hypothesis testing. Also, alternative procedures for model choice can be considered (e.g. Cox and Donnelly, 2011, p. 130-131).

`get_boot_response` has no `resp_testfn` argument so it cannot always reproduce the samples simulated by `spaMM_boot`. Note however that the latter function, with `simuland=function(y) {y}`, will return the samples simulated given the condition defined by `resp_testfn`.

`get_RLRTSim_args` extracts a list of arguments suitable for a call to `LRTSim::RLRTSim()` for a small-sample test of the presence of a random effect by an efficient simulation procedure. The test can be run by  
`do.call("RLRTSim",<get_RLRTSim_args return value>).`

## References

- Cox, D. R. and Donnelly C. A. (2011) Principles of Applied Statistics. Cambridge Univ. Press.
- McCullagh, P. and Nelder J. A. (1989) Generalized linear models. Second ed. Chapman & Hall: London.
- Ha, I. D., Lee, Y. and MacKenzie, G. (2007) Model selection for multi-component frailty models. *Statistics in Medicine* 26: 4790-4807.
- Lee, Y., Nelder, J. A. (2001) Hierarchical generalised linear models: A synthesis of generalised linear models, random-effect models and structured dispersions. *Biometrika* 88, 987-1006.
- Lee, Y., Nelder, J. A. and Pawitan, Y. (2006) Generalized linear models with random effects: unified analysis via h-likelihood. Chapman & Hall: London.
- Overholser R., and Xu R. (2104) Effective degrees of freedom and its application to conditional AIC for linear mixed-effects models with correlated error structures. *J. Multivariate Anal.* 132: 160-170.
- Vaida, F., and Blanchard, S. (2005) Conditional Akaike information for mixed-effects models. *Biometrika* 92, 351-370.

## Examples

```
data("wafers")
m1 <- HLfit(y ~X1+X2+(1|batch),
           resid.model = ~ 1 ,data=wafers,HLmethod="ML")
get_any_IC(m1)
fixef(m1)
vcov(m1)
ranef(m1)
```

fitme

*Fitting function for fixed- and mixed-effect models with GLM response.*

## Description

This is a common interface for fitting most models that spaMM can fit, from linear models to mixed models with non-gaussian random effects, therefore substituting to `corrHLfit`, `HLCor` and `HLfit`. By default, it uses ML rather than REML (differing in this respect from the other fitting functions). It may use “outer optimization”, i.e., generic optimization methods for estimating all dispersion parameters, rather than the iterative methods method in `HLfit`. The results of REML fits of non-gaussian mixed models by these different methods may (generally slightly) differ. Outer optimization should generally be faster than the alternative algorithms for large data sets when the residual variance model is a single constant term (no structured dispersion). For mixed models, `fitme` by default tries to select the fastest method when both can be applied, but precise decision criteria are subject to change in the future. `corrHLfit` (with non-default arguments to control the optimization method most suitable to a particular problem) may be used to ensure better consistency over successive versions of spaMM.

## Usage

```
fitme(formula, data, family = gaussian(), init = list(), fixed = list(),
      lower = list(), upper = list(), resid.model = ~1, init.HLfit = list(),
      control = list(), control.dist = list(), method = "ML",
      HLmethod = method, processed = NULL, nb_cores = NULL, objective = NULL,
      ...)
```

## Arguments

<code>formula</code>	Either a linear model <code>formula</code> (as handled by various fitting functions) or a predictor, i.e. a formula with attributes (see <a href="#">Predictor</a> and examples below). See Details in <a href="#">spaMM</a> for allowed terms in the formula.
<code>data</code>	A data frame containing the variables in the response and the model formula.
<code>family</code>	Either a <code>family</code> or a <code>multi</code> value.
<code>init</code>	An optional list of initial values for correlation and/or dispersion parameters and/or response family parameters, e.g. <code>list(rho=1, nu=1, lambda=1, phi=1)</code> where <code>rho</code> and <code>nu</code> are parameters of the Matérn family (see <a href="#">Matern</a> ), and <code>lambda</code> and <code>phi</code> are dispersion parameters (see Details in <a href="#">spaMM</a> for the meaning of these parameters). All are optional, but giving values for a dispersion parameter changes the ways it is estimated (see Details and Examples). <code>rho</code> may be a vector (see <a href="#">make_scaled_dist</a> ) and, in that case, it is possible that some or all of its elements are NA, for which <code>corrHLfit</code> substitute automatically determined values.
<code>fixed</code>	A list similar to <code>init</code> , but specifying fixed values of the parameters not estimated. See <a href="#">fixed</a> for further information.
<code>lower</code>	An optional list of values of parameters specified through <code>init.corrHLfit</code> , used as lower values in calls to <code>optim</code> . See Details for default values.



upper	Same as lower, but upper values.
resid.model	See identically named <a href="#">HLfit</a> argument.
init.HLfit	See identically named <a href="#">HLfit</a> argument.
control	A list of control parameters, with two possible elements: <ul style="list-style-type: none"> <li>• \$nloptr, itself a list of control parameters to be copied in the opts argument of <a href="#">nloptr</a>. Default controls are given by <code>spaMM.getOption('nloptr')</code></li> <li>• \$refit, a boolean, or a list of booleans with possible elements \$phi and \$lambda. If either element is set to TRUE, then the corresponding parameters are refitted by the internal <a href="#">HLfit</a> methods (see <a href="#">Details</a>). If \$refit is TRUE, all phi and lambda parameters are refit. By default none of them is refitted, but this default may change in the future.</li> </ul>
control.dist method, HLmethod	See <a href="#">control.dist</a> in <a href="#">HLCor</a> "ML" or "REML". "ML" is the default, in contrast to "REML" for the <a href="#">HLmethod</a> argument of other fitting functions. Other possible values of <a href="#">HLfit</a> 's <a href="#">HLmethod</a> argument are handled and should give results close to the other fitting methods with the same <a href="#">HLmethod</a> argument, the most notable exception being PQL/L for binary response data, as <a href="#">fitme</a> does not take into account the specific way leverages are computed in PQL/L.
nb_cores	<b>Not yet operative</b> , only for development purposes. Number of cores to use for parallel computations.
processed	For programming purposes, not documented.
objective	For development purpose, not documented.
...	Optional arguments passed to <a href="#">HLCor</a> , <a href="#">HLfit</a> or <a href="#">designL.from.Corr</a> , for example the <a href="#">distMatrix</a> argument of <a href="#">HLCor</a> .

## Details

For phi, lambda, and ranCoefs, [fitme](#) may or may not use the internal fitting methods of [HLfit](#). The latter methods are well suited for structured dispersion models, but require the computation of the so-called leverages, which can be slow for large datasets. Therefore, [fitme](#) tends to outer-optimize by default for large datasets, unless there is a non-trivial `resid.model`.

However, the internal fitting methods of [HLfit](#) also provide some more information such as the "cond. SE" (about which see warning in [Details](#) of [HLfit](#)). To force the evaluation of such information after an outer-optimization by a [fitme](#) call, use the `control$refit` argument (see [Example](#)). Alternatively (and possibly of limited use), one can force inner-optimization of lambda for a given random effect, or of phi, by setting it to NaN in `init`.

## Value

The return value of an [HLCor](#) or an [HLfit](#) call, with additional attributes. The [HLCor](#) call is evaluated at the estimated correlation parameter values. These values are included in the return object as its `$corrPars` member. The attributes added by [fitme](#) include the original call of the function (which can be retrived by `getCall(<fitted object>)`), and information about the optimization call within [fitme](#).

## Examples

```
## Contrasting different optimization methods:
# We simulate Gamma deviates with mean mu=3 and variance=2,
# ie. phi= var/mu^2= 2/9 in the (mu, phi) parametrization of a Gamma
# GLM; and shape=9/2, scale=2/3 in the parametrisation of rgamma().
# Note that phi is not equivalent to scale:
# shape = 1/phi and scale = mu*phi.
set.seed(123)
gr <- data.frame(y=rgamma(100,shape=9/2,scale=2/3))
# Here fitme uses HLfit methods which provide cond. SE for phi by default:
fitme(y~1,data=gr,family=Gamma(log))
# To force outer optimization of phi, use the init argument:
fitme(y~1,data=gr,family=Gamma(log),init=list(phi=1))
# To obtain cond. SE for phi after outer optimization, use the 'refit' control:
fitme(y~1,data=gr,family=Gamma(log),,init=list(phi=1),
      control=list(refit=list(phi=TRUE))) ## or ...refit=TRUE...

## see help("COMPOisson"), help("negbin"), help("Loaloo"), etc., for further examples.
```

---

fixed

*Fixing some parameters*

---

## Description

The fitting functions allow some parameters to be fixed rather than estimated, by way of `etaFix` (linear predictor coefficients) for all fitting functions, of the `fixed` argument for all (co-)variance parameters in `fitme`, of `ranFix` (random-effect and residual dispersion parameters) in `HLfit` and `corrHLfit`, and of `ranPars` in `HLCor`. The diversity of names may be confusing, but keep in mind that `ranFix` allows one to fix parameters that `HLfit` and `corrHLfit` would otherwise estimate, while `ranPars` can be used to set required parameters for `HLCor`, which it would otherwise be unable to estimate (e.g., Matern correlation parameters).

Each of these arguments is a list.

`ranFix` elements taken into account by `HLfit` include `phi` (variance of residual error, for gaussian and Gamma HGLMs), `lambda` (random-effect variances), and `ranCoefs` (variance-correlation information for random-coefficient terms). To assign values for only some random-effect terms, `lambda` and `ranCoefs` can be incomplete, e.g. `lambda=c(NA,1)` or `lambda=c("2"=1)` (note the name) to assign a value only to the variance of the second of two random effects. `ranCoefs` is a list of numeric vectors, each numeric vector specifying the variance and correlation parameters for a random-coefficient term. This input matches the printed summary of a fit. The elements must be given in the order of the `lower.tri` of a covariance matrix, as shown e.g. by

```
m2 <- matrix(NA, ncol=2,nrow=2); m2[lower.tri(m2,diag=TRUE)] <- seq(3); m2.
```

For example, to assign variances values 3 and 7, and correlation value -0.05, to a second random effect, one can use `ranCoefs=list("2"=c(3,-0.05,7))` (note the name).

Additional `ranFix` elements are taken into account by `corrHLfit`, as follows. For the Matern model, these are the correlation parameters `rho` (scale parameter(s)), `nu` (smoothness parameter), and (optional) `Nugget` (see [Matern](#)). The `rho` parameter can itself be a vector with different

values for different geographic coordinates. For the adjacency model, the only correlation parameter is a scalar  $\rho$  (see [adjacency](#)). For the AR1 model, the only correlation parameter is a scalar  $\text{AR}\phi$  (see [AR1](#)).

`ranPars` elements taken into account by `HLCor` include all the above `ranFix` elements.

The only `etaFix` element considered here is `beta`, which should be a vector of (a subset of) the coefficients ( $\beta$ ) of the fixed effects, with names as shown in a fit without such given values. In contrast to an offset specification, it affects by default the REML correction for estimation of dispersion parameters, which depends only on which  $\beta$  coefficients are estimated rather than given. This default behaviour will be overridden whenever a non-null REML formula is provided to `HLfit` or the other fitting functions (see [Example](#)).

fixed elements taken into account by `fitme` include all the above `ranFix` and `etaFix` elements.

## Details

`REMLformula` is the preferred way to control non-standard REML fits. Alternatively, with a non-NULL `etaFix$beta`, REML can also be performed as if all  $\beta$  coefficients were estimated, by adding attribute `keepInREML=TRUE` to `etaFix$beta`. Using an `REMLformula` will override such a specification.

## Examples

```
## Not run:
data("wafers")
# Fixing random-coefficient parameters:
HLfit(y~X1+(X2|batch), data=wafers, ranFix=list(ranCoefs=list("1"=c(2760, -0.1, 1844))))
# fixing coefficients of the linear predictor, but with REML as if they were not fixed:
HLfit(y ~X1+X2+X1*X3+X2*X3+I(X2^2)+(1|batch), data=wafers, family=Gamma(log),
      etaFix=list(beta=c("(Intercept)"=5.61208, X1=0.08818, X2=-0.21163, X3=-0.13948,
                        "I(X2^2)"=-0.10378, "X1:X3"=-0.08987, "X2:X3"=-0.08779)),
      REMLformula=y ~X1+X2+X1*X3+X2*X3+I(X2^2)+(1|batch))

data("Loaloe")
# Fixing some Matern correlation parameters, in corrHLfit:
corrHLfit(cbind(npos,ntot-npos)~elev1+elev2+elev3+elev4+maxNDVI1+seNDVI
          +Matern(1|longitude+latitude),
          data=Loaloe,family=binomial(),ranFix=list(nu=0.5,Nugget=2/7))
# Fixing all mandatory Matern correlation parameters, in HLCor:
HLCor(cbind(npos,ntot-npos)~elev1+elev2+elev3+elev4+maxNDVI1+seNDVI
      +Matern(1|longitude+latitude),
      data=Loaloe,family=binomial(),ranPars=list(nu=0.5,rho=0.7))

## End(Not run)
```

## Description

fixedLRT performs a likelihood ratio (LR) test between two models, the “full” and the “null” models, currently differing only in their fixed effects. Parametric bootstrap p-values can be computed, either using the raw bootstrap distribution of the likelihood ratio, or a bootstrap estimate of the Bartlett correction of the LR statistic. This function differ from LRT in its arguments (model fits for LRT, but all arguments required to fit the models for fixedLRT), and in the format of its return value.

## Usage

```
fixedLRT(null.formula, formula, data,
         method, HLMETHOD = method, REMLformula=NULL,
         boot.repl=0, control=list(), control.boot=list(),
         fittingFunction, nb_cores=NULL, ...)
```

## Arguments

null.formula	Either a formula (as in glm) or a predictor (see Predictor) for the null model.
formula	Either a formula or a predictor for the full model.
data	A data frame containing the variables in the model.
method	A method to fit the full and null models. See <code>HLfit</code> 's <code>HLMETHOD</code> argument for background information about such methods. The two most meaningful values of <code>method</code> in <code>fixedLRT</code> calls are: 'ML' for an LRT based on ML fits (generally recommended); and 'PQL/L' for an LRT based on PQL/L fits (recommended for spatial binary data). Also feasible, but more tricky, and not really recommended (see Rousset and Ferdy, 2014), is 'REML'. This will perform an LRT based on two REML fits of the data, *both* of which use the same conditional (or “restricted”) likelihood of residuals for estimating dispersion parameters $\lambda$ and $\phi$ (see <code>REMLformula</code> argument). Further, REML will not be effective on a given dispersion parameter if a non-trivial <code>init.corrHLfit</code> value is provided for this parameter.
HLMETHOD	Kept for back-compatibility. Same as <code>method</code> , but will work only for <code>fittingFunction=corrHLfit</code> .
REMLformula	a formula specifying the fixed effects which design matrix is used in the REML correction for the estimation of dispersion parameters, if these are estimated by REML. This formula is by default that for the *full* model.
boot.repl	the number of bootstrap replicates.
control	A set of control parameters for the fits of the data, mostly for development purposes. However, if an initial value is provided for a dispersion parameter, a better one may be sought if further <code>control=list(prefits=TRUE)</code> (the effect appears small, however).
control.boot	Same as <code>control</code> , but for the fits of the bootstrap replicates. Again, the option <code>control.boot=list(prefits=TRUE)</code> may yield a small improvement in the fits, at the expense of more computation time.
fittingFunction	Character string giving the function used to fit each model: either “ <code>corrHLfit</code> ” or “ <code>fitme</code> ”. Default is “ <code>corrHLfit</code> ” for small data sets (fewer than 300 observations), and “ <code>fitme</code> ” otherwise, but this may change in future versions.

nb_cores	Number of cores to use for parallel computation of bootstrap. The default is <code>spaMM.getOption("nb_cores")</code> , and 1 if the latter is NULL. <code>nb_cores=1</code> prevents the use of parallelisation procedures.
...	Further arguments passed to or from other methods; in particular, additional arguments passed to <code>corrHLfit</code> , including mandatory ones such as <code>data</code> and those ultimately passed to <code>designL.from.Corr</code> . With respect to the latter, note that <code>try.chol</code> affects the simulation of samples for the parametric bootstrap, and although ultimate differences in performance may be small, <code>try.chol=FALSE</code> may be slightly better.

### Details

Comparison of REML fits is a priori not suitable for performing likelihood ratio tests. Nevertheless, it is possible to contrive them for testing purposes (Wehler & Thompson 1997). This function generalizes some of Wehler & Thompson's methods to GLMMs.

See Details in [LRT](#) for details of the bootstrap procedures.

### Value

An object of class `fixedLRT`, actually a list with as-yet unstable format, but here with typical elements (depending on the options)

<code>fullfit</code>	the <code>HLfit</code> object for the full model;
<code>nullfit</code>	the <code>HLfit</code> object for the null model;
<code>LRTori</code>	A likelihood ratio chi-square statistic
<code>LRTprof</code>	Another likelihood ratio chi-square statistic, after a profiling step, if any.
<code>df</code>	the number of degrees of freedom of the test.
<code>trace.info</code>	Information on various steps of the computation.

and, if a bootstrap was performed, the additional elements described in [LRT](#).

### References

Rousset F., Ferdy, J.-B. (2014) Testing environmental and genetic effects in the presence of spatial autocorrelation. *Ecography*, 37: 781-790. <http://dx.doi.org/10.1111/ecog.00566>

Welham, S. J., and Thompson, R. (1997) Likelihood ratio tests for fixed model terms using residual maximum likelihood, *J. R. Stat. Soc. B* 59, 701-714.

### See Also

See also [corrHLfit](#) and [LRT](#).

### Examples

```
if (spaMM.getOption("example_maxtime")>1.6) {
  data("blackcap")
  ## result comparable to the corrHLfit examples based on blackcap
  fixedLRT(null.formula=migStatus ~ 1 + Matern(1|latitude+longitude),
```

```

        formula=migStatus ~ means + Matern(1|latitude+longitude),
        HLmethod='ML',data=blackcap)
    }
    if (spaMM.getOption("example_maxtime")>186) {
      ## longer version with bootstrap
      fixedLRT(null.formula=migStatus ~ 1 + Matern(1|latitude+longitude),
        formula=migStatus ~ means + Matern(1|latitude+longitude),
        HLmethod='ML',data=blackcap, boot.repl=100)
    }

```

---

 freight

*Freight dataset*


---

### Description

A set of data on airfreight breakage. Data are given on 10 air shipments, each carrying 1000 ampules of some substance. For each shipment, the number of ampules found broken upon arrival, and the number of times the shipments were transferred from one aircraft to another, are recorded.

### Usage

```
data("freight")
```

### Format

The data frame includes 10 observations on the following variables:

**broken** number of ampules found broken upon arrival.

**transfers** number of times the shipments were transferred from one aircraft to another.

**id** Shipment identifier.

### Source

The data set is reported by Kutner et al. (2003) and used by Sellers and Shmueli (2010) to illustrate COMPOisson analyses.

### References

Kutner MH, Nachtsheim CJ, Neter J, Li W (2005, p. 35). Applied Linear Regression Models, Fourth Edition. McGraw-Hill.

Sellers KF, Shmueli G (2010) A Flexible Regression Model for Count Data. Ann. Appl. Stat. 4: 943–961

### Examples

```
## see ?COMPOisson for examples
```

---

get\_ranPars *More functions to extract various components of a fit*

---

### Description

get\_ranPars returns various subsets of random-effect parameters.

### Usage

```
get_ranPars(object, which=NULL, ...)
```

### Arguments

object	A object of class <code>HLfit</code> , as returned by the fitting functions in <code>spaMM</code> .
which	For <code>get_ranPars</code> , the only non-null value is <code>"corrPars"</code> , to return correlation parameters of random effects.
...	Other arguments that may be needed by some method.

### Value

`get_ranPars(., which="corrPars")` returns a (possibly nested) list of correlation parameters (or `NULL` if there is no such parameter). Top-level elements correspond to the different random effects. The list has `"type"` attribute having the same nested-list structure and describing whether and how the parameters were fitted: `"fix"` means they were fixed, not fitted; `"var"` means they were fitted by `HLfit`'s specific algorithms; `"outer"` means they were fitted by a generic optimization method.

