Package 'georob'

September 2, 2015

Type Package

Title Robust Geostatistical Analysis of Spatial Data

Version 0.2-1

Date 2015-09-02

Depends R(>= 2.14.0), sp(>= 0.9-60)

Imports constrainedKriging(>= 0.2-1), lmtest, methods, nlme, nleqslv, quantreg, RandomFields(>= 3.0.10), robustbase(>= 0.90-2), snowfall(>= 1.84-6)

Suggests geoR

Description The georob package provides functions for efficiently fitting linear models with spatially correlated errors by robust and Gaussian (Restricted) Maximum Likelihood and for computing robust and customary point and block external-drift kriging predictions, along with utility functions for variogram modelling in ad-hoc geostatistical analyses, model building, model evaluation by cross-validation and for unbiased back-transformation of kriging predictions of log-transformed data.

License GPL (>= 2)

Author Andreas Papritz [cre, aut], Cornelia Schwierz [ctb]

Maintainer Andreas Papritz <andreas.papritz@env.ethz.ch>

NeedsCompilation no

Repository CRAN

Date/Publication 2015-09-02 15:52:13

R topics documented:

`georob-package` *The georob Package*

Description

This is a summary of the features and functionality of **georob**, a package in R for robust geostatistical analyses.

Details

georob is a package for robust analyses of geostatistical data. Such data, say $y_i = y(\boldsymbol{s}_i)$, are recorded at a set of locations, $\boldsymbol{s}_i$, $i = 1, 2, \ldots, n$, in a domain $G \in \mathbb{R}^d$, $d \in (1, 2, 3)$, along with covariate information $x_j(\boldsymbol{s}_i)$, $j = 1, 2, \ldots, p$.

Model: We use the following model for the data

$$Y(\boldsymbol{s}) = Z(\boldsymbol{s}) + \varepsilon(\boldsymbol{s}) = \boldsymbol{x}(\boldsymbol{s})^{\mathrm{T}}\boldsymbol{\beta} + B(\boldsymbol{s}) + \varepsilon(\boldsymbol{s}),$$

where $Z(\boldsymbol{s}) = \boldsymbol{x}(\boldsymbol{s})^{\mathrm{T}}\boldsymbol{\beta} + B(\boldsymbol{s})$ is the so-called signal, $\boldsymbol{x}(\boldsymbol{s})^{\mathrm{T}}\boldsymbol{\beta}$ is the external drift, $B(\boldsymbol{s})$ is a stationary or intrinsic spatial Gaussian random field with zero mean, and $\varepsilon(\boldsymbol{s})$ are *i.i.d* errors from a possibly long-tailed distribution with scale parameter τ (τ^2 is usually called nugget effect). In vector form the model leads to

$$\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{B} + \boldsymbol{\varepsilon},$$

where $\boldsymbol{X}$ is the model matrix consisting of the rows $\boldsymbol{x}(\boldsymbol{s}_i)^{\mathrm{T}}$.

The (generalized) covariance matrix of the vector of unobserved spatial Gaussian random effects $\boldsymbol{B}$ is denoted by

$$\mathrm{E}[\boldsymbol{B}\,\boldsymbol{B}^{\mathrm{T}}] = \boldsymbol{\Gamma}_\theta = \sigma^2_{\mathrm{n}}\boldsymbol{I} + \sigma^2\boldsymbol{V}_\alpha = \sigma^2_Z\boldsymbol{V}_{\alpha,\xi} = \sigma^2_Z(\xi\boldsymbol{I} + (1-\xi)\boldsymbol{V}_\alpha),$$

where σ^2_{n} is the variance of seemingly uncorrelated micro-scale variation in $B(\boldsymbol{s})$ that cannot be resolved with the chosen sampling design, σ^2 is the variance of the captured auto-correlated

variation in $B(\boldsymbol{s})$, $\sigma_Z^2 = \sigma_\mathrm{n}^2 + \sigma^2$ is the signal variance, and $\xi = \sigma_\mathrm{n}^2/\sigma_Z^2$. To estimate both σ_n^2 and τ^2 (and not only their sum), one needs replicated measurements for some of the $\boldsymbol{s}_i$.

We define $\boldsymbol{V}_{\alpha,\xi}$ to be the matrix with elements

$$(\boldsymbol{V}_{\alpha,\xi})_{ij} = \gamma_0 - \gamma(|\boldsymbol{A}\ (\boldsymbol{s}_i - \boldsymbol{s}_j)|, \xi),$$

where the constant γ_0 is chosen large enough so that $\boldsymbol{V}_{\alpha,\xi}$ is positive definite, $\gamma(\cdot)$ is (the sum of) a valid stationary or intrinsic variogram (and possibly a pure nugget variogram), and $\boldsymbol{A} = \boldsymbol{A}(\alpha, f_1, f_2; \omega, \phi, \zeta)$ is a matrix that is used to model geometrically anisotropic auto-correlation. In more detail, $\boldsymbol{A}$ maps an arbitrary point on an ellipsoidal surface with constant semivariance in $\mathbb{R}^d$, centred on the origin, and having lengths of semi-principal axes, $\boldsymbol{p}_1$, $\boldsymbol{p}_2$, $\boldsymbol{p}_3$, equal to $|\boldsymbol{p}_1| = \alpha$, $|\boldsymbol{p}_2| = f_1\,\alpha$ and $|\boldsymbol{p}_3| = f_2\,\alpha$, $0 < f_2 \le f_1 \le 1$, respectively, onto the surface of the unit ball centred on the origin.

The orientation of the ellipsoid is defined by the three angles ω, ϕ and ζ:

ω is the azimuth of $\boldsymbol{p}_1$ (= angle between north and the projection of $\boldsymbol{p}_1$ onto the x-y-plane, measured from north to south positive clockwise in degrees),

ϕ is 90 degrees minus the altitude of $\boldsymbol{p}_1$ (= angle between the zenith and $\boldsymbol{p}_1$, measured from zenith to nadir positive clockwise in degrees), and

ζ is the angle between $\boldsymbol{p}_2$ and the direction of the line, say y', defined by the intersection between the x-y-plane and the plane orthogonal to $\boldsymbol{p}_1$ running through the origin (ζ is measured from y' positive counterclockwise in degrees).

The transformation matrix is given by

$$\boldsymbol{A} = \begin{pmatrix} 1/\alpha & 0 & 0 \\ 0 & 1/(f_1\,\alpha) & 0 \\ 0 & 0 & 1/(f_2\,\alpha) \end{pmatrix} (\boldsymbol{C}_1, \boldsymbol{C}_2, \boldsymbol{C}_3,)$$

where

$$\boldsymbol{C}_1^\mathrm{T} = (\sin\phi\sin\omega, -\cos\zeta\cos\omega + \sin\zeta\cos\phi\sin\omega, -\cos\omega\sin\zeta - \cos\zeta\cos\phi\sin\omega)$$

$$\boldsymbol{C}_2^\mathrm{T} = (-\sin\phi\cos\omega, \cos\omega\sin\zeta\cos\phi + \cos\zeta\sin\omega, -\cos\zeta\cos\omega\cos\phi + \sin\zeta\sin\omega)$$

$$\boldsymbol{C}_3^\mathrm{T} = (\cos\phi, -\sin\phi\sin\zeta, \cos\zeta\sin\phi)$$

To model geometrically anisotropic variograms in $\mathbb{R}^2$ one has to set $\phi = 90$ and $f_2 = 1$, and for $f_1 = f_2 = 1$ one obtains the model for isotropic auto-correlation with range parameter α. Note that for isotropic auto-correlation d may exceed 3.

Depending on the context, the term "variogram parameters" denotes sometimes all parameters of a geometrically anisotropic variogram model, but in places only the parameters of an isotropic variogram model, i.e. $\sigma^2, \ldots, \alpha, \ldots$ and $f_1, \ldots, \zeta$ are denoted by the term "anisotropy parameters".

Estimation: The spatial random effects $\boldsymbol{B}$ and the model parameters $\boldsymbol{\beta}$ and $\boldsymbol{\theta}^\mathrm{T} = (\sigma^2, \sigma_\mathrm{n}^2, \tau^2, \alpha, \ldots, f_1, f_2, \omega, \phi, \zeta)$ are estimated by Gaussian or robust restricted maximum likelihood (REML) or Gaussian maximum likelihood (ML). Here $\ldots$ denote further parameters of the variogram such as the smoothness parameter of the Whittle-Matérn model. In brief, the robust REML method is based on the insight that the kriging predictions of $\boldsymbol{B}$ and the Gaussian maximum likelihood estimates of $\boldsymbol{\beta}$ and $\boldsymbol{\theta}$ can be obtained simultaneously by maximizing

$$-\log(\det(\tau^2\boldsymbol{I} + \boldsymbol{\Gamma}_\theta)) - \sum_i \left(\frac{y_i - \boldsymbol{x}(\boldsymbol{s}_i)^\mathrm{T}\boldsymbol{\beta} - B(\boldsymbol{s}_i)}{\tau}\right)^2 - \boldsymbol{B}^\mathrm{T}\boldsymbol{\Gamma}_\theta^{-1}\boldsymbol{B}$$

with respect to $\boldsymbol{B}$, $\boldsymbol{\beta}$, $\boldsymbol{\theta}$. The respective estimating equations can then by robustified by

- replacing the standardized residuals, say $\widehat{\varepsilon}/\widehat{\tau}$, by a bounded function, $\psi_c(\widehat{\varepsilon}/\widehat{\tau})$, of them and by
- introducting suitable bias correction terms for Fisher consistency at the Gaussian model,

see Künsch et al. (2011) for details. The robustified estimating equations are solved numerically by a combination of iterated re-weighted least squares (IRWLS) to estimate $\boldsymbol{B}$ and $\boldsymbol{\beta}$ for given $\boldsymbol{\theta}$ and nonlinear root finding by the function `nleqslv` of the R package **nleqslv** to get $\boldsymbol{\theta}$. The robustness of the procedure is controlled by the tuning parameter c of the ψ_c-function. For $c \geq 1000$ the algorithm computes Gaussian (RE)ML estimates and customary plug-in kriging predictions. Instead of solving the Gaussian (RE)ML estimating equations, our software then maximizes the Gaussian (restricted) loglikelihood using `nlminb` or `optim`.

georob uses variogram models implemented in the R package **RandomFields** (see `RMmodel`). Currently, estimation of the parameters of the following models is implemented: `"RMaskey"`, `"RMbessel"`, `"RMcauchy"`, `"RMcircular"`, `"RMcubic"`, `"RMdagum"`, `"RMdampedcos"`, `"RMdewijsian"`, `"RMexp"` (default), `"RMfbm"`, `"RMgauss"`, `"RMgencauchy"`, `"RMgenfbm"`, `"RMgengneiting"`, `"RMgneiting"`, `"RMlgd"`, `"RMmatern"`, `"RMpenta"`, `"RMqexp"`, `"RMspheric"`, `"RMstable"`, `"RMwave"`, `"RMwhittle"`. For most variogram parameters, closed-form expressions of $\partial\gamma/\partial\theta_i$ are used in the computations. However, for the parameter ν of the models `"RMbessel"`, `"RMmatern"` and `"RMwhittle"` $\partial\gamma/\partial\theta_i$ is evaluated numerically by the function `numericDeriv`, and this results in an increase in computing time when ν is estimated.

Prediction:

Robust plug-in external drift point kriging predictions can be computed for an unsampled location $\boldsymbol{s}_0$ from the covariates $\boldsymbol{x}(\boldsymbol{s}_0)$, the estimated parameters $\widehat{\boldsymbol{\beta}}$, $\widehat{\boldsymbol{\theta}}$ and the predicted random effects $\widehat{\boldsymbol{B}}$ by

$$\widehat{Y}(\boldsymbol{s}_0) = \widehat{Z}(\boldsymbol{s}_0) = \boldsymbol{x}(\boldsymbol{s}_0)^{\mathrm{T}}\widehat{\boldsymbol{\beta}} + \boldsymbol{\gamma}_{\widehat{\theta}}^{\mathrm{T}}(\boldsymbol{s}_0)\boldsymbol{\Gamma}_{\widehat{\theta}}^{-1}\widehat{\boldsymbol{B}},$$

where $\boldsymbol{\Gamma}_{\widehat{\theta}}$ is the estimated (generalized) covariance matrix of $\boldsymbol{B}$ and $\boldsymbol{\gamma}_{\widehat{\theta}}(\boldsymbol{s}_0)$ is the vector with the estimated (generalized) covariances between $\boldsymbol{B}$ and $B(\boldsymbol{s}_0)$. Kriging variances can be computed as well, based on approximated covariances of $\widehat{\boldsymbol{B}}$ and $\widehat{\boldsymbol{\beta}}$ (see Künsch et al., 2011, and Appendices of Nussbaum et al., 2012, for details).

The package **georob** provides in addition software for computing robust external drift *block* kriging predictions. The required integrals of the generalized covariance function are computed by functions of the R package **constrainedKriging**.

Main functionality: For the time being, the functionality of **georob** is limited to robust geostatistical analyses of *single* response variables. No software is currently available for robust multivariate geostatistical analyses. **georob** offers functions for:

1. Robustly fitting a spatial linear model to data that are possibly contaminated by independent errors from a long-tailed distribution by robust REML (see `georob`, which also fits such models efficiently by Gaussian (RE)ML, `profilelogLik` and `control.georob`).
2. Extracting estimated model components (see `residuals.georob`, `rstandard.georob`, `rstudent.georob`, `ranef.georob`).
3. Robustly estimating sample variograms and for fitting variogram model functions to them (see `sample.variogram` and `fit.variogram.model`).
4. Model building by forward and backward selection of covariates for the external drift (see `waldtest.georob`, `step.georob`, `add1.georob`, `drop1.georob`, `extractAIC.georob`,

logLik.georob, deviance.georob). For a robust fit, the loglikelihood is not defined. The function then computes the (restricted) loglikelihood of an equivalent Gaussian model with heteroscedastic nugget (see deviance.georob for details).

5. Assessing the goodness-of-fit and predictive power of the model by K-fold cross-validation (see cv.georob and validate.predictions).
6. Computing robust external drift point and block kriging predictions (see predict.georob).
7. Unbiased back-transformation of both point and block kriging predictions of log-transformed data to the original scale of the measurements (see lgnpp).

Author(s)

Andreas Papritz <andreas.papritz@env.ethz.ch>
http://www.step.ethz.ch/people/scientific-staff/andreas-papritz

References

Nussbaum, M., Papritz, A., Baltensweiler, A. and Walthert, L. (2012) *Organic Carbon Stocks of Swiss Forest Soils*, Institute of Terrestrial Ecosystems, ETH Zurich and Swiss Federal Institute for Forest, Snow and Landscape Research (WSL), pp. 51. http://e-collection.library.ethz.ch/eserv/eth:6027/eth-6027-01.pdf

Kuensch, H. R., Papritz, A., Schwierz, C. and Stahel, W. A. (in preparation) Robust Geostatistics.

Kuensch, H. R., Papritz, A., Schwierz, C. and Stahel, W. A. (2011) Robust estimation of the external drift and the variogram of spatial data. Proceedings of the ISI 58th World Statistics Congress of the International Statistical Institute. http://e-collection.library.ethz.ch/eserv/eth:7080/eth-7080-01.pdf

See Also

georob for (robust) fitting of spatial linear models; georobObject for a description of the class georob; plot.georob for display of RE(ML) variogram estimates; control.georob for controlling the behaviour of georob; cv.georob for assessing the goodness of a fit by georob; predict.georob for computing robust kriging predictions; and finally georobModelBuilding for stepwise building models of class georob; georobMethods for further methods for the class georob, sample.variogram and fit.variogram.model for robust estimation and modelling of sample variograms.

---

compress *Compact Storage of Symmetric and Triangular Matrices*

Description

The utility function compress stores symmetric or triangular matrices compactly by retaining only the diagonal and either the lower or upper off-diagonal elements. The function expand restores such compressed matrices again to a square form.

Usage

```
compress(m)

expand(object)
```

Arguments

m	either a single symmetric, lower or upper triangular matrix or a list of such matrices. The type of m (or of its component matrices) must be defined by the attribute struc with possible values "sym" (symmetric), "lt" (lower triangular) or "ut" (upper triangular).
object	a single compressed matrix or a list of such matrices generated by compress, see *Value*. The type of object (or of its components) must be defined by the attribute struc with possible values "sym" (symmetric), "lt" (lower triangular) or "ut" (upper triangular).

Value

If m is a single square matrix then compress generates a compressed matrix, which is a list with two components:

diag	a vector with the diagonal elements of m.
tri	a vector with non-redudant off-diagonal elements.

If m is a list of square matrices then the result is also a list of compressed matrices.

expand creates a square matrix if object is a list with components diag and tri and a list of square matrices if object is a list of such lists. If m or objects are lists that contain further components than squares or compressed matrices then these additional components are returned unchanged.

