# Package ‘PrevMap’

October 7, 2021

**Type** Package

**Title** Geostatistical Modelling of Spatially Referenced Prevalence Data

**Version** 1.5.4

**Date** 2021-10-06

**Author** Emanuele Giorgi, Peter J. Diggle

**Maintainer** Emanuele Giorgi <e.giorgi@lancaster.ac.uk>

**Imports** splancs, lme4, truncnorm, methods, numDeriv

**Depends** maxLik, raster, pdist, Matrix

**Description** Provides functions for both likelihood-based and Bayesian analysis of spatially referenced prevalence data. For a tutorial on the use of the R package, see Giorgi and Diggle (2017) <doi:10.18637/jss.v078.i08>.

**Encoding** UTF-8

**LazyData** true



**Suggests** geoR, R.rsp, INLA, knitr, rmarkdown

**Additional_repositories** https://inla.r-inla-download.org/R/testing/

**RoxygenNote** 7.1.1

**NeedsCompilation** no

**Repository** CRAN

**Date/Publication** 2021-10-07 14:30:02 UTC

## R topics documented:

`adjust.sigma2` *Adjustment factor for the variance of the convolution of Gaussian noise*

## Description

This function computes the multiplicative constant used to adjust the value of `sigma2` in the low-rank approximation of a Gaussian process.

## Usage

```
adjust.sigma2(knots.dist, phi, kappa)
```

## Arguments

| | |
|---|---|
| `knots.dist` | a matrix of the distances between the observed coordinates and the spatial knots. |
| `phi` | scale parameter of the Matern covariance function. |
| `kappa` | shape parameter of the Matern covariance function. |

## Details

Let $U$ denote the $n$ by $m$ matrix of the distances between the $n$ observed coordinates and $m$ pre-defined spatial knots. This function computes the following quantity

$$\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{m}K(u_{ij};\phi,\kappa)^2,$$

where $K(.;\phi,\kappa)$ is the Matern kernel (see `matern.kernel`) and $u_{ij}$ is the distance between the $i$-th sampled location and the $j$-th spatial knot.

## Value

A value corresponding to the adjustment factor for `sigma2`.

## Author(s)

Emanuele Giorgi <e.giorgi@lancaster.ac.uk>

Peter J. Diggle <p.diggle@lancaster.ac.uk>

## See Also

`matern.kernel`, `pdist`.

autocor.plot *Plot of the autocorrelgram for posterior samples*

## Description

Plots the autocorrelogram for the posterior samples of the model parameters and spatial random effects.

## Usage

```
autocor.plot(object, param, component.beta = NULL, component.S = NULL)
```

## Arguments

| | |
|---|---|
| `object` | an object of class 'Bayes.PrevMap'. |
| `param` | a character indicating for which component of the model the autocorrelation plot is required: `param="beta"` for the regression coefficients; `param="sigma2"` for the variance of the spatial random effect; `param="phi"` for the scale parameter of the Matern correlation function; `param="tau2"` for the variance of the nugget effect; `param="S"` for the spatial random effect. |
| `component.beta` | if `param="beta"`, `component.beta` is a numeric value indicating the component of the regression coefficients; default is `NULL`. |
| `component.S` | if `param="S"`, `component.S` can be a numeric value indicating the component of the spatial random effect, or set equal to `"all"` if the autocorrelgram should be plotted for all the components. Default is `NULL`. |

## Author(s)

Emanuele Giorgi <e.giorgi@lancaster.ac.uk>

Peter J. Diggle <p.diggle@lancaster.ac.uk>

binary.probit.Bayes *Bayesian estimation for the two-levels binary probit model*

## Description

This function performs Bayesian estimation for a geostatistical binary probit model. It also allows to specify a two-levels model so as to include individual-level and household-level (or any other unit comprising a group of individuals, e.g. village, school, compound, etc...) variables.

### Usage

```
binary.probit.Bayes(
  formula,
  coords,
  data,
  ID.coords,
  control.prior,
  control.mcmc,
  kappa,
  low.rank = FALSE,
  knots = NULL,
  messages = TRUE
)
```

### Arguments

| | |
|---|---|
| `formula` | an object of class `formula` (or one that can be coerced to that class): a symbolic description of the model to be fitted. |
| `coords` | an object of class `formula` indicating the geographic coordinates. |
| `data` | a data frame containing the variables in the model. |
| `ID.coords` | vector of ID values for the unique set of spatial coordinates obtained from `create.ID.coords`. These must be provided in order to specify spatial random effects at household-level. **Warning**: the household coordinates must all be distinct otherwise see `jitterDupCoords`. Default is `NULL`. |
| `control.prior` | output from `control.prior`. |
| `control.mcmc` | output from `control.mcmc.Bayes`. |
| `kappa` | value for the shape parameter of the Matern covariance function. |
| `low.rank` | logical; if `low.rank=TRUE` a low-rank approximation is required. Default is `low.rank=FALSE`. |
| `knots` | if `low.rank=TRUE`, `knots` is a matrix of spatial knots used in the low-rank approximation. Default is `knots=NULL`. |
| `messages` | logical; if `messages=TRUE` then status messages are printed on the screen (or output device) while the function is running. Default is `messages=TRUE`. |

### Details

This function performs Bayesian estimation for the parameters of the geostatistical binary probit model. Let $i$ and $j$ denote the indices of the $i$-th household and $j$-th individual within that household. The response variable $Y_{ij}$ is a binary indicator taking value 1 if the individual has been tested positive for the disease of interest and 0 otherwise. Conditionally on a zero-mean stationary Gaussian process $S(x_i)$, $Y_{ij}$ are mutually independent Bernoulli variables with probit link function $\Phi^{-1}(\cdot)$, i.e.

$$\Phi^{-1}(p_{ij}) = d_{ij}'\beta + S(x_i),$$

where $d_{ij}$ is a vector of covariates, both at individual- and household-level, with associated regression coefficients $\beta$. The Gaussian process $S(x)$ has isotropic Matern covariance function (see `matern`) with variance `sigma2`, scale parameter `phi` and shape parameter `kappa`.

**Priors definition.** Priors can be defined through the function `control.prior`. The hierarchical structure of the priors is the following. Let $\theta$ be the vector of the covariance parameters `c(sigma2,phi)`; each component of $\theta$ has independent priors that can be freely defined by the user. However, in `control.prior` uniform and log-normal priors are also available as default priors for each of the covariance parameters. The vector of regression coefficients `beta` has a multivariate Gaussian prior with mean `beta.mean` and covariance matrix `beta.covar`.

**Updating regression coefficents and random effects using auxiliary variables.** To update $\beta$ and $S(x_i)$, we use an auxiliary variable technique based on Rue and Held (2005). Let $V_{ij}$ denote a set of random variables that conditionally on $\beta$ and $S(x_i)$, are mutually independent Gaussian with mean $d_{ij}'\beta + S(x_i)$ and unit variance. Then, $Y_{ij} = 1$ if $V_{ij} > 0$ and $Y_{ij} = 0$ otherwise. Using this representation of the model, we use a Gibbs sampler to simulate from the full conditionals of $\beta$, $S(x_i)$ and $V_{ij}$. See Section 4.3 of Rue and Held (2005) for more details.

**Updating the covariance parameters with a Metropolis-Hastings algorithm.** In the MCMC algorithm implemented in `binary.probit.Bayes`, the transformed parameters

$$(\theta_1, \theta_2) = (\log(\sigma^2)/2, \log(\sigma^2/\phi^{2\kappa}))$$

are independently updated using a Metropolis Hastings algorithm. At the $i$-th iteration, a new value is proposed for each parameter from a univariate Gaussian distrubion with variance $h_i^2$. This is tuned using the following adaptive scheme

$$h_i = h_{i-1} + c_1 i^{-c_2}(\alpha_i - 0.45),$$

where $\alpha_i$ is the acceptance rate at the $i$-th iteration, 0.45 is the optimal acceptance rate for a univariate Gaussian distribution, whilst $c_1 > 0$ and $0 < c_2 < 1$ are pre-defined constants. The starting values $h_1$ for each of the parameters $\theta_1$ and $\theta_2$ can be set using the function `control.mcmc.Bayes` through the arguments `h.theta1`, `h.theta2` and `h.theta3`. To define values for $c_1$ and $c_2$, see the documentation of `control.mcmc.Bayes`.

**Low-rank approximation.** In the case of very large spatial data-sets, a low-rank approximation of the Gaussian spatial process $S(x)$ might be computationally beneficial. Let $(x_1, \dots, x_m)$ and $(t_1, \dots, t_m)$ denote the set of sampling locations and a grid of spatial knots covering the area of interest, respectively. Then $S(x)$ is approximated as $\sum_{i=1}^m K(\|x - t_i\|; \phi, \kappa)U_i$, where $U_i$ are zero-mean mutually independent Gaussian variables with variance `sigma2` and $K(.; \phi, \kappa)$ is the isotropic Matern kernel (see `matern.kernel`). Since the resulting approximation is no longer a stationary process (but only approximately), `sigma2` may take very different values from the actual variance of the Gaussian process to approximate. The function `adjust.sigma2` can then be used to (approximately) explore the range for `sigma2`. For example if the variance of the Gaussian process is `0.5`, then an approximate value for `sigma2` is `0.5/const.sigma2`, where `const.sigma2` is the value obtained with `adjust.sigma2`.