### Examples

```
data("wafers")
m1 <- HLfit(y ~X1+X2+(1|batch),
            resid.model = ~ 1 ,data=wafers,HLmethod="ML")
get_ranPars(m1,which="corrPars") # NULL since no correlated random effect
```

---

good-practice *Clear and trustworthy formulas*

---

### Description

Base fitting functions in R will seek variables in the global environment (or more generally in the environment where a call to ``~`` was made, defining the model formula) if they are not in the data. This easily leads to errors (see example in the discussion of `update.HLfit`). Indeed Chambers (2008, p.221), after describing how the environment is defined, comments that “Where clear and trustworthy software is a priority, I would personally avoid such tricks. Ideally, all the variables in the model frame should come from an explicit, verifiable data source...”. Hence, the main fitting functions in `spaMM` depart from the sloppy practice. They strip the formula environment from any

variable, and seek all variables from the formula in the data frame given by their data argument. **One never needs to specify the data in the formula.** The variables defining the prior weights should also be in the data. Variables used in other arguments such as `ranFix` are looked up neither in the data nor in the formula environment, but in the calling environment as usual.

## References

Chambers J.M. (2008) Software for data analysis: Programming with R. Springer-Verlag New York

---

HLCor	<i>Fits a (spatially) correlated mixed model, for given correlation parameters</i>
-------	--

---

## Description

A convenient interface for `HLfit`, constructing the correlation matrix of random effects from the arguments, then estimating fixed effects and dispersion parameters using `HLfit`.

## Usage

```
HLCor(formula, data, family = gaussian(), ranPars = NULL, distMatrix,
       uniqueGeo = NULL, adjMatrix, corrMatrix, covStruct=NULL,
       verbose = c(trace=FALSE),
       control.dist = list(), ...)
```

## Arguments

formula	A predictor, i.e. a formula with attributes (see <a href="#">Predictor</a> ), or possibly simply a simple formula if an offset is not required.
ranPars	A list of values for correlation parameters (some of which are mandatory), and possibly also dispersion parameters (optional, but passed to <code>HLfit</code> if present). See <a href="#">ranPars</a> for further information.
data	The data frame to be analyzed.
family	A family object describing the distribution of the response variable. See <a href="#">HLfit</a> for further information.
distMatrix	A distance matrix between geographic locations, forwarded to <code>MaternCorr</code>
uniqueGeo	A matrix of non-redundant geographic locations. Such a matrix is typically constructed automatically from the data and the model formula, but otherwise could be useful if further the <code>rho</code> parameter is a vector with different values for different coordinates, in which case a scaled distance matrix has to be reconstructed from <code>uniqueGeo</code> and <code>rho</code> .
adjMatrix	An adjacency matrix, used if a random effect of the form $y \sim \text{adjacency}(1 \langle \text{location index} \rangle)$ is present. See <a href="#">adjacency</a> for further details.



<code>corrMatrix</code>	A matrix <b>C</b> used if a random effect term of the form <code>corrMatrix(1 &lt;stuff&gt;)</code> is present. This allows to analyze non-spatial model by giving for example a matrix of genetic correlations. Each row corresponds to levels of a variable <code>&lt;stuff&gt;</code> . The covariance matrix of the random effects for each level is then $\lambda\mathbf{C}$ , where as usual $\lambda$ denotes a variance factor for the random effects (if <b>C</b> is a correlation matrix, then $\lambda$ is the variance, but other cases are possible). See <a href="#">corrMatrix</a> for further details.
<code>covStruct</code>	An interface for specifying correlation structures for different types of random effect ( <code>corrMatrix</code> or <code>adjacency</code> ). See <a href="#">covStruct</a> for details.
<code>verbose</code>	A vector of booleans. <code>trace</code> controls various diagnostic (possibly messy) messages about the iterations.
<code>control.dist</code>	A list of arguments that control the computation of the distance argument of the correlation functions. Possible elements are <b>rho.mapping</b> a set of indices controlling which elements of the rho scale vector scales which dimension(s) of the space in which (spatial) correlation matrices of random effects are computed. See same argument in <a href="#">make_scaled_dist</a> for details and examples. <b>dist.method</b> method argument of <code>proxy::dist</code> function (by default, "Euclidean", but see <a href="#">make_scaled_dist</a> for other distances such as spherical ones.)
<code>...</code>	Further parameters passed to <code>HLfit</code> or to <code>designL.from.Corr</code> .

## Details

The correlation matrix for random effects can be specified by various combination of formula terms and other arguments (see Examples):

**Basic Matérn model** `Matern(1|<...>)`, using the spatial coordinates in `<...>`. This will construct a correlation matrix according to the Matérn correlation function (see [MaternCorr](#));

**Matérn model with given distance matrix** `Matern(1|<...>)` with `distMatrix`;

**Given correlation matrix** `corrMatrix(1|<...>)` with `corrMatrix` argument. See [corrMatrix](#) for further details.

**CAR model with given adjacency matrix** `adjacency(1|<...>)` with `adjMatrix`. See [adjacency](#) for further details;

**AR1 model** `AR1(1|<...>)` See [AR1](#) for further details.

All these models except `corrMatrix` have additional parameters that must be specified by the `ranPars` argument.

## Value

The return value of an `HLfit` call, with the following additional attributes:

`HLCorcall` the `HLCor` call  
`info.uniqueGeo` Unique geographic locations.

## See Also

[autoregressive](#) for additional examples, [MaternCorr](#), [HLfit](#), and [corrHLfit](#)

## Examples

```
# Example with an adjacency matrix (autoregressive model):
# see 'adjacency' documentation page

#### Matérn correlation using only the Matern() syntax
if (spaMM.getOption("example_maxtime")>0.8) {
  data("Loaloe")
  HLCor(cbind(npos,ntot-npos)~elev1+elev2+elev3+elev4+maxNDVI1+seNDVI
        +Matern(1|longitude+latitude),data=Loaloe,
        family=binomial(),ranPars=list(nu=0.5,rho=1/0.7))
}

#### Matérn correlation using a distMatrix
data("blackcap")
MLdistMat <- as.matrix(proxy::dist(blackcap[,c("latitude","longitude")]))
HLCor(migStatus ~ means+ Matern(1|latitude+longitude),data=blackcap,
      distMatrix=MLdistMat,HLmethod="ML",ranPars=list(nu=0.6285603,rho=0.0544659))
```

---

HLfit

*Fit mixed models with given correlation matrix*


---

## Description

This function fits GLMMs as well as some hierarchical generalized linear models (HGLM; Lee and Nelder 2001). HLfit fits both fixed effects parameters, and dispersion parameters i.e. the variance of the random effects (full covariance for random-coefficient models), and the variance of the residual error. The linear predictor is of the standard form  $\text{offset} + X\beta + Zb$ , where  $X$  is the design matrix of fixed effects and  $Z$  is a design matrix of random effects. Models are fitted by an iterative algorithm alternating estimation of fixed effects and of dispersion parameters. The residual dispersion may follow a “structured-dispersion model” modeling heteroscedasticity. Estimation of the latter parameters is performed by a form of fit of debiased residuals, which allows fitting a structured-dispersion model (Smyth et al. 2001). However, evaluation of the debiased residuals can be slow in particular for large datasets. For models without structured dispersion, it is then worth using the `fitme` function (or the `corrHLfit` function with non-default arguments). These functions can optimize the likelihood of HLfit fits for different given values of the dispersion parameters (“outer optimization”), thereby avoiding the need to estimate debiased residuals.

## Usage

```
HLfit(formula, data, family = gaussian(), rand.family = gaussian(),
      resid.model = ~1, resid.formula, REMLformula = NULL,
      verbose = c(trace = FALSE), HLmethod = "HL(1,1)", control.HLfit = list(),
      control.glm = list(), init.HLfit = list(), ranFix = list(),
      etaFix = list(), prior.weights = NULL, processed = NULL)
## see 'rand.family' argument for inverse.Gamma
```

**Arguments**

formula	A <a href="#">formula</a> ; or a predictor, i.e. a formula with attributes created by <a href="#">Predictor</a> , if design matrices for random effects have to be provided. See Details in <a href="#">spaMM</a> for allowed terms in the formula (except spatial ones).
data	A data frame containing the variables named in the model formula.
family	A family object describing the distribution of the response variable. See Details in <a href="#">spaMM</a> for handled families.
rand.family	A family object describing the distribution of the random effect, or a list of family objects for different random effects (see Examples). Possible options are <code>gaussian()</code> , <code>Gamma(log)</code> , <code>Gamma(identity)</code> (see Details), <code>Beta(logit)</code> , <code>inverse.Gamma(-1/mu)</code> , and <code>inverse.Gamma(log)</code> . For discussion of these alternatives see Lee and Nelder 2001 or Lee et al. 2006, p. 178-. Here the family gives the distribution of a random effect $u$ and the link gives $v$ as function of $u$ (see Details). If there are several random effects and only one family is given, this family holds for all random effects.
resid.model	<b>Either</b> a formula (without left-hand side) for the dispersion parameter $\phi$ of the residual error. A log link is assumed by default; <b>or</b> a list, with at most two possible elements if its formula involves only fixed effects:  <b>formula</b> model formula as in formula-only case, without left-hand side <b>family</b> Always Gamma, with by default a log link. <code>Gamma(identity)</code> can be tried but may fail because only the log link ensures that the fitted $\phi$ is positive.  and additional possible elements (all named as <code>fitme</code> arguments) if its formula involves random effects: see <a href="#">phiHGLM</a> .
resid.formula	Obsolete, for back-compatibility; will be deprecated. Same as formula in <code>resid.model</code> .
REMLformula	A model formula that allows the estimation of dispersion parameters, and computation of restricted likelihood ( <code>p_bv</code> ) under a model different from the predictor formula.  For example, if only random effects are included in <code>REMLformula</code> , an ML fit is performed and <code>p_bv</code> equals the marginal likelihood (or its approximation), <code>p_v</code> . This ML fit can be performed more simply by setting <code>HLmethod="ML"</code> and leaving <code>REMLformula</code> at its default NULL value.
verbose	A vector of booleans. <code>trace</code> controls various diagnostic messages (possibly messy, and in bad need of revision) about the iterations. <code>TRACE</code> is most useful to follow the progress of a long computation, particularly in <code>fitme</code> or <code>corrHLfit</code> calls.
HLmethod	Allowed values are "REML", "ML", "EQL-" and "EQL+" for all models; "PQL" (= "REPQL") and "PQL/L" for GLMMs only; and further values for those curious to experiment (see Details). <b>The default is REML</b> (standard REML for LMMs, an extended definition for other models). REML can be viewed as a fom of conditional inference, and non-standard conditionings can be called as "REML" with a non-standard REMLformula. See Details for further information.
control.HLfit	A list of parameters controlling the fitting algorithms.

`resid.family` allows one to change the link for modeling of residual variance  $\phi$ , which is "log" by default. The family is always Gamma, so the non-default possible values of `resid.family` are `Gamma(identity)` or `Gamma(inverse)`. Only the default value ensures that the fitted  $\phi$  is positive.

Controls for the fitting algorithms should be ignored in routine use. They are:

`conv.threshold` and `spaMM_tol`: `spaMM_tol` is a list of tolerance values, with elements `Xtol_rel` and `Xtol_abs` that define thresholds for relative and absolute changes in parameter values in iterative algorithms (used in tests of the form " $d(\text{param}) < Xtol\_rel * \text{param} + Xtol\_abs$ ", so that `Xtol_abs` is operative only for small parameter values). `conv.threshold` is the older way to control `Xtol_rel`. Default values are given by `spaMM.getOption("spaMM_tol")`;

`break_conv_logL`, a boolean specifying whether the iterative algorithm should terminate when log-likelihood appears to have converged (roughly, when its relative variation over on iteration is lower than  $1e-8$ ). Default is FALSE (convergence is then assessed on the parameter estimates rather than on log-likelihood).

`iter.mean.dispFix`, the number of iterations of the iterative algorithm for coefficients of the linear predictor, if no dispersion parameters are estimated by the iterative algorithm. Defaults to 200;

`iter.mean.dispVar`, the number of iterations of the iterative algorithm for coefficients of the linear predictor, if some dispersion parameter(s) is estimated by the iterative algorithm. Defaults to 50;

`max.iter`, the number of iterations of the iterative algorithm for joint estimation of dispersion parameters and of coefficients of the linear predictor. Defaults to 200. This is typically much more than necessary, unless there is little information to separately estimate  $\lambda$  and  $\phi$  parameters.

<code>control.glm</code>	List of parameters controlling GLM fits, passed to <code>glm.control</code> ; e.g. <code>control.glm=list(maxit=100)</code> . See <a href="#">glm.control</a> for further details.
<code>init.HLfit</code>	A list of initial values for the iterative algorithm, with possible elements of the list are <code>fixef</code> for fixed effect estimates (beta), <code>v_h</code> for random effects vector $\mathbf{v}$ in the linear predictor, <code>lambda</code> for the parameter determining the variance of random effects $u$ as drawn from the <code>rand.family</code> distribution <code>phi</code> for the residual variance. However, this argument can be ignored in routine use.
<code>ranFix</code>	A list of fixed values of random effect parameters. See <a href="#">ranFix</a> for further information.
<code>etaFix</code>	A list of given values of the coefficients of the linear predictor. See <a href="#">etaFix</a> for further information.
<code>prior.weights</code>	An optional vector of prior weights as in <a href="#">glm</a> . This fits the data to a model with residual variance <code>phi/prior.weights</code> , so that increasing the weights by a constant factor $f$ will yield (Intercept) estimates of <code>phi</code> also increased by $f$ (this effect cannot be generally achieved if a non-trivial <code>resid.formula</code> with log link is used). This is not necessarily the way prior weights are interpreted in widely used packages, but this is consistent with what <code>glm</code> fits.
<code>processed</code>	A list of preprocessed arguments, for programming purposes only (as in <code>corrHLfit</code> ).

## Details

**I. Fitting methods:** Many approximations for likelihood have been defined to fit mixed models (e.g. Noh and Lee (2007) for some overview), and this function implements several of them, and some additional ones. In particular, PQL as originally defined by Breslow and Clayton (1993) uses REML to estimate dispersion parameters, but this function allows one to use an ML variant of PQL. Moreover, it allows some non-standard specification of the model formula that determines the conditional distribution used in REML.

EQL stands for the EQL method of Lee and Nelder (2001). The '+' version includes the  $d v / d \tau$  correction described p. 997 of that paper, and the '-' version ignores it. PQL can be seen as the version of EQL- for GLMMs. It estimates fixed effects by maximizing h-likelihood and dispersion parameters by an approximation of REML, i.e. by maximization of an approximation of restricted likelihood. PQL/L is PQL without the leverage corrections that define REML estimation of random-effect parameters. Thus, it estimates dispersion parameters by an approximation of marginal likelihood.

HLmethod also accepts values of the form "HL(<...>)", "ML(<...>)" and "RE(<...>)", e.g. HLmethod="RE(1,1)", which allow a more direct specification of the approximations used. HL and RE are equivalent (both imply an REML correction). The first '1' means that a first order Laplace approximation to the likelihood is used to estimate fixed effects (a '0' would instead mean that the h likelihood is used as the objective function). The second '1' means that a first order Laplace approximation to the likelihood or restricted likelihood is used to estimate dispersion parameters, this approximation including the  $dv/d \tau$  term specifically discussed by Lee & Nelder 2001, p. 997 (a '0' would instead mean that these terms are ignored).

It is possible to enforce the EQL approximation for estimation of dispersion parameter (i.e., Lee and Nelder's (2001) method) by adding a third index with value 0. "EQL+" is thus "HL(0,1,0)", while "EQL-" is "HL(0,0,0)". "PQL" is EQL- for GLMMs. "REML" is "HL(1,1)". "ML" is "ML(1,1)".

Some of these distinctions make sense for GLMs, and glm methods use approximations, which make a difference for Gamma GLMs. This means in particular that, (as stated in logLik) the logLik of a Gamma GLM fit by glm differs from the exact likelihood. Further, the dispersion estimate returned by summary.glm differs from the one implied by logLik, because summary.glm uses Pearson residuals instead of deviance residuals, and no HLmethod tries to reproduce this behaviour. logLik gives the approximation returned by an "ML(0,0,0)" fit. The dispersion estimate returned by an "HL(. , . , 0)" fit matches what can be computed from residual deviance and residual degrees of freedom of a glm fit, but this is not the estimate displayed by summary.glm. With a log link, the fixed effect estimates are unaffected by these distinctions.

**II. Random effects** are constructed in several steps. first, a vector  $\mathbf{u}$  of independent and identically distributed (iid) random effects is drawn from some distribution; second, a transformation  $\mathbf{v}=\mathbf{f}(\mathbf{u})$  is applied to each element (this defines  $\mathbf{v}$  which elements are still iid); third, correlated random effects are obtained as  $\mathbf{L}\mathbf{v}$  where  $\mathbf{L}$  is the "square root" of a correlation matrix (this may be meaningful only for Gaussian random effects). Coefficients in a random-coefficient model correspond to  $\mathbf{L}\mathbf{v}$ . Finally, a matrix  $\mathbf{Z}$  (or sometimes  $\mathbf{Z}\mathbf{A}$ , see Predictor) allows to specify how the correlated random effects affect the response values. In particular,  $\mathbf{Z}$  is the identity matrix if there is a single observation (response) for each location, but otherwise its elements  $z_{ji}$  are 1 for the  $j$ th observation in the  $i$ th location. The design matrix for  $\mathbf{v}$  is then of the form  $\mathbf{Z}\mathbf{L}$ .

The specification of the random effects  $\mathbf{u}$  and  $\mathbf{v}$  handles the following cases: **Gaussian** with zero mean, unit variance, and identity link; **Beta**-distributed, where  $u \sim B(1/(2\lambda), 1/(2\lambda))$  with mean=1/2, and var=  $\lambda/[4(1+\lambda)]$ ; and with logit link  $v=\text{logit}(u)$ ; **Gamma**-distributed random effects, where

$u$  Gamma(shape=1+1/ $\lambda$ ,scale=1/ $\lambda$ ): see [Gamma](#) for allowed links and further details; and **Inverse-Gamma**-distributed random effects, where  $u$  inverse-Gamma(shape=1+1/ $\lambda$ ,rate=1/ $\lambda$ ): see [inverse.Gamma](#) for allowed links and further details.

**III. The standard errors** reported may sometimes be misleading. For each set of parameters among  $\beta$ ,  $\lambda$ , and  $\phi$  parameters these are computed assuming that the other parameters are known without error. This is why they are labelled Cond. SE (conditional standard error). This is most uninformative in the unusual case where  $\lambda$  and  $\phi$  are not separately estimable parameters. Further, the SEs for  $\lambda$  and  $\phi$  are rough approximations as discussed in particular by Smyth et al. (2001;  $V_1$  method).

## Value

An object of class `HLfit`, which is a list with many elements, not all of which are documented.

A few extractor functions are available (see [extractors](#)), and should be used as far as possible as they should be backward-compatible from version 1.4 onwards, while the structure of the return object may still evolve. The following information will be useful for extracting further elements of the object.

Elements include **descriptors of the fit**:

<code>eta</code>	Fitted values on the linear scale (including the predicted random effects);
<code>fv</code>	Fitted values ( $\mu = \langle \text{inverse-link} \rangle (\eta)$ ) of the response variable (returned by the fitted function);
<code>fixef</code>	The fixed effects coefficients, $\beta$ (returned by the <code>fixef</code> function);
<code>ranef</code>	The random effects $u$ (returned by <code>ranef(*, type="uncorrelated")</code> );
<code>v_h</code>	The random effects on the linear scale, $v$ ;
<code>phi</code>	The residual variance $\phi$ ;
<code>phi.object</code>	A possibly more complex object describing $\phi$ ;
<code>lambda</code>	The random effects ( $u$ ) variance $\lambda$ ;
<code>lambda.object</code>	A possibly more complex object describing $\lambda$ ;
<code>corrPars</code>	Agglomerates information on correlation parameters, either fixed, or estimated by <code>HLfit</code> , <code>corrHLfit</code> or <code>fitme</code> ;
<code>APHLs</code>	A list which elements are various likelihood components, include conditional likelihood, h-likelihood, and the two adjusted profile h-likelihoods: the (approximate) marginal <b>likelihood</b> <code>p_v</code> and the (approximate) <b>restricted likelihood</b> <code>p_bv</code> (the latter two available through the <code>logLik</code> function). See the extractor function <a href="#">get_any_IC</a> for information criteria ("AIC") and effective degrees of freedom;  The covariance matrix of $\beta$ estimates is not included as such, but can be extracted by <a href="#">vcov</a> ;

**Information about the input** is contained in output elements named as `HLfit` or `corrHLfit` arguments (`data`, `family`, `resid.family`, `ranFix`, `prior.weights`), with the following notable exceptions or modifications:

<code>predictor</code>	The linear predictor, including the formula (possibly reformatted) and several attributes;
------------------------	--

resid.predictor            Analogous to predictor, for the residual variance;  
 rand.families    corresponding to the rand.family input;

**Further miscellaneous diagnostics and descriptors of model structure:**

X.pv                    The design matrix for fixed effects;  
 ZAlist, struList        Two lists of matrices, respectively the design matrices “**Z**”, and the “**L**” matrices, for the different random effects terms. The extractor `get_ZALMatrix` can be used to reconstruct a single “**ZL**” matrix for all terms.  
 fixef\_terms, fixef\_levels    Further information about fixed effect model;  
 weights                (binomial data only) the binomial denominators;  
 y                        the response vector; for binomial data, the frequency response.  
 models                 Additional information on model structure for  $\eta$ ,  $\lambda$  and  $\phi$ ;  
 HL                      A set of indices that characterize the approximations used for likelihood;  
 leve\_phi, lev\_lambda        Leverages;  
 dfs                     degrees of freedom for different components of the model;  
 warnings                A list of warnings for events that may have occurred during the fit.