Author(s)

Andreas Papritz <andreas.papritz@env.ethz.ch>

See Also

georob for (robust) fitting of spatial linear models.

Examples

```
## Not run:

data(meuse)

r.logzn.rob <- georob(log(zinc) ~ sqrt(dist) + ffreq, data = meuse, locations = ~ x + y,
    variogram.model = "RMexp",
    param = c( variance = 0.15, nugget = 0.05, scale = 200 ),
    tuning.psi = 1)

cov2cor( expand(r.logzn.rob[["cov"]][["cov.betahat"]]) )
```

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

`control.georob` *Tuning Parameters for georob*

Description

This page documents parameters used to control georob. It describes the arguments of the functions `control.georob`, `param.transf`, `fwd.transf`, `dfwd.transf`, `bwd.transf`, `control.rq`, `control.nleqslv`, `control.nlminb` and `control.optim`, which all serve to control the behaviour of georob.

Usage

```
control.georob(ml.method = c( "REML", "ML" ), reparam = TRUE,
    maximizer = c( "nlminb", "optim" ), initial.param = TRUE,
    initial.fixef = c("lmrob", "rq", "lm"), bhat = NULL,
    min.rweight = 0.25,
    param.tf = param.transf(), fwd.tf = fwd.transf(),
    deriv.fwd.tf = dfwd.transf(), bwd.tf = bwd.transf(),
    safe.param = 1.e12, psi.func = c("logistic", "t.dist", "huber"),
    tuning.psi.nr = 1000,
    irwls.initial = TRUE, irwls.maxiter = 50,
    irwls.ftol = 1.e-5, force.gradient = FALSE,
    min.condnum = 1.e-12, zero.dist = sqrt(.Machine[["double.eps"]]),
    error.family.estimation = c("gaussian", "long.tailed"),
    error.family.cov.effects = c("gaussian", "long.tailed"),
    error.family.cov.residuals = c("long.tailed", "gaussian"),
    cov.bhat = FALSE, full.cov.bhat = FALSE, cov.betahat = TRUE,
    cov.bhat.betahat = FALSE,
    cov.delta.bhat = TRUE, full.cov.delta.bhat = TRUE,
    cov.delta.bhat.betahat = TRUE,
    cov.ehat = TRUE, full.cov.ehat = FALSE,
    cov.ehat.p.bhat = FALSE, full.cov.ehat.p.bhat = FALSE,
    aux.cov.pred.target = FALSE,
    hessian = TRUE,
    rq = control.rq(), lmrob = lmrob.control(),
    nleqslv = control.nleqslv(),
    optim = control.optim(), nlminb = control.nlminb(),
    pmm = control.pmm(), ...)

param.transf(variance = "log", snugget = "log", nugget = "log", scale = "log",
    alpha = c(
     RMaskey = "log", RMdewijsian = "logit2", RMfbm = "logit2", RMgencauchy = "logit2",
     RMgenfbm = "logit2", RMlgd = "identity", RMqexp = "logit1", RMstable = "logit2"
    ),
```

```
    beta = c( RMdagum = "logit1", RMgencauchy = "log", RMlgd = "log" ),
    delta = "logit1", gamma = c( RMcauchy = "log", RMdagum = "logit1" ),
    kappa = "logit3", lambda = "log", mu = "log", nu = "log",
   f1 = "log", f2  ="log", omega = "identity", phi = "identity", zeta = "identity")

fwd.transf(...)

dfwd.transf(...)

bwd.transf(...)

control.rq(tau = 0.5, rq.method = "br", rq.alpha = 0.1, ci = FALSE, iid = TRUE,
    interp = TRUE, tcrit = TRUE, rq.beta = 0.99995, eps = 1e-06,
    Mm.factor = 0.8, max.bad.fixup = 3, ...)

control.nleqslv(method = c("Broyden", "Newton"),
    global = c( "dbldog", "pwldog", "qline", "gline", "none" ),
    xscalm = c( "fixed", "auto" ), control = list(ftol = 1e-04), ...)

control.optim(method = c("BFGS", "Nelder-Mead", "CG",
        "L-BFGS-B", "SANN", "Brent"), lower = -Inf, upper = Inf,
    control = list(reltol = 1e-05), ...)

control.nlminb(control = list( rel.tol = 1.e-5 ), lower = -Inf,
    upper = Inf, ...)
```

Arguments

`ml.method`	character keyword defining whether non-robust maximum likelihood (`ML`) or restricted maximum likelihood (`REML` default) estimates will be computed (ignored if `tuning.psi <= tuning.psi.nr`).
`reparam`	logical. If `TRUE` (default) the reparametrized variance parameters σ_Z^2, η and ξ are estimated by Gaussian (RE)ML, otherwise the original parameters τ^2, σ_{n}^2 and σ^2 (cf. subsection *Estimating variance parameters by Gaussian (RE)ML*, section *Details* of `georob`).
`maximizer`	character keyword defining the Gaussian (restricted) loglikelihood is maximized by `nlminb` (default) or `optim`.
`initial.param`	logical, controlling whether initial values of variogram parameters are computed for solving the estimating equations of the variogram and anisotropy parameters. If `initial.param = TRUE` (default) robust initial values of parameters are computed by discarding outlying observations based on the "robustness weights" of the initial fit of the regression model by `lmrob` and fitting the spatial linear model by Gaussian REML to the pruned data set. For `initial.param = FALSE` no initial parameter values are computed and the estimating equations are solved with the initial values passed by `param` and `aniso` to `georob` (see *Details* of `georob`.

`initial.fixef` character keyword defining whether the function `lmrob` or `rq` is used to compute robust initial estimates of the regression parameters $\boldsymbol{\beta}$ (default `"lmrob"`). If the fixed effects model matrix has not full columns rank, then `lm` is used to compute initial values of the regression coefficients.

`bhat` initial values for the spatial random effects $\widehat{\boldsymbol{B}}$, with $\widehat{\boldsymbol{B}} = \mathbf{0}$ if `bhat` is equal to `NULL` (default).

`min.rweight` positive numeric. "Robustness weight" of the initial `lmrob` fit that observations must exceed to be used for computing robust initial estimates of variogram parameters by setting `initial.param = TRUE` (see `georob`; default `0.25`).

`param.tf` a function such as `param.transf`, which returns a named vector of character strings that define the transformations to be applied to the variogram parameters for model fitting, see *Details*.

`fwd.tf` a function such as `fwd.transf`, which returns a named list of invertible functions to be used to transform variogram parameters, see *Details*.

`deriv.fwd.tf` a function such as `dfwd.transf`, which returns a named list of functions corresponding to the first derivatives of `fwd.tf`, see *Details*.

`bwd.tf` a function such as `bwd.transf`, which returns the named list of inverse functions corresponding to `fwd.tf`, see *Details*.

`safe.param` maximum acceptable value for any variogram parameter. If trial parameter values generated by `nlminb` `optim` or `nleqslv` exceed `safe.param` then an error is signalled to force `optim` or `nleqslv` to update the trial values (default `1.e12`).

`psi.func` character keyword defining what ψ_c-function should be used for robust model fitting. Possible values are `"logistic"` (a scaled and shifted logistic cdf, default), `"t.dist"` (re-descending ψ_c-function associated with Student t-distribution with c degrees of freedom) and `"huber"` (Huber's ψ_c-function).

`tuning.psi.nr` positive numeric. If `tuning.psi` is less than `tuning.psi.nr` then the model is fitted robustly by solving the robustified estimating equations, and for `tuning.psi` equal to or larger than `tuning.psi.nr` the Gaussian (restricted) loglikelihood is maximized (default `1000`).

`irwls.initial` logical. If `TRUE` (default) the estimating equations of $\boldsymbol{B}$ and $\boldsymbol{\beta}$ are always solved by IRWLS from the initial estimates of $\widehat{\boldsymbol{B}}$ and $\widehat{\boldsymbol{\beta}}$. If `FALSE` then IRWLS starts from respective estimates computed for the variogram parameter estimates of the previous iteration of `nleqslv` or `optim`.

`irwls.maxiter` positive integer equal to the maximum number of IRWLS iterations to solve the estimating equations of $\boldsymbol{B}$ and $\boldsymbol{\beta}$ (default `50`).

`irwls.ftol` numeric convergence criterion for IRWLS. Convergence is assumed if the objective function changes in one IRWLS iteration does not exceed `ftol`.

`force.gradient` logical controlling whether the estimating equations or the gradient of the Gaussian restricted loglikelihood are evaluated even if all variogram parameters are fixed (default `FALSE`).

`min.condnum` positive numeric. Minimum acceptable ratio of smallest to largest singular value of the model matrix $\boldsymbol{X}$ (default `1.e-12`).

`zero.dist` positive numeric equal to the maximum distance, separating two sampling locations that are still considered as being coincident.

`error.family.estimation`
: character keyword, defining the probability distribution for ε (default: `"gaussian"`) that is used to approximate the covariance of $\widehat{\boldsymbol{B}}$, see *Details*.

`error.family.cov.effects`
: character keyword, defining the probability distribution for ε (default: `"gaussian"`) that is used to approximate the covariances of $\widehat{\boldsymbol{\beta}}$, $\widehat{\boldsymbol{B}}$ and $\boldsymbol{B} - \widehat{\boldsymbol{B}}$, see *Details*.

`error.family.cov.residuals`
: character keyword, defining the probability distribution for ε (default: `"long.tailed"`) that is used to approximate the covariances of $\widehat{\boldsymbol{\varepsilon}} = \boldsymbol{Y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}} - \widehat{\boldsymbol{B}}$ and $\widehat{\boldsymbol{\varepsilon}} + \widehat{\boldsymbol{B}} = \boldsymbol{Y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}$, see *Details*.

`cov.bhat`
: logical controlling whether the covariances of $\widehat{\boldsymbol{B}}$ are returned by `georob` (default `FALSE`).

`full.cov.bhat`
: logical controlling whether the full covariance matrix (`TRUE`) or only the variance vector of $\widehat{\boldsymbol{B}}$ is returned (default `FALSE`).

`cov.betahat`
: logical controlling whether the covariance matrix of $\widehat{\boldsymbol{\beta}}$ is returned (default `TRUE`).

`cov.bhat.betahat`
: logical controlling whether the covariance matrix of $\widehat{\boldsymbol{B}}$ and $\widehat{\boldsymbol{\beta}}$ is returned (default `FALSE`).

`cov.delta.bhat`
: logical controlling whether the covariances of $\boldsymbol{B} - \widehat{\boldsymbol{B}}$ are returned (default `TRUE`).

`full.cov.delta.bhat`
: logical controlling whether the full covariance matrix (`TRUE`) or only the variance vector of $\boldsymbol{B} - \widehat{\boldsymbol{B}}$ is returned (default `TRUE`).

`cov.delta.bhat.betahat`
: logical controlling whether the covariance matrix of $\boldsymbol{B} - \widehat{\boldsymbol{B}}$ and $\widehat{\boldsymbol{\beta}}$ is returned (default `TRUE`).

`cov.ehat`
: logical controlling whether the covariances of $\widehat{\boldsymbol{\varepsilon}} = \boldsymbol{Y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}} - \widehat{\boldsymbol{B}}$ are returned (default `TRUE`).

`full.cov.ehat`
: logical controlling whether the full covariance matrix (`TRUE`) or only the variance vector of $\widehat{\boldsymbol{\varepsilon}} = \boldsymbol{Y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}} - \widehat{\boldsymbol{B}}$ is returned (default `FALSE`).

`cov.ehat.p.bhat`
: logical controlling whether the covariances of $\widehat{\boldsymbol{\varepsilon}} + \widehat{\boldsymbol{B}} = \boldsymbol{Y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}$ are returned (default `FALSE`).

`full.cov.ehat.p.bhat`
: logical controlling whether the full covariance matrix (`TRUE`) or only the variance vector of $\widehat{\boldsymbol{\varepsilon}} + \widehat{\boldsymbol{B}} = \boldsymbol{Y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}$ is returned (default `FALSE`).

`aux.cov.pred.target`
: logical controlling whether a covariance term required for the back-transformation of kriging predictions of log-transformed data is returned (default `FALSE`).

`hessian`
: logical scalar controlling whether for Gaussian (RE)ML the Hessian should be computed at the MLEs.

`rq`
: a list of arguments passed to `rq` or a function such as `control.rq` that generates such a list (see `rq` for allowed arguments).

`lmrob` a list of arguments passed to the `control` argument of `lmrob` or a function such as `lmrob.control` that generates such a list (see `lmrob.control` for allowed arguments).

`nleqslv` a list of arguments passed to `nleqslv` or a function such as `control.nleqslv` that generates such a list (see `nleqslv` for allowed arguments).

`nlminb` a list of arguments passed to `nlminb` or a function such as `control.nlminb` that generates such a list (see `nlminb` for allowed arguments).

`optim` a list of arguments passed to `optim` or a function such as `control.optim` that generates such a list (see `optim` for allowed arguments).

`pmm` a list of arguments, passed e.g. to `pmm` or a function such as `control.pmm` that generates such a list (see `control.pmm` for allowed arguments).

`...` for `fwd.transf`, `dfwd.transf` and `bwd.transf` a named vectors of functions, extending the definition of transformations for variogram parameters (see *Details*).

`variance, snugget, nugget, scale, alpha, beta, delta, gamma, kappa, lambda, mu, nu`
character strings with names of transformation functions of the variogram parameters.

`f1, f2, omega, phi, zeta`
character strings with names of transformation functions of the variogram parameters.

`tau, rq.method, rq.alpha, ci, iid, interp, tcrit`
arguments passed as `...` to `rq`.

`rq.beta, eps, Mm.factor, max.bad.fixup`
arguments passed as `...` to `rq`.

`method, global, xscalm, control, lower, upper, reltol, rel.tol`
arguments passed to related arguments of `nleqslv`, `nlminb` and `optim`, respectively.

Details

Parameter transformations:

The arguments `param.tf`, `fwd.tf`, `deriv.fwd.tf`, `bwd.tf` define the transformations of the variogram parameters for RE(ML) estimation. Implemented are currently `"log"`, `"logit1"`, `"logit2"`, `"logit3"` (various variants of logit-transformation, see code of function `fwd.transf`) and `"identity"` (= no) transformations. These are the possible values that the many arguments of the function `param.transf` accept (as quoted character strings) and these are the names of the list components returned by `fwd.transf`, `dfwd.transf` and `bwd.transf`. Additional transformations can be implemented by:

1. Extending the function definitions by arguments like
   `fwd.tf = fwd.transf(c(my.fun = function(x) your transformation))`,
   `deriv.fwd.tf = dfwd.transf(c(my.fun = function(x) your derivative))`,
   `bwd.tf = bwd.transf(c(my.fun = function(x) your back-transformation))`,
2. Assigning to a given argument of `param.transf` the name of the new function, e.g. `variance = "my.fun"`.

Note the values given for the arguments of `param.transf` must match the names of the functions returned by `fwd.transf`, `dfwd.transf` and `bwd.transf`.

Approximation of covariances of fixed and random efffects and residuals:
The robustified estimating equations of robust REML depend on the covariances of $\widehat{\boldsymbol{B}}$. These covariances (and the covariances of $\boldsymbol{B}-\widehat{\boldsymbol{B}}, \widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{\varepsilon}}, \widehat{\boldsymbol{\varepsilon}}+\widehat{\boldsymbol{B}}$) are approximated by expressions that in turn depend on the variances of ε, $\psi(\varepsilon/\tau)$ and the expectation of $\psi'(\varepsilon/\tau)(=\partial/\partial\varepsilon\psi(\varepsilon/\tau))$. The arguments `error.family.estimation`, `error.family.cov.effects` and `error.family.cov.residuals` control what parametric distribution for ε is used to compute the latter quantities. Possible options are: `"gaussian"` or `"long.tailed"`. In the latter case the pdf of ε is assumed to be proportional to $1/\tau exp(-\rho(\varepsilon/\tau))$ where $\psi(x) = \rho'(x)$.

Author(s)

Andreas Papritz <andreas.papritz@env.ethz.ch>

See Also

`georobIntro` for a description of the model and a brief summary of the algorithms; `georob` for (robust) fitting of spatial linear models; `georobObject` for a description of the class georob; `plot.georob` for display of RE(ML) variogram estimates; `predict.georob` for computing robust kriging predictions; and finally `georobMethods` for further methods for the class georob.

Examples

```
## Not run:
data( meuse )

r.logzn.rob <- georob(log(zinc) ~ sqrt(dist), data = meuse, locations = ~ x + y,
    variogram.model = "RMexp",
    param = c( variance = 0.15, nugget = 0.05, scale = 200 ),
    tuning.psi = 1, control = control.georob(cov.bhat = TRUE,
    cov.ehat.p.bhat = TRUE, initial.fixef = "rq"), verbose = 2)

qqnorm(rstandard(r.logzn.rob, level = 0)); abline(0, 1)
qqnorm(ranef(r.logzn.rob, standard = TRUE)); abline(0, 1)

## End(Not run)
```

---

cv *Generic Cross-validation*

Description

Generic function for cross-validating models.