## Value

An object of class "Bayes.PrevMap". The function `summary.Bayes.PrevMap` is used to print a summary of the fitted model. The object is a list with the following components:

`estimate`: matrix of the posterior samples of the model parameters.

`S`: matrix of the posterior samples for each component of the random effect.

`const.sigma2`: vector of the values of the multiplicative factor used to adjust the values of `sigma2` in the low-rank approximation.

y: binary observations.

D: matrix of covariarates.

coords: matrix of the observed sampling locations.

kappa: shape parameter of the Matern function.

ID.coords: set of ID values defined through the argument ID.coords.

knots: matrix of spatial knots used in the low-rank approximation.

h1: vector of values taken by the tuning parameter h.theta1 at each iteration.

h2: vector of values taken by the tuning parameter h.theta2 at each iteration.

call: the matched call.

## Author(s)

Emanuele Giorgi <e.giorgi@lancaster.ac.uk>

Peter J. Diggle <p.diggle@lancaster.ac.uk>

## References

Diggle, P.J., Giorgi, E. (2019). *Model-based Geostatistics for Global Public Health.* CRC/Chapman & Hall.

Giorgi, E., Diggle, P.J. (2017). *PrevMap: an R package for prevalence mapping.* Journal of Statistical Software. 78(8), 1-29. doi: 10.18637/jss.v078.i08

Rue, H., Held, L. (2005). *Gaussian Markov Random Fields: Theory and Applications.* Chapman & Hall, London.

Higdon, D. (1998). *A process-convolution approach to modeling temperatures in the North Atlantic Ocean.* Environmental and Ecological Statistics 5, 173-190.

## See Also

control.mcmc.Bayes, control.prior,summary.Bayes.PrevMap, matern, matern.kernel, create.ID.coords.

---

binomial.logistic.Bayes

*Bayesian estimation for the binomial logistic model*

---

## Description

This function performs Bayesian estimation for a geostatistical binomial logistic model.

### Usage

```
binomial.logistic.Bayes(
  formula,
  units.m,
  coords,
  data,
  ID.coords = NULL,
  control.prior,
  control.mcmc,
  kappa,
  low.rank = FALSE,
  knots = NULL,
  messages = TRUE,
  mesh = NULL,
  SPDE = FALSE
)
```

### Arguments

| | |
|---|---|
| formula | an object of class formula (or one that can be coerced to that class): a symbolic description of the model to be fitted. |
| units.m | an object of class formula indicating the binomial denominators. |
| coords | an object of class formula indicating the geographic coordinates. |
| data | a data frame containing the variables in the model. |
| ID.coords | vector of ID values for the unique set of spatial coordinates obtained from create.ID.coords. These must be provided if, for example, spatial random effects are defined at household level but some of the covariates are at individual level. **Warning**: the household coordinates must all be distinct otherwise see jitterDupCoords. Default is NULL. |
| control.prior | output from control.prior. |
| control.mcmc | output from control.mcmc.Bayes. |
| kappa | value for the shape parameter of the Matern covariance function. |
| low.rank | logical; if low.rank=TRUE a low-rank approximation is required. Default is low.rank=FALSE. |
| knots | if low.rank=TRUE, knots is a matrix of spatial knots used in the low-rank approximation. Default is knots=NULL. |
| messages | logical; if messages=TRUE then status messages are printed on the screen (or output device) while the function is running. Default is messages=TRUE. |
| mesh | an object obtained as result of a call to the function inla.mesh.2d. |
| SPDE | logical; if SPDE=TRUE the SPDE approximation for the Gaussian spatial model is used. Default is SPDE=FALSE. |

**Details**

This function performs Bayesian estimation for the parameters of the geostatistical binomial logistic model. Conditionally on a zero-mean stationary Gaussian process $S(x)$ and mutually independent zero-mean Gaussian variables $Z$ with variance `tau2`, the linear predictor assumes the form

$$\log(p/(1-p)) = d'\beta + S(x) + Z,$$

where $d$ is a vector of covariates with associated regression coefficients $\beta$. The Gaussian process $S(x)$ has isotropic Matern covariance function (see `matern`) with variance `sigma2`, scale parameter `phi` and shape parameter `kappa`.

**Priors definition.** Priors can be defined through the function `control.prior`. The hierarchical structure of the priors is the following. Let $\theta$ be the vector of the covariance parameters `c(sigma2,phi,tau2)`; then each component of $\theta$ has independent priors freely defined by the user. However, in `control.prior` uniform and log-normal priors are also available as default priors for each of the covariance parameters. To remove the nugget effect $Z$, no prior should be defined for `tau2`. Conditionally on `sigma2`, the vector of regression coefficients `beta` has a multivariate Gaussian prior with mean `beta.mean` and covariance matrix `sigma2*beta.covar`, while in the low-rank approximation the covariance matrix is simply `beta.covar`.

**Updating the covariance parameters with a Metropolis-Hastings algorithm.** In the MCMC algorithm implemented in `binomial.logistic.Bayes`, the transformed parameters

$$(\theta_1, \theta_2, \theta_3) = (\log(\sigma^2)/2, \log(\sigma^2/\phi^{2\kappa}), \log(\tau^2))$$

are independently updated using a Metropolis Hastings algorithm. At the $i$-th iteration, a new value is proposed for each from a univariate Gaussian distrubion with variance $h_i^2$ that is tuned using the following adaptive scheme

$$h_i = h_{i-1} + c_1 i^{-c_2}(\alpha_i - 0.45),$$

where $\alpha_i$ is the acceptance rate at the $i$-th iteration, 0.45 is the optimal acceptance rate for a univariate Gaussian distribution, whilst $c_1 > 0$ and $0 < c_2 < 1$ are pre-defined constants. The starting values $h_1$ for each of the parameters $\theta_1$, $\theta_2$ and $\theta_3$ can be set using the function `control.mcmc.Bayes` through the arguments `h.theta1`, `h.theta2` and `h.theta3`. To define values for $c_1$ and $c_2$, see the documentation of `control.mcmc.Bayes`.

**Hamiltonian Monte Carlo.** The MCMC algorithm in `binomial.logistic.Bayes` uses a Hamiltonian Monte Carlo (HMC) procedure to update the random effect $T = d'\beta + S(x) + Z$; see Neal (2011) for an introduction to HMC. HMC makes use of a postion vector, say $t$, representing the random effect $T$, and a momentum vector, say $q$, of the same length of the position vector, say $n$. Hamiltonian dynamics also have a physical interpretation where the states of the system are described by the position of a puck and its momentum (its mass times its velocity). The Hamiltonian function is then defined as a function of $t$ and $q$, having the form $H(t, q) = -\log\{f(t|y, \beta, \theta)\} + q'q/2$, where $f(t|y, \beta, \theta)$ is the conditional distribution of $T$ given the data $y$, the regression parameters $\beta$ and covariance parameters $\theta$. The system of Hamiltonian equations then defines the evolution of the system in time, which can be used to implement an algorithm for simulation from the posterior distribution of $T$. In order to implement the Hamiltonian dynamic on a computer, the Hamiltonian equations must be discretised. The *leapfrog method* is then used for this purpose, where two tuning parameters should be defined: the stepsize $\epsilon$ and the number of steps $L$. These respectively correspond to `epsilon.S.lim` and `L.S.lim` in the `control.mcmc.Bayes` function. However, it is advisable to let $epsilon$ and $L$ take different random values at each iteration of the HCM algorithm so as to account for the different variances amongst the components of the posterior of $T$. This can be done in control.mcmc.Bayes by defning epsilon.S.lim and L.S.lim as vectors of two elements, each of which represents the lower and upper limit of a uniform distribution used to generate values for epsilon.S.lim and L.S.lim, respectively.

**Using a two-level model to include household-level and individual-level information.** When analysing data from household sruveys, some of the avilable information information might be at household-level (e.g. material of house, temperature) and some at individual-level (e.g. age, gender). In this case, the Gaussian spatial process $S(x)$ and the nugget effect $Z$ are defined at hosuehold-level in order to account for extra-binomial variation between and within households, respectively.