Finally, the object includes programming tools: `call`, `spaMM.version`, `fit_time` and `envir`.

**References**

- Breslow, NE, Clayton, DG. (1993). Approximate Inference in Generalized Linear Mixed Models. *Journal of the American Statistical Association* 88, 9-25.
- Lee, Y., Nelder, J. A. (2001) Hierarchical generalised linear models: A synthesis of generalised linear models, random-effect models and structured dispersions. *Biometrika* 88, 987-1006.
- Lee, Y., Nelder, J. A. and Pawitan, Y. (2006). *Generalized linear models with random effects: unified analysis via h-likelihood*. Chapman & Hall: London.
- Noh, M., and Lee, Y. (2007). REML estimation for binary data in GLMMs, *J. Multivariate Anal.* 98, 896-915.
- Smyth GK, Huele AF, Verbyla AP (2001). Exact and approximate REML for heteroscedastic regression. *Statistical Modelling* 1, 161-175.

**See Also**

`HLCor` for estimation with given spatial correlation parameters; `corrHLfit` for joint estimation with spatial correlation parameters; `fitme` as an alternative to all these functions.

**Examples**

```

data("wafers")
## Gamma GLMM with log link

HLfit(y ~X1+X2+X1*X3+X2*X3+I(X2^2)+(1|batch), family=Gamma(log),
      resid.model = ~ X3+I(X3^2) ,data=wafers)

## Gamma - inverseGamma HGLM with log link
HLfit(y ~X1+X2+X1*X3+X2*X3+I(X2^2)+(1|batch), family=Gamma(log),
      HLmethod="HL(1,1)", rand.family=inverse.Gamma(log),
      resid.model = ~ X3+I(X3^2) ,data=wafers)

```

---

inverse.Gamma	<i>Distribution families for Gamma and inverse Gamma-distributed random effects</i>
---------------	---

---

**Description**

For dispersion parameter  $\lambda$ , Gamma means that random effects are distributed as  $u \text{ Gamma}(\text{shape}=1/\lambda, \text{scale}=\lambda)$ , so  $u$  has mean 1 and variance  $\lambda$ . Both the log ( $v = \log(u)$ ) and identity ( $v = u$ ) links are possible, though in the latter case the variance of  $u$  is constrained below 1 (otherwise Laplace approximations fail).

The two-parameter inverse Gamma distribution is the distribution of the reciprocal of a variable distributed according to the Gamma distribution Gamma with the same shape and scale parameters. `inverse.Gamma` implements the one-parameter inverse Gamma family with `shape=1+1/λ` and `rate=1/λ` (`rate=1/scale`). It is used to model the distribution of random effects. Its mean=1; and its variance  $=\lambda/(1-\lambda)$  if  $\lambda < 1$ , otherwise infinite. The default link is `"-1/mu"`, in which case  $v=-1/u$  is `"-Gamma"`-distributed with the same shape and rate, hence with mean  $-(\lambda+1)$  and variance  $\lambda(\lambda+1)$ , which is a different one-parameter Gamma family than the above-described Gamma. The other possible link is `v=log(u)` in which case  $v = \log(X \text{ Gamma}(1+1/\lambda, 1/\lambda))$ , with mean  $-(\log(1/\lambda)+\text{digamma}(1+1/\lambda))$  and variance  $\text{trigamma}(1+1/\lambda)$ .

**Usage**

```

inverse.Gamma(link = "-1/mu")
# Gamma(link = "inverse") using stats::Gamma

```

**Arguments**

link	For Gamma, allowed links are log and identity (the default link from <a href="#">Gamma</a> , <code>"inverse"</code> , cannot be used for the random effect specification). For <code>inverse.Gamma</code> , allowed links are <code>"-1/mu"</code> (default) and log.
------	---

**Examples**

```

# see help("HLfit") for fits using the inverse.Gamma distribution.

```



Loaloea

*Loa loa* prevalence in North Cameroon, 1991-2001**Description**

This data set describes prevalence of infection by the nematode *Loa loa* in North Cameroon, 1991-2001. This is a superset of the data discussed by Diggle and Ribeiro (2007) and Diggle et al. (2007). The study investigated the relationship between altitude, vegetation indices, and prevalence of the parasite.

**Usage**

```
data("Loaloea")
```

**Format**

The data frame includes 197 observations on the following variables:

**latitude** latitude, in degrees.

**longitude** longitude, in degrees.

**ntot** sample size per location

**npos** number of infected individuals per location

**maxNDVI** maximum normalised-difference vegetation index (NDVI) from repeated satellite scans

**seNDVI** standard error of NDVI

**elev1** altitude, in m.

**elev2,elev3,elev4** Additional altitude variables derived from the previous one, provided for convenience: respectively, positive values of altitude-650, positive values of altitude-1000, and positive values of altitude-1300

**maxNDVI1** a copy of maxNDVI modified as `maxNDVI1[maxNDVI1>0.8] <- 0.8`

**Source**

The data were last retrieved on March 1, 2013 from P.J. Ribeiro's web resources at [www.leg.ufpr.br/doku.php/pessoais:paulojus:mbgbook:datasets](http://www.leg.ufpr.br/doku.php/pessoais:paulojus:mbgbook:datasets).

**References**

Diggle, P., and Ribeiro, P. 2007. Model-based geostatistics, Springer series in statistics, Springer, New York.

Diggle, P. J., Thomson, M. C., Christensen, O. F., Rowlingson, B., Obsomer, V., Gardon, J., Wanji, S., Takougang, I., Enyong, P., Kamgno, J., Remme, J. H., Boussinesq, M., and Molyneux, D. H. 2007. Spatial modelling and the prediction of *Loa loa* risk: decision making under uncertainty, *Ann. Trop. Med. Parasitol.* 101, 499-509.

## Examples

```

data("Loaloa")

### Variations on the model fit by Diggle et al.
###   on a subset of the Loaloa data
### In each case this shows the slight differences in syntax,
###   and the difference in 'typical' computation times,
###   when fit using corrHLfit() or fitme().

if (spaMM.getOption("example_maxtime")>4.2) {
  corrHLfit(cbind(npos,ntot-npos)~elev1+elev2+elev3+elev4+maxNDVI1+seNDVI
            +Matern(1|longitude+latitude),HLmethod="HL(0,1)",
            data=Loaloa,family=binomial(),ranFix=list(nu=0.5))
}
if (spaMM.getOption("example_maxtime")>1.8) {
  fitme(cbind(npos,ntot-npos)~elev1+elev2+elev3+elev4+maxNDVI1+seNDVI
        +Matern(1|longitude+latitude),method="HL(0,1)",
        data=Loaloa,family=binomial(),fixed=list(nu=0.5))
}

if (spaMM.getOption("example_maxtime")>7.4) {
  corrHLfit(cbind(npos,ntot-npos)~elev1+elev2+elev3+elev4+maxNDVI1+seNDVI
            +Matern(1|longitude+latitude),
            data=Loaloa,family=binomial(),ranFix=list(nu=0.5))
}
if (spaMM.getOption("example_maxtime")>3.9) {
  fitme(cbind(npos,ntot-npos)~elev1+elev2+elev3+elev4+maxNDVI1+seNDVI
        +Matern(1|longitude+latitude),
        data=Loaloa,family=binomial(),fixed=list(nu=0.5),method="REML")
}

## Diggle and Ribeiro (2007) assumed (in this package notation) Nugget=2/7:
if (spaMM.getOption("example_maxtime")>7.2) {
  corrHLfit(cbind(npos,ntot-npos)~elev1+elev2+elev3+elev4+maxNDVI1+seNDVI
            +Matern(1|longitude+latitude),
            data=Loaloa,family=binomial(),ranFix=list(nu=0.5,Nugget=2/7))
}
if (spaMM.getOption("example_maxtime")>2.8) {
  fitme(cbind(npos,ntot-npos)~elev1+elev2+elev3+elev4+maxNDVI1+seNDVI
        +Matern(1|longitude+latitude),method="REML",
        data=Loaloa,family=binomial(),fixed=list(nu=0.5,Nugget=2/7))
}

## with nugget estimation:
if (spaMM.getOption("example_maxtime")>18.1) {
  corrHLfit(cbind(npos,ntot-npos)~elev1+elev2+elev3+elev4+maxNDVI1+seNDVI
            +Matern(1|longitude+latitude),
            data=Loaloa,family=binomial(),
            init.corrHLfit=list(Nugget=0.1),ranFix=list(nu=0.5))
}
if (spaMM.getOption("example_maxtime")>7.9) {

```

```

fitme(cbind(npos,ntot-npos)~elev1+elev2+elev3+elev4+maxNDVI1+seNDVI
      +Matern(1|longitude+latitude),
      data=Loaloe,family=binomial(),method="REML",
      init=list(Nugget=0.1),fixed=list(nu=0.5))
}

```

LRT

*Likelihood ratio test of fixed effects.***Description**

LRT performs a likelihood ratio (LR) test between two model fits, the “full” and the “null” model fits, currently differing only in their fixed effects. Parametric bootstrap p-values can be computed, either using the raw bootstrap distribution of the likelihood ratio, or a bootstrap estimate of the Bartlett correction of the LR statistic. This function differ from `fixedLRT` in its arguments (model fits for LRT, but all arguments required to fit the models for `fixedLRT`), and in the format of its return value. The function will stop or return possibly incorrect results for models differing beyond their fixed effects. By conceptual drift, `anova` works as an alias for LRT.

**Usage**

```

## S3 method for class 'HLfit'
anova(object, object2, ..., method="")
LRT(object, object2, boot.repl=0, nb_cores=NULL, ...)

```

**Arguments**

<code>object,object2</code>	Two models fits being compared (their order does not matter).
<code>boot.repl</code>	the number of bootstrap replicates.
<code>nb_cores</code>	Number of cores to use for parallel computation of bootstrap. The default is <code>spaMM.getOption("nb_cores")</code> , and 1 if the latter is NULL. <code>nb_cores=1</code> prevents the use of parallelisation procedures.
<code>method</code>	For development purposes, not documented.
<code>...</code>	Further arguments passed to or from other methods.

**Details**

A raw bootstrap p-value can be computed from the simulated distribution as  $(1 + \sum(t \geq t_0)) / (N + 1)$  where  $t_0$  is the original likelihood ratio,  $t$  the vector of bootstrap replicates and  $N$  its length. See Davison & Hinkley (1997, p. 141) for discussion of the adjustments in this formula. However, a computationally more economical use of the bootstrap is to provide a Bartlett correction for the likelihood ratio test in small samples. According to this correction, the mean value  $m$  of the likelihood ratio statistic under the null hypothesis is computed (here estimated by a parametric bootstrap) and the original LR statistic is multiplied by  $n/m$  where  $n$  is the number of degrees of freedom of the test.

**Value**

An object of class `fixedLRT`, actually a list with as-yet unstable format, but here with typical elements (depending on the options)

`fullfit`            the HLfit object for the full model;  
`nullfit`            the HLfit object for the null model;  
`basicLRT`           A data frame including values of the likelihood ratio chi2 statistic, its degrees of freedom, and the p-value;

and, if a bootstrap was performed:

`rawBootLRT`        A data frame including values of the likelihood ratio chi2 statistic, its degrees of freedom, and the raw bootstrap p-value;  
`BartBootLRT`        A data frame including values of the Bartlett-corrected likelihood ratio chi2 statistic, its degrees of freedom, and its p-value;  
`bootInfo`           a list with the following elements:  
                   **bootreps** A table of fitted likelihoods for bootstrap replicates;  
                   **meanbootLRT** The mean likelihood ratio chi-square statistic for bootstrap replicates;

**References**

Bartlett, M. S. (1937) Properties of sufficiency and statistical tests. *Proceedings of the Royal Society (London) A* 160: 268-282.

Davison A.C., Hinkley D.V. (1997) *Bootstrap methods and their applications*. Cambridge Univ. Press, Cambridge, UK.

**See Also**

See also [fixedLRT](#).

**Examples**

```
data("wafers")
## Gamma GLMM with log link
m1 <- HLfit(y ~X1+X2+X1*X3+X2*X3+I(X2^2)+(1|batch),family=Gamma(log),
            resid.model = ~ X3+I(X3^2) ,data=wafers,HLmethod="ML")
m2 <- update(m1,formula.= ~ . -I(X2^2))
anova(m1,m2)
```

---

make_scaled_dist	<i>Scaled distances between unique locations</i>
------------------	--

---

### Description

This function computes scaled distances from whichever relevant argument it can use (see Details). The result can directly be used as input for computation of the Matérn correlation matrix. It is usually called internally by HLCor, so that users may ignore it, except if they wish to control the distance used through `control.dist$method`, or the parametrization of the scaling through `control.dist$rho.mapping`.

### Usage

```
make_scaled_dist(uniqueGeo, uniqueGeo2=NULL, distMatrix, rho,
                 rho.mapping=seq_len(length(rho)),
                 dist.method="Euclidean",
                 return_matrix=FALSE)
```

### Arguments

<code>uniqueGeo</code>	A matrix of geographical coordinates (e.g. 2 columns for latitude and longitude), without replicates of the same location.
<code>uniqueGeo2</code>	NULL, or a second matrix of geographical coordinates, without replicates of the same location. If NULL, scaled distances among <code>uniqueGeo</code> locations are computed. Otherwise, scaled distances between locations in the two input matrices are computed.
<code>distMatrix</code>	A distance matrix.
<code>rho</code>	A scalar or vector of positive values. Scaled distance is computed as <code>&lt;distances in each coordinate&gt;</code> unless a non-trivial <code>rho.mapping</code> is used.
<code>rho.mapping</code>	A set of indices controlling which elements of the <code>rho</code> scale vector scales which dimension(s) of the space in which (spatial) correlation matrices of random effects are computed. Scaled distance is generally computed as <code>&lt;distances in each coordinate&gt; * rho</code> . As shown in the Example, if one wishes to combine isotropic geographical distance and some environmental distance, the coordinates being latitude, longitude and one environmental variable, the scaled distance may be computed as (say) <code>(lat, long, env) * rho[c(1, 1, 2)]</code> so that the same scaling <code>rho[1]</code> applies for both geographical coordinates. In this case, <code>rho</code> should have length 2 and <code>rho.mapping</code> should be <code>c(1, 1, 2)</code> .
<code>dist.method</code>	method argument of <code>proxy::dist</code> function (by default, "Euclidean", but other distances are possible (see Details)).
<code>return_matrix</code>	Whether to return a matrix rather than a <code>proxy::dist</code> or <code>proxy::crossdist</code> object.

## Details

The function uses the `distMatrix` argument if provided, in which case `rho` must be a scalar. Vectorial `rho` (i.e., different scaling of different dimensions) is feasible only by providing `uniqueGeo`.

The `dist.method` argument gives access to distances implemented in the `proxy` package, or to user-defined ones that are made accessible to `proxy` through its database. Of special interest for spatial analyses are spherical distances. `proxy` implements the "Geodesic" distance which does not use longitude and latitude as coordinates. For this reason, `spaMM` implements an "Earth" method for `proxy::dist`, which uses longitude and latitude (in that order: see Examples) as coordinates for computing orthodromic distances in km. The Earth radius used for spherical approximation is 6371.009 km.

## Value

A matrix or `dist` object. If there are two input matrices, rows of the return value correspond to rows of the first matrix.

## Examples

```
data("blackcap")
## a biologically not very meaningful, but syntactically correct example of rho.mapping
corrHLfit(migStatus ~ 1 + Matern(1|latitude+longitude+means),data=blackcap,
          HLmethod="ML",ranFix=list(nu=0.5,phi=1e-6),
          init.corrHLfit=list(rho=c(1,1)),
          control.dist=list(rho.mapping=c(1,1,2)))
## Using orthodromic distances: order of variables in Matern(.|longitude+latitude) matters
corrHLfit(migStatus ~ 1 + Matern(1|longitude+latitude),data=blackcap,
          HLmethod="ML",ranFix=list(nu=0.5,phi=1e-6),
          control.dist=list(dist.method="Earth"))
```

---

mapMM

*Colorful plots of predicted responses in two-dimensional space.*

---

## Description

These functions provide either a map of predicted response in analyzed locations, or a predicted surface. `mapMM` is a straightforward representation of the analysis of the data, while `filled.mapMM` copes with the fact that all predictor variables may not be known in all locations on a fine spatial grid, but may involve questionable choices as a result (see `map.formula` argument). Both functions takes an `HLfit` object as input. `mapMM` calls `spaMMplot2D`, which is similar but takes a more conventional `(x,y,z)` input.

## Usage

```
spaMMplot2D(x, y, z,xrange=range(x, finite = TRUE),
            yrange=range(y, finite = TRUE),
            margin=1/20,add.map= FALSE, nlevels = 20,
            color.palette = spaMM.colors,map.asp=NULL,
```

```
col = color.palette(length(levels) - 1),
plot.title=NULL, plot.axes=NULL, decorations=NULL,
key.title=NULL, key.axes=NULL, xaxs = "i",
yaxs = "i", las = 1, axes = TRUE, frame.plot = axes, ...)
```

```
mapMM(fitobject,Ztransf=NULL,coordinates,
      add.points,decorations=NULL,plot.title=NULL,plot.axes=NULL,envir=-3, ...)
```

```
filled.mapMM(fitobject, Ztransf = NULL, coordinates, xrange = NULL,
             yrange = NULL, margin = 1/20, map.formula, phi =
             1e-05, gridSteps = 41, decorations =
             quote(points(pred[, coordinates], cex = 1, lwd = 2)),
             add.map = FALSE, axes = TRUE, plot.title = NULL,
             plot.axes = NULL, map.asp = NULL, variance = NULL,
             var.contour.args = list(), smoothObject = NULL, ...)
```

## Arguments

fitobject	The return object of a corrHLfit call.
x,y,z	Three vectors of coordinates, with z being expectedly the response.
Ztransf	A transformation of the predicted response, given as a function whose only required argument can be a one-column matrix. The name of this argument must be Z (not x), as is appropriate for use in <code>do.call(Ztransf, list(Z=Zvalues))</code> .
coordinates	The geographical coordinates. By default they are deduced from the model formula. For example if this formula is $\text{resp} \sim 1 + \text{Matern}(1   x + y)$ the default coordinates are <code>c("x","y")</code> . If this formula is $\text{resp} \sim 1 + \text{Matern}(1   x + y + z)$ , the user must choose two of the three coordinates.
xrange	The x range of the plot (a vector of length 2); by default defined to cover all analyzed points.
yrange	The y range of the plot (a vector of length 2); by default defined to cover all analyzed points.
margin	This controls how far (in relative terms) the plot extends beyond the x and y ranges of the analyzed points, and is overridden by explicit xrange and yrange arguments.
map.formula	Plotting a filled contour generally requires prediction in non-observed locations, where predictor variables used in the original data analysis may be missing. In that case, the original model formula cannot be used and an alternative map.formula must be used to interpolate (not smooth) the predicted values in observed locations (these predictions still resulting from the original analysis based on predictor variables). As a result (1) filled.mapMM will be slower than a mere plotting function, since it involves the analysis of spatial data; (2) the results may have little useful meaning if the effects of the original predictor variables is not correctly represented by this interpolation step. For example, it may involve biases analogous to predicting temperature in non-observed locations while ignoring effect of variation in altitude in such locations.

phi	This controls the phi value assumed in the interpolation step. Ideally phi would be zero, but problems with numerically singular matrices may arise when phi is too small.
gridSteps	The number of levels of the grid of x and y values
variance	Either NULL, or the name of a component of prediction variance to be plotted. Must name one of the components that can be returned by <code>predict.HLfit</code> . <code>variance="predVar"</code> is suitable for uncertainty in point prediction.
<code>var.contour.args</code>	A list of control parameters for rendering of prediction variances. See <a href="#">contour</a> for possible arguments (except x, y, z and add).
<code>add.map</code>	Either a boolean or an explicit expression, enclosed in quote (see Examples). If TRUE, the <code>map</code> function from the <code>maps</code> package (which much therefore the loaded) is used to add a map from its default <code>world</code> database. <code>xrange</code> and <code>yrange</code> are used to select the area, so it is most convenient if the coordinates are longitude and latitude (in this order and in standard units). An explicit expression can also be used for further control.
levels	a set of levels which are used to partition the range of z. Must be strictly increasing (and finite). Areas with z values between consecutive levels are painted with the same color.
nlevels	if <code>levels</code> is not specified, the range of z, values is divided into *approximately* this many levels (a call to <a href="#">pretty</a> determines the actual number of levels).
<code>color.palette</code>	a color palette function to be used to assign colors in the plot.
<code>map.asp</code>	the y/x aspect ratio of the 2D plot area (not of the full figure including the scale). By default, the scales for x and y are identical unless the x and y ranges are too different. Namely, the scales are identical if $(\text{plotted y range})/(\text{plotted x range})$ is $1/4 < . < 4$ , and <code>map.asp</code> is 1 otherwise.
col	an explicit set of colors to be used in the plot. This argument overrides any palette function specification. There should be one less color than levels
<code>plot.title</code>	statements which add titles to the main plot. See Details for differences between functions.
<code>plot.axes</code>	statements which draw axes (and a box) on the main plot. See Details for differences between functions.
decorations	Either NULL or Additional graphic statements (points, polygon, etc.), enclosed in quote (the default value illustrates the latter syntax). .
<code>add.points</code>	Obsolete, use <code>decorations</code> instead.
envir	Controls the environment in which <code>plot.title</code> , <code>plot.axes</code> , and <code>decorations</code> are evaluated. <code>mapMM</code> calls <code>spaMM2Dplot</code> from where these graphic arguments are evaluated, and the default value -3 means that they are evaluated within the environment from where <code>mapMM</code> was called.
<code>key.title</code>	statements which add titles for the plot key.
<code>key.axes</code>	statements which draw axes on the plot key.
xaxs	the x axis style. The default is to use internal labeling.
yaxs	the y axis style. The default is to use internal labeling.