Usage

```
cv(object, ...)
```

Arguments

object	any model object.
...	additional arguments as required by the methods.

Value

will depend on the method function used; see the respective documentation.

Author(s)

Andreas Papritz <andreas.papritz@env.ethz.ch>

See Also

georob for (robust) fitting of spatial linear models; cv.georob for assessing the goodness of a model fitted by georob.

---

cv.georob *Cross-Validating a Spatial Linear Model Fitted by* georob

---

Description

This function assesses the goodness-of-fit of a spatial linear model by K-fold cross-validation. In more detail, the model is re-fitted K times by robust (or Gaussian) (RE)ML, excluding each time *1/K*th of the data. The re-fitted models are used to compute robust (or customary) external kriging predictions for the omitted observations. If the response variable is log-transformed then the kriging predictions can be optionally transformed back to the orginal scale of the measurements. S3methods for evaluating and plotting diagnostic summaries of the cross-validation errors are decribed for the function validate.predictions.

Usage

```
## S3 method for class 'georob'
cv(object, formula = NULL, subset = NULL,
    method = c("block", "random"), nset = 10,
    seed = NULL, sets = NULL, duplicates.in.same.set = TRUE,
    re.estimate = TRUE, param = object[["param"]],
    fit.param = object[["initial.objects"]][["fit.param"]],
    aniso = object[["aniso"]][["aniso"]],
    fit.aniso = object[["initial.objects"]][["fit.aniso"]],
    return.fit = FALSE, reduced.output = TRUE, lgn = FALSE,
    mfl.action = c("offset", "stop"),
    ncores = min(nset, detectCores()), verbose = 0, ...)
```

Arguments

object	an object of class of "georob", see georobObject.
formula	an optional formula for the regression model passed by update to georob, see *Details*.
subset	an optional vector specifying a subset of observations to be used in the fitting process, see *Details*.
method	keyword, controlling whether subsets are formed by partitioning data set into blocks by kmeans (default) or randomly. Ignored if sets is non-NULL.
nset	positive integer defining the number K of subsets into which the data set is partitioned (default: nset = 10). Ignored if sets is non-NULL.
seed	optional integer seed to initialize random number generation, see set.seed. Ignored if sets is non-NULL.
sets	an optional vector of the same length as the response vector of the fitted model and with positive integers taking values in $(1, 2, \ldots, K)$, defining in this way the K subsets into which the data set is split. If sets = NULL (default) the partition is randomly generated by kmeans or runif (using possibly seed).
duplicates.in.same.set	logical controlling whether replicated observations at a given location are assigned to the same subset when partitioning the data (default TRUE).
re.estimate	logical controlling whether the model is re-fitted to the reduced data sets before computing the kriging predictions (TRUE, default) or whether the model passed in object is used to compute the predictions for the omitted observations, see *Details*.
param	an optional named numeric vector or a matrix or data frame with variogram parameters passed by update to georob, see *Details*. If param is a matrix (or a data frame) then it must have nset rows and length(object[["param"]]) columns with initial values of variogram parameters for the nset cross-validation sets and colnames(param) must match names(object[["param"]]).
fit.param	an optional named logical vector or a matrix or data frame defining which variogram parameters should be adjusted when passed by update to georob, see *Details*. If fit.param is a matrix (or a data frame) then it must have nset rows and length(object[["param"]]) columns with variogram parameter fitting flags for the nset cross-validation sets and colnames(param) must match names(object[["param"]]).
aniso	an optional named numeric vector or a matrix or data frame with anisotropy parameters passed by update to georob, see *Details*. If aniso is a matrix (or a data frame) then it must have nset rows and length(object[["aniso"]][["aniso"]]) columns with initial values of anisotropy parameters for the nset cross-validation sets and colnames(aniso) must match names(object[["aniso"]][["aniso"]]).
fit.aniso	an optional named logical vector or a matrix or data frame defining which anisotropy parameters should be adjusted when passed by update to georob, see *Details*. If fit.aniso is a matrix (or a data frame) then it must have nset rows and length(object[["aniso"]][["aniso"]]) columns with anisotropy parameter fitting flags for the nset cross-validation sets and colnames(aniso) must match names(object[["aniso"]][["aniso"]]).

return.fit	logical controlling whether information about the fit should be returned when re-estimating the model with the reduced data sets (default FALSE).
reduced.output	logical controlling whether the complete fitted model objects, fitted to the reduced data sets, are returned (FALSE) or only some components (TRUE, default, see *Value*). Ignored if return.fit = FALSE.
lgn	logical controlling whether kriging predictions of a log-transformed response should be transformed back to the original scale of the measurements (default FALSE).
mfl.action	character controlling what is done when some levels of factor(s) are not present in any of the subsets used to fit the model. The function either stops ("stop") or treats the respective factors as model offset ("offset", default).
ncores	positive integer controlling how many cores are used for parallelized computations, see *Details*.
verbose	positive integer controlling logging of diagnostic messages to the console during model fitting. Passed by update to georob, see *Details*.
...	additional arguments passed by update to georob, see *Details*.

Details

Note that *the dataframe passed as* data *argument to* georob *must exist in the user workspace when calling* cv.georob.

cv.georob then uses the package **parallel** for parallelized computations. By default, the function uses K CPUs but not more than are physically available (as returned by detectCores).

cv.georob uses the function update to re-estimated the model with the reduced data sets. Therefore, any argument accepted by georob can be changed when re-fitting the model. Some of them (e.g. formula, subset, etc.) are explicit arguments of cv.georob, but also the remaining ones can be passed by ... to the function.

Practitioners in geostatistics commonly cross-validate a fitted model without re-estimating the model parameters with the reduced data sets. This is clearly an unsound practice (see Hastie et al., 2009, sec. 7.10). Therefore, the argument re.estimate should always be set to TRUE. The alternative is provided only for historic reasons.

Value

An object of class cv.georob, which is a list with the two components pred and fit.

pred is a data frame with the coordinates and the cross-validation prediction results with the following variables:

subset	an integer vector defining to which of the K subsets an observation was assigned.
data	the values of the (possibly log-transformed) response.
pred	the kriging predictions.
se	the kriging standard errors.

If lgn = TRUE then pred has the additional variables:

lgn.data	the untransformed response.
lgn.pred	the unbiasedly back-transformed predictions of a log-transformed response.
lgn.se	the kriging standard errors of the back-transformed predictions of a log-transformed response.

The second component fit contains either the full outputs of georob, fitted for the K reduced data sets (reduced.output = FALSE), or K lists with the components tuning.psi, converged, convergence.code, gradient, variogram.model, param, aniso[["aniso"]], coefficients along with the standard errors of $\widehat{\boldsymbol{\beta}}$, see georobObject.

Author(s)

Andreas Papritz <andreas.papritz@env.ethz.ch>

References

Hastie, T., Tibshirani, R. and Friedman, J. (2009) *The Elements of Statistical Learning; Data Mining, Inference and Prediction*. New York: Springer-Verlag.

See Also

validate.predictions for computing statistics of the cross-validation errors; georob for (robust) fitting of spatial linear models; georobObject for a description of the class georob; predict.georob for computing robust kriging predictions.

Examples

```
## Not run:
data( meuse )

r.logzn <- georob(log(zinc) ~ sqrt(dist), data = meuse, locations = ~ x + y,
    variogram.model = "RMexp",
    param = c( variance = 0.15, nugget = 0.05, scale = 200 ),
    tuning.psi = 1)

r.logzn.cv.1 <- cv(r.logzn, seed = 1, lgn = TRUE )
r.logzn.cv.2 <- cv(r.logzn, formula = .~. + ffreq, seed = 1, lgn = TRUE )

plot(r.logzn.cv.1, type = "bs")
plot(r.logzn.cv.2, type = "bs", add = TRUE, col = "red")

legend("topright", lty = 1, col = c( "black", "red"), bty = "n",
    legend = c("log(Zn) ~ sqrt(dist)", "log(Zn) ~ sqrt(dist) + ffreq"))
## End(Not run)
```

`fit.variogram.model` *Fitting Model Functions to Sample Variograms*

Description

The function `fit.variogram.model` fits a variogram model to a sample variogram by (weighted) non-linear least squares. There are `print`, `summary` and `lines` methods for summarizing and displaying fitted variogram models.

Usage

```
fit.variogram.model(sv,
    variogram.model = c("RMexp", "RMaskey", "RMbessel", "RMcauchy",
        "RMcircular", "RMcubic", "RMdagum", "RMdampedcos", "RMdewijsian",
        "RMfbm", "RMgauss", "RMgencauchy", "RMgenfbm", "RMgengneiting",
        "RMgneiting", "RMlgd", "RMmatern", "RMpenta", "RMqexp",
        "RMspheric", "RMstable", "RMwave", "RMwhittle"),
    param,
    fit.param = c(variance = TRUE, snugget = FALSE, nugget = TRUE,
        scale = TRUE, alpha = FALSE, beta = FALSE, delta = FALSE,
        gamma = FALSE, kappa = FALSE, lambda = FALSE, mu = FALSE,
        nu = FALSE)[names(param)],
    aniso = c(f1 = 1, f2 = 1, omega = 90, phi = 90, zeta = 0),
    fit.aniso = c(f1 = FALSE, f2 = FALSE, omega = FALSE,
        phi = FALSE, zeta = FALSE),
    max.lag = max(sv[["lag.dist"]]), min.npairs = 30,
    weighting.method = c("cressie", "equal", "npairs"), hessian = TRUE,
    verbose = 0, ...)

## S3 method for class 'fitted.variogram'
print(x, digits = max(3, getOption("digits") - 3), ...)

## S3 method for class 'fitted.variogram'
summary(object, correlation = FALSE, signif = 0.95, ...)

## S3 method for class 'fitted.variogram'
lines(x, what = c("variogram", "covariance", "correlation"),
    from = 1.e-6, to, n = 501, xy.angle = 90, xz.angle = 90,
    col = 1:length(xy.angle), pch = 1:length(xz.angle), lty = "solid", ...)
```

Arguments

`sv`	an object of class `sample.variogram`, see `sample.variogram`.
`variogram.model`	a character keyword defining the variogram model to be fitted. Currently, most basic variogram models provided by the package **RandomFields** can be fitted

(see *Details* of georob and RMmodel).

param — a named numeric vector with initial values of the variogram parameters. The following parameter names are allowed (see *Details* of georob and georobIntro for information about the parametrization of variogram models):

- variance: variance (sill σ^2) of the auto-correlated component of the Gaussian random field $B(\boldsymbol{s})$.
- snugget: variance (spatial nugget σ^2_{n}) of the seemingly spatially uncorrelated component of $B(\boldsymbol{s})$ (micro-scale spatial variation; default value snugget = 0).
- nugget: variance (nugget τ^2) of the independent errors $\varepsilon(\boldsymbol{s})$.
- scale: range parameter (α) of the variogram.
- names of additional variogram parameters such as the smoothness parameter ν of the Whittle-Mat\'ern model (see RMmodel and param.names).

fit.param — a named logical vector with the same names as used for param, defining which parameters are adjusted (TRUE) and which are kept fixed at their initial values (FALSE) when fitting the model.

aniso — a named numeric vector with initial values for fitting geometrically anisotropic variogram models. The following parameter names are allowed (see *Details* of georob and georobIntro for information about the parametrization of variogram models):

- f1: ratio f_1 of lengths of second and first second semi-principal axes of an ellipsoidal surface with constant semivariance in $\mathbb{R}^3$ (default f1 = 1).
- f2: ratio f_2 of lengths of third and first semi-principal axes of the semivariance ellipsoid (default f2 = 1).
- omega: azimuth in degrees of first semi-principal axis of the semivariance ellipsoid (default omega = 90).
- phi: 90 degrees minus altitude of first semi-principal axis of the semivariance ellipsoid (default phi = 90).
- zeta: angle in degrees between the second semi-principal axis and the direction of the line defined by the intersection between the x-y-plane and the plane orthogonal to the first semi-principal axis of the semivariance ellipsoid through the origin (default zeta = 0).

fit.aniso — a named logical vector with the same names as used for aniso, defining which parameters are adjusted (TRUE) and which are kept fixed at their initial values (FALSE) when fitting the model.

max.lag — a positive numeric defining the maximum lag distance to be used for fitting or plotting variogram models (default all lag classes).

min.npairs — a positive integer defining the minimum number of data pairs required so that a lag class is used for fitting a variogram model (default 30).

weighting.method — a character keyword defining the weights for non-linear least squares. Possible values are:

- "equal": no weighting ,
- "npairs": weighting by number of data pairs in a lag class,

- "cressie": "Cressie's weights" (default, see Cressie, 1993, sec. 2.6.2).

hessian	logical controlling whether the hessian is computed by optim.
verbose	positive integer controlling logging of diagnostic messages to the console during model fitting.
object, x	an object of class fitted.variogram.
digits	positive integer indicating the number of decimal digits to print.
correlation	logical controlling whether the correlation matrix of the fitted variogram parameters is computed (default FALSE).
signif	confidence level for computing confidence intervals for variogram parameters (default 0.95).
what	the quantity that should be displayed (default "variogram").
from	numeric, minimal lag distance used in plotting variogram models.
to	numeric, maximum lag distance used in plotting variogram models (default: largest lag distance of current plot).
n	positive integer specifying the number of equally spaced lag distances for which semivariances are evaluated in plotting variogram models (default 501).
xy.angle	numeric (vector) with azimuth angles (in degrees, clockwise positive from north) in x-y-plane for which semivariances should be plotted.
xz.angle	numeric (vector) with angles in x-z-plane (in degrees, clockwise positive from zenith to south) for which semivariances should be plotted.
col	color of curves to distinguish curves relating to different azimuth angles in x-y-plane.
pch	type of plotting symbols added to lines to distinguish curves relating to different angles in x-z-plane.
lty	line type for plotting variogram models.
...	additional arguments passed to optim or to methods.

Details

The parametrization of geometrically anisotropic variograms is described in detail in georobIntro, and the section *Details* of georob describes how the parameter estimates are constrained to permissible ranges. The same mechanisms are used in fit.variogram.model.

Value

The function fit.variogram.model generates an object of class fitted.variogram which is a list with the following components:

sse	the value of the object function (weighted residual sum of squares) evaluated at the solution.
variogram.model	the name of the fitted parametric variogram model.
param	a named vector with the (estimated) variogram parameters of the fitted model.

`fit.param` logical vector indicating which variogram parameters were fitted.

`aniso` a list with the following components:

- `isotropic`: logical indicating whether an isotropic variogram was fitted.
- `aniso`: a named numeric vector with the (estimated) anisotropy parameters of the fitted model.
- `fit.aniso`: logical vector indicating which anisotropy parameters were fitted.
- `sincos`: a list with `sin` and `cos` of the angles ω, ϕ and ζ that define the orientation of the anisotropy ellipsoid.
- `rotmat`: the matrix $(\boldsymbol{C}_1, \boldsymbol{C}_2, \boldsymbol{C}_3)$ (see `georobIntro`).
- `sclmat`: a vector with the elements 1, $1/f_1$, $1/f_2$ (see `georobIntro`).

`param.tf` a character vector listing the transformations of the variogram parameters used for model fitting.

`fwd.tf` a list of functions for variogram parameter transformations.

`bwd.tf` a list of functions for *inverse* variogram parameter transformations.

`converged` logical indicating whether numerical maximization by `optim` converged.

`convergence.code` a diagnostic integer issued by `optim` (component `convergence`) about convergence.

`call` the matched call.

`residuals` a numeric vector with the residuals, that is the sample semivariance minus the fitted values.

`fitted` a numeric vector with the modelled semivariances.

`weights` a numeric vector with the weights used for fitting.

`hessian` a symmetric matrix giving an estimate of the Hessian at the solution (missing if `hessian` is false).

Author(s)

Andreas Papritz <andreas.papritz@env.ethz.ch>.

References

Cressie, N. A. C. (1993) *Statistics for Spatial Data*. New York: John Wiley & Sons.

See Also

`georobIntro` for a description of the model and a brief summary of the algorithms; `georob` for (robust) fitting of spatial linear models; `sample.variogram` for computing sample variograms.