**Low-rank approximation.** In the case of very large spatial data-sets, a low-rank approximation of the Gaussian spatial process $S(x)$ might be computationally beneficial. Let $(x_1, \ldots, x_m)$ and $(t_1, \ldots, t_m)$ denote the set of sampling locations and a grid of spatial knots covering the area of interest, respectively. Then $S(x)$ is approximated as $\sum_{i=1}^m K(\|x - t_i\|; \phi, \kappa) U_i$, where $U_i$ are zero-mean mutually independent Gaussian variables with variance sigma2 and $K(.; \phi, \kappa)$ is the isotropic Matern kernel (see matern.kernel). Since the resulting approximation is no longer a stationary process (but only approximately), sigma2 may take very different values from the actual variance of the Gaussian process to approximate. The function adjust.sigma2 can then be used to (approximately) explore the range for sigma2. For example if the variance of the Gaussian process is 0.5, then an approximate value for sigma2 is 0.5/const.sigma2, where const.sigma2 is the value obtained with adjust.sigma2.

## Value

An object of class "Bayes.PrevMap". The function summary.Bayes.PrevMap is used to print a summary of the fitted model. The object is a list with the following components:

estimate: matrix of the posterior samples of the model parameters.

S: matrix of the posterior samples for each component of the random effect.

const.sigma2: vector of the values of the multiplicative factor used to adjust the values of sigma2 in the low-rank approximation.

y: binomial observations.

units.m: binomial denominators.

D: matrix of covariarates.

coords: matrix of the observed sampling locations.

kappa: shape parameter of the Matern function.

ID.coords: set of ID values defined through the argument ID.coords.

knots: matrix of spatial knots used in the low-rank approximation.

h1: vector of values taken by the tuning parameter h.theta1 at each iteration.

h2: vector of values taken by the tuning parameter h.theta2 at each iteration.

h3: vector of values taken by the tuning parameter h.theta3 at each iteration.

acc.beta.S: empirical acceptance rate for the regression coefficients and random effects (only if SPDE=TRUE).

mesh: the mesh used in the SPDE approximation.

call: the matched call.

## Author(s)

Emanuele Giorgi <e.giorgi@lancaster.ac.uk>

Peter J. Diggle <p.diggle@lancaster.ac.uk>

## References

Diggle, P.J., Giorgi, E. (2019). *Model-based Geostatistics for Global Public Health.* CRC/Chapman & Hall.

Giorgi, E., Diggle, P.J. (2017). *PrevMap: an R package for prevalence mapping.* Journal of Statistical Software. 78(8), 1-29. doi: 10.18637/jss.v078.i08

Neal, R. M. (2011) *MCMC using Hamiltonian Dynamics*, In: Handbook of Markov Chain Monte Carlo (Chapter 5), Edited by Steve Brooks, Andrew Gelman, Galin Jones, and Xiao-Li Meng Chapman & Hall / CRC Press.

Higdon, D. (1998). *A process-convolution approach to modeling temperatures in the North Atlantic Ocean.* Environmental and Ecological Statistics 5, 173-190.

## See Also

`control.mcmc.Bayes`, `control.prior`,`summary.Bayes.PrevMap`, `matern`, `matern.kernel`, `create.ID.coords`.

---

`binomial.logistic.MCML` *Monte Carlo Maximum Likelihood estimation for the binomial logistic model*

---

## Description

This function performs Monte Carlo maximum likelihood (MCML) estimation for the geostatistical binomial logistic model.

## Usage

```
binomial.logistic.MCML(
  formula,
  units.m,
  coords,
  times = NULL,
  data,
  ID.coords = NULL,
  par0,
  control.mcmc,
  kappa,
  kappa.t = NULL,
  sst.model = NULL,
  fixed.rel.nugget = NULL,
  start.cov.pars,
```

```
  method = "BFGS",
  low.rank = FALSE,
  SPDE = FALSE,
  knots = NULL,
  mesh = NULL,
  messages = TRUE,
  plot.correlogram = TRUE
)
```

## Arguments

formula
: an object of class formula (or one that can be coerced to that class): a symbolic description of the model to be fitted.

units.m
: an object of class formula indicating the binomial denominators in the data.

coords
: an object of class formula indicating the spatial coordinates in the data.

times
: an object of class formula indicating the times in the data, used in the spatio-temporal model.

data
: a data frame containing the variables in the model.

ID.coords
: vector of ID values for the unique set of spatial coordinates obtained from create.ID.coords. These must be provided if, for example, spatial random effects are defined at household level but some of the covariates are at individual level. **Warning**: the household coordinates must all be distinct otherwise see jitterDupCoords. Default is NULL.

par0
: parameters of the importance sampling distribution: these should be given in the following order c(beta,sigma2,phi,tau2), where beta are the regression coefficients, sigma2 is the variance of the Gaussian process, phi is the scale parameter of the spatial correlation and tau2 is the variance of the nugget effect (if included in the model).

control.mcmc
: output from control.mcmc.MCML.

kappa
: fixed value for the shape parameter of the Matern covariance function.

kappa.t
: fixed value for the shape parameter of the Matern covariance function in the separable double-Matern spatio-temporal model.

sst.model
: a character value that specifies the spatio-temporal correlation function.

  - sst.model="DM" separable double-Matern.
  - sst.model="GN1" separable correlation functions. Temporal correlation: $f(x) = 1/(1 + x/\psi)$; Spatial correaltion: Matern function.

  Deafault is sst.model=NULL, which is used when a purely spatial model is fitted.

fixed.rel.nugget
: fixed value for the relative variance of the nugget effect; fixed.rel.nugget=NULL if this should be included in the estimation. Default is fixed.rel.nugget=NULL.

start.cov.pars
: a vector of length two with elements corresponding to the starting values of phi and the relative variance of the nugget effect nu2, respectively, that are used in the optimization algorithm. If nu2 is fixed through fixed.rel.nugget, then start.cov.pars represents the starting value for phi only.

- `method` — method of optimization. If `method="BFGS"` then the `maxBFGS` function is used; otherwise `method="nlminb"` to use the `nlminb` function. Default is `method="BFGS"`.
- `low.rank` — logical; if `low.rank=TRUE` a low-rank approximation of the Gaussian spatial process is used when fitting the model. Default is `low.rank=FALSE`.
- `SPDE` — logical; if `SPDE=TRUE` the SPDE approximation for the Gaussian spatial model is used. Default is `SPDE=FALSE`.
- `knots` — if `low.rank=TRUE`, `knots` is a matrix of spatial knots that are used in the low-rank approximation. Default is `knots=NULL`.
- `mesh` — an object obtained as result of a call to the function `inla.mesh.2d`.
- `messages` — logical; if `messages=TRUE` then status messages are printed on the screen (or output device) while the function is running. Default is `messages=TRUE`.
- `plot.correlogram` — logical; if `plot.correlogram=TRUE` the autocorrelation plot of the samples of the random effect is displayed after completion of conditional simulation. Default is `plot.correlogram=TRUE`.

## Details

This function performs parameter estimation for a geostatistical binomial logistic model. Conditionally on a zero-mean stationary Gaussian process $S(x)$ and mutually independent zero-mean Gaussian variables $Z$ with variance `tau2`, the observations `y` are generated from a binomial distribution with probability $p$ and binomial denominators `units.m`. A canonical logistic link is used, thus the linear predictor assumes the form

$$\log(p/(1-p)) = d'\beta + S(x) + Z,$$

where $d$ is a vector of covariates with associated regression coefficients $\beta$. The Gaussian process $S(x)$ has isotropic Matern covariance function (see `matern`) with variance `sigma2`, scale parameter `phi` and shape parameter `kappa`. In the `binomial.logistic.MCML` function, the shape parameter is treated as fixed. The relative variance of the nugget effect, `nu2=tau2/sigma2`, can also be fixed through the argument `fixed.rel.nugget`; if `fixed.rel.nugget=NULL`, then the relative variance of the nugget effect is also included in the estimation.

**Monte Carlo Maximum likelihood.** The Monte Carlo maximum likelihood method uses conditional simulation from the distribution of the random effect $T(x) = d(x)'\beta + S(x) + Z$ given the data `y`, in order to approximate the high-dimensiional intractable integral given by the likelihood function. The resulting approximation of the likelihood is then maximized by a numerical optimization algorithm which uses analytic epression for computation of the gradient vector and Hessian matrix. The functions used for numerical optimization are `maxBFGS` (`method="BFGS"`), from the **maxLik** package, and `nlminb` (`method="nlminb"`).

**Using a two-level model to include household-level and individual-level information.** When analysing data from household sruveys, some of the avilable information information might be at household-level (e.g. material of house, temperature) and some at individual-level (e.g. age, gender). In this case, the Gaussian spatial process $S(x)$ and the nugget effect $Z$ are defined at hosuehold-level in order to account for extra-binomial variation between and within households, respectively.

**Low-rank approximation.** In the case of very large spatial data-sets, a low-rank approximation of the Gaussian spatial process $S(x)$ might be computationally beneficial. Let $(x_1, \dots, x_m)$ and

$(t_1, \ldots, t_m)$ denote the set of sampling locations and a grid of spatial knots covering the area of interest, respectively. Then $S(x)$ is approximated as $\sum_{i=1}^m K(\|x - t_i\|; \phi, \kappa) U_i$, where $U_i$ are zero-mean mutually independent Gaussian variables with variance sigma2 and $K(.; \phi, \kappa)$ is the isotropic Matern kernel (see matern.kernel). Since the resulting approximation is no longer a stationary process (but only approximately), the parameter sigma2 is then multiplied by a factor constant.sigma2 so as to obtain a value that is closer to the actual variance of $S(x)$.