<code>las</code>	the style of labeling to be used. The default is to use horizontal labeling.
<code>axes, frame.plot</code>	logicals indicating if axes and a box should be drawn, as in <code>plot.default</code> .
<code>smoothObject</code>	Either NULL, or an object inheriting from class <code>HLfit</code> (hence, an object on which <code>predict.HLfit</code> can be called), predicting the response surface in any coordinates. See Details for typical usages.
<code>...</code>	further arguments passed to or from other methods. For <code>mapMM</code> , all such arguments are passed to <code>spaMMplot2D</code> ; for <code>spaMMplot2D</code> , currently only additional graphical parameters passed to <code>title()</code> (see Details). For <code>filled.mapMM</code> , these parameters are those that can be passed to <code>spaMM.filled.contour</code> .

## Details

The `smoothObject` argument may be used to redraw a figure faster by recycling the predictor of the response surface returned invisibly by a previous call to `filled.mapMM`.

For `smoothObject=NULL` (the default), `filled.mapMM` interpolates the predicted response, with sometimes unpleasant effects. For example, if one interpolates probabilities, the result may not be within  $[0,1]$ , and then (say) a logarithmic `Ztransf` may generate NaN values that would otherwise not occur. The `smoothObject` argument may be used to overcome the default behaviour, by providing an alternative predictor.

If you have values for all predictor variables in all locations of a fine spatial grid, `filled.mapMM` may not be a good choice, since it will ignore that information (see `map.formula` argument). Rather, one should use `predict(<fitobject>, newdata= <all predictor variables >)` to generate all predictions, and then either `spaMM.filled.contour` or some other raster functions.

The different functions are (currently) inconsistent among themselves in the way they handle the `plot.title` and `plot.axes` argument:

**spaMM.filled.contour** behaves like `graphics::filled.contour`, which (1) handles arguments which are calls such as `title(.)` or `{axis(1);axis(2)}`; (2) ignores `...` arguments if `plot.title` is missing; and (3) draws axes by default when `plot.axes` is missing, given `axes = TRUE`.

By contrast, **filled.mapMM** handles arguments which are language expressions such as produced by `quote(.)` or `substitute(.)` (see Examples).

**mapMM** can handles language expressions, but also accepts at least some calls.

## Value

`filled.mapMM` returns invisibly a predictor of the response surface. `mapMM` has no return value. Plots are produced as side-effects.

## See Also

[http://kimura.univ-montp2.fr/~rousset/spaMM/example\\_raster.html](http://kimura.univ-montp2.fr/~rousset/spaMM/example_raster.html) for more elaborate plot procedures.

## Examples

```

data("blackcap")
bfit <- corrHLfit(migStatus ~ means+ Matern(1|longitude+latitude),data=blackcap,
                 HLmethod="ML",
                 ranFix=list(lambda=0.5537,phi=1.376e-05,rho=0.0544740,nu=0.6286311))
if (requireNamespace("maps")) { ## required for add.map=TRUE
  mapMM(bfit,color.palette = function(n){spaMM.colors(n,redshift=1/2)},add.map=TRUE)
}

if (spaMM.getOption("example_maxtime")>0.8) {
  ## filled.mapMM takes a bit longer
  # showing 'add.map', 'nlevels', and contour lines for 'variances'
  if (requireNamespace("maps")) { ## required for add.map=TRUE
    filled.mapMM(bfit,nlevels=30,add.map=TRUE,plot.axes=quote({axis(1);axis(2)}),
                 variance="respVar",
                 plot.title=title(main="Inferred migration propensity of blackcaps",
                                   xlab="longitude",ylab="latitude"))
  }
}

if (spaMM.getOption("example_maxtime")>2.5) {
  data("Loaloa")
  lfit <- corrHLfit(cbind(npos,ntot-npos)~elev1+elev2+elev3+elev4+maxNDVI1+seNDVI
                   +Matern(1|longitude+latitude),HLmethod="HL(0,1)",data=Loaloa,
                   family=binomial(),ranFix=list(nu=0.5,rho=2.255197,lambda=1.075))

  ## longer computation requiring interpolation of 197 points
  if (requireNamespace("maps")) { ## required for add.map=TRUE
    filled.mapMM(lfit,add.map=TRUE,plot.axes=quote({axis(1);axis(2)}),
                 decorations=quote(points(pred[,coordinates],pch=15,cex=0.3)),
                 plot.title=title(main="Inferred prevalence, North Cameroon",
                                   xlab="longitude",ylab="latitude"))
  }
}

```

---

MaternCorr

*Matern correlation function and Matern formula term.*

---

## Description

The Matérn correlation function describes realizations of Gaussian spatial processes with different smoothnesses (i.e. either smooth or rugged surfaces). It also includes a scaling and a 'nugget' parameter. It can be invoked in two ways. First, the `MaternCorr` function evaluates these correlations, using distances as input. Second, a term of the form `Matern(1|<...>)` in a formula specifies a random effect with Matérn correlation function, using coordinates found in a data frame as input. In the latter case, the correlations between realizations of the random effect for any two observations in the data will be the value of the Matérn function at the scaled Euclidean distance between coordinates specified in `<...>`, using "+" as separator (e.g., `Matern(1|latitude + longitude)`).

**Usage**

```
## Default S3 method:
MaternCorr(d, rho = 1, smoothness, nu = smoothness, Nugget = NULL)
# Matern(1|...)
```

**Arguments**

d	A distance or a distance matrix.
rho	A scaling factor for distance. The 'range' considered in some formulations is the reciprocal of this scaling factor
smoothness	The smoothness parameter, >0. $\nu = 0.5$ corresponds to the exponential correlation function, and the limit function when $\mu$ goes to $\infty$ is the squared exponential function (as in a Gaussian).
nu	Same as smoothness
Nugget	(Following the jargon of Kriging) a parameter describing a discontinuous decrease in correlation at zero distance. Correlation will always be 1 at $d = 0$ , and from which it immediately drops to (1-Nugget)
...	Names of coordinates, using "+" as separator (e.g., Matern(1 latitude + longitude)

**Details**

The correlation at distance  $d > 0$  is

$$(1 - \text{Nugget}) \frac{(\rho d)^\nu K_\nu(\rho d)}{2^{(\nu-1)} \Gamma(\nu)}$$

where  $K_\nu$  is the [besselK](#) function of order  $\nu$ .

**Value**

Scalar/vector/matrix depending on input.

**References**

Stein, M.L. (1999) Statistical Interpolation of Spatial Data: Some Theory for Kriging. Springer, New York.

**See Also**

See [corMatern](#) for an implementation of this correlation function as a `corSpatial` object for use with `lme` or `glmmPQL`.

By default the `Nugget` is set to 0. See one of the examples on data set [LoaLoa](#) for a fit including the estimation of the `Nugget`.

**Examples**

```
## See examples in help("spaMM"), help("HLCor"), help("Loaloo"), etc.
## The Matérn family can be used in Euclidean spaces of any dimension:
set.seed(123)
randpts <- matrix(rnorm(20),nrow=5)
distMatrix <- as.matrix(proxy::dist(randpts))
MaternCorr(distMatrix,nu=2)
```

---

multinomial

---

*Analyzing multinomial data*


---

**Description**

These functions facilitate the conversion and analysis of multinomial data as a series of nested binomial data. The main function is `multi`, to be used in the `family` argument of the fitting functions. It calls `binomialize`, which can be called directly to check how the data are converted to nested binomial data. The `fitted.HLfitlist` method of the fitted generic function returns a matrix of fitted multinomial probabilities. The `logLik.HLfitlist` method of the `logLik` generic function returns a log-likelihood for the joint fits.

**Usage**

```
multi(binResponse=c("npos", "neg"), binfamily=binomial(), input="types", ...)
binomialize(data, responses, sortedTypes=NULL, binResponse=c("npos", "neg"),
            depth=Inf, input="types")
## S3 method for class 'HLfitlist'
fitted(object, ...)
## S3 method for class 'HLfitlist'
logLik(object, which, ...)
```

**Arguments**

<code>data</code>	The data frame to be analyzed.
<code>object</code>	A list of binomial fits returned by a multinomial analysis
<code>responses</code>	column names of the data, such that <code>&lt;data&gt;[, &lt;responses&gt;]</code> contain the multinomial response data, as levels of factor variables.
<code>sortedTypes</code>	Names of multinomial types, i.e. levels of the multinomial response factors. Their order determines which types are taken first to define the nested binomial samples. By default, the most common types are considered first.
<code>binResponse</code>	The names to be given to the number of “success” and “failures” in the binomial response.
<code>depth</code>	The maximum number of nested binomial responses to be generated from the multinomial data.
<code>binfamily</code>	The family applied to each binomial response.

input	If input="types", then the responses columns must contain factor levels of the binomial response. If input="counts", then the responses columns must contain counts of different factor levels, and the column names are the types.
which	Which element of the APHLs list to return. The default depends on the fitting method. In particular, if it was REML or one of its variants, the function returns the log restricted likelihood (exact or approximated).
...	Other arguments passed from or to other functions.

### Details

A multinomial response, say counts 17, 13, 25, 8, 3, 1 for types type1 to type6 can be represented as a series of nested binomials e.g. type1 against others (17 vs 50) then among these 50 others, type2 versus others (13 vs 37), etc. The `binomialize` function generates such a representation. By default the representation considers types in decreasing order of the number of positives, i.e. first type3 against others (25 vs 42), then type1 against others within these 42, etc. It stops if it has reached depth nested binomial responses. This can be modified by the `sortedTypes` argument, e.g. `sortedTypes=c("type6", "type4", "type2")`. `binomialize` returns a list of data frames which can be directly provided as a data argument for the fitting functions, with binomial response.

Alternatively, one can provide the multinomial response data frame, which will be internally converted to nested binomial data if the `family` argument is a call to `multinomial` (see examples).

For mixed models, the multinomial data can be fitted to a model with the same correlation parameters, and either the same or different variances of random effects, for all binomial responses. Which analysis is performed depends on the `init.corrHLfit` argument (see `corrHLfit` and the Examples).

### Value

`binomialize` returns a list of data frames appropriate for analysis as binomial response. Each data frame contains the original one plus Two columns named according to `binResponse`. `multi` returns a list.

### Examples

```
## An example considering pseudo-data at one diploid locus for 50 individuals
set.seed(123)
genecopy1 <- sample(4, size=50, prob=c(1/2, 1/4, 1/8, 1/8), replace=TRUE)
genecopy2 <- sample(4, size=50, prob=c(1/2, 1/4, 1/8, 1/8), replace=TRUE)
alleles <- c("122", "124", "126", "128")
genotypes <- data.frame(type1=alleles[genecopy1], type2=alleles[genecopy2])
## Columns "type1", "type2" each contains an allele type => input is "types" (the default)
datalist <- binomialize(genotypes, responses=c("type1", "type2"))

## two equivalent fits:
f1 <- HLfit(cbind(npos, nneg)~1, data=datalist, family=binomial())
f2 <- HLfit(cbind(npos, nneg)~1, data=genotypes, family=multi(responses=c("type1", "type2")))
fitted(f2)

## distinct fits for spatial data
## Not run:
```

```

genoInSpace <- data.frame(type1=alleles[genecopy1],type2=alleles[genecopy2],x=runif(50),y=runif(50))
## Fitting distinct variances of random effects for each binomial response
corrHLfit(cbind(npos,nneg)~1+Matern(1|x+y),data=genoInSpace,
          family=multi(responses=c("type1","type2")),
          ranFix=list(rho=1,nu=0.5))
## Fitting the same variance for all binomial responses
corrHLfit(cbind(npos,nneg)~1+Matern(1|x+y),data=genoInSpace,
          family=multi(responses=c("type1","type2")),
          ranFix=list(rho=1,nu=0.5),init.corrHLfit=list(lambda=1))

## End(Not run)

```

---

negbin	<i>Family function for GLMs and mixed models with negative binomial and zero-truncated negative binomial response.</i>
--------	--

---

### Description

`family` object that specifies the information required to fit a negative binomial generalized linear model, with known or unknown underlying Gamma shape parameter. The zero-truncated variant can be specified either as `Tnegbin(.)` or as `negbin(., trunc = 0L)`.

### Usage

```

negbin(shape = stop("negbin's 'shape' must be specified"), link = "log", trunc = -1L)
Tnegbin(shape = stop("negbin's 'shape' must be specified"), link = "log")
# (the shape parameter is actually not requested unless this is used in a glm() call)

```

### Arguments

shape	Shape parameter of the underlying Gamma distribution, given that the negbin family can be represented as a Poisson-Gamma mixture, where the conditional Poisson mean is $\mu$ times a Gamma random variable with mean 1 and shape shape (as produced by <code>rgamma(., shape=shape, scale=1/shape)</code> ).
link	log, sqrt or identity link, specified by the several available ways for GLM links (name, character string, one-element character vector, or object of class <code>link-glm</code> as returned by <code>make.link</code> ).
trunc	Either <code>0L</code> for zero-truncated distribution, or <code>-1L</code> for default untruncated distribution.

### Details

shape is the  $k$  parameter of McCullagh and Nelder (1989, p.373) and the theta parameter of Venables and Ripley (2002, section 7.4). The latent Gamma variable has mean 1 and variance  $1/\text{shape}$ , and the negbin with mean  $\mu$  has variance  $\mu + \mu^2/\text{shape}$ . The negbin family is sometimes called the NegBin1 model in the literature on negative binomial models.

The name `NB_shape` should be used to set values of shape in control arguments of the fitting functions (e.g., `fitme(., init=list(NB_shape=1))`).

**Value**

A family object.

**References**

McCullagh, P. and Nelder, J.A. (1989) *Generalized Linear Models*, 2nd edition. London: Chapman & Hall.

Venables, W. N. and Ripley, B. D. (2002) *Modern Applied Statistics with S-PLUS*. Fourth Edition. Springer.

**Examples**

```
## Fitting negative binomial model with estimated scale parameter:
data("scotlip")
fitme(cases~I(prop.ag/10)+offset(log(expec)),family=negbin(), data=scotlip)
negfit <- fitme(I(1+cases)~I(prop.ag/10)+offset(log(expec)),family=Tnegbin(), data=scotlip)
simulate(negfit,nsim=3)
```

---

options	<i>spaMM options settings</i>
---------	-------------------------------

---

**Description**

Allow the user to set and examine a variety of *options* which affect operations of the spaMM package.

**Usage**

```
spaMM.options(...)
```

```
spaMM.getOption(x)
```

**Arguments**

x	a character string holding an option name.
...	A named value or a list of named values. The following values, with their defaults, are used in spaMM:  COMP_maxn: Number of terms for truncation of infinite sums that are evaluated in the fitting of <b>COMPoisson</b> models. MESSAGES.FULL.STACK=TRUE: Whether to give information on the stack of calls in some warning messages. QRmethod: A character string, to control whether dense matrix or sparse matrix methods are used in intensive matrix computations, overcoming the defaults choices made by spaMM in this respect. Possible values are "dense" and "sparse".

`matrix_method`: A character string, to control the factorization of dense model matrices. Default value is "def\_sXaug\_EigenDense\_QR\_scaled". "def\_sXaug\_EigenDense\_QRP\_" is sometimes useful (see [arabidopsis](#)). The source code should be consulted for further information.

`Matrix_method`: A character string, to control the factorization of sparse model matrices. Default value is "def\_sXaug\_Matrix\_QRP\_scaled". The source code should be consulted for further information.

`LevenbergM=NULL`: NULL or boolean. Whether to use a Levenberg-Marquardt algorithm (see [Details](#)) by default in most computations. But it is advised to use instead `control.HLfit=list(LevenbergM=...)` to control this on a case-by-case basis. The joint default behaviour is that Levenberg-Marquardt is used by default for binomial response data that takes only extreme values (in particular, for binary 0/1 response), and that for other models the fitting algorithm switches to it if divergence is suspected. FALSE inhibits its use; TRUE forces its use for all iterative least-square fits, except when 'confint()' is called.

`USEEIGEN=TRUE`: Whether to use the Eigen C++ library for some matrix computations.. The source code should be consulted for further information.

`wRegularization=FALSE`: Whether to warn about the use of regularization in some operations on nearly singular matrices.

`maxLambda=1e10`: The maximum value of lambda: higher fitted lambda values in HLfit are reduced to this.

`example_maxtime=0.7`: Used in the documentation to control whether the longer examples should be run. The approximate running time of given examples on one author's laptop is compared to this value.

`optimizer1D="optimize"`: Optimizer for one-dimensional optimization. If you want to control the initial value, you should select another optimizer.

`optimizer="nloptr"`: Optimizer for optimization in several dimensions. Use `optimizer="nloptr"` to call `nloptr` with method "NLOPT\_LN\_BOBYQA"; use `optimizer="bobyqa"` to call `bobyqa`; and use `optimizer="L-BFGS-B"` to call `optim` with method "L-BFGS-B". The optimizer can also be specified on a fit-by-fit basis as the value of `control$optimizer` in a `fitme` call, or as the value of `control.corrHLfit$optimizer`.

`nloptr`: Default control values of `nloptr` calls.

`CMP_asympto_cond`: Condition for applying an approximation or the COM-Poisson response family, as detailed in [COMPoisson](#).

and possibly other undocumented values for development purposes. Additional options without default values can also be used (e.g., see [sparse\\_precision](#)).

## Details

`spaMM.options()` provides an interface for changing maximal values of parameters of the Matérn correlation function. However, it is not recommended to change these values unless a `spaMM` message specifically suggests so. Errors may occur if too low values are chosen as these may conflict with default initial values for the parameters.