Examples

```
data(wolfcamp, package = "geoR")

## fitting an isotropic IRF(0) model
r.sv.iso <- sample.variogram(wolfcamp[["data"]], locations = wolfcamp[[1]],
    lag.dist.def = seq(0, 200, by = 15))

r.irf0.iso <- fit.variogram.model(r.sv.iso, variogram.model = "RMfbm",
    param = c(variance = 100, nugget = 1000, scale = 1., alpha = 1.),
    fit.param = c( variance = TRUE, nugget = TRUE, scale = FALSE, alpha = TRUE),
    method = "Nelder-Mead", hessian = FALSE, control = list(maxit = 5000))
summary(r.irf0.iso, correlation = TRUE)

## Not run:
plot( r.sv.iso, type = "l")
lines( r.irf0.iso, line.col = "red")
## End(Not run)

## fitting an anisotropic IRF(0) model
r.sv.aniso <- sample.variogram(wolfcamp[["data"]],
    locations = wolfcamp[[1]], lag.dist.def = seq(0, 200, by = 15),
    xy.angle.def = c(0., 22.5, 67.5, 112.5, 157.5, 180.))
## Not run:
plot(r.sv.aniso, type = "l")
## End(Not run)

r.irf0.aniso <- fit.variogram.model(r.sv.aniso, variogram.model = "RMfbm",
    param = c(variance = 100, nugget = 1000, scale = 1., alpha = 1.5),
    fit.param = c(variance = TRUE, nugget = TRUE, scale = FALSE, alpha = TRUE),
    aniso = c(f1 = 0.4, f2 = 1., omega = 135, phi = 90., zeta = 0.),
    fit.aniso = c(f1 = TRUE, f2 = FALSE, omega = TRUE, phi = FALSE, zeta = FALSE),
    method = "Nelder-Mead", hessian = TRUE, control = list(maxit = 5000))
summary(r.irf0.aniso, correlation = TRUE)

## Not run:
lines(r.irf0.aniso, xy.angle = seq( 0, 135, by = 45))
## End(Not run)
```

---

georob *Robust Fitting of Spatial Linear Models*

Description

The function georob fits a linear model with spatially correlated errors to geostatistical data that are possibly contaminated by independent outliers. The regression coefficients and the parameters of the variogram model are estimated by robust or Gaussian restricted maximum likelihood (REML) or by Gaussian maximum likelihood (ML).

Usage

```
georob(formula, data, subset, weights, na.action, model = TRUE,
    x = FALSE, y = FALSE, contrasts = NULL, offset, locations,
    variogram.model = c("RMexp", "RMaskey", "RMbessel", "RMcauchy",
        "RMcircular", "RMcubic", "RMdagum", "RMdampedcos", "RMdewijsian",
        "RMfbm", "RMgauss", "RMgencauchy", "RMgenfbm", "RMgengneiting",
        "RMgneiting", "RMlgd", "RMmatern", "RMpenta", "RMqexp",
        "RMspheric", "RMstable", "RMwave", "RMwhittle"),
    param,
    fit.param = c(variance = TRUE, snugget = FALSE, nugget = TRUE,
        scale = TRUE, alpha = FALSE, beta = FALSE, delta = FALSE,
        gamma = FALSE, kappa = FALSE, lambda = FALSE, mu = FALSE,
        nu = FALSE)[names(param)],
    aniso = c(f1 = 1, f2 = 1, omega = 90, phi = 90, zeta = 0),
    fit.aniso = c(f1 = FALSE, f2 = FALSE, omega = FALSE,
        phi = FALSE, zeta = FALSE),
    tuning.psi = 2, control = control.georob(),
    verbose = 0, ...)
```

Arguments

formula	a symbolic description of the regression model to be fit. See lm and formula for more details.
data	an optional data frame, a SpatialPointsDataFrame, list or environment (or another object coercible by as.data.frame to a data frame) containing the variables in the model and the coordinates where the data was recorded. If not found in data, the variables are taken from environment(formula), typically the environment from which georob is called.
subset	an optional vector specifying a subset of observations to be used in the fitting process.
weights	an optional vector of weights to be used in the fitting process, currently ignored.
na.action	a function which indicates what should happen when the data contain NAs. The default is set by the na.action argument of options, and is na.fail if that is unset. The "factory-fresh" default is na.omit. Another possible value is NULL, no action. Value na.exclude can be useful.
model, x, y	logicals. If TRUE the corresponding components of the fit (the model frame, the model matrix, the response) are returned. The model frame is augmented by the coordinates.
contrasts	an optional list. See the contrasts.arg of model.matrix.default.
offset	this can be used to specify an *a priori* known component to be included in the linear predictor during fitting. An offset term can be included in the formula instead or as well, and if both are specified their sum is used.
locations	a one-sided formula defining the variables that are used as coordinates of the locations were the data was recorded.

`variogram.model`
: a character keyword defining the variogram model to be fitted. Currently, most basic variogram models provided by the package **RandomFields** can be fitted (see *Details* and `RMmodel`).

`param`
: a named numeric vector with initial values of the variogram parameters. The names of `param` are matched against the following names (see *Details* and `georobIntro` for information about the parametrization of variogram models):

- `variance`: variance (sill σ^2) of the auto-correlated component of the Gaussian random field $B(\boldsymbol{s})$.
- `snugget`: variance (spatial nugget σ^2_{n}) of the seemingly spatially uncorrelated component of $B(\boldsymbol{s})$ (micro-scale spatial variation; default value `snugget = 0`).
- `nugget`: variance (nugget τ^2) of the independent errors $\varepsilon(\boldsymbol{s})$.
- `scale`: range parameter (α) of the variogram.
- names of additional variogram parameters such as the smoothness parameter ν of the Whittle-Matérn model (see `RMmodel` and `param.names`).

`fit.param`
: a named logical vector with the same names as used for `param`, defining which parameters are adjusted (`TRUE`) and which are kept fixed at their initial values (`FALSE`) when fitting the model.

`aniso`
: a named numeric vector with initial values for fitting geometrically anisotropic variogram models. The names of `aniso` are matched against the following names (see *Details* and `georobIntro` for information about the parametrization of variogram models):

- `f1`: ratio f_1 of lengths of second and first semi-principal axes of an ellipsoidal surface with constant semivariance in $\mathbb{R}^3$ (default `f1 = 1`).
- `f2`: ratio f_2 of lengths of third and first semi-principal axes of the semivariance ellipsoid (default `f2 = 1`).
- `omega`: azimuth in degrees of the first semi-principal axis of the semivariance ellipsoid (default `omega = 90`).
- `phi`: 90 degrees minus altitude of the first semi-principal axis of the semivariance ellipsoid (default `phi = 90`).
- `zeta`: angle in degrees between the second semi-principal axis and the direction of the line defined by the intersection between the x-y-plane and the plane orthogonal to the first semi-principal axis of the semivariance ellipsoid through the origin (default `zeta = 0`).

`fit.aniso`
: a named logical vector with the same names as used for `aniso`, defining which parameters are adjusted (`TRUE`) and which are kept fixed at their initial values (`FALSE`) when fitting the model.

`tuning.psi`
: positive numeric. The tuning constant c of the ψ_c-function of the robust REML algorithm.

`control`
: a list specifying parameters that control the behaviour of `georob`. Use the function `control.georob` and see its help page for the components of `control`.

`verbose`
: positive integer controlling logging of diagnostic messages to the console during model fitting. `verbose = 0` largely suppresses such messages and `verbose = 4` asks for most verbose output (see `control` arguments of `nleqslv`, `nlminb` and

optim and control.georob for information how to fine tuning diagnostic output generated by nleqslv, nlminb and optim).

... further arguments passed to function (e.g. object. used internally for updating georob objects).

Details

georob fits a spatial linear model by robust or Gaussian RE(ML) (Kuensch et al., 2011, Kuensch et al., in preparation). georobIntro describes the employed model and briefly sketches the robust REML estimation and the robust external-drift kriging method. Here, we describe further details of georob.

Implemented variogram models:

Currently, most basic variogram models provided by the package **RandomFields** can be fitted by georob (see argument variogram.model for a list of implemented models). Some of these models have in addition to variance, snugget, nugget and scale further parameters. Initial values of these parameters (param) and fitting flags (fit.param) must be passed to georob by the same names as used by the functions RM... of the package **RandomFields** (see RMmodel. Use the function param.names to list additional parameters of a given variogram.model.

Estimation of variance of micro-scale variation:

Simultaneous estimation of the variance of the micro-scale variation (snugget, σ_n^2), which appears as seemingly uncorrelated with a given sampling design, and of the variance (nugget, τ^2) of the independent errors requires that for some locations $\boldsymbol{s}_i$ replicated observations are available. Locations less or equal than zero.dist apart are thereby considered as being coincient (see control.georob).

Fitting intrinsic variogram models:

The intrinsic variogram model RMfbm is overparametrized when both the variance (plus possibly snugget) and the scale are fitted. Therefore, to estimate the parameters of this model scale must be kept fixed at an arbitray value by using fit.param["scale"] = FALSE.

Fitting geometrically anisotropic variogram models:

The subsection **Model** of georobIntro describes how such models are parametrized and gives definitions the various elements of aniso. Some additional remarks might be helpful:

- The first semi-principal axis points into the direction with the farthest reaching auto-correlation, which is described by the range parameter scale (α).
- The ranges in the direction of the second and third semi-principal axes are given by $f_1\alpha$ and $f_2\alpha$, with $0 < f_2 \leq f_1 \leq 1$.
- The default values for aniso ($f_1 = 1$, $f_2 = 1$) define an isotropic variogram model.
- Valid ranges for the angles characterizing the orientation of the semivariance ellipsoid are (in degrees): ω [0, 180], ϕ [0, 180], ζ [-90, 90].

Constraining estimates of variogram parameters:

Parameters of variogram models can vary only within certain bounds (see param.bounds and RMmodel for allowed ranges). georob uses three mechanisms to constrain parameter estimates to permissible ranges:

1. *Parameter transformations*: By default, all variance (`variance`, `snugget`, `nugget`), the range `scale` and the anisotropy parameters `f1` and `f2` are log-transformed before solving the estimating equations or maximizing the restricted loglikelihood and this warrants that the estimates are always positive (see `control.georob` for controlling parameter transformations).
2. *Checking permissible ranges*: The additional parameters of the variogram models such as the smoothness parameter ν of the Whittle-Mat\'ern model are forced to stay in the permissible ranges by signalling an error to `nleqslv` or `optim` if the current trial values are invalid. These functions then graciously update the trial values of the parameters and carry their task on. However, it is clear that such a procedure likely gets stuck at a point on the boundary of the parameter space and is therefore just a workaround for avoiding runtime errors due to invalid parameter values.
3. *Exploiting the functionality of* `nlminb` *and* `optim`: If a spatial model is fitted non-robustly, then the arguments `lower`, `upper` (and `method` of `optim`) can be used to constrain the parameters (see `control.optim` how to pass them to `optim`). For `optim` one has to use the arguments `method = "L-BFGS-B"`, `lower = l`, `upper = u`, where *l* and *u* are numeric vectors with the lower and upper bounds of the *transformed* parameters in the order as they appear in
`c( c(variance, snugget, nugget, scale, ...)[fit.param], aniso[fit.aniso])`,
where `...` are additional parameters of isotropic variogram models (use `param.names(variogram.model)` to display the names and the order of the additional parameters for `variogram.model`).

Computing robust initial estimates of parameters for robust REML:

To solve the robustified estimating equations for $\boldsymbol{B}$ and $\boldsymbol{\beta}$ the following initial estimates are used:

- $\widehat{\boldsymbol{B}} = \mathbf{0}$, if this turns out to be unfeasible, initial values can be passed to `georob` by the argument `bhat` of `control.georob`.
- $\widehat{\boldsymbol{\beta}}$ is either estimated robustly by the function `lmrob`, `rq` or non-robustly by `lm` (see argument `initial.fixef` of `control.georob`).

Finding the roots of the robustified estimating equations of the variogram and anisotropy parameters is more sensitive to a good choice of initial values than maximizing the Gaussian (restricted) loglikelihood with respect to the same parameters. Setting `initial.param = TRUE` allows one to find initial values that are often sufficiently close to the roots so that `nleqslv` converges. This is achieved by:

1. Initial values of the regression parameters are computed by `lmrob` irrespective of the choice for `initial.fixef` (see `control.georob`).
2. Observations with "robustness weights" of the `lmrob` fit, satisfying $\psi_c(\widehat{\varepsilon}_i/\widehat{\tau})/(\widehat{\varepsilon}_i/\widehat{\tau}) \leq$ `min.rweight`, are discarded (see `control.georob`).
3. The model is fit to the pruned data set by Gaussian REML using `optim`.
4. The resulting estimates of the variogram parameters (`param`, `aniso`) are used as initial estimates for the subsequent robust fit of the model by `nleqslv`.

Estimating variance parameters by Gaussian (RE)ML:

Unlike robust REML, where robustified estimating equations are solved for the variance parameters `nugget` (τ^2), `variance` (σ^2), and possibly `snugget` (σ^2_{n}), for Gaussian (RE)ML the variances can be reparametrized to

- the signal variance $\sigma^2_Z = \sigma^2 + \sigma^2_{\mathrm{n}}$,

- the inverse relative nugget $\eta = \sigma_Z^2/\tau^2$ and
- the relative spatial nugget $\xi = \sigma_n^2/\sigma_Z^2$.

georob maximizes then a (restricted) *profile loglikelihood* that depends only on η, ξ, α, ..., and σ_Z^2 is estimated by an explicit expression that depends on these parameters (e.g. Diggle and Ribeiro, 2006, p. 113). This is ususally more efficient than maximizing the (restricted) loglikelihood with respect to the original variance parameters τ^2, σ_n^2 and σ^2. georob chooses the parametrization automatically but the user can control it by the argument reparam of the function control.georob.

Value

An object of class georob representing a robust (or Gaussian) (RE)ML fit of a spatial linear model. See georobObject for the components of the fit.

Author(s)

Andreas Papritz <andreas.papritz@env.ethz.ch>
http://www.step.ethz.ch/people/scientific-staff/andreas-papritz
with contributions by Cornelia Schwierz.

References

Diggle, P. J. and Ribeiro, P. J. R. (2006) Model-based Geostatistics. Springer.

Kuensch, H. R., Papritz, A., Schwierz, C. and Stahel, W. A. (in preparation) Robust Geostatistics.

Kuensch, H. R., Papritz, A., Schwierz, C. and Stahel, W. A. (2011) Robust estimation of the external drift and the variogram of spatial data. Proceedings of the ISI 58th World Statistics Congress of the International Statistical Institute. http://e-collection.library.ethz.ch/eserv/eth:7080/eth-7080-01.pdf

See Also

georobIntro for a description of the model and a brief summary of the algorithms; georobObject for a description of the class georob; plot.georob for display of RE(ML) variogram estimates; control.georob for controlling the behaviour of georob; cv.georob for assessing the goodness of a fit by georob; predict.georob for computing robust kriging predictions; and finally georobModelBuilding for stepwise building models of class georob; georobMethods for further methods for the class georob.

Examples

```
## Not run:
################
## meuse data ##
################
data( meuse )

## Gaussian REML fit
r.logzn.reml <- georob(log(zinc) ~ sqrt(dist), data = meuse, locations = ~ x + y,
    variogram.model = "RMexp",
```

```
    param = c( variance = 0.15, nugget = 0.05, scale = 200 ),
    tuning.psi = 1000)
summary(r.logzn.reml, correlation = TRUE)

## robust REML fit
r.logzn.rob <- update(r.logzn.reml, tuning.psi = 1)

summary(r.logzn.rob, correlation = TRUE)

plot(r.logzn.reml, lag.dist.def = seq( 0, 2000, by = 100 ))
lines(r.logzn.rob, col = "red")


###################
## wolfcamp data ##
###################
data(wolfcamp, package = "geoR")
d.wolfcamp <- data.frame(x = wolfcamp[[1]][,1], y = wolfcamp[[1]][,2],
    pressure = wolfcamp[[2]])

## fitting isotropic IRF(0) model

r.irf0.iso <- georob(pressure ~ 1, data = d.wolfcamp, locations = ~ x + y,
    variogram.model = "RMfbm",
    param = c( variance = 10, nugget = 1500, scale = 1, alpha = 1.5 ),
    fit.param = c( variance = TRUE, nugget = TRUE, scale = FALSE, alpha = TRUE),
    tuning.psi = 1000)

summary(r.irf0.iso)

## fitting isotropic IRF(0) model

r.irf0.aniso <- georob(pressure ~ 1, data = d.wolfcamp, locations = ~ x + y,
    variogram.model = "RMfbm",
    param = c( variance = 5.9, nugget = 1450, scale = 1, alpha = 1 ),
    fit.param = c( variance = TRUE, nugget = TRUE, scale = FALSE, alpha = TRUE),
    aniso = c( f1 = 0.51, f2 = 1, omega = 148, phi = 90, zeta = 0 ),
    fit.aniso = c( f1 = TRUE, f2 = FALSE, omega = TRUE, phi = FALSE, zeta = FALSE ),
    tuning.psi = 1000)
summary(r.irf0.aniso)

plot(r.irf0.iso, lag.dist.def = seq(0, 200, by = 7.5))
plot(r.irf0.aniso, lag.dist.def = seq(0, 200, by = 7.5),
    xy.angle.def = c(0, 22.5, 67.5, 112.5, 157.5, 180.),
    add = TRUE, col = 2:5)

pchisq( 2*(r.irf0.aniso[["loglik"]] - r.irf0.iso[["loglik"]]), 2, lower = FALSE )
## End(Not run)
```

---

georob-S3methods *Common S3 Methods for Class* georob

Description

This page documents the methods fixef, fixed.effects, model.frame, model.matrix, nobs, print, ranef, random.effects, resid, residuals, rstandard, rstudent, summary and vcov for the class georob.