## Value

An object of class "PrevMap". The function summary.PrevMap is used to print a summary of the fitted model. The object is a list with the following components:

estimate: estimates of the model parameters; use the function coef.PrevMap to obtain estimates of covariance parameters on the original scale.

covariance: covariance matrix of the MCML estimates.

log.lik: maximum value of the log-likelihood.

y: binomial observations.

units.m: binomial denominators.

D: matrix of covariates.

coords: matrix of the observed sampling locations.

method: method of optimization used.

ID.coords: set of ID values defined through the argument ID.coords.

kappa: fixed value of the shape parameter of the Matern function.

kappa.t: fixed value for the shape parameter of the Matern covariance function in the separable double-Matern spatio-temporal model.

knots: matrix of the spatial knots used in the low-rank approximation.

mesh: the mesh used in the SPDE approximation.

const.sigma2: adjustment factor for sigma2 in the low-rank approximation.

h: vector of the values of the tuning parameter at each iteration of the Langevin-Hastings MCMC algorithm; see Laplace.sampling, or Laplace.sampling.lr if a low-rank approximation is used.

samples: matrix of the random effects samples from the importance sampling distribution used to approximate the likelihood function.

fixed.rel.nugget: fixed value for the relative variance of the nugget effect.

call: the matched call.

## Author(s)

Emanuele Giorgi <e.giorgi@lancaster.ac.uk>

Peter J. Diggle <p.diggle@lancaster.ac.uk>

### References

Diggle, P.J., Giorgi, E. (2019). *Model-based Geostatistics for Global Public Health.* CRC/Chapman & Hall.

Giorgi, E., Diggle, P.J. (2017). *PrevMap: an R package for prevalence mapping.* Journal of Statistical Software. 78(8), 1-29. doi: 10.18637/jss.v078.i08

Christensen, O. F. (2004). *Monte carlo maximum likelihood in model-based geostatistics.* Journal of Computational and Graphical Statistics 13, 702-718.

Higdon, D. (1998). *A process-convolution approach to modeling temperatures in the North Atlantic Ocean.* Environmental and Ecological Statistics 5, 173-190.

### See Also

`Laplace.sampling`, `Laplace.sampling.lr`, `summary.PrevMap`, `coef.PrevMap`, `matern`, `matern.kernel`, `control.mcmc.MCML`, `create.ID.coords`.

---

## `coef.PrevMap` *Extract model coefficients*

---

### Description

`coef` extracts parameters estimates from models fitted with the functions `linear.model.MLE` and `binomial.logistic.MCML`.

### Usage

```
## S3 method for class 'PrevMap'
coef(object, ...)
```

### Arguments

| | |
|---|---|
| `object` | an object of class "PrevMap". |
| `...` | other arguments. |

### Value

coefficients extracted from the model object `object`.

### Author(s)

Emanuele Giorgi <e.giorgi@lancaster.ac.uk>

Peter J. Diggle <p.diggle@lancaster.ac.uk>

coef.PrevMap.ps — *Extract model coefficients from geostatistical linear model with preferentially sampled locations*

### Description

coef extracts parameters estimates from models fitted with the functions lm.ps.MCML.

### Usage

```
## S3 method for class 'PrevMap.ps'
coef(object, ...)
```

### Arguments

| | |
|---|---|
| object | an object of class "PrevMap.ps". |
| ... | other arguments. |

### Value

a list of coefficients extracted from the model in object.

### Author(s)

Emanuele Giorgi <e.giorgi@lancaster.ac.uk>

continuous.sample — *Spatially continuous sampling*

### Description

Draws a sample of spatial locations within a spatially continuous polygonal sampling region.

### Usage

```
continuous.sample(poly, n, delta, k = 0, rho = NULL)
```

### Arguments

| | |
|---|---|
| poly | boundary of a polygon. |
| n | number of events. |
| delta | minimum permissible distance between any two events in preliminary sample. |
| k | number of locations in preliminary sample to be replaced by near neighbours of other preliminary sample locations in final sample (must be between 0 and n/2) |
| rho | maximum distance between close pairs of locations in final sample. |

### Details

To draw a sample of size n from a spatially continuous region $A$, with the property that the distance between any two sampled locations is at least `delta`, the following algorithm is used.

- Step 1. Set $i = 1$ and generate a point $x_1$ uniformly distributed on $A$.
- Step 2. Increase $i$ by 1, generate a point $x_i$ uniformly distributed on $A$ and calculate the minimum, $d_{\min}$, of the distances from $x_i$ to all $x_j : j < i$.
- Step 3. If $d_{\min} \geq \delta$, increase $i$ by 1 and return to step 2 if $i \leq n$, otherwise stop;
- Step 4. If $d_{\min} < \delta$, return to step 2 without increasing $i$.

**Sampling close pairs of points.** For some purposes, it is desirable that a spatial sampling scheme include pairs of closely spaced points. In this case, the above algorithm requires the following additional steps to be taken. Let k be the required number of close pairs. Choose a value `rho` such that a close pair of points will be a pair of points separated by a distance of at most `rho`.

- Step 5. Set $j = 1$ and draw a random sample of size 2 from the integers $1, 2, \ldots, n$, say $(i_1; i_2)$;
- Step 6. Replace $x_{i_1}$ by $x_{i_2} + u$ , where $u$ is uniformly distributed on the disc with centre $x_{i_2}$ and radius `rho`, increase $i$ by 1 and return to step 5 if $i \leq k$, otherwise stop.

### Value

A matrix of dimension n by 2 containing event locations.

### Author(s)

Emanuele Giorgi <`e.giorgi@lancaster.ac.uk`>

Peter J. Diggle <`p.diggle@lancaster.ac.uk`>

---

## `contour.pred.PrevMap` *Contour plot of a predicted surface*

---

### Description

`plot.pred.PrevMap` displays contours of predictions obtained from `spatial.pred.linear.MLE`, `spatial.pred.linear.Bayes`,`spatial.pred.binomial.MCML` and `spatial.pred.binomial.Bayes`.

### Usage

```
## S3 method for class 'pred.PrevMap'
contour(x, type = NULL, summary = "predictions", ...)
```

### Arguments

| | |
|---|---|
| `x` | an object of class "pred.PrevMap". |
| `type` | a character indicating the type of prediction to display: 'prevalence', 'odds', 'logit' or 'probit'. |
| `summary` | character indicating which summary to display: 'predictions','quantiles', 'standard.errors' or 'exceedance.prob'; default is 'predictions'. If `summary="exceedance.prob"`, the argument `type` is ignored. |
| `...` | further arguments passed to `contour`. |

### Author(s)

Emanuele Giorgi <`e.giorgi@lancaster.ac.uk`>

Peter J. Diggle <`p.diggle@lancaster.ac.uk`>

---

## `control.mcmc.Bayes` *Control settings for the MCMC algorithm used for Bayesian inference*

---

### Description

This function defines the different tuning parameter that are used in the MCMC algorithm for Bayesian inference.

### Usage

```
control.mcmc.Bayes(
  n.sim,
  burnin,
  thin,
  h.theta1 = 0.01,
  h.theta2 = 0.01,
  h.theta3 = 0.01,
  L.S.lim = NULL,
  epsilon.S.lim = NULL,
  start.beta = "prior mean",
  start.sigma2 = "prior mean",
  start.phi = "prior mean",
  start.S = "prior mean",
  start.nugget = "prior mean",
  c1.h.theta1 = 0.01,
  c2.h.theta1 = 1e-04,
  c1.h.theta2 = 0.01,
  c2.h.theta2 = 1e-04,
  c1.h.theta3 = 0.01,
  c2.h.theta3 = 1e-04,
  linear.model = FALSE,
  binary = FALSE
)
```