By default `spaMM` use Iteratively Reweighted Least Squares (IRLS) methods to estimate fixed effect parameters. However, a Levenberg-Marquardt algorithm, as described by Nocedal & Wright



(1999, p. 266), is also implemented. The Levenberg-Marquardt algorithm is well suited to compute a PQL fit, which is based on maximization of a single function, the h-likelihood. By contrast, in a fit of a mixed model by (RE)ML, one computes jointly fixed-effect estimates that maximizes marginal likelihood, and random-effect values that maximize h-likelihood given the fixed-effect estimates. The Levenberg-Marquardt algorithm is not directly applicable in this case, as it may produce random-effect values that it will accept as increasing marginal likelihood rather than h-likelihood. The (RE)ML variant of the algorithm therefore uses additional nested h-likelihood-maximizing steps for correcting random-effect values.

### Value

For `spaMM.getOption`, the current value set for option `x`, or `NULL` if the option is unset.

For `spaMM.options()`, a list of all set options. For `spaMM.options(name)`, a list of length one containing the set value, or `NULL` if it is unset. For uses setting one or more options, a list with the previous values of the options changed (returned invisibly).

### References

Jorge Nocedal and Stephen J. Wright (1999) Numerical Optimization. Springer-Verlag, New York.

### Examples

```
spaMM.options()
spaMM.getOption("example_maxtime")
## Not run:
spaMM.options(maxLambda=1e06)

## End(Not run)
```

---

pedigree

*Fit mixed-effects models incorporating pedigrees*

---

### Description

This illustrates how to use `spaMM` for quantitative genetic analyses. `spaMM` appears competitive in terms of speed for GLMMs with large data sets, particularly when using the PQL method, which may be a quite good approximation in such cases. For large pedigrees it may be useful to compute the inverse of the relationship matrix using some efficient ad hoc algorithm, then to provide it as argument of the fit using the `covStruct(list(precision=...))` syntax.

### Examples

```
## Not run:
if(requireNamespace("pedigreemm", quietly=TRUE)) {
## derived from help("pedigreemm")
p1 <- new("pedigree",
         sire = as.integer(c(NA,NA,1, 1,4,5)),
         dam  = as.integer(c(NA,NA,2,NA,3,2)),
```

```

        label = as.character(1:6))
A <- pedigreeemm::getA(p1) ## relationship matrix
## data simulation
cholA <- chol(A)
varU <- 0.4; varE <- 0.6; rep <- 20
n <- rep*6
set.seed(108)
bStar <- rnorm(6, sd=sqrt(varU))
b <- crossprod(as.matrix(cholA),bStar)
ID <- rep(1:6, each=rep)
e0 <- rnorm(n, sd=sqrt(varE))
y <- b[ID]+e0
obs <- data.frame(y=y,IDgen=ID,IDenv=ID) ## two copies of ID for readability of GLMM results
## fits
fitme(y ~ 1+ corrMatrix(1|IDgen) , corrMatrix=A,data=obs,method="REML")
obs$y01 <- ifelse(y<1.3,0,1)
fitme(y01 ~ 1+ corrMatrix(1|IDgen)+(1|IDenv), corrMatrix=A,data=obs,
      family=binomial(), method="REML")
prec_mat <- solve(A)
colnames(prec_mat) <- rownames(prec_mat) <- rownames(A) # important
fitme(y01 ~ 1+ corrMatrix(1|IDgen)+(1|IDenv) , covStruct=list(precision=prec_mat),
      data=obs,family=binomial(), method="REML")
}

## End(Not run)

```

---

 phiHGLM

*Fitting random effects in the residual dispersion model*


---

## Description

$\phi$  parameters are estimated by fitting a Gamma HGLM to response values computed by the parent fitting function (e.g., by `HLfit` in the Examples). The `fitme` function is used to perform this fit. The `resid.model` of the parent call is used to control the arguments of this `fitme` call.

## Usage

```
# 'resid.model' argument of main fitting functions
```

## Arguments

`resid.model` is **either** a formula (without left-hand side) for the dispersion parameter  $\phi$  of the residual error (a log link is assumed);  
**or** a list, with possible elements:

- `model` formula as in formula-only case, without left-hand side
- `family` The family is always Gamma. The default link is log. The identity link can be tried but may fail because only the log link ensures that the fitted  $\phi$  is positive.
- `fixed` fixed values of parameters. Same usage as documented in [fitme](#)

control.dist	A list of arguments that control the computation of the distance argument of the correlation functions. Same usage as documented in <a href="#">HLCor</a>
rand.family	A family object or a list of family objects describing the distribution of the random effect(s). Same usage as documented for <a href="#">HLfit</a> resid.model with random effects is still experimental and complex combinations of arguments could give unexpected results. In particular, the functionality of <code>init.HLfit</code> , <code>lower</code> , <code>upper</code> , <code>control</code> has not been tested. The list should not contain the following elements:
init	Currently ignored;
method	which is constrained to be identical to the method from the parent call;
control.HLfit, control.glm	constrained to be identical to the same-named controls from the parent call;
resid.model	(constrained: no resid.model for a resid.model);
REMLformula	(constrained to NULL);
data	identical to data from the parent call, which must therefore include all the variables required for the resid.model;
prior.weights	constrained: no prior weights;
verbose	constrained: will display a progress line summarizing the results of the resid.model fit at each iteration of main loop of the parent <code>HLfit</code> call.

## References

Lee, Y., Nelder, J. A. and Pawitan, Y. (2006) Generalized linear models with random effects: unified analysis via h-likelihood. Chapman & Hall: London.

## Examples

```
if (spaMM.getOption("example_maxtime")>4.9) {
  data("crack") # crack data, Lee et al. 2006 chapter 11 etc
  hlfit <- HLfit(y~crack0+(1|specimen), family=Gamma(log),
               data=crack, HLmethod="REML",
               rand.family=inverse.Gamma(log),
               resid.model=list(formula=~cycle+(1|specimen)) )
}
```

---

plot.HLfit

*Model checking plots for mixed models*

---

## Description

This function provides diagnostic plots for residual errors from the mean model and for random effects. Plots for the mean models are similar to those for GLMs, as described in Lee et al. 2006. Plots for residual errors consider the *standardized* deviance residuals (Lee et al. 2006, p.52), and plots for random effects likewise consider standardized values, i.e. each random deviate divided by  $\sqrt{(1 - q)}$  where  $q$  is the corresponding leverage for  $\lambda$ .

**Usage**

```
## S3 method for class 'HLfit'
plot(x, which = c("mean", "ranef"),
     titles = list(
       meanmodel=list(outer="Mean model",devres="Deviance residuals",
         absdevres="|Deviance residuals|", resq="Residual quantiles",
         devreshist="Deviance residuals"),
       ranef=list(outer="Random effects and leverages",qq="Random effects Q-Q plot",
         levphi=expression(paste("Leverages for ",phi)),
         levlambda=expression(paste("Leverages for ",lambda)))
     ),
     control= list(), ask=TRUE, ...)
```

**Arguments**

x	The return object of an HLCor / HLfit / corrHLfit call.
which	A vector of keywords for different types of plots. By default, two types of plots are presented on different devices: diagnostic plots for mean values, and diagnostic plots for random effects. Either one can be selected using this argument. Use keyword "predict" for a plot of predicted response against actual response.
titles	A list of the main (inner and outer) titles of the plots. See the default value for the format.
control	A list of default options for the plots. Defaults are pch="+" and pcol="blue" for points, and lcol="red" for curves.
ask	Logical; passed to devAskNewPage which is run when a new device is opened by code.HLfit.
...	Options passed from plot.HLfit to par.

**Details**

The standardized deviance residuals are defined as the deviance residuals divided by  $\phi\sqrt{(1-q)}$ , where  $q$  is the corresponding leverage for  $\phi$ , and the deviance residuals are defined as for a GLM. The leverages are zero for ML methods. Otherwise, they depend on the fitting method used, as defined in the Details of [HLfit](#). The PQL and EQL- method use leverages obtained as diagonal elements of the "hat" matrix; more elaborate methods will introduce corrections for non-Gaussian response and for non-Gaussian random effects; and "(.,1)" methods will add another correction taking into account the variation of the GLM weights in the logdet Hessian term of restricted likelihood.

In principle the deviance residuals for the mean model should have a nearly Gaussian distribution hence form a nearly straight line on a Q-Q plot. However this is (trivially) not so for well-specified (nearly-)binary response data nor even for well-specified Poisson response data with moderate expectations. Hence this plot is not so useful.

**Value**

Returns the input object invisibly.

## References

Lee, Y., Nelder, J. A. and Pawitan, Y. (2006). Generalized linear models with random effects: unified analysis via h-likelihood. Chapman & Hall: London.

## Examples

```
## see example for data("scotlip")
```

---

Poisson	<i>Family function for GLMs and mixed models with Poisson and zero-truncated Poisson response.</i>
---------	--

---

## Description

Poisson (with a capital P) is a [family](#) that specifies the information required to fit a Poisson generalized linear model. Differs from the base version `stats::poisson` only in that it handles the zero-truncated variant, which can be specified either as `Tpoisson(<link>)` or as `Poisson(<link>, trunc = 0L)`.

## Usage

```
Poisson(link = "log", trunc = -1L)
Tpoisson(link="log")
```

## Arguments

link	log, sqrt or identity link, specified by the several available ways for GLM links (name, character string, one-element character vector, or object of class <code>link-glm</code> as returned by <a href="#">make.link</a> ).
trunc	Either <code>0L</code> for zero-truncated distribution, or <code>-1L</code> for default untruncated distribution.

## Value

A family object.

## References

McCullagh, P. and Nelder, J.A. (1989) Generalized Linear Models, 2nd edition. London: Chapman & Hall.

## Examples

```
data("scotlip")
logLik(glm(I(1+cases)~1,family=Tpoisson(),data=scotlip))
logLik(fitme(I(1+cases)~1+(1|id),family=Tpoisson(),fixed=list(lambda=1e-8),data=scotlip))
```

predict

*Prediction from a model fit.***Description**

Prediction of the response variable by its expected value obtained as (the inverse link transformation of) the linear predictor ( $\eta$ ) and more generally for terms of the form  $\mathbf{X}_n\beta + \mathbf{Z}_n\mathbf{L}\mathbf{v}$ , for new design matrices  $\mathbf{X}_n$  and  $\mathbf{Z}_n$ . Various components of prediction variances and predictions intervals can also be computed using predict. The `get_...` functions are convenient extractors for such components. `get_predCov_var_fix` extracts a block of a prediction covariance matrix. It was conceived for the specific purpose of computing the spatial prediction covariances between two “new” sets of geographic locations, without computing the full covariance matrix for both the new locations and the original (fitted) locations. `preprocess_fix_corr` was conceived to perform this covariance computation more efficiently when one of the two sets of new locations is fixed while the other varies.

**Usage**

```
## S3 method for class 'HLfit'
predict(object, newdata = newX, newX = NULL, re.form = NULL,
        variances=list(), binding = FALSE, intervals = NULL,
        level = 0.95, blockSize = 2000L, ...)
get_predCov_var_fix(object, newdata = NULL, fix_X_ZAC.object, fixdata,
        variances=list(dispatch=TRUE,residVar=FALSE), re.form = NULL, ...)
preprocess_fix_corr(object, fixdata, re.form = NULL,
        variances=list(residVar=FALSE))

get_fixefVar(...)
get_predVar(...)
get_residVar(...)
get_respVar(...)
get_intervals(...)
```

**Arguments**

object	The return object of fitting functions <code>HLfit</code> , <code>corrHLfit</code> , <code>HLCor...</code> returning an object inheriting from <code>HLfit</code> class.
newdata	<p><b>Either</b> <code>NULL</code>, a matrix or data frame, or a numeric vector.</p> <p>If <code>NULL</code>, the original data are reused. Otherwise, all variables required to evaluate model formulas must be included. Which variables are required may depend on other arguments: see “prediction with given phi’s” example, also illustrating the syntax when formulas include an offset.</p> <p><b>or</b> a numeric vector, which names (if any) are ignored. This makes it easier to use <code>predict</code> as an objective function for an optimization procedure such as <code>optim</code>, which calls the objective function on unnamed vectors. However, one must make sure that the order of elements in the vector is the order of first occurrence of the variables in the model formula. This order can be checked in</p>

	the error message returned when calling predict on a newX vector of clearly wrong size, e.g. <code>predict(&lt;object&gt;, newdata=numeric(0))</code> .
newX	equivalent to newdata, available for back-compatibility
re.form	formula for random effects to include. By default, it is NULL, in which case all random effects are included. If it is NA, no random effect is included. If it is a formula, only the random effects it contains are retained. The other variance components are removed from both point prediction and variances calculations. If you want to retain only the spatial effects in the point prediction, but all variances, either use re.form and add missing variances (on linear predictor scale) manually, or ignore this argument and see Details and Examples for different ways of controlling variances.
variances	<p>A list which elements control the computation of different estimated variances. In particular, <code>list(linPred=TRUE, disp=TRUE)</code> is suitable for uncertainty in point prediction.</p> <p>predict can return four components of prediction variance: <code>fixefVar</code>, <code>predVar</code>, <code>residVar</code> and <code>respVar</code>, detailed below. They are all returned as attributes of the point predictions. By default, each component is a vector of variances. However, if <code>variances\$cov=TRUE</code>, a covariance matrix is returned when applicable (i.e. not for "residVar").</p> <p><code>fixefVar</code> is the (co)variance of fixed effects (<math>\mathbf{X}\beta</math>) due to uncertainty in <math>\beta</math>. It is called by <code>variances\$fixefVar=TRUE</code>.</p> <p><code>predVar</code> is the (co)variance of the linear predictor <math>\eta</math>. It is the most common measure of uncertainty in point prediction. It accounts for uncertainty in fixed effects (<math>\mathbf{X}\beta</math>) and random effects (<math>\mathbf{ZL}\mathbf{v}</math>) for given dispersion parameters (see Details), but it can also accounts for uncertainty in dispersion parameters (<math>\lambda</math> and <math>\phi</math>) estimates if <code>variances\$disp=TRUE</code>, for models in which the effect of uncertainty in dispersion parameters can be computed. Currently, this effect can be computed for a scalar residual variance (<math>\phi</math>) and a single random effect with a scalar variance (<math>\lambda</math>). <code>variances\$predVar=TRUE</code> will return the sum of the two components, if available; otherwise it returns only the (co)variance for given <math>\lambda</math> and <math>\phi</math>. The latter component can be requested by <code>variances\$linPred=TRUE</code>.</p> <p><code>residVar</code> provides the residual variances (for Gaussian or Gamma responses). It is called by <code>variances\$residVar=TRUE</code>.</p> <p><code>respVar</code> is the variance of the response (see Details). It is called by <code>variances\$respVar=TRUE</code>. Calling for one (co)variance implies that some of its components may be also returned.</p>
intervals	NULL or character string or vector of strings. Provides prediction intervals with nominal level <code>level</code> , deduced from the given prediction variance term, e.g. <code>intervals="predVar"</code> . Currently only intervals from <code>fixefVar</code> and <code>predVar</code> (and for LMMs <code>respVar</code> including the residual variance) may have a probabilistic meaning. Intervals returned in other cases are (currently) meaningless.
level	Coverage of the intervals.
binding	If <code>binding</code> is a character string, the predicted values are bound with the newdata and the result is returned as a data frame. The predicted values column name is the given <code>binding</code> , or a name based on it if the newdata already include a variable with this name. If <code>binding</code> is FALSE, The predicted values are returned

	as a one-column matrix and the data frame used for prediction is returned as an attribute (unless it was NULL). If <code>binding</code> is NA, a vector is returned, without the previous attributes.
<code>fixdata</code>	A data frame describing reference data which covariances with variable <code>newdata</code> may be requested.
<code>fix_X_ZAC.object</code>	The return value of calling <code>preprocess_fix_corr</code> (see trivial Example). This is a more efficient way of providing information about the <code>fixdata</code> for repeated calls to <code>get_predCov_var_fix</code> with variable <code>newdata</code> .
<code>blockSize</code>	Mainly for development purposes. For original or new data with many rows, it may be more efficient to split these data in small blocks, and this gives the maximum number or rows of the blocks. However, this will be ignored if a prediction covariance matrix is requested.
<code>...</code>	further arguments passed to or from other methods. For the <code>get_...</code> functions, they are passed to <code>predict</code> .

## Details

If `newdata` is NULL, `predict` returns the fitted responses, including random effects, from the object. Otherwise it computes new predictions including random effects as far as possible. For spatial random effects it constructs a correlation matrix  $\mathbf{C}$  between new locations and locations in the original fit. Then it infers the random effects in the new locations as  $\mathbf{C}(\mathbf{L}')^{-1}\mathbf{v}$  (see `spaMM` for notation). For non-spatial random effects, it checks whether any group (i.e., level of a random effect) in the new data was represented in the original data, and it adds the inferred random effect for this group to the prediction for individuals in this group.

`fixefVar` is the (co)variance of  $\mathbf{X}\beta$  (or  $\mathbf{X}_n\beta$ ), deduced from the asymptotic covariance matrix of  $\beta$  estimates.

`predVar` is the prediction (co)variance of  $\eta=\mathbf{X}\beta+\mathbf{Z}\mathbf{v}$  (see `HLfit` Details for notation), or more generally of  $\mathbf{X}_n\beta+\mathbf{Z}_n\mathbf{L}\mathbf{v}$ , by default computed for given dispersion parameters.

For levels of the random effects present in the original data, `predVar` computation takes into account the joint uncertainty in estimation of  $\beta$  and prediction of  $\mathbf{v}$ .

For new levels of the random effects, `predVar` computation additionally takes into account uncertainty in prediction of  $\mathbf{v}$  for these new levels. For **prediction covariance** with a new  $\mathbf{Z}_n$ , it matters whether a single or multiple new levels are used: see Examples.

If `variances$disp` is TRUE, prediction variance may also include a term accounting for uncertainty in  $\phi$  and  $\lambda$ , computed following Booth and Hobert (1998, eq. 19). This computation ignores uncertainties in spatial correlation parameters.

`respVar` is the sum of `predVar` (pre- and post-multiplied by  $\partial\mu/\partial\eta$  for models with non-identity link) and of `residVar`.

These variance calculations are approximate except for LMMs, and cannot be guaranteed to give accurate results.

In the **point prediction** of the linear predictor, the unconditional expected value of  $u$  is assigned to the realizations of  $u$  for unobserved levels of non-spatial random effects (it is zero in GLMMs but not for non-gaussian random effects), and the inferred value of  $u$  is assigned in all other cases. Corresponding values of  $v$  are then deduced. This computation yields the classical “BLUP” or



empirical Bayes predictor in LMMs, but otherwise it may yield less well characterized predictors, where “unconditional”  $v$  may not be its expected value when the `rand.family` link is not identity.

**Intervals** computations use the relevant variance estimates plugged in a Gaussian approximation, except for the simple linear model where it uses Student’s  $t$  distribution.

## Value

For `predict`, a matrix or data frame (according to the binding argument), with optional attributes `frame`, `intervals`, `predVar`, `fixefVar`, `residVar`, and/or `respVar`, the last four holding one or more variance vector or covariance matrices. The further attribute `fittedName` contains the binding name, if any.

The `get_...` extractor functions call `predict` and extract from its result the attribute implied by the name of the extractor. By default, `get_intervals` will return prediction intervals using `predVar`.

## References

Booth, J.G., Hobert, J.P. (1998) Standard errors of prediction in generalized linear mixed models. *J. Am. Stat. Assoc.* 93: 262-272.