Usage

```
## S3 method for class 'georob'
fixef(object, ...)

## S3 method for class 'georob'
fixed.effects(object, ...)

## S3 method for class 'georob'
model.frame(formula, ...)

## S3 method for class 'georob'
model.matrix(object, ...)

## S3 method for class 'georob'
nobs(object, ...)

## S3 method for class 'georob'
print(x, digits = max(3, getOption("digits") - 3), ...)

## S3 method for class 'georob'
ranef(object, standard = FALSE, ...)

## S3 method for class 'georob'
random.effects(object, standard = FALSE, ...)

## S3 method for class 'georob'
resid(object,
    type = c("working", "response", "deviance", "pearson", "partial" ),
    terms = NULL,
    level = 1, ... )

## S3 method for class 'georob'
residuals(object,
    type = c("working", "response", "deviance", "pearson", "partial" ),
    terms  = NULL,
    level = 1, ... )

## S3 method for class 'georob'
rstandard(model, level = 1, ...)

## S3 method for class 'georob'
rstudent(model, ...)
```

```
## S3 method for class 'georob'
summary(object, correlation = FALSE, signif = 0.95, ...)

## S3 method for class 'georob'
vcov(object, ...)
```

Arguments

object, model, x	an object of class georob, see georobObject.
formula	a model formula or terms object or an object of class georob, see georobObject.
correlation	logical controlling whether the correlation matrix of the estimated regression coefficients and of the fitted variogram parameters (only for non-robust fits) is computed (default FALSE).
digits	positive integer indicating the number of decimal digits to print.
level	an optional integer giving the level for extracting the residuals from object. level = 0 extracts the regression residuals $\widehat{B}(\boldsymbol{s})+\widehat{\varepsilon}(\boldsymbol{s})$ and level = 1 (default) only the estimated errors $\widehat{\varepsilon}(\boldsymbol{s})$.
signif	confidence level for computing confidence intervals for variogram parameters (default 0.95).
standard	logical controlling whether the spatial random effects $\boldsymbol{B}$ should be standardized (default FALSE).
type	character keyword indicating the type of residuals to compute, see residuals.lm. type = "huber" computes 'huberized' residuals $\widehat{\sigma}/\gamma_1\psi(\widehat{\varepsilon}(\boldsymbol{s})/\widehat{\sigma})$.
terms	If type = "terms", which terms (default is all terms).
...	additional arguments passed to methods.

Details

For robust REML fits deviance returns (possibly with a warning) the deviance of the Gaussian REML fit of the equivalent Gaussian spatial linear model with heteroscedastic nugget.

The methods model.frame, model.matrix and nobs extract the model frame, model matrix and the number of observations, see help pages of respective generic functions.

The methods residuals (and resid) extract either the estimated independent errors $\widehat{\varepsilon}(\boldsymbol{s})$ or the sum of the latter quantities and the spatial random effects $\widehat{B}(\boldsymbol{s})$. rstandard does the same but standardizes the residuals to unit variance. ranef (random.effects) extracts the spatial random effects with the option to standardize them as well, and fixef (fixed.effects) extracts the fitted regression coefficients, which may of course also be obtained by coef.

Besides, the default methods of the generic functions coef, confint, df.residual, fitted, formula, termplot and update can be used for objects of class georob.

Author(s)

Andreas Papritz <andreas.papritz@env.ethz.ch>

See Also

georobIntro for a description of the model and a brief summary of the algorithms; georob for (robust) fitting of spatial linear models; georobModelBuilding for stepwise building models of class georob; georobObject for a description of the class georob.

Examples

```
## Not run:

data(meuse)

## Gaussian REML fit
r.logzn.reml <- georob(log(zinc) ~ sqrt(dist), data = meuse, locations = ~ x + y,
    variogram.model = "RMexp",
    param = c( variance = 0.15, nugget = 0.05, scale = 200 ),
    tuning.psi = 1000,
    control = control.georob(cov.bhat = TRUE, cov.ehat.p.bhat = TRUE))
summary(r.logzn.reml, correlation = TRUE)

## robust REML fit
r.logzn.rob <- update(r.logzn.reml, tuning.psi = 1)

summary(r.logzn.rob, correlation = TRUE)

## residual diagnostics
old.par <- par(mfrow = c(2,3))

plot(fitted(r.logzn.reml), rstandard(r.logzn.reml))
abline(h = 0, lty = "dotted")
qqnorm(rstandard(r.logzn.reml))
abline(0, 1)
qqnorm(ranef(r.logzn.reml, standard = TRUE))
abline(0, 1)
plot(fitted(r.logzn.rob), rstandard(r.logzn.rob))
abline(h = 0, lty = "dotted")
qqnorm(rstandard(r.logzn.rob))
abline(0, 1)
qqnorm(ranef(r.logzn.rob, standard = TRUE))
abline(0, 1)

par(old.par)

## End(Not run)
```

---

georobModelBuilding *S3 Methods for Stepwise Building Fixed-Effects Models for Class* georob

---

Description

This page documents the methods deviance, logLik, extractAIC, add1, drop1, step and waldtest for the class georob. The package georob provides a generic step function and a default method which is identical with the (non-generic) function step.

Usage

```
## S3 method for class 'georob'
deviance(object, warn = TRUE, REML = FALSE, ...)

## S3 method for class 'georob'
logLik(object, warn = TRUE, REML = FALSE, ...)

## S3 method for class 'georob'
extractAIC(fit, scale = 0, k = 2, ...)

## S3 method for class 'georob'
add1(object, scope, scale = 0, test = c("none", "Chisq"), k = 2,
   trace = FALSE, data = NULL, fixed = TRUE, use.fitted.param = TRUE, verbose = 0,
    ncores = 1, ...)

## S3 method for class 'georob'
drop1(object, scope, scale = 0, test = c("none", "Chisq"), k = 2,
   trace = FALSE, data = NULL, fixed = TRUE, use.fitted.param = TRUE, verbose = 0,
    ncores = 1, ...)

step(object, ...)

## Default S3 method:
step(object, scope, scale = 0,
    direction = c("both", "backward", "forward"), trace = 1,
    keep = NULL, steps = 1000, k = 2, ...)

## S3 method for class 'georob'
step(object, scope, scale = 0,
    direction = c("both", "backward", "forward"), trace = 1,
    keep = NULL, steps = 1000, k = 2, data = NULL, fixed = TRUE,
    use.fitted.param = TRUE, verbose = 0, ncores = 1, ...)

## S3 method for class 'georob'
waldtest(object, ..., vcov = NULL, test = c("F", "Chisq"),
    name = NULL)
```

Arguments

object, fit	an object of class georob, see georobObject.

data	an optional data frame.
direction	the mode of stepwise search, see step.
fixed	logical controlling whether the variogram parameters are *not* adjusted when fitting alternative regression models for adding or dropping model terms (default TRUE).
k	numeric specifying the 'weight' of the equivalent degrees of freedom (=: edf) part in the AIC formula, see extractAIC.
keep	a filter function whose input is a fitted model object and the associated AIC statistic, and whose output is arbitrary, see step.
name	a function for extracting a suitable name/description from a fitted model object. By default the name is queried by calling formula, see waldtest.
ncores	integer specifying the number of cores used for parallelized execution of add1 and drop1. If larger than one then the minimum of ncores, detectCores() and the number of terms to be added or dropped determines the number of cores that is actually used.
REML	logical controlling whether the restricted loglikelihood should be extracted (default TRUE).
scale	numeric, currently not used, see extractAIC.
scope	defines the range of models examined in the stepwise search. This should be either a single formula, or a list containing components upper and lower, both formulae, see step for details.
steps	the maximum number of steps to be considered (default is 1000), see step.
test	character keyword specifying whether to compute the large sample Chi-squared statistic (with asymptotic Chi-squared distribution) or the finite sample F statistic (with approximate F distribution), see waldtest.
trace	if positive, information is printed during the running of step, see step.
use.fitted.param	logical scalar controlling whether fitted values of param (and aniso are used as initial values when variogram parameters are fitted afresh for adding and dropping terms from the model (default TRUE).
vcov	a function for estimating the covariance matrix of the regression coefficients, see waldtest.
verbose	positive integer controlling logging of diagnostic messages to the console during model fitting, see georob (default 0).
warn	logical scalar controlling whether warnings should be suppressed.
...	additional arguments passed to methods (see in particular waldtest.default).

Details

For a non-robust fit the function deviance returns the residual deviance

$$(\boldsymbol{Y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}})^{\mathrm{T}}(\widehat{\tau}^2\boldsymbol{I} + \boldsymbol{\Gamma}_{\widehat{\theta}})^{-1}(\boldsymbol{Y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}})$$

(see georob-package for an explanation of the notation). For a robust fit the deviance is not defined. The function then computes with a warning the deviance of an equivalent Gaussian model with heteroscedastic nugget $\tau^2/\boldsymbol{w}$ where $\boldsymbol{w}$ are the "robustness weights" rweights, see georobObject.

logLik returns the the maximized (restricted) loglikelihood. For a robust fit, the loglikelihood is not defined. The function then computes the (restricted) loglikelihood of an equivalent Gaussian model with heteroscedastic nugget (see above).

The methods extractAIC, add1, drop1 and step are used for stepwise model building. If fixed==TRUE (default) then the variogram parameters are kept fixed at the values fitted for fixed-effects model in object. For fixed==FALSE the variogram parameters are fitted for each model tested by add1 and drop1. Then either the variogram parameters in object$initial.objects (use.fitted.param==FALSE) or the fitted parameters of object (use.fitted.param==TRUE) are used as initial values.

In addition, the functions of the R package **multcomp** can be used to test general linear hypotheses about the fixed effects of the model.

Author(s)

Andreas Papritz <andreas.papritz@env.ethz.ch>

See Also

georobIntro for a description of the model and a brief summary of the algorithms; georob for (robust) fitting of spatial linear models; georobObject for a description of the class georob; georobMethods for further methods for the class georob.

Examples

```
## Not run:

data(meuse)

## Gaussian REML fit
r.logzn.reml <- georob(log(zinc) ~ sqrt(dist), data = meuse, locations = ~ x + y,
    variogram.model = "RMexp",
    param = c( variance = 0.15, nugget = 0.05, scale = 200 ),
    tuning.psi = 1000,
    control = control.georob(cov.bhat = TRUE, cov.ehat.p.bhat = TRUE))
summary(r.logzn.reml, correlation = TRUE)

deviance(r.logzn.reml)
logLik(r.logzn.reml)

waldtest(r.logzn.reml, .~. + ffreq)

step(r.logzn.reml, ~ sqrt(dist) + ffreq + soil)

## robust REML fit
r.logzn.rob <- update(r.logzn.reml, tuning.psi = 1)

deviance(r.logzn.rob)
logLik(r.logzn.rob)
```

```
logLik(r.logzn.rob, REML=TRUE)

step(r.logzn.rob, ~ sqrt(dist) + ffreq + soil, fixed=FALSE, trace=2)

## End(Not run)
```

---

georobObject *Fitted georob Object*

Description

An object of class georob as returned by georob and representing a (robustly) fitted spatial linear model. Objects of this class have methods for model building (see georobModelBuilding) and cross-validation (see cv.georob), for computing (robust) kriging predictions (see predict.georob), for plotting (see plot.georob) and for common generic functions (see georobMethods).

Value

A georob object is a list with following components:

loglik	the maximized (restricted) Gaussian loglikelihood of a non-robust (RE)ML fit or NA for a robust fit if tuning.psi is less than tuning.psi.nr.
variogram.model	the name of the fitted parametric variogram model.
param	a named numeric vector with the (estimated) variogram parameters.
aniso	a list with the following components: • isotropic: logical indicating whether an isotropic variogram was fitted. • aniso: a named numeric vector with the (estimated) anisotropy parameters. • sincos: a list with sin and cos of the angles ω, ϕ and ζ that define the orientation of the anisotropy ellipsoid. • rotmat: the matrix $(\boldsymbol{C}_1, \boldsymbol{C}_2, \boldsymbol{C}_3)$ (see georobIntro). • sclmat: a vector with the elements 1, $1/f_1$, $1/f_2$ (see georobIntro).
gradient	a named numeric vector with the estimating equations (robust REML) or the gradient of the maximized (restricted) loglikelihood (Gaussian (RE)ML) evaluated at the solution .
tuning.psi	the value of the tuning constant c of the ψ_c-function.
coefficients	a named vector with the estimated regression coefficients.
fitted.values	a named vector with the fitted values of the external drift $\boldsymbol{X}\widehat{\boldsymbol{\beta}}$.
bhat	a named vector with the predicted spatial random effects $\widehat{\boldsymbol{B}}$ at the data locations.
residuals	a named vector with the residuals $\widehat{\boldsymbol{\varepsilon}} = \boldsymbol{Y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}} - \widehat{\boldsymbol{B}}$.
rweights	a named numeric vector with the "robustness weights" $\psi(\widehat{\varepsilon}_i/\widehat{\tau})/(\widehat{\varepsilon}_i/\widehat{\tau})$.
converged	logical indicating whether numerical maximization of the (restricted) loglikelihood by nlminb or optim or root finding by nleqslv converged.

`convergence.code`
: a diagnostic integer issued by `nlminb`, `optim` (component `convergence`) or `nleqslv` (component `termcd`) about convergence.

`iter`
: a named integer vector of length two, indicating either

  - the number of function and gradient evaluations when maximizing the (restricted) Gaussian loglikelihood by `nlminb` or `optim`, or
  - the number of function and Jacobian evaluations when solving the robustified estimating equations by `nleqslv`.

`Tmat`
: the compressed design matrix for replicated observations at coincident locations (integer vector that contains for each observation the row index of the respective unique location).

`cov`
: a list with covariance matrices (or diagonal variance vectors). Covariance matrices are stored in *compressed form* (see `compress`) and can be expanded to square matrices by `expand`. What `cov` actually contains depends on the flags passed to `georob` for computing covariances (see `control.georob`). Possible components are:

  - `cov.bhat`: the covariances of $\widehat{\boldsymbol{B}}$.
  - `cov.betahat`: the covariances of $\widehat{\boldsymbol{\beta}}$.
  - `cov.bhat.betahat`: the covariances of $\widehat{\boldsymbol{B}}$ and $\widehat{\boldsymbol{\beta}}$.
  - `cov.delta.bhat`: the covariances of $\boldsymbol{B} - \widehat{\boldsymbol{B}}$.
  - `cov.delta.bhat.betahat`: the covariances of $\boldsymbol{B} - \widehat{\boldsymbol{B}}$ and $\widehat{\boldsymbol{\beta}}$.
  - `cov.ehat`: the covariances of $\widehat{\boldsymbol{\varepsilon}} = \boldsymbol{Y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}} - \widehat{\boldsymbol{B}}$.
  - `cov.ehat.p.bhat`: the covariances of $\widehat{\boldsymbol{\varepsilon}} + \widehat{\boldsymbol{B}} = \boldsymbol{Y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}$.
  - `cov.pred.target`: a covariance term required for the back-transformation of kriging predictions of log-transformed data.

`expectations`
: a named numeric vector with the expectations of $\partial\psi_c(x)/\partial x$ (`dpsi`) and $\psi_c^2(x)$ (`psi2`) with respect to a standard normal distribution.

`Valphaxi.objects`
: a list of matrices in *compressed form* with (among others) the following components:

  - `gcr.constant`: the constant γ_0 (see expression for $\boldsymbol{V}_{\alpha,\xi}$ in section **Model** of `georobIntro`).
  - `Valphaxi`: the correlation matrix $\boldsymbol{V}_{\alpha,\xi} = \boldsymbol{\Gamma}_{\alpha,\xi}/(\sigma_{\mathrm{n}}^2 + \sigma^2)$ that includes the spatial nugget effect.
  - `Valphaxi.inverse`: the inverse of $\boldsymbol{V}_{\alpha,\xi}$.
  - `log.det.Valphaxi`: $\log(\det(\boldsymbol{V}_{\alpha,\xi}))$.