**Arguments**

| | |
|---|---|
| n.sim | total number of simulations. |
| burnin | initial number of samples to be discarded. |
| thin | value used to retain only evey thin-th sampled value. |
| h.theta1 | starting value of the tuning parameter of the proposal distribution for $\theta_1 = \log(\sigma^2)/2$. See 'Details' in binomial.logistic.Bayes or linear.model.Bayes. |
| h.theta2 | starting value of the tuning parameter of the proposal distribution for $\theta_2 = \log(\sigma^2/\phi^{2\kappa})$. See 'Details' in binomial.logistic.Bayes or linear.model.Bayes. |
| h.theta3 | starting value of the tuning parameter of the proposal distribution for $\theta_3 = \log(\tau^2)$. See 'Details' in binomial.logistic.Bayes or linear.model.Bayes. |
| L.S.lim | an atomic value or a vector of length 2 that is used to define the number of steps used at each iteration in the Hamiltonian Monte Carlo algorithm to update the spatial random effect; if a single value is provided than the number of steps is kept fixed, otherwise if a vector of length 2 is provided the number of steps is simulated at each iteration as floor(runif(1,L.S.lim[1],L.S.lim[2]+1)). |
| epsilon.S.lim | an atomic value or a vector of length 2 that is used to define the stepsize used at each iteration in the Hamiltonian Monte Carlo algorithm to update the spatial random effect; if a single value is provided than the stepsize is kept fixed, otherwise if a vector of length 2 is provided the stepsize is simulated at each iteration as runif(1,epsilon.S.lim[1],epsilon.S.lim[2]). |
| start.beta | starting value for the regression coefficients beta. |
| start.sigma2 | starting value for sigma2. |
| start.phi | starting value for phi. |
| start.S | starting value for the spatial random effect. |
| start.nugget | starting value for the variance of the nugget effect; default is NULL if the nugget effect is not present. |
| c1.h.theta1 | value of $c_1$ used to adaptively tune the variance of the Gaussian proposal for the transformed parameter log(sigma2)/2; see 'Details' in binomial.logistic.Bayes or linear.model.Bayes. |
| c2.h.theta1 | value of $c_2$ used to adaptively tune the variance of the Gaussian proposal for the transformed parameter log(sigma2)/2; see 'Details' in binomial.logistic.Bayes or linear.model.Bayes. |
| c1.h.theta2 | value of $c_1$ used to adaptively tune the variance of the Gaussian proposal for the transformed parameter log(sigma2.curr/(phi.curr^(2*kappa))); see 'Details' in binomial.logistic.Bayes or linear.model.Bayes. |
| c2.h.theta2 | value of $c_2$ used to adaptively tune the variance of the Gaussian proposal for the transformed parameter log(sigma2.curr/(phi.curr^(2*kappa))); see 'Details' in binomial.logistic.Bayes or linear.model.Bayes. |
| c1.h.theta3 | value of $c_1$ used to adaptively tune the variance of the Gaussian proposal for the transformed parameter log(tau2); see 'Details' in binomial.logistic.Bayes or linear.model.Bayes. |

`c2.h.theta3` value of $c_2$ used to adaptively tune the variance of the Gaussian proposal for the transformed parameter `log(tau2)`; see 'Details' in `binomial.logistic.Bayes` or `linear.model.Bayes`.

`linear.model` logical; if `linear.model=TRUE`, the control parameters are set for the geostatistical linear model. Default is `linear.model=FALSE`.

`binary` logical; if `binary=TRUE`, the control parameters are set the binary geostatistical model. Default is `binary=FALSE`.

## Value

an object of class "mcmc.Bayes.PrevMap".

## Author(s)

Emanuele Giorgi <e.giorgi@lancaster.ac.uk>

Peter J. Diggle <p.diggle@lancaster.ac.uk>

---

# control.mcmc.Bayes.SPDE

*Control settings for the MCMC algorithm used for Bayesian inference using SPDE*

---

## Description

This function defines the different tuning parameter that are used in the MCMC algorithm for Bayesian inference using a SPDE approximation for the spatial Gaussian process.

## Usage

```
control.mcmc.Bayes.SPDE(
  n.sim,
  burnin,
  thin,
  h.theta1 = 0.01,
  h.theta2 = 0.01,
  start.beta = "prior mean",
  start.sigma2 = "prior mean",
  start.phi = "prior mean",
  start.S = "prior mean",
  n.iter = 1,
  h = 1,
  c1.h.theta1 = 0.01,
  c2.h.theta1 = 1e-04,
  c1.h.theta2 = 0.01,
  c2.h.theta2 = 1e-04
)
```

### Arguments

| | |
|---|---|
| `n.sim` | total number of simulations. |
| `burnin` | initial number of samples to be discarded. |
| `thin` | value used to retain only evey `thin`-th sampled value. |
| `h.theta1` | starting value of the tuning parameter of the proposal distribution for $\theta_1 = \log(\sigma^2)/2$. See 'Details' in `binomial.logistic.Bayes` or `linear.model.Bayes`. |
| `h.theta2` | starting value of the tuning parameter of the proposal distribution for $\theta_2 = \log(\sigma^2/\phi^{2\kappa})$. See 'Details' in `binomial.logistic.Bayes` or `linear.model.Bayes`. |
| `start.beta` | starting value for the regression coefficients `beta`. If not provided the prior mean is used. |
| `start.sigma2` | starting value for `sigma2`. If not provided the prior mean is used. |
| `start.phi` | starting value for `phi`. If not provided the prior mean is used. |
| `start.S` | starting value for the spatial random effect. If not provided the prior mean is used. |
| `n.iter` | number of iteration of the Newton-Raphson procedure used to compute the mean and coviariance matrix of the Gaussian proposal in the MCMC; defaut is `n.iter=1`. |
| `h` | tuning parameter for the covariance matrix of the Gaussian proposal. Default is `h=1`. |
| `c1.h.theta1` | value of $c_1$ used to adaptively tune the variance of the Gaussian proposal for the transformed parameter `log(sigma2)/2`; see 'Details' in `binomial.logistic.Bayes` or `linear.model.Bayes`. |
| `c2.h.theta1` | value of $c_2$ used to adaptively tune the variance of the Gaussian proposal for the transformed parameter `log(sigma2)/2`; see 'Details' in `binomial.logistic.Bayes` or `linear.model.Bayes`. |
| `c1.h.theta2` | value of $c_1$ used to adaptively tune the variance of the Gaussian proposal for the transformed parameter `log(sigma2.curr/(phi.curr^(2*kappa)))`; see 'Details' in `binomial.logistic.Bayes` or `linear.model.Bayes`. |
| `c2.h.theta2` | value of $c_2$ used to adaptively tune the variance of the Gaussian proposal for the transformed parameter `log(sigma2.curr/(phi.curr^(2*kappa)))`; see 'Details' in `binomial.logistic.Bayes` or `linear.model.Bayes`. |

### Value

an object of class "mcmc.Bayes.PrevMap".

### Author(s)

Emanuele Giorgi <e.giorgi@lancaster.ac.uk>

Peter J. Diggle <p.diggle@lancaster.ac.uk>

control.mcmc.MCML *Control settings for the MCMC algorithm used for classical inference on a binomial logistic model*

## Description

This function defines the options for the MCMC algorithm used in the Monte Carlo maximum likelihood method.

## Usage

```
control.mcmc.MCML(n.sim, burnin, thin = 1, h = NULL, c1.h = 0.01, c2.h = 1e-04)
```

## Arguments

| | |
|---|---|
| n.sim | number of simulations. |
| burnin | length of the burn-in period. |
| thin | only every thin iterations, a sample is stored; default is thin=1. |
| h | tuning parameter of the proposal distribution used in the Langevin-Hastings MCMC algorithm (see Laplace.sampling and Laplace.sampling.lr); default is h=NULL and then set internally as $1.65/n^(1/6)$, where $n$ is the dimension of the random effect. |
| c1.h | value of $c_1$ used in the adaptive scheme for h; default is c1.h=0.01. See also 'Details' in binomial.logistic.MCML |
| c2.h | value of $c_2$ used in the adaptive scheme for h; default is c1.h=0.01. See also 'Details' in binomial.logistic.MCML |

## Value

A list with processed arguments to be passed to the main function.

## Author(s)

Emanuele Giorgi <e.giorgi@lancaster.ac.uk>

Peter J. Diggle <p.diggle@lancaster.ac.uk>

## Examples

```
control.mcmc <- control.mcmc.MCML(n.sim=1000,burnin=100,thin=1,h=0.05)
str(control.mcmc)
```

control.prior *Priors specification*

## Description

This function is used to define priors for the model parameters of a Bayesian geostatistical model.

## Usage

```
control.prior(
  beta.mean,
  beta.covar,
  log.prior.sigma2 = NULL,
  log.prior.phi = NULL,
  log.prior.nugget = NULL,
  uniform.sigma2 = NULL,
  log.normal.sigma2 = NULL,
  uniform.phi = NULL,
  log.normal.phi = NULL,
  uniform.nugget = NULL,
  log.normal.nugget = NULL
)
```

## Arguments

`beta.mean` mean vector of the Gaussian prior for the regression coefficients.

`beta.covar` covariance matrix of the Gaussian prior for the regression coefficients.

`log.prior.sigma2` a function corresponding to the log-density of the prior distribution for the variance `sigma2` of the Gaussian process. **Warning:** if a low-rank approximation is used, then `sigma2` corresponds to the variance of the iid zero-mean Gaussian variables. Default is `NULL`.

`log.prior.phi` a function corresponding to the log-density of the prior distribution for the scale parameter of the Matern correlation function; default is `NULL`.

`log.prior.nugget` optional: a function corresponding to the log-density of the prior distribution for the variance of the nugget effect; default is `NULL` with no nugget incorporated in the model; default is `NULL`.

`uniform.sigma2` a vector of length two, corresponding to the lower and upper limit of the uniform prior on `sigma2`. Default is `NULL`.