## Examples

```
data("blackcap")
fitobject <- corrHLfit(migStatus ~ 1 + Matern(1|latitude+longitude),data=blackcap,
                      ranFix=list(nu=4,rho=0.4,phi=0.05))
predict(fitobject)
getDistMat(fitobject)

#### multiple controls of prediction variances
## (1) fit with an additional random effect
grouped <- cbind(blackcap,grp=c(rep(1,7),rep(2,7)))
fitobject <- corrHLfit(migStatus ~ 1 + (1|grp) +Matern(1|latitude+longitude),
                      data=grouped, ranFix=list(nu=4,rho=0.4,phi=0.05))

## (2) re.form usage to remove a random effect from point prediction and variances:
predict(fitobject,re.form= ~ 1 + Matern(1|latitude+longitude))

## (3) comparison of covariance matrices for two types of new data
moregroups <- grouped[1:5,]
rownames(moregroups) <- paste0("newloc",1:5)
moregroups$grp <- rep(3,5) ## all new data belong to an unobserved third group
cov1 <- get_predVar(fitobject,newdata=moregroups,
                   variances=list(linPred=TRUE,cov=TRUE))
moregroups$grp <- 3:7 ## all new data belong to distinct unobserved groups
cov2 <- get_predVar(fitobject,newdata=moregroups,
                   variances=list(linPred=TRUE,cov=TRUE))
cov1-cov2 ## the expected off-diagonal covariance due to the common group in the first fit.

## Not run:
## prediction with distinct given phi's in different locations:
varphi <- cbind(blackcap,logphi=runif(14))
vphifit <- corrHLfit(migStatus ~ 1 + Matern(1|latitude+longitude),
```

```

        resid.model = list(formula=~0+offset(logphi)),
        data=varphi, ranFix=list(nu=4,rho=0.4))
# for respVar computation, one needs the resid.model formula to specify phi:
get_respVar(vphifit,newdata=data.frame(latitude=1,longitude=1,logphi=1))
# for predVar computation, phi is not needed
# (and could have been specified through ranFix):
get_predVar(vphifit,newdata=data.frame(latitude=1,longitude=1))

## Effects of numerically singular correlation matrix C:
fitobject <- corrHLfit(migStatus ~ 1 + Matern(1|latitude+longitude),data=blackcap,
        ranFix=list(nu=10,rho=0.001)) ## numerically singular C
predict(fitobject) ## predicted mu computed as X beta + L v
predict(fitobject,newdata=blackcap) ## predicted mu computed as X beta + C
#
fix_X_ZAC.object <- preprocess_fix_corr(fitobject,fixdata=blackcap)
get_predCov_var_fix(fitobject,newdata=blackcap[14,],fix_X_ZAC.object=fix_X_ZAC.object)

## point predictions and variances with new X and Z
if(requireNamespace("rsae", quietly = TRUE)) {
  data("landsat")
  fitobject <- HLfit(HACorn ~ PixelsCorn + PixelsSoybeans + (1|CountyName),
        data=landsat[-33,],HLmethod="ML")
  newXandZ <- unique(data.frame(PixelsCorn=landsat$MeanPixelsCorn,
        PixelsSoybeans=landsat$MeanPixelsSoybeans,
        CountyName=landsat$CountyName))
  predict(fitobject,newdata=newXandZ,variances = list(predVar=TRUE))
  get_predVar(fitobject,newdata=newXandZ,variances = list(predVar=TRUE))
}

## End(Not run)

```

---

 Predictor

*Interface for model formulas*


---

## Description

`Predictor(...)` performs some minimal syntax checking, and returns a formula with attributes. It serves as a unified interface for the set of descriptors for a linear predictor, including design matrices for random effects.

In the current version this function could be ignored by users.

## Usage

```
Predictor(formula, offset=NULL, LMatrix = NULL, AMatrix = NULL, ZALMatrix = NULL)
```

**Arguments**

formula	a <a href="#">formula</a> , which can include fixed effects, random effects, and offsets.
offset	a <a href="#">offset</a> can be provided in this way, as a numeric vector. However, it <b>may be better</b> to provide the offset as an offset formula term (see <a href="#">scotlip</a> example), in particular for later use with <a href="#">predict</a> where the formula can be reevaluated on new data.
LMatrix	The “square root” of the correlation matrix between unique locations, see Details.
AMatrix	A matrix that relates observed (unique) locations to unobserved locations, see Details.
ZALMatrix	Deprecated

**Details**

In a spatial model a vector of correlated random effects  $\mathbf{Lv}$  can be constructed from uncorrelated ones,  $\mathbf{v}$ , for some matrix  $\mathbf{L}$  (this may be meaningful only for Gaussian random effects). Typically  $\mathbf{L}$  is the Cholesky “square root” of a correlation matrix determined by the random effect specification (e.g., `Matern(...)`), or given as the `corrMatrix` argument of `HLCor`.

If there is one realized random effect per response value, the linear predictor contains  $\mathbf{Lv}$ , where  $\mathbf{L}$  is a square matrix which dimension is the number of observations.

Several observations may be taken in the same location, and a matrix  $\mathbf{Z}$  (usually automatically constructed) tells which element of  $\mathbf{Lv}$  affects each observation. The linear predictor then contains  $\mathbf{ZLv}$ , where  $\dim(\mathbf{Z})$  is (number of observations, number of locations).

Finally, in some applications the realized random effects in response locations may be viewed as linear combinations  $\mathbf{ALv}$  of random effects  $\mathbf{Lv}$  in distinct locations. In that case the dimension of  $\mathbf{L}$  is the number of such distinct locations,  $\mathbf{A}$  maps them to the observed locations, and  $\mathbf{Z}$  again maps them to possibly repeated observations in observed locations.

Thus, in general the random term in the linear predictor is written  $\mathbf{Mv}$ , where  $\mathbf{M}=\mathbf{ZAL}$  is reconstructed from the element matrices (usually automatically constructed if needed).

**Value**

A formula with attributes. This return object has classes `formula` and `predictor`.

**Examples**

```
# In the current version this function can be ignored by users,
# so examples are not required.
# (Use of AMatrix could perhaps be shown)
```

---

`rankinfo`*Checking the rank of the fixed-effects design matrix*

---

## Description

By default, fitting functions in `spaMM` check the rank of the design matrix for fixed effects, as `stats::lm` or `stats::glm` do (but not, say, by `nlme::lme`). This computation can be quite long. To save time when fitting different models with the same fixed-effect terms to the same data, the result of the check can be extracted from a return object by `get_rankinfo()`, and can be provided as argument `control.HLfit$rankinfo` to another fit. Alternatively, the check will not be performed if `control.HLfit$rankinfo` is set to `NA`.

## Usage

```
get_rankinfo(object)
```

## Arguments

`object`            A object of class `HLfit`, as returned by the fitting functions in `spaMM`.

## Details

The check is performed by a call to `qr()` methods for either dense or sparse matrices. If the design matrix is singular, a set of columns from the design matrix that define a non-singular matrix is identified. Note that different sets may be identified by sparse- and dense-matrix `qr` methods.

## Value

A list with elements `rank`, `whichcols` (a set of columns that define a non-singular matrix), and `method` (identifying the algorithm used).

## Examples

```
##---- Should be DIRECTLY executable !! ----
##-- ==> Define data, use random,
##--or do help(data=index) for the standard data sets.

## The function is currently defined as
function (x)
{
}
```

---

salamander

*Salamander mating data*

---

### Description

Data from a salamander mating experiment discussed by McCullagh and Nelder (1989, Ch. 14). Twenty males and twenty females from two populations (Rough Butt and Whiteside) were each paired with 6 individuals from their own or from the other population. The experiments were later published by Arnold et al. (1996).

### Usage

```
data("salamander")
```

### Format

The data frame includes 360 observations on the following variables:

**Female** Index of the female;

**Male** Index of the male;

**Mate** Whether the pair successfully mated or not;

**TypeF** Population of origin of female;

**TypeM** Population of origin of male;

**Cross** Interaction term between TypeF and TypeM;

**Season** A factor with levels Summer and Fall;

**Experiment** Index of experiment

### Source

The data frame was borrowed from the HGLMMM package (Molas and Lesaffre, 2011), version 0.1.2.

### References

Arnold, S.J., Verrell, P.A., and Tilley S.G. (1996) The evolution of asymmetry in sexual isolation: a model and a test case. *Evolution* 50, 1024-1033.

McCullagh, P. and Nelder, J.A. (1989). *Generalized Linear Models*, 2nd edition. London: Chapman & Hall.

Molas, M., Lesaffre, E. (2011) Hierarchical Generalized Linear Models: The R Package HGLMMM. *Journal of Statistical Software* 39, 1-20.

**Examples**

```

data("salamander")

## Not run:

if (spaMM.getOption("example_maxtime")>0.7) {
  HLfit(cbind(Mate,1-Mate)~TypeF+TypeM+TypeF*TypeM+(1|Female)+(1|Male),
        family=binomial(),data=salamander,HLmethod="ML",control.HLfit=list(LevenbergM=FALSE))
}

## End(Not run)
if (spaMM.getOption("example_maxtime")>0.7) {
  fitme(cbind(Mate,1-Mate)~TypeF+TypeM+TypeF*TypeM+(1|Female)+(1|Male),
        family=binomial(),data=salamander,control.HLfit=list(LevenbergM=FALSE))
}

```

---

scotlip

*Lip cancer in Scotland 1975 - 1980*


---

**Description**

This data set provides counts of lip cancer diagnoses made in Scottish districts from 1975 to 1980, and additional information relative to these data from Clayton and Kaldor (1987) and Breslow and Clayton (1993). The data set contains (for each district) counts of disease events and estimates of the fraction of the population involved in outdoor industry (agriculture, fishing, and forestry) which exposes it to sunlight.

`data("scotlip")` actually loads a data frame, `scotlip`, and an adjacency matrix, `Nmatrix`, between 56 Scottish districts, as given by Clayton and Kaldor (1987, Table 1).

**Usage**

```
data("scotlip")
```

**Format**

The data frame includes 56 observations on the following 7 variables:

**gridcode** alternative district identifier.

**id** numeric district identifier (1 to 56).

**district** district name.

**cases** number of lip cancer cases diagnosed 1975 - 1980.

**population** total person years at risk 1975 - 1980.

**prop.ag** percent of the population engaged in outdoor industry.

**exp** offsets considered by Breslow and Clayton (1993, Table 6, 'Exp' variable)

The rows are ordered according to `gridcode`, so that they match the rows of `Nmatrix`.

## References

Clayton D, Kaldor J (1987). Empirical Bayes estimates of age-standardized relative risks for use in disease mapping. *Biometrics*, 43: 671 - 681.

Breslow, NE, Clayton, DG. (1993). Approximate Inference in Generalized Linear Mixed Models. *Journal of the American Statistical Association*: 88 9-25.

## Examples

```
## see 'help(autoregressive)' for several exampels involving 'scotlip'.
```

---

seaMask	<i>Masks of seas or lands</i>
---------	-------------------------------

---

## Description

These convenient masks can be added to maps of (parts of) the world to mask map information for these areas.

## Usage

```
data("seaMask")
data("landMask")
data("worldcountries")
data("oceanmask")
```

## Format

seaMask and landMask are data frames with two variables, x and y for longitude and latitude. Its contents are suitable for use with [polypath](#): they define different polygons, each separated by a row of NAs.

worldcountries and oceanmask are SpatialPolygonsDataFrame objects.

## Details

A land mask can be produced out of worldcountries by filling the countries (i.e. fill="black" in the code for country.layer in the Examples in [http://kimura.univ-montp2.fr/~rousset/spaMM/example\\_raster.html](http://kimura.univ-montp2.fr/~rousset/spaMM/example_raster.html)).

worldcountries and oceanmask were created from public domain shapefiles downloaded from [www.naturalearth.com](http://www.naturalearth.com) on 2015/10/21. These are suitable for plots involving geographical projections not available through map, and more generally for raster plots. Only the lowest-resolution data are included in spaMM, to minimize the size of the package archive, but higher-resolution files are available on [www.naturalearth.com](http://www.naturalearth.com), from where they can be loaded as shown in the examples. worldcountries had to be edited for non-ASCII characters before inclusion in spaMM: worldcountries@data\$formal\_fr was removed and the "Côte d'Ivoire" level of some factor variables was renamed.

seaMask and landMask were created from the world map in the maps package. polypath requires polygons, while map(interior=FALSE,plot=FALSE) returns small segments. landMask is the result of reconnecting the segments into full coastlines of all land blocks.

### See Also

[http://kimura.univ-montp2.fr/~rousset/spaMM/example\\_raster.html](http://kimura.univ-montp2.fr/~rousset/spaMM/example_raster.html) for uses of worldcountries and oceanmask

### Examples

```
if (spaMM.getOption("example_maxtime")>1.1) {

data("seaMask")
## plot of predictions of behaviour for a land bird:
if (requireNamespace("maps")){
  data("blackcap")
  bfit <- corrHLfit(migStatus ~ means+ Matern(1|longitude+latitude),data=blackcap,
                    HLmethod="ML",
                    ranFix=list(lambda=0.5537,phi=1.376e-05,rho=0.0544740,nu=0.6286311))
  ## We add small masks to the points on small islands to see the predictions there
  ll <- blackcap[,c("longitude","latitude")]
  pointmask <- function(xy,r=1,npts=12) {
    theta <- 2*pi/npts *seq(npts)
    hexas <- lapply(seq(nrow(xy)), function(li){
      p <- as.numeric(xy[li,])
      hexa <- cbind(x=p[1]+r*cos(theta),y=p[2]+r*sin(theta))
      rbind(rep(NA,2),hexa) ## initial NA before each polygon
    })
    do.call(rbind,hexas)
  }
  pmasks <- pointmask(ll[c(2,4,5,6,7),],r=0.8) ## small islands only
  filled.mapMM(bfit,add.map=TRUE,
               plot.title=title(main="Inferred migration propensity of blackcaps",
                                xlab="longitude",ylab="latitude"),
               decorations=quote(points(pred[,coordinates],cex=1,pch="+")),
               plot.axes=quote({axis(1);axis(2);
                                polypath(rbind(seaMask,pmasks),border=FALSE,
                                                col="grey", rule="evenodd")
                                })))
}
}

## Not run:
# All shape files can be found here: http://www.naturalearthdata.com/downloads/
# Once downloaded, they can be loaded into R by
if (requireNamespace("rgdal", quietly = TRUE)) {
  worldcountries <- readOGR("ne_110m_admin_0_countries_lakes.shp",
                            layer="ne_110m_admin_0_countries_lakes")
}

## End(Not run)
```



---

seeds

*Seed germination data*

---

### Description

A classic toy data set, “from research conducted by microbiologist Dr P. Whitney of Surrey University. A batch of tiny seeds is brushed onto a plate covered with a certain extract at a given dilution. The numbers of germinated and ungerminated seeds are subsequently counted” (Crowder, 1978). Two seed types and two extracts are here considered in a 2x2 factorial design.

### Usage

```
data("seeds")
```

### Format

The data frame includes 21 observations on the following variables:

**plate** Factor for replication;

**seed** Seed type, a factor with two levels O73 and O75;

**extract** Root extract, a factor with two levels Bean and Cucumber;

**r** Number of seeds that germinated;

**n** Total number of seeds tested

### Source

Crowder (1978), Table 3.

### References

Crowder, M.J., 1978. Beta-binomial anova for proportions. *Appl. Statist.*, 27, 34-37.

Y. Lee and J. A. Nelder. 1996. Hierarchical generalized linear models (with discussion). *J. R. Statist. Soc. B*, 58: 619-678.

### Examples

```
data("seeds")
## An extended quasi-likelihood (EQL) fit as considered by Lee and Nelder (1996):
HLfit(cbind(r,n-r)~seed*extract+(1|plate),family=binomial(),
      rand.family=Beta(),
      HLmethod="HL(0,0)",
      data=seeds)
```

---

separation	<i>Checking separation in binomial-response models</i>
------------	--

---

### Description

Separation occurs in binomial response models when a combination of the predictor variables perfectly predict a level of the response. In such a case the estimates of the coefficients for these variables diverge to (+/-)infinity, and the numerical algorithms typically fail. To anticipate such a problem, the fitting functions in spaMM try to check for separation by default, using the lpSolveAPI package which can also detect some borderline cases (“quasi-separation”). If this package is not available, spaMM tries to use the e1071 package (in a way which will not detect quasi-separation), except for large data sets where this may take time, in which case a message notifies this to the user. The threshold size of the data is set by `spaMM.options(separation_max=<.>)`

---

simulate.HLfit	<i>Simulate realizations of a fitted model.</i>
----------------	---

---

### Description

From an HLfit object, simulate.HLfit function generates new samples given the estimated fixed effects and dispersion parameters. Simulation may be conditional on the predicted values of random effects, or unconditional. Simulations may be run for the original values of fixed-effect predictor variables and of random effect levels (spatial locations for spatial random effects), or for new values of these. However, conditional simulation in mixed models with new such values is not yet handled.

### Usage

```
## S3 method for class 'HLfit'
simulate(object, nsim = 1, seed = NULL, newdata=NULL,
         type = "marginal", conditional=NULL, verbose=TRUE,
         sizes = NULL, resp_testfn=NULL, ...)
## S3 method for class 'HLfitlist'
simulate(object, nsim = 1, seed = NULL,
         newdata=object[[1]]$data, sizes=NULL, ...)
```

### Arguments

object	The return object of HLfit or similar function.
nsim	number of response vectors to simulate. Defaults to '1'.
seed	A seed for <a href="#">set.seed</a> . If such a value is provided, the initial state of the random number generator at a global level is restored on exit from simulate.
newdata	A data frame closely matching the original data, except that response values are not needed. May provide new values of fixed predictor variables, new spatial locations, or new individuals within a block.

type	character string specifying the type of simulation for mixed models. "marginal" is for simulation from the marginal distribution of the random effect; "residual" accounts only for the residual variation of the fitted model; and "(ranef response)" accounts both for residual variation and for the conditional distribution of the random effects given the response and the point estimates of model parameters. This distribution is known exactly in LMMs, and otherwise approximated as a Gaussian distribution with mean vector and covariance matrix given as per the Laplace approximation.
conditional	Obsolete and will be deprecated. Boolean; TRUE and FALSE are equivalent to type="residual" and type="marginal", respectively.
verbose	Boolean; whether to print some information or not.
sizes	A vector of sample sizes to simulate in the case of a binomial fit. Defaults to the sizes in the original data.
resp_testfn	NULL, or a function that tests a condition which simulated samples should satisfy. This function takes a response vector as argument and return a boolean (TRUE indicating that the sampel satisfies the condition).
...	further arguments passed to or from other methods.

### Value

For the `HLfitlist` method (i.e., the result of a multinomial fit), a list of simulated responses. Otherwise, a vector (if `nsim=1`) or a matrix with `nsim` columns, each containing a simulated response.

### Examples

```
data("Loaloe")
HLC <- HLCor(cbind(npos,ntot-npos)~Matern(1|longitude+latitude),
            data=Loaloe,family=binomial(),
            ranPars=list(lambda=1,nu=0.5,rho=1/0.7))
simulate(HLC,nsim=2)
```

### Description

Fits a range of mixed models, including those with spatially correlated random effects. The random effects are either Gaussian (which defines GLMMs), or other distributions (which defines the wider class of hierarchical GLMs), or simply absent (which makes a GLM).

### Details

The standard response families gaussian, binomial, poisson, and Gamma are handled, as well as negative binomial (see [negbin](#)), zero-truncated poisson and negative binomial, and Conway-Maxwell-Poisson response (see [Tpoisson](#), [Tnegbin](#) and [COMPoisson](#)). A multi family look-alike is also available for [multinomial](#) response, with some constraints.

The package fits models including several nested or crossed random effects, one of which can be following a Matérn correlation model (see [Matern](#)), or an adjacency matrix model (see [adjacency](#)), or an AR1 model, or a given `corrMatrix`. GLMMs and HGLMs are fit via Laplace approximations for (1) the marginal likelihood with respect to random effects and (2) the restricted likelihood (as in REML), i.e the likelihood of random effect parameters given the fixed effect estimates.

The variance(s) of random effects ( $u$ ) is (are) denoted  $\lambda$  (lambda in input and output). The variance parameter of residual error is denoted  $\phi$  (phi): this is the residual variance for gaussian response, but for Gamma-distributed response, the residual variance is  $\phi\mu^2$  where  $\mu$  is expected response. A fixed-effects linear predictor for  $\phi$ , modeling heteroscedasticity, can be considered (see [Examples](#)). Fixed effects are described in the standard form  $\mathbf{X}\beta$  where  $\mathbf{X}$  is the design matrix of fixed effects and  $\beta$  (beta) is a vector of fixed effect parameters.

The structure of the random effects can generally be described by the following steps. First, independent and identically distributed (iid) random effects  $u$  are drawn from one of the following distributions: gaussian, Beta-distributed, Gamma and inverse-Gamma distributed random effects, implemented as detailed in the [HLfit](#) documentation. Second, a transformation  $\mathbf{v} = f(\mathbf{u})$  is applied ( $\mathbf{v}$  elements are still iid). Third, correlated random effects are obtained as  $\mathbf{M}\mathbf{v}$ , where the matrix  $\mathbf{M}$  can describe spatial correlation between observed locations, block effects (or repeated observations in given locations), and correlations involving unobserved locations. See [Details](#) in [Predictor](#) for the general form of  $\mathbf{M}$ . In most cases  $\mathbf{M}$  is determined from the model formula, but it can also be input directly (e.g., to describe genetic correlations).