`zhat.objects`
: a list of matrices in (partly) *compressed form* with the following components:

  - `Aalphaxi`: the matrix $(\boldsymbol{X}^T\boldsymbol{V}_{\alpha,\xi}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{V}_{\alpha,\xi}^{-1}$.
  - `Palphaxi`: the matrix $\boldsymbol{I} - \boldsymbol{X}\boldsymbol{A}_{\alpha,\xi}$.
  - `Valphaxi.inverse.Palphaxi`: the matrix $\boldsymbol{V}_{\alpha,\xi}^{-1}\boldsymbol{P}_{\alpha,\xi}$.

`locations.object`
: a list with 3 components:

- `locations`: a formula indicating the coordinates of the measurement locations.
- `locations.coords`: a numeric matrix with the coordinates of the measurement locations.
- `lag.vectors`: a numeric matrix with the lag vectors between any distinct pairs of measurement locations.

`initial.objects`
: a list with 5 components:

- `coefficients`: initial estimates of $\boldsymbol{\beta}$ computed either by `lmrob` or `rq`.
- `bhat`: initial predictions of $\boldsymbol{B}$.
- `param`: numeric vector with initial estimates of the variogram parameters, either computed (`initial.param = TRUE`) or as passed to `georob` (`initial.param = FALSE`).
- `fit.param`: logical vector indicating which variogram parameters were fitted.
- `aniso`: numeric vector with initial estimates of the anisotropy parameters, either either computed (`initial.param = TRUE`) or as passed to `georob` (`initial.param = FALSE`).
- `fit.aniso`: logical vector indicating which anisotropy parameters were fitted.

`hessian`
: a symmetric matrix giving an estimate of the Hessian at the solution if the model was fitted non-robustly with the argument `hessian = TRUE` (see `control.georob`). Missing otherwise.

`control`
: a list with control parameters generated by `control.georob`.

`MD`
: optionally a matrix of robust distances in the space spanned by $\boldsymbol{X}$ (see argument `compute.rd` of `lmrob.control` and `control.georob`).

`model, x, y`
: if requested the model frame, the model matrix and the response, respectively.

`na.action, offset, contrasts, xlevels, rank, df.residual, call, terms`
: further components of the fit as described for an object of class `lm`.

Author(s)

Andreas Papritz <andreas.papritz@env.ethz.ch>

See Also

`georobIntro` for a description of the model and a brief summary of the algorithms; `georob` for (robust) fitting of spatial linear models; `control.georob` for controlling the behaviour of `georob`; `plot.georob` for display of (RE)ML variogram estimates; `cv.georob` for assessing the goodness of a fit by `georob`; `predict.georob` for computing robust kriging predictions; and finally `georobModelBuilding` for stepwise building models of class `georob`; `georobMethods` for further methods for the class `georob`.

lgnpp *Unbiased Back-Transformations for Lognormal Kriging*

Description

The function lgnpp back-transforms point or block kriging predictions of a log-transformed response variable computed by predict.georob. Alternatively, the function averages lognormal point kriging predictions for a block and approximates the mean squared prediction error of the block mean.

Usage

```
lgnpp(object, newdata, locations, is.block = FALSE, all.pred = NULL,
    extended.output = FALSE)
```

Arguments

object	an object with kriging predictions of a log-transformed response variable as obtained by predict(*georob-object, ...*).
newdata	an optional object as passed as argument newdata to predict.georob, see *Details*.
locations	an optional one-sided formula specifying what variables of newdata are the coordinates of the prediction points, see predict.georob.
is.block	an optional logical (default FALSE) specifying whether point predictions contained in object are considered to belong to a single block and should be averaged after back-transformation. Ignored if object contains block kriging predictions, see *Details*.
all.pred	an optional positive integer or an object as obtained by lgnpp(predict(*georob-object, ...*)), see *Details*.
extended.output	logical controlling whether the covariance matrix of the errors of the back-transformed point predictions is added as an attribute to the result, see *Details*.

Details

The function lgnpp performs three tasks:

1. Back-transformation of point kriging predictions of a log-transformed response:
The usual, marginally unbiased back-transformation for lognormal point kriging is used:

$$\widehat{U}(\boldsymbol{s}) = \exp(\widehat{Z}(\boldsymbol{s}) + 1/2(\mathrm{Var}_{\hat{\theta}}[Z(\boldsymbol{s})] - \mathrm{Var}_{\hat{\theta}}[\widehat{Z}(\boldsymbol{s})])),$$

$$\mathrm{Cov}_{\hat{\theta}}[U(\boldsymbol{s}_i) - \widehat{U}(\boldsymbol{s}_i), U(\boldsymbol{s}_j) - \widehat{U}(\boldsymbol{s}_j)] = \mu_{\hat{\theta}}(\boldsymbol{s}_i)\mu_{\hat{\theta}}(\boldsymbol{s}_j)$$

$$\times\{\exp(\mathrm{Cov}_{\hat{\theta}}[Z(\boldsymbol{s}_i), Z(\boldsymbol{s}_j)]) - 2\exp(\mathrm{Cov}_{\hat{\theta}}[\widehat{Z}(\boldsymbol{s}_i), Z(\boldsymbol{s}_j)]) + \exp(\mathrm{Cov}_{\hat{\theta}}[\widehat{Z}(\boldsymbol{s}_i), \widehat{Z}(\boldsymbol{s}_j)])\},$$

where $\widehat{Z}$ and $\widehat{U}$ denote the log- and back-transformed predictions of the signal, and

$$\mu_{\hat{\theta}}(\boldsymbol{s}) \approx \exp(\boldsymbol{x}(\boldsymbol{s})^{\mathrm{T}}\widehat{\beta} + 1/2\mathrm{Var}_{\hat{\theta}}[Z(\boldsymbol{s})]).$$

The expressions for the required covariance terms can be found in the Appendices of Nussbaum et al. (2012). Instead of the signal $Z(\boldsymbol{s})$, predictions of the log-transformed response $Y(\boldsymbol{s})$ or the estimated trend $\boldsymbol{x}(\boldsymbol{s})^{\mathrm{T}}\widehat{\boldsymbol{\beta}}$ of the log-transformed data can be back-transformed (see georobIntro). The above transformations are used if `object` contains point kriging predictions (see `predict.georob`, *Value*) and if `is.block = FALSE` and `all.pred` is missing.

2. Back-transformation of block kriging predictions of a log-transformed response:
Block kriging predictions of a log-transformed response variable are back-transformed by the approximately unbiased transformation proposed by Cressie (2006)

$$\widehat{U}(B) = \exp(\widehat{Z}(B) + 1/2\{\mathrm{Var}_{\hat{\theta}}[Z(\boldsymbol{s})] + \widehat{\boldsymbol{\beta}}^{\mathrm{T}}\boldsymbol{M}(B)\widehat{\boldsymbol{\beta}} - \mathrm{Var}_{\hat{\theta}}[\widehat{Z}(B)]\}),$$

$$\mathrm{E}_{\hat{\theta}}[\{U(B) - \widehat{U}(B))^2] = \mu_{\hat{\theta}}(B)^2\{\exp(\mathrm{Var}_{\hat{\theta}}[Z(B)]) - 2\exp(\mathrm{Cov}_{\hat{\theta}}[\widehat{Z}(B), Z(B)]) + \exp(\mathrm{Var}_{\hat{\theta}}[\widehat{Z}(B)])\}$$

where $\widehat{Z}(B)$ and $\widehat{U}(B)$ are the log- and back-transformed predictions of the block mean $U(B)$, respectively, $\boldsymbol{M}(B)$ is the spatial covariance matrix of the covariates

$$\boldsymbol{M}(B) = 1/|B| \int_B (\boldsymbol{x}(\boldsymbol{s}) - \boldsymbol{x}(B))(\boldsymbol{x}(\boldsymbol{s}) - \boldsymbol{x}(B))^{\mathrm{T}}\, d\boldsymbol{s}$$

with

$$\boldsymbol{x}(B) = 1/|B| \int_B \boldsymbol{x}(\boldsymbol{s})\, d\boldsymbol{s}$$

and

$$\mu_{\hat{\theta}}(B) \approx \exp(\boldsymbol{x}(B)^{\mathrm{T}}\widehat{\boldsymbol{\beta}} + 1/2\mathrm{Var}_{\hat{\theta}}[Z(B)]).$$

This back-transformation is based on the assumption that both the point data $U(\boldsymbol{s})$ and the block means $U(B)$ follow lognormal laws, which strictly cannot hold. But for small blocks the assumption works well as the bias and the loss of efficiency caused by this assumption are small (Cressie, 2006; Hofer et al., 2013).
The above formulae are used by `lgnpp` if `object` contains block kriging predictions in the form of a SpatialPolygonsDataFrame. To approximate $\boldsymbol{M}(B)$, one needs the covariates on a fine grid within the block B. The covariates are passed to `lgnpp` as argument `newdata`, where `newdata` can be any spatial data frame accepted by `predict.georob`. For evaluating $\boldsymbol{M}(B)$ the geometry of the blocks is taken from the `polygons` slot of the `SpatialPolygonsDataFrame` passed as `object` to `lgnpp`.

3. Back-transformation and averaging of point kriging predictions of a log-transformed response:
`lgnpp` allows as a further option to back-transform and *average* point kriging predictions passed as `object` to the function. One then assumes that the predictions refer to points that lie in *a single* block. Hence, one uses the approximation

$$\widehat{U}(B) \approx \frac{1}{K} \sum_{s_i \in B} \widehat{U}(\boldsymbol{s}_i)$$

to predict the block mean $U(B)$, where K is the number of points in B. The mean squared error can be approximated by

$$\mathrm{E}_{\hat{\theta}}[\{U(B) - \widehat{U}(B)\}^2] \approx \frac{1}{K^2} \sum_{s_i \in B} \sum_{s_j \in B} \mathrm{Cov}_{\hat{\theta}}[U(\boldsymbol{s}_i) - \widehat{U}(\boldsymbol{s}_i), U(\boldsymbol{s}_j) - \widehat{U}(\boldsymbol{s}_j)].$$

In most instances, the evaluation of the above double sum is not feasible because a large number of points is used to discretize the block B. lgnpp then uses the following approximation for the mean squared error (see also Appendix E of Nussbaum et al., 2012):

- Prediction results are passed as `object` to `lgnpp` only for a *random sample of points in* B (of size k), for which the evaluation of the above double sum is feasible.
- The prediction results for the *complete set of points* within the block are passed as argument `all.pred` to `lgnpp`. These results are used to compute $\widehat{U}(B)$.
- The mean squared error is then approximated by

$$\mathrm{E}_{\hat{\theta}}[\{U(B) - \widehat{U}(B)\}^2] \approx \frac{1}{K^2} \sum_{s_i \in B} \mathrm{E}_{\hat{\theta}}[\{U(\boldsymbol{s}_i) - \widehat{U}(\boldsymbol{s}_i)\}^2]$$

$$+ \frac{K-1}{Kk(k-1)} \sum_{s_i \in \text{sample}} \sum_{s_j \in \text{sample}, s_j \neq s_i} U(\boldsymbol{s}_i) - \widehat{U}(\boldsymbol{s}_i), U(\boldsymbol{s}_j) - \widehat{U}(\boldsymbol{s}_j)].$$

  The first term of the RHS can be computed from the point kriging results contained in `all.pred`, and the double sum is evaluated from the full covariance matrices of the predictions and the respective targets, passed to `lgnpp` as `object`.
- If the prediction results are not available for the complete set of points in B then `all.pred` may be equal to K. The block mean is then approximated by

$$\widehat{U}(B) \approx \frac{1}{k} \sum_{s_i \in \text{sample}} \widehat{U}(\boldsymbol{s}_i)$$

  and the first term of the RHS of the expression for the mean squared error by

$$\frac{1}{kK} \sum_{s_i \in \text{sample}} \mathrm{E}_{\hat{\theta}}[\{U(\boldsymbol{s}_i) - \widehat{U}(\boldsymbol{s}_i)\}^2].$$

- By drawing samples repeatedly and passing the related kriging results as `object` to `lgnpp`, one can reduce the error of the approximation of the mean squared error.

Value

If `is.block` is `FALSE` and `all.pred` is equal to `NULL` an updated object of the same class as `object` (see section *Value* of `predict.georob`). The data frame with the point or block kriging predictions is complemented by `lgnpp` with the following new components:

- `lgn.pred`: the back-transformed kriging predictions of a log-transformed response.
- `lgn.se`: the standard errors of the back-transformed predictions.
- `lgn.lower`, `lgn.upper`: the bounds of the back-transformed prediction intervals.

If `is.block` is `TRUE` or `all.pred` not equal to `NULL` a named numeric vector with two elements:

- `mean`: the back-transformed block kriging estimate, see *Details*.
- `mse`: the (approximated) block kriging variance, see *Details*.

If `extended.output` is `TRUE` then the vector is supplemented with the attribute `mse.lgn.pred` that contains the full covariance matrix of the back-transformed point prediction errors.

Author(s)

Andreas Papritz <andreas.papritz@env.ethz.ch>.

References

Cressie, N. (2006) Block Kriging for Lognormal Spatial Processes. *Mathematical Geology*, **38**, 413–443.

Hofer, C., Borer, F., Bono, R., Kayser, A. and Papritz, A. 2013. Predicting topsoil heavy metal content of parcels of land: An empirical validation of customary and constrained lognormal block kriging and conditional simulations. *Geoderma*, **193–194**, 200–212.

Nussbaum, M., Papritz, A., Baltensweiler, A. and Walthert, L. (2012) *Organic Carbon Stocks of Swiss Forest Soils*, Institute of Terrestrial Ecosystems, ETH Zurich and Swiss Federal Institute for Forest, Snow and Landscape Research (WSL), pp. 51. http://e-collection.library.ethz.ch/eserv/eth:6027/eth-6027-01.pdf

See Also

`georobIntro` for a description of the model and a brief summary of the algorithms; `georob` for (robust) fitting of spatial linear models; `predict.georob` for computing robust kriging predictions.

Examples

```
## Not run:
data(meuse )

data(meuse.grid)
coordinates(meuse.grid) <- ~x+y
meuse.grid.pixdf <- meuse.grid
gridded(meuse.grid.pixdf) <- TRUE

library(constrainedKriging)
data(meuse.blocks)

r.logzn.rob <- georob(log(zinc) ~ sqrt(dist), data = meuse, locations = ~ x + y,
    variogram.model = "RMexp", param = c( variance = 0.15, nugget = 0.05, scale = 200 ),
    tuning.psi = 1., control = control.georob(cov.bhat = TRUE, full.cov.bhat = TRUE,
        cov.bhat.betahat = TRUE, aux.cov.pred.target = TRUE))
```

```
## point predictions of log(Zn)
r.pred.points <- predict(r.logzn.rob, newdata = meuse.grid.pixdf,
    control = control.predict.georob(extended.output = TRUE, full.covmat = TRUE))
str(r.pred.points$pred@data)

## back-transformation of point predictions
r.backtf.pred.points <- lgnpp(r.pred.points)
str(r.backtf.pred.points$pred@data)

spplot(r.backtf.pred.points[["pred"]], zcol = "lgn.pred", main = "Zn content")

## predicting mean Zn content for whole area
r.block <- lgnpp(r.pred.points, is.block = TRUE, all.pred = r.backtf.pred.points[["pred"]])
r.block

## block predictions of log(Zn)
r.pred.block <- predict(r.logzn.rob, newdata = meuse.blocks,
    control = control.predict.georob( extended.output = TRUE,
        pwidth = 75, pheight = 75))
r.backtf.pred.block <- lgnpp(r.pred.block, newdata = meuse.grid)

spplot(r.backtf.pred.block, zcol = "lgn.pred", main = "block means Zn content")
## End(Not run)
```

`param.names` *Names and Permissible Ranges of Variogram Parameters*

Description

Helper functions to query names and permissble ranges of variogram parameters.

Usage

```
param.names(model)

param.bounds(model, d)
```

Arguments

`model`	character keyword denoting a valid variogram, see `georob` and `georobIntro`.
`d`	integer equal number of dimensions of the survey domain.

Value

Either a character vector with the names of the additional variogram parameters such as the smoothness parameter of the Whittle-Matérn model (`param.names`) or a named list with the lower and upper bounds of permissble parameter ranges.

Author(s)

Andreas Papritz <andreas.papritz@env.ethz.ch>

See Also

georobIntro for a description of the model and a brief summary of the algorithms; georob for (robust) fitting of spatial linear models.

Examples

```
param.names("RMgengneiting")
param.bounds("RMgengneiting", d = 2)
```

---

plot.georob *Plot Methods for Class* georob

Description

The plot and lines methods for class georob plot the variogram model, estimated by (robust) restricted maximum likelihood. plot.georob computes and plots in addition the sample variogram of the (robust) regression residuals.