`log.normal.sigma2` a vector of length two, corresponding to the mean and standard deviation of the distribution on the log scale for the log-normal prior on `sigma2`. Default is `NULL`.

`uniform.phi` a vector of length two, corresponding to the lower and upper limit of the uniform prior on `phi`. Default is `NULL`.

log.normal.phi — a vector of length two, corresponding to the mean and standard deviation of the distribution on the log scale for the log-normal prior on phi. Default is NULL.

uniform.nugget — a vector of length two, corresponding to the lower and upper limit of the uniform prior on tau2. Default is NULL.

log.normal.nugget — a vector of length two, corresponding to the mean and standard deviation of the distribution on the log scale for the log-normal prior on tau2. Default is NULL.

## Value

a list corresponding the prior distributions for each model parameter.

## Author(s)

Emanuele Giorgi <e.giorgi@lancaster.ac.uk>

Peter J. Diggle <p.diggle@lancaster.ac.uk>

## See Also

See "Priors definition" in the Details section of the binomial.logistic.Bayes function.

---

# control.profile — *Auxliary function for controlling profile log-likelihood in the linear Gaussian model*

---

## Description

Auxiliary function used by loglik.linear.model. This function defines whether the profile-loglikelihood should be computed or evaluation of the likelihood is required by keeping the other parameters fixed.

## Usage

```
control.profile(
  phi = NULL,
  rel.nugget = NULL,
  fixed.beta = NULL,
  fixed.sigma2 = NULL,
  fixed.phi = NULL,
  fixed.rel.nugget = NULL
)
```

### Arguments

`phi` a vector of the different values that should be used in the likelihood evalutation for the scale parameter `phi`, or `NULL` if a single value is provided either as first argument in `start.par` (for profile likelihood maximization) or as fixed value in `fixed.phi`; default is `NULL`.

`rel.nugget` a vector of the different values that should be used in the likelihood evalutation for the relative variance of the nugget effect `nu2`, or `NULL` if a single value is provided either in `start.par` (for profile likelihood maximization) or as fixed value in `fixed.nu2`; default is `NULL`.

`fixed.beta` a vector for the fixed values of the regression coefficients `beta`, or `NULL` if profile log-likelihood is to be performed; default is `NULL`.

`fixed.sigma2` value for the fixed variance of the Gaussian process `sigma2`, or `NULL` if profile log-likelihood is to be performed; default is `NULL`.

`fixed.phi` value for the fixed scale parameter `phi` in the Matern function, or `NULL` if profile log-likelihood is to be performed; default is `NULL`.

`fixed.rel.nugget` value for the fixed relative variance of the nugget effect; `fixed.rel.nugget=NULL` if profile log-likelihood is to be performed; default is `NULL`.

### Value

A list with components named as the arguments.

### Author(s)

Emanuele Giorgi <e.giorgi@lancaster.ac.uk>

Peter J. Diggle <p.diggle@lancaster.ac.uk>

### See Also

loglik.linear.model

---

## create.ID.coords *ID spatial coordinates*

---

### Description

Creates ID values for the unique set of coordinates.

### Usage

```
create.ID.coords(data, coords)
```

### Arguments

| | |
|---|---|
| data | a data frame containing the spatial coordinates. |
| coords | an object of class formula indicating the geographic coordinates. |

### Value

a vector of integers indicating the corresponding rows in data for each distinct coordinate obtained with the unique function.

### Author(s)

Emanuele Giorgi <e.giorgi@lancaster.ac.uk>

Peter J. Diggle <p.diggle@lancaster.ac.uk>

### Examples

```
x1 <- runif(5)
x2 <- runif(5)
data <- data.frame(x1=rep(x1,each=3),x2=rep(x2,each=3))
ID.coords <- create.ID.coords(data,coords=~x1+x2)
data[,c("x1","x2")]==unique(data[,c("x1","x2")])[ID.coords,]
```

---

## data_sim *Simulated binomial data-set over the unit square*

---

### Description

This binomial data-set was simulated by generating a zero-mean Gaussian process over a 30 by 30 grid covering the unit square. The parameters used in the simulation are sigma2=1, phi=0.15 and kappa=2. The nugget effect was not included, hence tau2=0. The variables are as follows:

- y binomial observations.
- units.m binomial denominators.
- x1 horizontal coordinates.
- x2 vertical coordinates.
- S simulated values of the Gaussian process.

### Usage

```
data(data_sim)
```

### Format

A data frame with 900 rows and 5 variables dens.plot *Density plot for posterior samples*

## Description

Plots the autocorrelogram for the posterior samples of the model parameters and spatial random effects.

## Usage

```
dens.plot(
  object,
  param,
  component.beta = NULL,
  component.S = NULL,
  hist = TRUE,
  ...
)
```

## Arguments

| | |
|---|---|
| `object` | an object of class 'Bayes.PrevMap'. |
| `param` | a character indicating for which component of the model the density plot is required: `param="beta"` for the regression coefficients; `param="sigma2"` for the variance of the spatial random effect; `param="phi"` for the scale parameter of the Matern correlation function; `param="tau2"` for the variance of the nugget effect; `param="S"` for the spatial random effect. |
| `component.beta` | if `param="beta"`, `component.beta` is a numeric value indicating the component of the regression coefficients; default is `NULL`. |
| `component.S` | if `param="S"`, `component.S` can be a numeric value indicating the component of the spatial random effect. Default is `NULL`. |
| `hist` | logical; if `TRUE` a histrogram is added to density plot. |
| `...` | additional parameters to pass to `density`. |

## Author(s)

Emanuele Giorgi <e.giorgi@lancaster.ac.uk>

Peter J. Diggle <p.diggle@lancaster.ac.uk>

discrete.sample *Spatially discrete sampling*

## Description

Draws a sub-sample from a set of units spatially located irregularly over some defined geographical region by imposing a minimum distance between any two sampled units.

## Usage

```
discrete.sample(xy.all, n, delta, k = 0)
```

## Arguments

| | |
|---|---|
| `xy.all` | set of locations from which the sample will be drawn. |
| `n` | size of required sample. |
| `delta` | minimum distance between any two locations in preliminary sample. |
| `k` | number of locations in preliminary sample to be replaced by nearest neighbours of other preliminary sample locations in final sample (must be between 0 and `n/2`). |

## Details

To draw a sample of size `n` from a population of spatial locations $X_i : i = 1, \ldots, N$, with the property that the distance between any two sampled locations is at least `delta`, the function implements the following algorithm.

- Step 1. Draw an initial sample of size `n` completely at random and call this $x_i : i = 1, \ldots, n$.
- Step 2. Set $i = 1$ and calculate the minimum, $d_{\min}$, of the distances from $x_i$ to all other $x_j$ in the initial sample.
- Step 3. If $d_{\min} \geq \delta$, increase $i$ by 1 and return to step 2 if $i \leq n$, otherwise stop.
- Step 4. If $d_{\min} < \delta$, draw an integer $j$ at random from $1, 2, \ldots, N$, set $x_i = X_j$ and return to step 3.

Samples generated in this way will exhibit a more regular spatial arrangement than would a random sample of the same size. The degree of regularity achievable will be influenced by the spatial arrangement of the population $X_i : i = 1, \ldots, N$, the specified value of `delta` and the sample size `n`. For any given population, if `n` and/or `delta` are too large, a sample of the required size with the distance between any two sampled locations at least `delta` will not be achievable; the suggested solution is then to run the algorithm with a smaller value of `delta`.

**Sampling close pairs of points**. For some purposes, it is desirable that a spatial sampling scheme include pairs of closely spaced points. In this case, the above algorithm requires the following additional steps to be taken. Let `k` be the required number of close pairs.

- Step 5. Set $j = 1$ and draw a random sample of size 2 from the integers $1, 2, \ldots, n$, say $(i_1, i_2)$.

- Step 6. Find the integer $r$ such that the distances from $x_{i_1}$ to $X_r$ is the minimum of all $N-1$ distances from $x_{i_1}$ to the $X_j$.
- Step 7. Replace $x_{i_2}$ by $X_r$, increase $i$ by 1 and return to step 5 if $i \leq k$, otherwise stop.

## Value

A matrix of dimension n by 2 containing the final sampled locations.