The package has been extensively tested mainly for analysis of spatial GLMMs (Rousset and Ferdy 2014 and ongoing work), where the random effects are Gaussian. Other models have been checked against literature results and a few simulations.

### Author(s)

François Rousset and Jean-Baptiste Ferdy.

The syntax of formula terms mostly matches the one in the `lme4` package, so bits of code for parsing formulas has been recycled in `spaMM` from various versions of `lme4`.

### References

Lee, Y., Nelder, J. A. and Pawitan, Y. (2006). Generalized linear models with random effects: unified analysis via h-likelihood. Chapman & Hall: London.

Rousset F., Ferdy, J.-B. (2014) Testing environmental and genetic effects in the presence of spatial autocorrelation. *Ecography*, 37: 781-790. <http://dx.doi.org/10.1111/ecog.00566>

### See Also

`spaMM` is designed to be used through the high-level functions [corrHLfit](#), [HLCor](#), [HLfit](#), and [fixedLRT](#)

### Examples

```
## Fit a Poisson GLMM with adjacency (CAR) correlation model
# see ?adjacency for how to fit efficiently such model models
data("scotlip") ## loads 'scotlip' data frame, but also 'Nmatrix'
HLCor(cases~I(prop.ag/10) +adjacency(1|gridcode)+offset(log(expec)),
```

```

adjMatrix=Nmatrix,family=poisson(),data=scotlip)

if (spaMM.getOption("example_maxtime")>2.1) {
  ## Adding a Gamma random effect to fit a negative-binomial response:
  HLCor(cases~I(prop.ag/10) +(1|gridcode)+adjacency(1|gridcode)
        +offset(log(expec)),
        data=scotlip,family=poisson(),rand.family=list(Gamma(log),gaussian()),
        adjMatrix=Nmatrix)
}

## Not run:
## fit non-spatial crossed random effects with distinct families
data("salamander")
HLfit(cbind(Mate,1-Mate)~1+(1|Female)+(1|Male),family=binomial(),
      rand.family=list(gaussian(),Beta(logit)),data=salamander,HLmethod="ML")

## End(Not run)

## Nested effects

## Not run:
# lmer syntax allowing several degrees of nesting
HLfit(cbind(Mate,1-Mate)~1+(1|Female/Male),
      family=binomial(),rand.family=Beta(logit),data=salamander,HLmethod="ML")
# [ also allowed is cbind(Mate,1-Mate)~1+(1|Female)+(1|Male %in% Female) ]

## End(Not run)

## fit a non-spatial, Gamma GLMM:
data("wafers")
HLfit(y ~X1*X3+X2*X3+I(X2^2)+(1|batch),family=Gamma(log),
      data=wafers)

## Same with fixed-effects predictor for residual variance
## (= structured-dispersion model):
HLfit(y ~X1*X3+X2*X3+I(X2^2)+(1|batch),family=Gamma(log),
      resid.model = ~ X3+I(X3^2) ,data=wafers)

## Random-slope model (mind the output!)
if (spaMM.getOption("example_maxtime")>1) {
  HLfit(y~X1+(X2|batch),data=wafers)
}

## fit a GLM (not mixed) with structured dispersion:
HLfit( y ~X1+X2+X1*X3+X2*X3+I(X2^2),family=Gamma(log),
      resid.model = ~ X3+I(X3^2) ,data=wafers)

## Fit of binary data using PQL/L. See ?arabidopsis
## Not run:
data("arabidopsis")
HLCor(cbind(pos1046738,1-pos1046738)~seasonal+Matern(1|LAT+LONG),

```

```
ranPars=list(rho=0.129,lambda=4.28,nu=0.291),
family=binomial(),HLmethod="PQL/L",data=arabidopsis)
```

```
## End(Not run)
```

---

spaMM-conventions

*spaMM conventions and differences from related fitting procedures*


---

## Description

**input arguments** are generally similar to those of `glm` and `(g)lmer`, in particular for the `spaMM::fitme` function, with the exception of the `prior.weights` argument, which is simply `weights` in the other packages. The name `prior.weights` seems more consistent, since e.g. `glm` returns its input `weights` as output `prior.weights`, while its output `weights` are instead the weights in the final iteration of an iteratively weighted least-square fit.

The `bolddefault` likelihood target for dispersion parameters is restricted likelihood (REML estimation) for `corrHLfit` and (marginal) likelihood (ML estimation) for `fitme`. Model fits may provide restricted likelihood values (`ReL`) even if restricted likelihood is not used as an objective function at any step in the analysis.

See [good-practice](#) for advice about the proper syntax of formula.

**Computation times** depend on control parameters given by `spaMM.getOption("spaMM_tol")` parameters (for iterative algorithms), and `spaMM.getOption("nloptr")` parameters for the default optimizer. Do not use `spaMM.options()` to control them globally, unless you know what you are doing. Rather control them locally by the `control.HLfit` argument to control `spaMM_tol`, and by the control arguments of `corrHLfit` and `fitme` to control `nloptr`. If `nloptr$Xtol_rel` is set above  $5e-06$ , `fitme` will by default refit the fixed effects and dispersion parameters (but not other correlation parameters estimated by `nloptr`) by the iterative algorithm after `nloptr` convergence. Increasing `nloptr$Xtol_rel` value may therefore switches the bulk of computation time from the optimizer to the iterative algorithm, and may increase or decrease computation time depending on which algorithm is faster for a given input. Use `control$refit` if you wish to inhibit this, but note that by default it provides a rescue to a poor `nloptr` result due to a too large `Xtol_rel`.

## References

Chambers J.M. (2008) Software for data analysis: Programming with R. Springer-Verlag New York

---

spaMM.colors

*A flashy color palette.*


---

## Description

`spaMM.colors` is the default color palette for some color plots in `spaMM`.

**Usage**

```
spaMM.colors(n = 64, redshift = 1, adjustcolor_args=NULL)
```

**Arguments**

**n** Number of color levels returned by the function. A calling graphic function with argument `nlevels` will typically take the first (i.e., bluest) `nlevels` color levels. If `n < nlevels`, the color levels are recycled

**redshift** The higher it is, the more the palette blushes....

**adjustcolor\_args** Either NULL or a list of arguments for `adjustcolor`, in which case `adjustcolor` is called to modify `spaMM.colors`'s default vector of colors. See the documentation of the latter function for further information. All arguments except `col` are possible.

**Details**

If you don't like this color palette, have a look at the various ones provided by the `fields` package.

**Value**

A vector giving the colors in a hexadecimal format.

**Examples**

```
## see mapMM examples
```

---

`spaMM.filled.contour` *Level (Contour) Plots with better aspect ratio control (for geographical maps, at least)*

---

**Description**

This function is derived from `filled.contour` in the `graphics` package, and this documentation is likewise heavily based on that of `filled.contour`.

This function likewise produces a contour plot with the areas between the contours filled in solid color, and a key showing how the colors map to `z` values is likewise shown to the right of the plot. The only difference is the way the aspect ratio is determined and can be controlled (using the `map.asp` parameter instead of `asp`), They thus easily provide nice-looking maps with meaningful latitude/longitude ratio (see Examples). However, this does not work well with `rstudio`.

**Usage**

```
spaMM.filled.contour(x = seq(0, 1, length.out = nrow(z)),
                    y = seq(0, 1, length.out = ncol(z)),
                    z,
                    xrange = range(x, finite = TRUE),
                    yrange = range(y, finite = TRUE),
                    zrange = range(z, finite = TRUE),
                    margin=1/20,
                    levels = pretty(zrange, nlevels), nlevels = 20,
                    color.palette = spaMM.colors,
                    col = color.palette(length(levels) - 1),
                    plot.title, plot.axes, key.title=NULL, key.axes=NULL,
                    map.asp = NULL, xaxs = "i", yaxs = "i", las = 1,
                    axes = TRUE, frame.plot = axes, ...)
```

**Arguments**

x, y	locations of grid lines at which the values in z are measured. These must be in ascending order. (The rest of this description does not apply to <code>.filled.contour</code> .) By default, equally spaced values from 0 to 1 are used. If x is a list, its components <code>x\$x</code> and <code>x\$y</code> are used for x and y, respectively. If the list has component z this is used for z.
z	a numeric matrix containing the values to be plotted.. Note that x can be used instead of z for convenience.
xrange	x range of the plot.
yrange	y range of the plot.
zrange	z range of the plot.
margin	This controls how far (in relative terms) the plot extends beyond the x and y ranges of the analyzed points, and is overridden by explicit xrange and yrange arguments.
levels	a set of levels which are used to partition the range of z. Must be <b>strictly</b> increasing (and finite). Areas with z values between consecutive levels are painted with the same color.
nlevels	if levels is not specified, the range of z, values is divided into approximately this many levels.
color.palette	a color palette function to be used to assign colors in the plot.
col	an explicit set of colors to be used in the plot. This argument overrides any palette function specification. There should be one less color than levels
plot.title	statements which add titles to the main plot.
plot.axes	statements which draw axes (and a <code>box</code> ) on the main plot. This overrides the default axes.
key.title	statements which add titles for the plot key.
key.axes	statements which draw axes on the plot key. This overrides the default axis.



map.asp	the y/x aspect ratio of the 2D plot area (not of the full figure including the scale). Default is (plotted y range)/(plotted x range) (i.e., scales for x are identical).
xaxs	the x axis style. The default is to use internal labeling.
yaxs	the y axis style. The default is to use internal labeling.
las	the style of labeling to be used. The default is to use horizontal labeling.
axes, frame.plot	logicals indicating if axes and a box should be drawn, as in <a href="#">plot.default</a> .
...	additional <a href="#">graphical parameters</a> , currently only passed to <a href="#">title()</a> .

### Details

The values to be plotted can contain NAs. Rectangles with two or more corner values are NA are omitted entirely: where there is a single NA value the triangle opposite the NA is omitted.

Values to be plotted can be infinite: the effect is similar to that described for NA values.

### Note

Builds heavily on `filled.contour` by Ross Ihaka and R-core. `spaMM.filled.contour` uses the [layout](#) function and so is restricted to a full page display.

The output produced by `spaMM.filled.contour` is actually a combination of two plots; one is the filled contour and one is the legend. Two separate coordinate systems are set up for these two plots, but they are only used internally – once the function has returned these coordinate systems are lost. If you want to annotate the main contour plot, for example to add points, you can specify graphics commands in the `plot.axes` argument. See the Examples.

### References

Cleveland, W. S. (1993) *Visualizing Data*. Summit, New Jersey: Hobart.

### See Also

[contour](#), [image](#), [palette](#); [contourplot](#) and [levelplot](#) from package `lattice`.

### Examples

```
spaMM.filled.contour(volcano, color = spaMM.colors) # simple

## Comparing the layout with that of filled.contour:
# (except that it does not always achieve the intended effect
# in RStudio Plots pane).

x <- 10*1:nrow(volcano)
y <- 10*1:ncol(volcano)
spaMM.filled.contour(x, y, volcano, color = terrain.colors,
  plot.title = title(main = "The Topography of Maunga Whau",
    xlab = "Meters North", ylab = "Meters West"),
  plot.axes = { axis(1, seq(100, 800, by = 100))
    axis(2, seq(100, 600, by = 100)) },
  key.title = title(main = "Height\n(meters)"),
```

```

    key.axes = axis(4, seq(90, 190, by = 10)) # maybe also asp = 1
mtext(paste("spaMM.filled.contour(.) from", R.version.string),
      side = 1, line = 4, adj = 1, cex = .66)

## compare with

filled.contour(x, y, volcano, color = terrain.colors,
  plot.title = title(main = "The Topography of Maunga Whau",
    xlab = "Meters North", ylab = "Meters West"),
  plot.axes = { axis(1, seq(100, 800, by = 100))
    axis(2, seq(100, 600, by = 100)) },
  key.title = title(main = "Height\n(meters)"),
  key.axes = axis(4, seq(90, 190, by = 10)) # maybe also asp = 1
mtext(paste("filled.contour(.) from", R.version.string),
      side = 1, line = 4, adj = 1, cex = .66)

```

---

spaMM\_boot

*Parametric bootstrap*


---

## Description

This simulates samples from a fit object inheriting from class "HLfit", as produced by spaMM's fitting function, and applies a given function to each simulated sample. Parallelization is supported (see Details). A typical usage of the parametric bootstrap is to fit by one model samples produced under another model (see Example). spaMM\_boot provides more control on what is computed on each bootstrap replicate than (say) fixedLRT.

## Usage

```

spaMM_boot(object, simuland, nsim, nb_cores = NULL, resp_testfn=NULL,
  control.foreach=list(), ...)

```

## Arguments

object	The fit object to simulate from.
simuland	The function to apply to each simulated sample. See Details for requirements of this function.
nsim	Number of samples to simulate and analyze.
nb_cores	Number of cores to use for parallel computation. The default is spaMM.getOption("nb_cores"), and 1 if the latter is NULL. nb_cores=1 prevents the use of parallelisation procedures.
resp_testfn	Passed to simulate.HLfit; NULL, or a function that tests a condition which simulated samples should satisfy. This function takes a response vector as argument and return a boolean (TRUE indicating that the sampel satisfies the condition).

```
control.foreach
```

list of control arguments for foreach. These include in particular `.combine` (with default value `"rbind"`), and `.errorhandling` (with default value `"remove"`, but `"pass"` is quite useful for debugging).

```
...
```

Further arguments passed to the `simuland` function.

## Details

The `simuland` function must take as first argument a vector of response values, and must have a `...` argument. `spaMM_boot` calls `simulate.HLfit` on the fit object and applies `simuland` on each column of the matrix returned by this call.

This function handles parallel backends with different features. `pbapply::pbapply` has a very simple interface (essentially equivalent to `apply`) and provides progress bars, but (currently: version 1.3.4) does not have efficient load-balancing. `doSNOW` also provides a progress bar and allows more efficient load-balancing, but it requires `foreach`, whose handling of `'...'` arguments is tortuous. `foreach` will be used if `doSNOW` is loaded; then, some of the `'...'` arguments may need to be quoted (see Example).

## Value

A list with two elements: `bootreps`, `nsim` return values in the format returned either by `apply` or `parallel::parApply` or by `foreach::%dopar%` as controlled by `control.foreach$.combine` (these formats typically differ from each other); and `RNGstate`, the state of `.Random.seed` at the beginning of the simulation.

## Examples

```
if (spaMM.getOption("example_maxtime")>10) {
  data("blackcap")

  # Generate fits of null and full models:
  lrt <- fixedLRT(null.formula=migStatus ~ 1 + Matern(1|latitude+longitude),
                 formula=migStatus ~ means + Matern(1|latitude+longitude),
                 HLmethod='ML',data=blackcap)

  # The 'simuland' argument:
  myfun <- function(y, what=NULL, lrt, ...) {
    data <- lrt$fullfit$data
    data$migStatus <- y ## replaces original response (! more complicated for binomial fits)
    full_call <- getCall(lrt$fullfit) ## call for full fit
    full_call$data <- data
    res <- eval(full_call) ## fits the full model on the simulated response
    if (! deparse(substitute(what))=="NULL") ## post-process the fit
      res <- eval(eval(substitute(what),list(res=res)))
    # : when only pbapply is used, the following code suffices:
    # if (!is.null(what)) res <- eval(what)
    return(res) ## the fit, or anything produced by evaluating 'what'
  }
  # where the 'what' argument (not required) of myfun() allows one to control
  # what the function returns without redefining the function.
```

```

# Call myfun() with no 'what' argument: returns a list of fits
spaMM_boot(lrt$nullfit, simuland = myfun, nsim=1, lrt=lrt)[["bootreps"]]

# Return only a model coefficient for each fit:
spaMM_boot(lrt$nullfit, simuland = myfun, nsim=7,
           what=quote(fixef(res)[2L]), lrt=lrt)[["bootreps"]]
}

```

---

spaMM_glm.fit	<i>Fitting generalized linear models without initial-value or divergence headaches</i>
---------------	--

---

## Description

spaMM\_glm.fit is a stand-in replacement for glm.fit, which can be called through glm by using glm(<>, method="spaMM\_glm.fit"). Input and output structure are exactly as for glm.fit. It uses a Levenberg-Marquardt algorithm to prevent divergence of estimates. If the rcdd package is installed, the function can automatically find valid starting values or else indicate that no parameter value is feasible. spaMM\_glm is a convenient wrapper, calling glm with default method glm.fit, then calling method spaMM\_glm.fit, with possibly different initial values, if glm.fit failed.

## Usage

```

spaMM_glm.fit(x, y, weights = rep(1, nobs), start = NULL, etastart = NULL,
             mustart = NULL, offset = rep(0, nobs), family = gaussian(),
             control = list(maxit=200), intercept = TRUE, singular.ok = TRUE)
spaMM_glm(formula, family = gaussian, data, weights, subset,
          na.action, start = NULL, etastart, mustart, offset,
          control = list(...), model = TRUE, method = c("glm.fit", "spaMM_glm.fit"),
          x = FALSE, y = TRUE, singular.ok = TRUE, contrasts = NULL, strict=FALSE, ...)

```

## Arguments

All arguments except strict are common to these functions and their stats package equivalents, glm and glm.fit. Most arguments operate as for the latter functions, whose documentation is repeated below. The control argument may operate differently.

an object of class "formula" (or one that can be coerced to that class): a symbolic description of the model to be fitted. The details of model specification are given in the 'Details' section of glm.

**family** a description of the error distribution and link function to be used in the model. For spaMM\_glm this can be a character string naming a family function, a family function or the result of a call to a family function. For spaMM\_glm.fit only the third option is supported. (See family for details of family functions.)

**data** an optional data frame, list or environment (or object coercible by as.data.frame to a data frame) containing the variables in the model. If not found in data, the variables are taken from environment(formula), typically the environment from which glm is called.

weights	an optional vector of ‘prior weights’ to be used in the fitting process. Should be NULL or a numeric vector.
subset	an optional vector specifying a subset of observations to be used in the fitting process.
na.action	a function which indicates what should happen when the data contain NAs. The default is set by the na.action setting of <code>options</code> , and is <code>na.fail</code> if that is unset. The ‘factory-fresh’ default is <code>na.omit</code> . Another possible value is NULL, no action. Value <code>na.exclude</code> can be useful.
start	starting values for the parameters in the linear predictor.
etastart	starting values for the linear predictor.
mustart	starting values for the vector of means.
offset	this can be used to specify an <i>a priori</i> known component to be included in the linear predictor during fitting. This should be NULL or a numeric vector of length equal to the number of cases. One or more <code>offset</code> terms can be included in the formula instead or as well, and if more than one is specified their sum is used. See <code>model.offset</code> .
control	a list of parameters for controlling the fitting process. This is passed to <code>glm.control</code> , as for <code>glm.fit</code> . Because one can assume that <code>spaMM_glm.fit</code> will converge in many cases where <code>glm.fit</code> does not, <code>spaMM_glm.fit</code> allows more iterations (200) by default. However, if <code>spaMM_glm.fit</code> is called through <code>glm(. . . , method="spaMM_glm.fit")</code> then the number of iterations is controlled by the <code>glm.control</code> call within <code>glm</code> , so that it is 25 by default, overriding the <code>spaMM_glm.fit</code> default.
model	a logical value indicating whether <i>model frame</i> should be included as a component of the returned value.
method	A 2-elements vector specifying first the method to be used by <code>spaMM_glm</code> in the first attempt at fitting the model, second the method to be used in a second attempt if the first failed. Possible methods include those shown in the default, <code>"model.frame"</code> , which returns the model frame and does no fitting, or user-supplied fitting functions. These functions can be supplied either as a function or a character string naming a function, with a function which takes the same arguments as <code>glm.fit</code> .
x, y	For <code>spaMM_glm</code> : x is a design matrix of dimension $n * p$ , and y is a vector of observations of length n. For <code>spaMM_glm.fit</code> : x is a design matrix of dimension $n * p$ , and y is a vector of observations of length n.
singular.ok	logical; if FALSE a singular fit is an error.
contrasts	an optional list. See the <code>contrasts.arg</code> of <code>model.matrix.default</code> .
intercept	logical. Should an intercept be included in the <i>null</i> model?
strict	logical. Whether to perform a fit by <code>spaMM_glm.fit</code> if <code>glm.fit</code> returned the warning "glm.fit: algorithm did not converge".
...	arguments to be used to form the default <code>control</code> argument if it is not supplied directly.

**Value**

An object inheriting from class `glm`. See [glm](#) for details.

**Note**

The source and documentation is derived in large part from those of `glm.fit`.