Usage

```
## S3 method for class 'georob'
plot(x, type, what = c("variogram", "covariance", "correlation"),
    plot.sv = TRUE, add = FALSE, lag.dist.def,
    xy.angle.def = c(0, 180), xz.angle.def = c(0, 180),
    max.lag = Inf, estimator = c("mad", "qn", "ch", "matheron"),
    mean.angle = TRUE, col, pch, lty, ...)

## S3 method for class 'georob'
lines(x, what = c("variogram", "covariance", "correlation"),
    from = 1.e-6, to, n = 501, xy.angle = 90, xz.angle = 90,
    col = 1:length(xy.angle), pch = 1:length(xz.angle), lty = "solid", ...)
```

Arguments

x	an object of class georob, see georobObject.
type	the type of plot for display of the sample variogram, see plot.
what	the quantity that should be displayed. Note that plot.sv will be set to FALSE unless what == "variogram" (default).
plot.sv	logical controlling if the sample variogram of the regression residuals, $\widehat{B}(\boldsymbol{s}) + \widehat{\varepsilon}(\boldsymbol{s})$ should be added to the plot (default TRUE).
add	logical controlling whether a new plot should be generated (FALSE, default) or whether the information should be added to the current plot (TRUE).

lag.dist.def	a numeric scalar defining a constant bin width for grouping the lag distances or a numeric vector with the upper bounds of a set of contiguous bins.
xy.angle.def	an numeric vector defining angular classes in the horizontal plane for computing directional variograms. xy.angle.def must contain an ascending sequence of azimuth angles in degrees from north (positive clockwise to south), see sample.variogram. Omnidirectional variograms are computed with the default c(0,180).
xz.angle.def	an numeric vector defining angular classes in the x-z-plane for computing directional variograms. xz.angle.def must contain an ascending sequence of angles in degrees from zenith (positive clockwise to nadir), see sample.variogram. Omnidirectional variograms are computed with the default c(0,180).
max.lag	positive numeric defining the largest lag distance for which semivariances should be computed (default no restriction).
estimator	character keyword defining the estimator for computing the sample variogram. Possible values are: • "qn": Genton's robust Qn-estimator (default, Genton, 1998), • "mad": Dowd's robust MAD-estimator (Dowd, 1984), • "matheron": non-robust method-of-moments estimator, • "ch": robust Cressie-Hawkins estimator (Cressie and Hawkins, 1980).
mean.angle	logical controlling whether the mean lag vector (per combination of lag distance and angular class) is computed from the mean angles of all the lag vectors falling into a given class (TRUE, default) or from the mid-angles of the respective angular classes (FALSE).
from	numeric, minimal lag distance for plotting variogram models.
to	numeric, maximum lag distance for plotting variogram models (default: largest lag distance of current plot).
n	positive integer specifying the number of equally spaced lag distances for which semivariances are evaluated in plotting variogram models (default 501).
xy.angle	numeric (vector) with azimuth angles (in degrees, clockwise positive from north) in x-y-plane for which semivariances should be plotted.
xz.angle	numeric (vector) with angles in x-z-plane (in degrees, clockwise positive from zenith to south) for which semivariances should be plotted.
col	color of curves to distinguish curves relating to different azimuth angles in x-y-plane.
pch	type of plotting symbols added to lines to distinguish curves relating to different angles in x-z-plane.
lty	line type for plotting variogram models.
...	additional arguments passed to plot.sample.variogram and methods.

Author(s)

Andreas Papritz <andreas.papritz@env.ethz.ch>.

See Also

georobIntro for a description of the model and a brief summary of the algorithms; georob for (robust) fitting of spatial linear models; georobObject for a description of the class georob; sample.variogram for computing sample variograms.

Examples

```
## Not run:
################
## meuse data ##
################
data( meuse )

## Gaussian REML fit
r.logzn.reml <- georob(log(zinc) ~ sqrt(dist), data = meuse, locations = ~ x + y,
    variogram.model = "RMexp",
    param = c( variance = 0.15, nugget = 0.05, scale = 200 ),
    tuning.psi = 1000)
summary(r.logzn.reml, correlation = TRUE)

## robust REML fit
r.logzn.rob <- update(r.logzn.reml, tuning.psi = 1)

summary(r.logzn.rob, correlation = TRUE)

plot(r.logzn.reml, lag.dist.def = seq( 0, 2000, by = 100 ))
lines(r.logzn.rob, col = "red")
## End(Not run)
```

---

pmm *Parallelized Matrix Multiplication*

Description

This page documents the functions pmm for parallelized matrix multiplication and the function control.pmm, which controls the behaviour of pmm and other functions that execute tasks in parallel.

Usage

```
pmm(A, B, control = control.pmm())

control.pmm(ncores = 1, max.ncores = detectCores(),
    f = 2, sfstop = FALSE, allow.recursive = TRUE, ...)
```

Arguments

A, B	matrices to be mulitplied.
control	a list of with the arguments ncores, f, and sfstop or a function such as control.pmm that generates such a list.
ncores	number (integer, default 1) of cores used for parallelized matrix multiplication.
max.ncores	maximum number of cores (integer, default all cores of a machine) used for parallelized computations.
f	number (integer, default 2) of tasks assigned on non-Windows OS to each core in parallelized matrix multiplication.
sfstop	logical controlling whether the SNOW socket cluster is stopped after each parallelized matrix multiplication on windows OS (default FALSE).
allow.recursive	logical controlling whether nested parallelized computation should be allowed (default TRUE).
...	further arguments, currently not used.

Details

Parallelized matrix multiplication shortens computing time for large data sets ($n > 1000$). However, the spawning of child processes requires itself ressources and increasing the number of cores for parallel matrix multiplication and evaluation of estimating equations/gradient does not always result in reduced computing time. A sensible default for the number of cores assigned to both tasks is likely ncores=2.

max.ncores controls how many child processes are spawned in total for both tasks. This can be used to prevent that child processes spawn themselves children which may result in a considerable number of child processes.

Value

pmm: the matrix product A %*% B,

control.pmm: a list with components ncores, max.ncores, f, sfstop, allow.recursive.

Author(s)

Andreas Papritz <andreas.papritz@env.ethz.ch>

Examples

```
## Not run:
A <- as.matrix(dist(rnorm(2000)))
B <- as.matrix(dist(rnorm(2000)))
system.time(C <- pmm(A, B, control = control.pmm(ncores = 1)))
system.time(C <- pmm(A, B, control = control.pmm(ncores = 4)))

## End(Not run)
```

predict.georob *Predict Method for Robustly Fitted Spatial Linear Models*

Description

Robust and customary external drift kriging prediction based on a spatial linear models fitted by georob. The predict method for the class georob computes fitted values, point and block kriging predictions as well as model terms for display by termplot.

Usage

```
## S3 method for class 'georob'
predict(object, newdata, type =  c("signal", "response", "trend", "terms"),
    terms = NULL, se.fit = TRUE, signif = 0.95, locations,
    control = control.predict.georob(), verbose = 0, ...)

control.predict.georob(full.covmat = FALSE, extended.output = FALSE,
    mmax = 10000, ncores = pmm[["max.ncores"]], pwidth = NULL, pheight = NULL,
    napp = 1, pmm = control.pmm())
```

Arguments

object	an object of class "georob", see georobObject.
newdata	an optional data frame, SpatialPointsDataFrame, SpatialPixelsDataFrame, SpatialGridDataFrame or SpatialPolygonsDataFrame in which to look for variables with which to compute fitted values or kriging predictions. If newdata is a SpatialPolygonsDataFrame then block kriging predictions are computed, otherwise point kriging predictions.
type	character keyword defining what target quantity should be predicted (computed). Possible values are • "signal": the "signal" $Z(\boldsymbol{s}) = \boldsymbol{x}(\boldsymbol{s})^{\mathrm{T}}\boldsymbol{\beta} + B(\boldsymbol{s})$ of the process (default), • "response": the observations $Y(\boldsymbol{s}) = Z(\boldsymbol{s}) + \varepsilon(\boldsymbol{s})$, • "trend": the external drift $\boldsymbol{x}(\boldsymbol{s})^{\mathrm{T}}\boldsymbol{\beta}$, • "terms": the model terms.
terms	If type = "terms", which terms (default is all terms).
se.fit	logical, only used if type is equal to "terms", see predict.lm.
signif	positive numeric equal to the tolerance or confidence level for computing respective intervals. If NULL no intervals are computed.
locations	an optional one-sided formula specifying what variables of newdata are the coordinates of the prediction points (default: object[["locatons.objects"]]$locations).
control	a list with the arguments full.covmat, extended.output, mmax, ncores, pwidth, pheight, napp and pmm or a function such as control.predict.georob that generates such a list.

`full.covmat` logical controlling whether the full covariance matrix of the prediction errors is returned (`TRUE`) or only the vector with its diagonal elements (`FALSE`, default), see *Value* for an explanation of the effect of `full.covmat`.

`extended.output` logical controlling whether the covariance matrices of the kriging predictions and of the data should be computed, see *Details* (default `FALSE`).

`mmax` integer equal to the maximum number (default `10000`) of prediction items, computed in a sub-task, see *Details*.

`ncores` positive integer controlling how many cores are used for parallelized computations, see *Details*.

`pwidth, pheight, napp` numeric scalars, used to tune numeric integration of semivariances for block kriging, see `preCKrige`.

`pmm` a list of arguments passed to `pmm` or a function such as `control.pmm` that generates such a list (see `control.pmm` for allowed arguments).

`verbose` positive integer controlling logging of diagnostic messages to the console. `verbose = 0` (default) largely suppresses such messages.

`...` arguments passed to `control.predict.georob`.

Details

The `predict` method for class `georob` uses the package **parallel** for parallelized computation of kriging predictions. If there are m items to predict, the task is split into `ceiling(m/mmax)` sub-tasks that are then distributed to `ncores` CPUs. Evidently, `ncores = 1` suppresses parallel execution. By default, the function uses all available CPUs as returned by `detectCores`.
Note that if `full.covmat` is `TRUE` `mmax` must exceed m (and parallel execution is not possible).

The argument `extended.output = TRUE` is used to compute all quantities required for (approximately) unbiased back-transformation of kriging predictions of log-transformed data to the original scale of the measurements by `lgnpp`. In more detail, the following items are computed:

- `trend`: the fitted values, $\boldsymbol{x}(\boldsymbol{s})^{\mathrm{T}}\widehat{\boldsymbol{\beta}}$,
- `var.pred`: the variances of the kriging predictions, $\mathrm{Var}_{\hat{\theta}}[\widehat{Y}(\boldsymbol{s})]$ or $\mathrm{Var}_{\hat{\theta}}[\widehat{Z}(\boldsymbol{s})]$,
- `cov.pred.target`: the covariances between the predictions and the prediction targets, $\mathrm{Cov}_{\hat{\theta}}[\widehat{Y}(\boldsymbol{s}), Y(\boldsymbol{s})]$ or $\mathrm{Cov}_{\hat{\theta}}[\widehat{Z}(\boldsymbol{s}), Z(\boldsymbol{s})]$,
- `var.target`: the variances of the prediction targets $\mathrm{Var}_{\hat{\theta}}[Y(\boldsymbol{s})]$ or $\mathrm{Var}_{\hat{\theta}}[Z(\boldsymbol{s})]$.

Note that the component `var.pred` is also present if `type` is equal to `"trend"`, irrespective of the choice for `extended.output`. This component contains then the variances of the fitted values.

Value

If `type` is equal to `"terms"` then a vector, a matrix, or a list with prediction results along with bounds and standard errors, see `predict.lm`. Otherwise, the structure and contents of the output generated by `predict.georob` are determined by the class of `newdata` and the logical flags `full.covmat` and `extended.output`:

If full.covmat is FALSE then the result is an object of the same class as newdata (data frame, SpatialPointsDataFrame, SpatialPixelsDataFrame SpatialGridDataFrame, SpatialPolygonsDataFrame). The data frame or the data slot of the Spatial...DataFrame objects have the following components:

- the coordinates of the prediction points (only present if newdata is a data frame).
- pred: the kriging predictions (or fitted values).
- se: the root mean squared prediction errors (kriging standard errors).
- lower, upper: the limits of tolerance/confidence intervals,
- trend, var.pred, cov.pred.target, var.target: only present if extend.output is TRUE, see *Details*.

If full.covmat is TRUE then predict.georob returns a list with the following components:

- pred: a data frame or a Spatial...DataFrame object as described above for full.covmat = FALSE.
- mse.pred: the full covariance matrix of the prediction errors, $Y(\boldsymbol{s}) - \widehat{Y}(\boldsymbol{s})$ or $Z(\boldsymbol{s}) - \widehat{Z}(\boldsymbol{s})$ see *Details*.
- var.pred: the full covariance matrix of the kriging predictions, see *Details*.
- cov.pred.target: the full covariance matrix of the predictions and the prediction targets, see *Details*.
- var.target: the full covariance matrix of the prediction targets, see *Details*.

Author(s)

Andreas Papritz <andreas.papritz@env.ethz.ch>

References

Nussbaum, M., Papritz, A., Baltensweiler, A. and Walthert, L. (2012) *Organic Carbon Stocks of Swiss Forest Soils*, Institute of Terrestrial Ecosystems, ETH Zurich and Swiss Federal Institute for Forest, Snow and Landscape Research (WSL), pp. 51. http://e-collection.library.ethz.ch/eserv/eth:6027/eth-6027-01.pdf

Kuensch, H. R., Papritz, A., Schwierz, C. and Stahel, W. A. (2011) Robust estimation of the external drift and the variogram of spatial data. Proceedings of the ISI 58th World Statistics Congress of the International Statistical Institute. http://e-collection.library.ethz.ch/eserv/eth:7080/eth-7080-01.pdf

See Also

georobIntro for a description of the model and a brief summary of the algorithms; georob for (robust) fitting of spatial linear models; georobObject for a description of the class georob.

Examples

```
## Not run:
data(meuse )

data(meuse.grid)
coordinates(meuse.grid) <- ~x+y
meuse.grid.pixdf <- meuse.grid
gridded(meuse.grid.pixdf) <- TRUE

library(constrainedKriging)
data(meuse.blocks)

r.logzn.rob <- georob(log(zinc) ~ sqrt(dist), data = meuse, locations = ~ x + y,
    variogram.model = "RMexp", param = c( variance = 0.15, nugget = 0.05, scale = 200 ),
    tuning.psi = 1., control = control.georob(cov.bhat = TRUE, full.cov.bhat = TRUE,
        cov.bhat.betahat = TRUE, aux.cov.pred.target = TRUE))

## point predictions of log(Zn)
r.pred.points <- predict(r.logzn.rob, newdata = meuse.grid.pixdf,
    control = control.predict.georob(extended.output = TRUE, full.covmat = TRUE))
str(r.pred.points$pred@data)

## back-transformation of point predictions
r.backtf.pred.points <- lgnpp(r.pred.points)
str(r.pred.points$pred@data)

spplot(r.backtf.pred.points[["pred"]], zcol = "lgn.pred", main = "Zn content")

## predicting mean Zn content for whole area
r.block <- lgnpp(r.pred.points, is.block = TRUE, all.pred = r.backtf.pred.points[["pred"]])
r.block

## block predictions of log(Zn)
r.pred.block <- predict(r.logzn.rob, newdata = meuse.blocks,
    control = control.predict.georob( extended.output = TRUE,
        pwidth = 75, pheight = 75))
r.backtf.pred.block <- lgnpp(r.pred.block, newdata = meuse.grid)

spplot(r.backtf.pred.block, zcol = "lgn.pred", main = "block means Zn content")
## End(Not run)
```

`profilelogLik` *Profile Likelihood*

Description

The function `profilelogLik` computes for an array of fixed variogram parameters the profile log-likelihood by maximizing the (restricted) loglikelihood with respect to the remaining variogram parameters, the fixed and random effects.

Usage

```
profilelogLik(object, values, use.fitted = TRUE, verbose = 0,
  ncores = min(detectCores(), NROW(values)))
```

Arguments

object	an object of class "georob", see georobObject.
values	a data.frame or a matrix with the values of the variogram and anisotropy parameters that should be kept fixed (see georob and georobIntro for information about the parametrization of variogram models). The names of the columns of values must match the names of variogram and anisotropy parameters.
use.fitted	logical scalar controlling whether the fitted variogram parameters of object should be used as initial values (default TRUE) when maximizing the profile log-likelihood or the initial values used to generate object.
verbose	positive integer controlling logging of diagnostic messages to the console during model fitting, see georob.
ncores	positive integer controlling how many cores are used for parallelized computations, see *Details*.

Details

For robust REML fits profilelogLik returns (possibly with a warning) the loglikelihood of the Gaussian (RE)ML fit of the equivalent Gaussian spatial linear model with heteroscedastic nugget.