## Author(s)

Emanuele Giorgi <e.giorgi@lancaster.ac.uk>

Peter J. Diggle <p.diggle@lancaster.ac.uk>

## Examples

```
x<-0.015+0.03*(1:33)
xall<-rep(x,33)
yall<-c(t(matrix(xall,33,33)))
xy<-cbind(xall,yall)+matrix(-0.0075+0.015*runif(33*33*2),33*33,2)
par(pty="s",mfrow=c(1,2))
plot(xy[,1],xy[,2],pch=19,cex=0.25,xlab="Easting",ylab="Northing",
   cex.lab=1,cex.axis=1,cex.main=1)

set.seed(15892)
# Generate spatially random sample
xy.sample<-xy[sample(1:dim(xy)[1],50,replace=FALSE),]
points(xy.sample[,1],xy.sample[,2],pch=19,col="red")
points(xy[,1],xy[,2],pch=19,cex=0.25)
plot(xy[,1],xy[,2],pch=19,cex=0.25,xlab="Easting",ylab="Northing",
   cex.lab=1,cex.axis=1,cex.main=1)

set.seed(15892)
# Generate spatially regular sample
xy.sample<-discrete.sample(xy,50,0.08)
points(xy.sample[,1],xy.sample[,2],pch=19,col="red")
points(xy[,1],xy[,2],pch=19,cex=0.25)
```

---

# galicia *Heavy metal biomonitoring in Galicia*

---

## Description

This data-set relates to two studies on lead concentration in moss samples, in micrograms per gram dry weight, collected in Galicia, norther Spain. The data are from two surveys, one conducted in October 1997 and on in July 2000. The variables are as follows:

- x x-coordinate of the spatial locations.
- y y-coordinate of the spatial locations.

- lead lead concentration.
- survey year of the survey (either 1997 or 2000).

## Usage

```
data(galicia)
```

## Format

A data frame with 195 rows and 4 variables

## Source

Diggle, P.J., Menezes, R. and Su, T.-L. (2010). Geostatistical analysis under preferential sampling (with Discussion). Applied Statistics, 59, 191-232.

---

`galicia.boundary` *Boundary of Galicia*

---

## Description

This data-set contains the geographical coordinates of the boundary of the Galicia region in northern Spain.

The variables are as follows:

- x x-coordinate of the spatial locations.
- y y-coordinate of the spatial locations.

## Usage

```
data(galicia.boundary)
```

## Format

A data frame with 42315 rows and 2 variables glgm.LA *Maximum Likelihood estimation for generalised linear geostatistical models via the Laplace approximation*

## Description

This function performs the Laplace method for maximum likelihood estimation of a generalised linear geostatistical model.

## Usage

```
glgm.LA(
  formula,
  units.m = NULL,
  coords,
  times = NULL,
  data,
  ID.coords = NULL,
  kappa,
  kappa.t = 0.5,
  fixed.rel.nugget = NULL,
  start.cov.pars,
  method = "nlminb",
  messages = TRUE,
  family,
  return.covariance = TRUE
)
```

## Arguments

| | |
|---|---|
| formula | an object of class formula (or one that can be coerced to that class): a symbolic description of the model to be fitted. |
| units.m | an object of class formula indicating the binomial denominators in the data. |
| coords | an object of class formula indicating the spatial coordinates in the data. |
| times | an object of class formula indicating the times in the data, used in the spatio-temporal model. |
| data | a data frame containing the variables in the model. |
| ID.coords | vector of ID values for the unique set of spatial coordinates obtained from create.ID.coords. These must be provided if, for example, spatial random effects are defined at household level but some of the covariates are at individual level. **Warning**: the household coordinates must all be distinct otherwise see jitterDupCoords. Default is NULL. |
| kappa | fixed value for the shape parameter of the Matern covariance function. |
| kappa.t | fixed value for the shape parameter of the Matern covariance function in the separable double-Matern spatio-temporal model. |

fixed.rel.nugget
: fixed value for the relative variance of the nugget effect; fixed.rel.nugget=NULL if this should be included in the estimation. Default is fixed.rel.nugget=NULL.

start.cov.pars
: a vector of length two with elements corresponding to the starting values of phi and the relative variance of the nugget effect nu2, respectively, that are used in the optimization algorithm. If nu2 is fixed through fixed.rel.nugget, then start.cov.pars represents the starting value for phi only.

method
: method of optimization. If method="BFGS" then the maxBFGS function is used; otherwise method="nlminb" to use the nlminb function. Default is method="BFGS".

messages
: logical; if messages=TRUE then status messages are printed on the screen (or output device) while the function is running. Default is messages=TRUE.

family
: character, indicating the conditional distribution of the outcome. This should be "Gaussian", "Binomial" or "Poisson".

return.covariance
: logical; if return.covariance=TRUE then a numerical estimation of the covariance function for the model parameters is returned. Default is return.covariance=TRUE.

## Details

This function performs parameter estimation for a generealized linear geostatistical model. Conditionally on a zero-mean stationary Gaussian process $S(x)$ and mutually independent zero-mean Gaussian variables $Z$ with variance tau2, the observations y are generated from a GLM with link function $g(.)$ and linear predictor

$$\eta = d'\beta + S(x) + Z,$$

where $d$ is a vector of covariates with associated regression coefficients $\beta$. The Gaussian process $S(x)$ has isotropic Matern covariance function (see matern) with variance sigma2, scale parameter phi and shape parameter kappa. The shape parameter is treated as fixed. The relative variance of the nugget effect, nu2=tau2/sigma2, can also be fixed through the argument fixed.rel.nugget; if fixed.rel.nugget=NULL, then the relative variance of the nugget effect is also included in the estimation.

**Laplace Approximation** The Laplace approximation (LA) method uses a second-order Taylor expansion of the integrand expressing the likelihood function. The resulting approximation of the likelihood is then maximized by a numerical optimization as defined through the argument method.

**Using a two-level model to include household-level and individual-level information.** When analysing data from household sruveys, some of the avilable information information might be at household-level (e.g. material of house, temperature) and some at individual-level (e.g. age, gender). In this case, the Gaussian spatial process $S(x)$ and the nugget effect $Z$ are defined at hosuehold-level in order to account for extra-binomial variation between and within households, respectively.

## Value

An object of class "PrevMap". The function summary.PrevMap is used to print a summary of the fitted model. The object is a list with the following components:

estimate: estimates of the model parameters; use the function coef.PrevMap to obtain estimates of covariance parameters on the original scale.

`covariance`: covariance matrix of the MCML estimates.

`log.lik`: maximum value of the log-likelihood.

`y`: binomial observations.

`units.m`: binomial denominators.

`D`: matrix of covariates.

`coords`: matrix of the observed sampling locations.

`times`: vector of the time points used in a spatio-temporal model.

`method`: method of optimization used.

`ID.coords`: set of ID values defined through the argument `ID.coords`.

`kappa`: fixed value of the shape parameter of the Matern function.

`kappa.t`: fixed value for the shape parameter of the Matern covariance function in the separable double-Matern spatio-temporal model.

`fixed.rel.nugget`: fixed value for the relative variance of the nugget effect.

`call`: the matched call.

## Author(s)

Emanuele Giorgi <`e.giorgi@lancaster.ac.uk`>

Peter J. Diggle <`p.diggle@lancaster.ac.uk`>

## References

Diggle, P.J., Giorgi, E. (2019). *Model-based Geostatistics for Global Public Health.* CRC/Chapman & Hall.

Giorgi, E., Diggle, P.J. (2017). *PrevMap: an R package for prevalence mapping.* Journal of Statistical Software. 78(8), 1-29. doi: 10.18637/jss.v078.i08

Christensen, O. F. (2004). *Monte carlo maximum likelihood in model-based geostatistics.* Journal of Computational and Graphical Statistics 13, 702-718.

Higdon, D. (1998). *A process-convolution approach to modeling temperatures in the North Atlantic Ocean.* Environmental and Ecological Statistics 5, 173-190.

## See Also

`Laplace.sampling`, `Laplace.sampling.lr`, `summary.PrevMap`, `coef.PrevMap`, `matern`, `matern.kernel`, `control.mcmc.MCML`, `create.ID.coords`.

Laplace.sampling *Langevin-Hastings MCMC for conditional simulation*

## Description

This function simulates from the conditional distribution of a Gaussian random effect, given binomial or Poisson observations y.

## Usage

```
Laplace.sampling(
  mu,
  Sigma,
  y,
  units.m,
  control.mcmc,
  ID.coords = NULL,
  messages = TRUE,
  plot.correlogram = TRUE,
  poisson.llik = FALSE
)
```

## Arguments

| | |
|---|---|
| mu | mean vector of the marginal distribution of the random effect. |
| Sigma | covariance matrix of the marginal distribution of the random effect. |
| y | vector of binomial/Poisson observations. |
| units.m | vector of binomial denominators, or offset if the Poisson model is used. |
| control.mcmc | output from control.mcmc.MCML. |
| ID.coords | vector of ID values for the unique set of spatial coordinates obtained from create.ID.coords. These must be provided if, for example, spatial random effects are defined at household level but some of the covariates are at individual level. **Warning**: the household coordinates must all be distinct otherwise see jitterDupCoords. Default is NULL. |
| messages | logical; if messages=TRUE then status messages are printed on the screen (or output device) while the function is running. Default is messages=TRUE. |
| plot.correlogram | logical; if plot.correlogram=TRUE the autocorrelation plot of the conditional simulations is displayed. |
| poisson.llik | logical; if poisson.llik=TRUE a Poisson model is used or, if poisson.llik=FALSE, a binomial model is used. |

## Details

**Binomial model.** Conditionally on the random effect $S$, the data y follow a binomial distribution with probability $p$ and binomial denominators `units.m`. The logistic link function is used for the linear predictor, which assumes the form

$$\log(p/(1-p)) = S.$$

**Poisson model.** Conditionally on the random effect $S$, the data y follow a Poisson distribution with mean $m\lambda$, where $m$ is an offset set through the argument `units.m`. The log link function is used for the linear predictor, which assumes the form

$$\log(\lambda) = S.$$

The random effect $S$ has a multivariate Gaussian distribution with mean `mu` and covariance matrix `Sigma`.