**Examples**

```
# glm() failures last checked in R-devel version (2018-01-06 r74073)
x <- c(8.752,20.27,24.71,32.88,27.27,19.09)
y <- c(5254,35.92,84.14,641.8,1.21,47.2)
## Not run: # fails:
  glm(y~ x,data=data.frame(x,y),family=Gamma(log))

## End(Not run)
spaMM_glm(y~ x,data=data.frame(x,y),family=Gamma(log))

## Gamma(inverse) examples
x <- c(43.6,46.5,21.7,18.6,17.3,16.7)
y <- c(2420,708,39.6,16.7,46.7,10.8)
## Not run: # fails (can't find starting value)
  glm(y~ x,data=data.frame(x,y),family=Gamma())

## End(Not run)
if (requireNamespace("rcdd",quietly=TRUE)) {
  spaMM_glm(y~ x,data=data.frame(x,y),family=Gamma())
}
```

---

sparse\_precision

*Sparse\_precision algorithm*

---

**Description**

A fitting algorithm efficient for random effects with sparse precision matrix (i.e. inverse covariance matrix) is implemented. It is used by default only in two cases: for conditional autoregressive models (with a random effect of the form `adjacency(1|<grouping factor>)`), and when the `covStruct` syntax is used to provide a fixed precision matrix (see [pedigree](#) for an example). A non-default choice of fitting algorithm can be selected in this and other models by using `spaMM.options(sparse_precision= <TRUE|FALSE>)` with often poor results.

A precision matrix is meaningful for a Gaussian random effect, but beyond this the algorithm works for HGLMs, i.e. the model may include another random effect with non-Gaussian distribution.

---

stripHLfit	<i>Reduce the size of fitted objects</i>
------------	--

---

### Description

Large matrices and other memory-expensive objects may be stored in a fit object. This function removes them in order to reduce the size of the object, particularly when stored on disk. In principle, the removed objects can be regenerated automatically when needed (e.g., for a predict()).

### Usage

```
stripHLfit(object, ...)
```

### Arguments

object	The result of a fit (an object of class HLfit).
...	Further arguments, not currently used.

### Value

The input fit objects with some elements removed.

### Note

The effect may change without notice between versions as the efficiency of the operation is highly sensitive to implementation details.

### Examples

```
## Not run:
## rather unconvincing example : quantitative effect is small.

# measure size of saved object:
saveSize <- function (object,...) {
  tf <- tempfile(fileext = ".RData")
  on.exit(unlink(tf))
  save(object, file = tf,...)
  file.size(tf)
}
data("Loaloo")
lfit <- fitme(cbind(npos,ntot-npos)~elev1+elev2+elev3+elev4+maxNDVI1+seNDVI
             +Matern(1|longitude+latitude), method="HL(0,1)",
             data=Loaloo, family=binomial(), fixed=list(nu=0.5,rho=1,lambda=0.5))
saveSize(lfit)
pfit <- predict(lfit,newdata=Loaloo,variances=list(cov=TRUE)) # increases size!
saveSize(lfit)
lfit <- stripHLfit(lfit)
saveSize(lfit)

## End(Not run)
```

---

summary.HLfit

*Summary and print methods for fit and test results.*


---

### Description

Summary and print methods for results from HLfit or related functions.

### Usage

```
## S3 method for class 'HLfit'
summary(object, details=FALSE, max.print=100L, ...)
## S3 method for class 'HLfitlist'
summary(object, ...)
## S3 method for class 'fixedLRT'
summary(object, verbose=TRUE, ...)
## S3 method for class 'HLfit'
print(x,...)
## S3 method for class 'HLfitlist'
print(x,...)
## S3 method for class 'fixedLRT'
print(x,...)
```

### Arguments

object	The return object of HLfit or related functions.
x	The return object of HLfit or related functions.
verbose	for summary.fixedLRT, whether to print the model fits or not.
max.print	Controls options("max.print") locally.
details	whether to print some obscure details. Currently affects only random-coefficients models (see Details).
...	further arguments passed to or from other methods.

### Details

The random effect terms of the linear predictor are of the form  $\mathbf{ZL}\mathbf{v}$ . In particular, for **random-coefficients models** (i.e., including random-effect terms such as (z|group) specifying a random-slope component), correlated random effects are represented as  $\mathbf{b} = \mathbf{L}\mathbf{v}$  for some matrix  $\mathbf{L}$ , and where the elements of  $\mathbf{v}$  are uncorrelated. In the output of the fit, the Var. column gives the variances of the correlated effects,  $\mathbf{b}=\mathbf{L}\mathbf{v}$ . The Corr. column(s) give their correlation(s). If details is TRUE, estimates and SEs of the (log) variances of the elements of  $\mathbf{v}$  are reported as for other random effects in the Estimate and cond. SE. columns of the table of lambda coefficients. However, this non-default output is potentially misleading as the elements of  $\mathbf{v}$  cannot generally be assigned to specific terms (such as intercept and slope) of the random-effect formula, and the representation of  $\mathbf{b}$  as  $\mathbf{L}\mathbf{v}$  is not unique.



**Value**

These methods return the object invisibly. They print details of the (lower level) HLfit results in a convenient form.

**Examples**

```
## see examples of corrHLfit usage
```

---

 sym\_eigen

*Singular Value Decomposition of a Symmetric Matrix*


---

**Description**

Computes the symmetric eigenvalue decomposition of a symmetric matrix,  $X = U.D.U'$ , where  $U$  is a matrix of orthogonal eigenvectors and  $D$  is a diagonal matrix of eigenvalues. The result is similar to that of `eigen(., symmetric=TRUE)` but its evaluation uses the `SelfAdjointEigenSolver` function from the Eigen C++ library. No symmetry check is performed.

**Usage**

```
sym_eigen(X)
```

**Arguments**

`X` A symmetric matrix (possibly in `sparseMatrix` format).

**Value**

A list with members `u` (matrix of eigenvectors) and `d` (vector of eigenvalues).

**Examples**

```
hilbert <- function(n) { i <- 1:n; 1 / outer(i - 1, i, "+") }
X <- hilbert(9)
s <- sym_eigen(X)
range(s$u %*% diag(s$d) %*% t(s$u) - X) # X=U D U'
```

---

update.HLfit	<i>Updates an HLCor or HLfit fit</i>
--------------	--------------------------------------

---

### Description

update will update and (by default) re-fit a model. It does this mostly by extracting the call stored in the object, updating the call and evaluating that call. (however, currently the predictor argument is processed differently). Using update is a risky programming style (see Note).

### Usage

```
## S3 method for class 'HLfit'
update(object, formula., ..., evaluate = TRUE)
```

### Arguments

object	A return object from an HLfit call.
formula.	Changes to the formula. Beware of the syntax: see <a href="#">update.formula</a> for details.
...	Additional arguments to the call, or arguments with changed values. Use name = NULL to remove the argument name.
evaluate	If TRUE, evaluate the new call else return the call.

### Value

An HLCor or HLfit fit of the same type as the input object.

### Note

update, as a general rule, is tricky. For example with `m1 <- glm(y~x + offset(<...>), <...>)`; `m2 <- update(m1, .~.~)` `m2` will lose the intercept term, which would be retained in the absence of an offset. `update.HLfit` has the same behaviour. Further, update is easily affected in a non-transparent way by changes in variables used in the original call. For example `foo <- rep(1,10)` `m <- lm(rnorm(10)~1, weights=foo)` `rm(foo)` `update(m, .~.)` # Error To avoid such problems, spaMM tries to avoid references to variables in the global environment, by enforcing that the data are explicitly provided by the data argument, and that prior weights, if not constant, are in the data.

### See Also

See also [HLCor](#), [HLfit](#).

## Examples

```

data("wafers")
## First the fit to be updated:
wFit <- HLFit(y ~X1*X3+X2*X3+I(X2^2)+(1|batch),family=Gamma(log),
             resid.model = ~ X3+I(X3^2) ,data=wafers)

# For estimates given by Lee et al., Appl. Stochastic Models Bus. Ind. (2011) 27: 315-328:
# Refit with given beta or/and phi values:

betavals <- c(5.55,0.08,-0.14,-0.21,-0.08,-0.09,-0.09)
# reconstruct fitted phi value from predictor for log(phi)
Xphi <- with(wafers,cbind(1,X3,X3^2)) ## design matrix
phifit <- exp(Xphi %*% c(-2.90,0.1,0.95))
update(wFit,formula.= . ~ offset(wFit$`X.pv` %*% betavals)+(1|batch),
       ranFix=list(lambda=exp(-3.67),phi=phifit))

## There are subtlety in performing REML fits of constrained models,
## illustrated by thefact thatthe following fir does not recover
## the original likelihood values, because dispersion parameters are
## estimated but the REML correction changes with the formula:
update(wFit,formula.= . ~ offset(wFit$`X.pv` %*% fixef(wFit))+(1|batch))
## To maintain the original REML correction, Consider instead
update(wFit,formula.= . ~ offset(wFit$`X.pv` %*% fixef(wFit))+(1|batch),
       REMLformula=wFit$predictor) ## recover original p_v and p_bv
## Alternatively, show original wFit as differences from betavals:
update(wFit,formula.= . ~ . +offset(wFit$`X.pv` %*% betavals))

```

---

wafers

*Data from a resistivity experiment for semiconductor materials.*

---

## Description

This data set was reported and analyzed by Robinson et al. (2006) and reanalyzed by Lee et al. (2011). The data “deal with wafers in a single etching process in semiconductor manufacturing. Wafers vary through time since there are some variables that are not perfectly controllable in the etching process. For this reason, wafers produced on any given day (batch) may be different from those produced on another day (batch). To measure variation over batch, wafers are tested by choosing several days at random. In this data, resistivity is the response of interest. There are three variables, gas flow rate (x1), temperature (x2), and pressure (x3) and one random effect (batch or day).” (Lee et al 2011).

## Usage

```
data("wafers")
```

**Format**

The data frame includes 198 observations on the following variables:

**y** resistivity.

**batch** batch, indeed.

**X1** gas flow rate.

**X2** temperature.

**X3** pressure.

**Source**

This data set was manually pasted from Table 3 of Lee et al. (2011). Transcription errors may have occurred.

**References**

Robinson TJ, Wulff SS, Montgomery DC, Khuri AI. 2006. Robust parameter design using generalized linear mixed models. *Journal of Quality Technology* 38: 38–65.

Lee, Y., Nelder, J.A., and Park, H. 2011. HGLMs for quality improvement. *Applied Stochastic Models in Business and Industry* 27, 315-328.

**Examples**

```
## see examples in the main Documentation page for the package.
```

---

welding

*Welding data set*

---

**Description**

The data give the results of an unreplicated experiment for factors affecting welding quality conducted by the National Railway Corporation of Japan (Taguchi and Wu, 1980, cited in Smyth et al., 2001). It is a toy example for heterocedastic models and is also suitable for illustrating fit of overparameterized models.

**Usage**

```
data("welding")
```

**Format**

The data frame includes 16 observations on 10 variables:

**Strength** response variable;

... nine two-level factors.

**Source**

The data were downloaded from <http://www.statsci.org/data/general/welding.txt> on 2014/08/19 and are consistent with those shown in table 5 of Bergman and Hynén (1997).

**References**

Bergman B, Hynén A (1997) Dispersion effects from unreplicated designs in the  $2^{k-p}$  series. *Technometrics*, 39, 191–98.

Smyth GK, Huele AF, Verbyla AP (2001). Exact and approximate REML for heteroscedastic regression. *Statistical Modelling* 1, 161-175.

Taguchi G, Wu Y (1980) Introduction to off-line quality control. Nagoya, Japan: Central Japan Quality Control Association.

**Examples**

```
data("welding")
## toy example from Smyth et al.
HLfit(Strength ~ Drying + Material, resid.model = ~ Material+Preheating ,data=welding)
## toy example of overparameterized model
HLfit(Strength ~ Rods+Thickness*Angle+(1|Rods), resid.model = ~ Rods+Thickness*Angle ,data=welding)
```

# Index

## \*Topic **\textasciitildekwd1**

rankinfo, [68](#)  
stripHLfit, [87](#)  
sym\_eigen, [89](#)

## \*Topic **\textasciitildekwd2**

rankinfo, [68](#)  
stripHLfit, [87](#)  
sym\_eigen, [89](#)

## \*Topic **datagen**

simulate.HLfit, [74](#)

## \*Topic **datasets**

adjlg, [3](#)  
arabidopsis, [4](#)  
blackcap, [7](#)  
freight, [30](#)  
Loaloa, [41](#)  
salamander, [69](#)  
scotlip, [70](#)  
seaMask, [71](#)  
seeds, [73](#)  
wafers, [91](#)  
welding, [92](#)

## \*Topic **family**

multinomial, [52](#)

## \*Topic **hplot**

mapMM, [46](#)  
plot.HLfit, [59](#)

## \*Topic **htest**

fixedLRT, [27](#)  
LRT, [43](#)  
spaMM\_boot, [82](#)

## \*Topic **manip**

multinomial, [52](#)

## \*Topic **models**

COMPoisson, [10](#)  
negbin, [54](#)  
Poisson, [61](#)  
spaMM\_glm.fit, [84](#)

## \*Topic **model**

corrHLfit, [15](#)

fitme, [24](#)

HLCor, [32](#)

HLfit, [34](#)

make\_scaled\_dist, [45](#)

multinomial, [52](#)

## \*Topic **package**

spaMM, [75](#)

## \*Topic **print**

summary.HLfit, [88](#)

## \*Topic **regression**

COMPoisson, [10](#)

negbin, [54](#)

Poisson, [61](#)

spaMM\_glm.fit, [84](#)

## \*Topic **spatial**

spaMM, [75](#)

adjacency, [27](#), [32](#), [33](#), [76](#)

adjacency (autoregressive), [6](#)

adjlg, [3](#)

adjlgMat (adjlg), [3](#)

adjustcolor, [79](#)

AIC (extractors), [21](#)

anova (LRT), [43](#)

AR1, [27](#), [33](#)

AR1 (autoregressive), [6](#)

arabidopsis, [4](#), [56](#)

as.data.frame, [84](#)

autoregressive, [6](#), [33](#)

besselK, [51](#)

Beta (HLfit), [34](#)

Beta-distribution-random-effects  
(HLfit), [34](#)

binomialize (multinomial), [52](#)

blackcap, [7](#)

bobyqa, [56](#)

box, [80](#)

- CAR (autoregressive), 6
- Cauchy (CauchyCorr), 9
- CauchyCorr, 9
- chol, 20
- coef.corMatern (corMatern), 13
- coef<- .corMatern (corMatern), 13
- COMPOisson, 10, 55, 56, 75
- confint (confint.HLfit), 12
- confint.HLfit, 12
- contour, 48, 81
- contourplot, 81
- corFactor.corMatern (corMatern), 13
- corMatern, 13, 51
- corMatrix.corMatern (corMatern), 13
- Corr (extractors), 21
- corrHLfit, 15, 29, 33, 34, 39, 53, 76
- corrMatrix, 17, 33
- covStruct, 18, 33, 57, 86
  
- designL.from.Corr, 16, 19, 25
- dev\_resids (extractors), 21
- deviance (extractors), 21
- dist, 46
  
- Earth (make\_scaled\_dist), 45
- eigen, 20
- etaFix, 36
- etaFix (fixed), 26
- extractors, 21, 38
  
- family, 10, 16, 24, 54, 61, 84
- filled.mapMM (mapMM), 46
- fitme, 24, 34, 39, 58
- fitted (extractors), 21
- fitted.HLfitlist (multinomial), 52
- fitted.values, 21
- fixed, 24, 26
- fixedLRT, 16, 27, 44, 76
- fixef (extractors), 21
- formula, 15, 24, 35, 67, 84
- freight, 30
  
- Gamma, 38, 40
- Gamma (inverse.Gamma), 40
- geometric (COMPOisson), 10
- get\_any\_IC, 38
- get\_any\_IC (extractors), 21
- get\_boot\_response (extractors), 21
- get\_fixefVar (predict), 62
- get\_intervals (predict), 62
- get\_predCov\_var\_fix (predict), 62
- get\_predVar (predict), 62
- get\_rankinfo (rankinfo), 68
- get\_ranPars, 31
- get\_residVar (predict), 62
- get\_respVar (predict), 62
- get\_RLRTSim\_args (extractors), 21
- get\_ZALMatrix, 39
- get\_ZALMatrix (extractors), 21
- getCovariate.corMatern (corMatern), 13
- getDistMat (extractors), 21
- glm, 10, 36, 37, 84, 86
- glm.control, 36, 85
- glmmPQL, 14
- good-practice, 31
- graphical parameters, 81
  
- HLCor, 16, 25, 32, 39, 59, 76, 90
- HLfit, 15, 16, 21, 22, 25, 28, 32, 33, 34, 59, 60, 64, 76, 90
  
- image, 81
- Initialize.corMatern (corMatern), 13
- intervals (predict), 62
- inverse.Gamma, 38, 40
  
- landMask (seaMask), 71
- layout, 81
- levelplot, 81
- LevenbergM (options), 55
- lme, 14
- Loaloo, 16, 41, 51
- logDet.corMatern (corMatern), 13
- logLik, 37
- logLik (extractors), 21
- logLik.HLfitlist (multinomial), 52
- lower.tri, 26
- LRT, 29, 43
  
- make.link, 54, 61
- make\_scaled\_dist, 15, 24, 33, 45
- mapMM, 46
- Matern, 13, 15, 24, 26, 76
- Matern (MaternCorr), 50
- MaternCorr, 13, 14, 33, 50
- model.offset, 85
- multi, 16, 24
- multi (multinomial), 52

- multinomial, [52, 75](#)
- na.exclude, [85](#)
- na.fail, [85](#)
- na.omit, [85](#)
- negbin, [54, 75](#)
- nloptr, [25, 56](#)
- Nmatrix (scotlip), [70](#)
- oceanmask (seaMask), [71](#)
- offset, [67, 85](#)
- optim, [56](#)
- options, [55, 85](#)
- palette, [81](#)
- pedigree, [19, 57, 86](#)
- phiHGLM, [35, 58](#)
- plot (plot.HLfit), [59](#)
- plot.default, [81](#)
- plot.HLfit, [59](#)
- Poisson, [61](#)
- polypath, [71](#)
- predict, [62, 67](#)
- Predictor, [15, 24, 32, 35, 37, 66, 76](#)
- preprocess\_fix\_corr (predict), [62](#)
- pretty, [48](#)
- print (summary.HLfit), [88](#)
- print.ranef (extractors), [21](#)
- ranCoefs (fixed), [26](#)
- ranef (extractors), [21](#)
- ranFix, [15, 36](#)
- ranFix (fixed), [26](#)
- rankinfo, [68](#)
- ranPars, [32](#)
- ranPars (fixed), [26](#)
- recalc.corMatern (corMatern), [13](#)
- residuals (extractors), [21](#)
- residuals.glm, [22](#)
- rho.mapping (make\_scaled\_dist), [45](#)
- salamander, [69](#)
- scotlip, [67, 70](#)
- seaMask, [71](#)
- seeds, [73](#)
- separation, [74](#)
- set.seed, [74](#)
- simulate (simulate.HLfit), [74](#)
- simulate.HLfit, [74](#)
- spaMM, [15, 24, 35, 64, 75](#)
- spaMM-conventions, [78](#)
- spaMM-package (spaMM), [75](#)
- spaMM.colors, [78](#)
- spaMM.filled.contour, [49, 79](#)
- spaMM.getOption (options), [55](#)
- spaMM.options, [16](#)
- spaMM.options (options), [55](#)
- spaMM\_boot, [82](#)
- spaMM\_glm (spaMM\_glm.fit), [84](#)
- spaMM\_glm.fit, [84](#)
- spaMMplot2D (mapMM), [46](#)
- sparse\_precision, [56, 86](#)
- sparseMatrix, [89](#)
- stripHLfit, [87](#)
- summary (summary.HLfit), [88](#)
- summary.HLfit, [88](#)
- svd, [20](#)
- sym\_eigen, [89](#)
- title, [81](#)
- Tnegbin, [75](#)
- Tnegbin (negbin), [54](#)
- Tpoisson, [75](#)
- Tpoisson (Poisson), [61](#)
- update (update.HLfit), [90](#)
- update.formula, [90](#)
- update.HLfit, [31, 90](#)
- Variogram.corMatern (corMatern), [13](#)
- vcov, [38](#)
- vcov (extractors), [21](#)
- wafers, [91](#)
- welding, [92](#)
- worldcountries (seaMask), [71](#)