Note that *the dataframe passed as* data *argument to* georob *must exist in the user workspace when calling* profilelogLik.

profilelogLik uses the package **parallel** for parallelized computation of the profile likelihood. By default, the function uses NROW(values) CPUs but not more than are physically available (as returned by detectCores).

profilelogLik uses the function update to re-estimated the model with partly fixed variogram parameters. Therefore, any argument accepted by georob can be changed when re-fitting the model. Some of them (e.g. values, verbose) are explicit arguments of profilelogLik, but also the remaining ones can be passed by ... to the function.

Value

A data.frame with the columns of values, a column loglik (contains the maximized [restricted] loglikelihood), columns with the estiamted variogram and fixed effect parameters and the column converged indicating whether convergence has occurred converged ==1 when fitting the respective model.

Author(s)

Andreas Papritz <andreas.papritz@env.ethz.ch>

See Also

georob for (robust) fitting of spatial linear models.

Examples

```
## Not run:

data(meuse)

r.logzn.ml <- georob(log(zinc)~sqrt(dist), data=meuse, locations=~x+y,
    variogram.model="RMexp", param=c(variance=0.15, nugget=0.05, scale=200),
    tuning.psi=1000, control=control.georob(ml.method="ML"))

r.prflik <- profilelogLik(r.logzn.ml, values=expand.grid(scale=seq(75, 600, by=25)))
plot(loglik~scale, r.prflik, type="l")
abline(v=r.logzn.ml$param["scale"], lty="dotted")
abline(h=r.logzn.ml$loglik-0.5*qchisq(0.95, 1), lty="dotted")


## End(Not run)
```

---

sample.variogram *Computing (Robust) Sample Variograms of Spatial Data*

Description

The function sample.variogram computes the sample (empirical) variogram of a spatial variable by the method-of-moment and three robust estimators. Both omnidirectional and direction-dependent variograms can be computed, the latter for observation locations in a three-dimensionsal domain. There are summary and plot methods for summarizing and displaying sample variograms.

Usage

```
sample.variogram(response, locations, lag.dist.def,
    xy.angle.def = c(0, 180), xz.angle.def = c(0, 180), max.lag = Inf,
    estimator = c("qn", "mad", "matheron", "ch"), mean.angle = TRUE)

## S3 method for class 'sample.variogram'
summary(object, ...)

## S3 method for class 'sample.variogram'
plot(x, type = "p", add = FALSE, xlim = c(0, max(x[["lag.dist"]])),
    ylim = c(0, 1.1 * max(x[["gamma"]])), col, pch, cex = 0.8,
    xlab = "lag distance", ylab = "semivariance",
    annotate.npairs = FALSE, npairs.pos = 3, npairs.cex = 0.7,
    legend = nlevels(x[["xy.angle"]]) > 1 || nlevels(x[["xz.angle"]]) > 1,
    legend.pos = "topleft", ...)
```

Arguments

response	a numeric vector with the values of the response for which the sample variogram should be computed.
locations	a numeric matrix with the coordinates of the locations where the response was observed. May have an arbitrary number of columns for an omnidirectional variogram, but at most 3 columns if a directional variogram is computed.
lag.dist.def	a numeric scalar defining a constant bin width for grouping the lag distances or a numeric vector with the upper bounds of a set of contiguous bins.
xy.angle.def	an numeric vector defining angular classes in the horizontal plane for computing directional variograms. xy.angle.def must contain an ascending sequence of azimuth angles in degrees from north (positive clockwise to south), see *Details*. Omnidirectional variograms are computed with the default c(0,180).
xz.angle.def	an numeric vector defining angular classes in the x-z-plane for computing directional variograms. xz.angle.def must contain an ascending sequence of angles in degrees from zenith (positive clockwise to nadir), see *Details*. Omnidirectional variograms are computed with the default c(0,180).
max.lag	positive numeric defining the largest lag distance for which semivariances should be computed (default no restriction).
estimator	character keyword defining the estimator for computing the sample variogram. Possible values are: • "qn": Genton's robust Qn-estimator (default, Genton, 1998), • "mad": Dowd's robust MAD-estimator (Dowd, 1984), • "matheron": non-robust method-of-moments estimator, • "ch": robust Cressie-Hawkins estimator (Cressie and Hawkins, 1980).
mean.angle	logical controlling whether the mean lag vector (per combination of lag distance and angular class) is computed from the mean angles of all the lag vectors falling into a given class (TRUE, default) or from the mid-angles of the respective angular classes (FALSE).
object, x	an object of class sample.variogram.
type, xlim, ylim, xlab, ylab	see respective arguments of plot.default.
add	logical controlling whether a new plot should be generated (FALSE, default) or whether the information should be added to the current plot (TRUE).
col	the color of plotting symbols for distinguishing semivariances for angular classes in the x-y-plane.
pch	the type of plotting symbols for distinguishing semivariances for angular classes in the x-z-plane.
cex	character expansion factor for plotting symbols.
annotate.npairs	logical controlling whether the plotting symbols should be annotated by the number of data pairs per lag class.
npairs.pos	integer defining the position where text annotation about number of pairs should be plotted, see text.

npairs.cex	numeric defining the character expansion for text annotation about number of pairs.
legend	logical controlling whether a legend should be plotted.
legend.pos	a character keyword defining where to place the legend, see legend for possible values.
...	additional arguments passed to plot.sample.variogram and lines.georob.

Details

The angular classes in the x-y- and x-z-plane are defined by vectors of ascending angles on the half circle. The ith angular class is defined by the vector elements, say l and u, with indices i and $i + 1$. A lag vector belongs to the ith angular class if its azimuth (or angle from zenith), say φ, satisfies $l < \varphi \leq u$. If the first and the last element of xy.angle.def or xz.angle.def are equal to 0 and 180 degrees, respectively, then the first and the last angular class are "joined", i.e., if there are K angles, there will be only $K - 2$ angular classes and the first class is defined by the interval (*xy.angle.def*[K-1]-180, *xy.angle.def*[2]] and the last class by (*xy.angle.def*[K-2], *xy.angle.def*[K-1]].

Value

An object of class sample.variogram, which is a data frame with the following components:

lag.dist	the mean lag distance of the lag class,
xy.angle	the angular class in the x-y-plane,
xz.angle	the angular class in the x-z-plane,
gamma	the estimated semivariance of the lag class,
npairs	the number of data pairs in the lag class,
lag.x	the x-component of the mean lag vector of the lag class,
lag.x	the y-component of the mean lag vector of the lag class,
lag.z	the z-component of the mean lag vector of the lag class.

Author(s)

Andreas Papritz <andreas.papritz@env.ethz.ch>.

References

Cressie, N. and Hawkins, D. M. (1980) Robust Estimation of the Variogram: I. *Mathematical Geology*, **12**, 115–125.

Dowd, P. A. (1984) The variogram and kriging: Robust and resistant estimators. In *Geostatistics for Natural Resources Characterization*, Verly, G., David, M., Journel, A. and Marechal, A. (Eds.) Dordrecht: D. Reidel Publishing Company, Part I, 1, 91–106.

Genton, M. (1998) Highly Robust Variogram Estimation. *Mathematical Geology*, **30**, 213–220.

See Also

georobIntro for a description of the model and a brief summary of the algorithms; georob for

(robust) fitting of spatial linear models; fit.variogram.model for fitting variogram models to sample variograms.

Examples

```
data(wolfcamp, package = "geoR")

## fitting an isotropic IRF(0) model
r.sv.iso <- sample.variogram(wolfcamp[["data"]], locations = wolfcamp[[1]],
    lag.dist.def = seq(0, 200, by = 15))

## Not run:
plot( r.sv.iso, type = "l")
## End(Not run)

## fitting an anisotropic IRF(0) model
r.sv.aniso <- sample.variogram(wolfcamp[["data"]],
    locations = wolfcamp[[1]], lag.dist.def = seq(0, 200, by = 15),
    xy.angle.def = c(0., 22.5, 67.5, 112.5, 157.5, 180.))
## Not run:
plot(r.sv.aniso, type = "l", add = TRUE, col = 2:5)
## End(Not run)
```

---

validate.predictions *Summary Statistics of (Cross-)Validation Prediction Errors*

Description

Functions to compute and plot summary statistics of prediction errors to (cross-)validate fitted spatial linear models by the criteria proposed by Gneiting et al. (2007) for assessing probabilistic forecasts.

Usage

```
validate.predictions(data, pred, se.pred,
    statistic = c( "crps", "pit", "mc", "bs", "st" ), ncutoff = NULL)

## S3 method for class 'cv.georob'
plot(x,
    type = c( "sc", "lgn.sc", "ta", "qq", "hist.pit", "ecdf.pit", "mc", "bs" ),
    smooth = TRUE, span = 2/3, ncutoff = NULL, add = FALSE,
    col, pch, lty, main, xlab, ylab, ...)

## S3 method for class 'cv.georob'
print(x, digits = max(3, getOption("digits") - 3), ...)
```

```
## S3 method for class 'cv.georob'
rstudent(model, ...)

## S3 method for class 'cv.georob'
summary(object, se = FALSE, ...)
```

Arguments

data	a numeric vector with observations about a response.
pred	a numeric vector with predictions for the response.
se.pred	a numeric vector with prediction standard errors.
statistic	character keyword defining what statistic of the prediction errors should be computed. Possible values are (see *Details*): • "crps": continuous ranked probability score (default), • "pit": probability integral transform, • "mc": average predictive distribution (marginal calibration), • "bs": Brier score, • "st": mean and dispersion statistics of (standardized) prediction errors.
ncutoff	positive integer (N) giving the number of quantiles, for which CDFs are evaluated (type = "mc"), or the number of thresholds for which the Brier score is computed (type = "bs"), see *Details* (default: min(500, length(data))).
x, model, object	objects of class cv.georob.
digits	positive integer indicating the number of decimal digits to print.
type	character keyword defining what type of plot is created by the plot.cv.georob. Possible values are: • "sc": a scatterplot of the (possibly log-transformed) response vs. the respective predictions (default). • "lgn.sc": a scatterplot of the untransformed response against back-transformed predictions of the log-transformed response. • "ta": Tukey-Anscombe plot (plot of standardized prediction errors vs. predictions). • "qq": normal QQ plot of standardized prediction errors. • "hist.pit": histogram of probability integral transform, see *Details*. • "ecdf.pit": empirical cdf of probability integral transform, see *Details*. • "mc": a marginal calibration plot, see *Details*, • "bs": a plot of Brier score vs. threshold, see *Details*.
smooth	control whether scatter plots of data vs. predictions should be smoothed by loess.smooth.
span	smoothness parameter for loess (see loess.smooth).
add	logical controlling wether the current high-level plot is added to an existing graphics without cleaning the frame before (default: FALSE).

`main, xlab, ylab`
: title and axes labels of plot.

`col, pch, lty`
: color, symbol and line type.

`se`
: logical controlling if the standard errors of the averaged continuous ranked probability score and of the mean and dispersion statistics of the prediction errors (see *Details*) are computed from the respective values computed for the K cross-validation subsets (default: `FALSE`).

`...`
: additional arguments passed to the methods.

Details

`validate.predictions` computes the items required to evaluate (and plot) the diagnostic criteria proposed by Gneiting et al. (2007) for assessing the *calibration* and the *sharpness* of probabilistic predictions. To this aim, `validate.predictions` uses the assumption that the prediction errors $Y(\boldsymbol{s}) - \widehat{Y}(\boldsymbol{s})$ follow normal distributions with zero mean and standard deviations equal to the kriging standard errors. This assumption is used to compute

- the *probability integral transform* (PIT),

$$\mathrm{PIT}_i = F_i(y_i),$$

  where $F_i(y_i)$ denotes the predictive CDF evaluted at y_i, the value of the ith (cross-)validation datum,

- the *average predictive CDF*

$$\bar{F}_n(y) = 1/n \sum_{i=1}^{n} F_i(y),$$

  where n is the number of (cross-)validation observations and the F_i are evaluated at N quantiles equal to the set of distinct y_i (or a subset of size N of them),

- the *Brier Score*

$$\mathrm{BS}(y) = 1/n \sum_{i=1}^{n} \left(F_i(y) - I(y_i \leq y)\right)^2,$$

  where $I(x)$ is the indicator function for the event x, and the Brier score is again evaluated at the unique values of the (cross-)validation observations (or a subset of size N of them),

- the *averaged continuous ranked probability score*, CRPS, a strictly proper scoring criterion to rank predictions, which is related to the Brier score by

$$\mathrm{CRPS} = \int_{-\infty}^{\infty} \mathrm{BS}(y)\, dy.$$

Gneiting et al. (2007) proposed the following plots to validate probabilistic predictions:

- A histogram of the PIT values. For ideal predictions, with observed coverages of prediction intervals matching nominal coverages, the PIT values have an uniform distribution.

- Plots of $\bar{F}_n(y)$ and of the empirical CDF of the data, say $\widehat{G}_n(y)$, and of their difference, $\bar{F}_n(y) - \widehat{G}_n(y)$ vs y. The forecasts are said to be *marginally calibrated* if $\bar{F}_n(y)$ and $\widehat{G}_n(y)$ match.
- A plot of $\mathrm{BS}(y)$ vs. y. Probabilistic predictions are said to be *sharp* if the area under this curve, which equals CRPS, is minimized.

The `plot` method for class `cv.georob` allows to create these plots, along with scatterplots of observations and predictions, Tukey-Anscombe and normal QQ plots of the standardized prediction errors.

`summary.cv.georob` computes the mean and dispersion statistics of the (standardized) prediction errors (by a call to `validate.prediction` with argument `statistic = "st"`, see *Value*) and the averaged continuous ranked probability score (`crps`). If present in the `cv.georob` object, the error statistics are also computed for the errors of the unbiasedly back-transformed predictions of a log-transformed response. If `se` is `TRUE` then these statistics are evaluated separately for the K cross-validation subsets and the standard errors of the means of these statistics are returned as well.

The `print` method for class `cv.georob` returns the mean and dispersion statistics of the (standardized) prediction errors.

The method `rstudent` returns for class `cv.georob` the standardized prediction errors.

Value

Depending on the argument `statistic`, the function `validate.predictions` returns

- a numeric vector of PIT values if `statistic` is equal to `"pit"`,
- a named numeric vector with summary statistics of the (standardized) prediction errors if `statistic` is equal to `"st"`. The following statistics are computed:

`me`	mean prediction error
`mede`	median prediction error
`rmse`	root mean squared prediction error
`made`	median absolute prediction error
`qne`	Qn dispersion measure of prediction errors (see `Qn`)
`msse`	mean squared standardized prediction error
`medsse`	median squared standardized prediction error

- a data frame if `statistic` is equal to `"mc"` or `"bs"` with the components (see *Details*):

`z`	the sorted unique (cross-)validation observations (or a subset of size `ncutoff` of them)
`ghat`	the empirical CDF of the (cross-)validation observations $\widehat{G}_n(y)$
`fbar`	the average predictive distribution $\bar{F}_n(y)$
`bs`	the Brier score $\mathrm{BS}(y)$

The function `rstudent.cv.georob` returns a numeric vector with the standardized cross-validation prediction errors.

Author(s)

Andreas Papritz <andreas.papritz@env.ethz.ch>

References

Gneiting, T., Balabdaoui, F. and Raftery, A. E.(2007) Probabilistic forecasts, calibration and sharpness. *Journal of the Royal Statistical Society Series B* **69**, 243–268.

See Also

georob for (robust) fitting of spatial linear models; cv.georob for assessing the goodness of a fit by georob.

Examples

```
## Not run:
data( meuse )

r.logzn <- georob(log(zinc) ~ sqrt(dist), data = meuse, locations = ~ x + y,
    variogram.model = "RMexp",
    param = c( variance = 0.15, nugget = 0.05, scale = 200 ),
    tuning.psi = 1)

r.logzn.cv.1 <- cv(r.logzn, seed = 1, lgn = TRUE )
r.logzn.cv.2 <- cv(r.logzn, formula = .~. + ffreq, seed = 1, lgn = TRUE )

summary(r.logzn.cv.1, se = TRUE)
summary(r.logzn.cv.2, se = TRUE)

op <- par(mfrow = c(2,2))
plot(r.logzn.cv.1, type = "lgn.sc")
plot(r.logzn.cv.2, type = "lgn.sc", add = TRUE, col = "red")
abline(0, 1, lty= "dotted")
plot(r.logzn.cv.1, type = "ta")
plot(r.logzn.cv.2, type = "ta", add = TRUE, col = "red")
abline(h=0, lty= "dotted")
plot(r.logzn.cv.2, type = "mc", add = TRUE, col = "red")
plot(r.logzn.cv.1, type = "bs")
plot(r.logzn.cv.2, type = "bs", add = TRUE, col = "red")
legend("topright", lty = 1, col = c( "black", "red"), bty = "n",
    legend = c("log(Zn) ~ sqrt(dist)", "log(Zn) ~ sqrt(dist) + ffreq"))
par(op)
## End(Not run)
```

Index