**Laplace sampling.** This function generates samples from the distribution of $S$ given the data y. Specifically a Langevin-Hastings algorithm is used to update $\tilde{S} = \tilde{\Sigma}^{-1/2}(S - \tilde{s})$ where $\tilde{\Sigma}$ and $\tilde{s}$ are the inverse of the negative Hessian and the mode of the distribution of $S$ given y, respectively. At each iteration a new value $\tilde{s}_{prop}$ for $\tilde{S}$ is proposed from a multivariate Gaussian distribution with mean

$$\tilde{s}_{curr} + (h/2)\nabla \log f(\tilde{S}|y),$$

where $\tilde{s}_{curr}$ is the current value for $\tilde{S}$, $h$ is a tuning parameter and $\nabla \log f(\tilde{S}|y)$ is the the gradient of the log-density of the distribution of $\tilde{S}$ given y. The tuning parameter $h$ is updated according to the following adaptive scheme: the value of $h$ at the $i$-th iteration, say $h_i$, is given by

$$h_i = h_{i-1} + c_1 i^{-c_2}(\alpha_i - 0.547),$$

where $c_1 > 0$ and $0 < c_2 < 1$ are pre-defined constants, and $\alpha_i$ is the acceptance rate at the $i$-th iteration ($0.547$ is the optimal acceptance rate for a multivariate standard Gaussian distribution). The starting value for $h$, and the values for $c_1$ and $c_2$ can be set through the function `control.mcmc.MCML`.

**Random effects at household-level.** When the data consist of two nested levels, such as households and individuals within households, the argument `ID.coords` must be used to define the household IDs for each individual. Let $i$ and $j$ denote the $i$-th household and the $j$-th person within that household; the logistic link function then assumes the form

$$\log(p_{ij}/(1-p_{ij})) = \mu_{ij} + S_i$$

where the random effects $S_i$ are now defined at household level and have mean zero. **Warning:** this modelling option is available only for the binomial model.

## Value

A list with the following components

`samples`: a matrix, each row of which corresponds to a sample from the predictive distribution.

`h`: vector of the values of the tuning parameter at each iteration of the Langevin-Hastings MCMC algorithm.

### Author(s)

Emanuele Giorgi <e.giorgi@lancaster.ac.uk>

Peter J. Diggle <p.diggle@lancaster.ac.uk>

### See Also

control.mcmc.MCML, create.ID.coords.

---

## Laplace.sampling.lr *Langevin-Hastings MCMC for conditional simulation (low-rank approximation)*

---

### Description

This function simulates from the conditional distribution of the random effects of binomial and Poisson models.

### Usage

```
Laplace.sampling.lr(
  mu,
  sigma2,
  K,
  y,
  units.m,
  control.mcmc,
  messages = TRUE,
  plot.correlogram = TRUE,
  poisson.llik = FALSE
)
```

### Arguments

| | |
|---|---|
| mu | mean vector of the linear predictor. |
| sigma2 | variance of the random effect. |
| K | random effect design matrix, or kernel matrix for the low-rank approximation. |
| y | vector of binomial/Poisson observations. |
| units.m | vector of binomial denominators, or offset if the Poisson model is used. |
| control.mcmc | output from control.mcmc.MCML. |
| messages | logical; if messages=TRUE then status messages are printed on the screen (or output device) while the function is running. Default is messages=TRUE. |
| plot.correlogram | logical; if plot.correlogram=TRUE the autocorrelation plot of the conditional simulations is displayed. |
| poisson.llik | logical; if poisson.llik=TRUE a Poisson model is used or, if poisson.llik=FALSE, a binomial model is used. |

## Details

**Binomial model.** Conditionally on $Z$, the data `y` follow a binomial distribution with probability $p$ and binomial denominators `units.m`. Let $K$ denote the random effects design matrix; a logistic link function is used, thus the linear predictor assumes the form

$$\log(p/(1-p)) = \mu + KZ$$

where $\mu$ is the mean vector component defined through `mu`. **Poisson model.** Conditionally on $Z$, the data `y` follow a Poisson distribution with mean $m\lambda$, where $m$ is an offset set through the argument `units.m`. Let $K$ denote the random effects design matrix; a log link function is used, thus the linear predictor assumes the form

$$\log(\lambda) = \mu + KZ$$

where $\mu$ is the mean vector component defined through `mu`. The random effect $Z$ has iid components distributed as zero-mean Gaussian variables with variance `sigma2`.

**Laplace sampling.** This function generates samples from the distribution of $Z$ given the data `y`. Specifically, a Langevin-Hastings algorithm is used to update $\tilde{Z} = \tilde{\Sigma}^{-1/2}(Z - \tilde{z})$ where $\tilde{\Sigma}$ and $\tilde{z}$ are the inverse of the negative Hessian and the mode of the distribution of $Z$ given `y`, respectively. At each iteration a new value $\tilde{z}_{prop}$ for $\tilde{Z}$ is proposed from a multivariate Gaussian distribution with mean

$$\tilde{z}_{curr} + (h/2)\nabla \log f(\tilde{Z}|y),$$

where $\tilde{z}_{curr}$ is the current value for $\tilde{Z}$, $h$ is a tuning parameter and $\nabla \log f(\tilde{Z}|y)$ is the the gradient of the log-density of the distribution of $\tilde{Z}$ given `y`. The tuning parameter $h$ is updated according to the following adaptive scheme: the value of $h$ at the $i$-th iteration, say $h_i$, is given by

$$h_i = h_{i-1} + c_1 i^{-c_2}(\alpha_i - 0.547),$$

where $c_1 > 0$ and $0 < c_2 < 1$ are pre-defined constants, and $\alpha_i$ is the acceptance rate at the $i$-th iteration ($0.547$ is the optimal acceptance rate for a multivariate standard Gaussian distribution). The starting value for $h$, and the values for $c_1$ and $c_2$ can be set through the function `control.mcmc.MCML`.

## Value

A list with the following components

`samples`: a matrix, each row of which corresponds to a sample from the predictive distribution.

`h`: vector of the values of the tuning parameter at each iteration of the Langevin-Hastings MCMC algorithm.

## Author(s)

Emanuele Giorgi <`e.giorgi@lancaster.ac.uk`>

Peter J. Diggle <`p.diggle@lancaster.ac.uk`>

## See Also

`control.mcmc.MCML`.

## Laplace.sampling.SPDE *Independence sampler for conditional simulation of a Gaussian process using SPDE*

### Description

This function simulates from the conditional distribution of a Gaussian process given binomial y. The Guassian process is also approximated using SPDE.

### Usage

```
Laplace.sampling.SPDE(
  mu,
  sigma2,
  phi,
  kappa,
  y,
  units.m,
  coords,
  mesh,
  control.mcmc,
  messages = TRUE,
  plot.correlogram = TRUE,
  poisson.llik
)
```

### Arguments

| | |
|---|---|
| mu | mean vector of the Gaussian process to approximate. |
| sigma2 | variance of the Gaussian process to approximate. |
| phi | scale parameter of the Matern function for the Gaussian process to approximate. |
| kappa | smothness parameter of the Matern function for the Gaussian process to approximate. |
| y | vector of binomial observations. |
| units.m | vector of binomial denominators. |
| coords | matrix of two columns corresponding to the spatial coordinates. |
| mesh | mesh object set through inla.mesh.2d. |
| control.mcmc | control parameters of the Independence sampler set through control.mcmc.MCML. |
| messages | logical; if messages=TRUE then status messages are printed on the screen (or output device) while the function is running. Default is messages=TRUE. |
| plot.correlogram | logical; if plot.correlogram=TRUE the autocorrelation plot of the conditional simulations is displayed. |
| poisson.llik | logical: if poisson.llik=TRUE then conditional conditional distribution of the data is Poisson; poisson.llik=FALSE then conditional conditional distribution of the data is Binomial. |

## Details

**Binomial model.** Conditionally on the random effect $S$, the data y follow a binomial distribution with probability $p$ and binomial denominators units.m. The logistic link function is used for the linear predictor, which assumes the form

$$\log(p/(1-p)) = S.$$

The random effect $S$ has a multivariate Gaussian distribution with mean mu and covariance matrix Sigma.

## Value

A list with the following components

samples: a matrix, each row of which corresponds to a sample from the predictive distribution.

## Author(s)

Emanuele Giorgi <e.giorgi@lancaster.ac.uk>

Peter J. Diggle <p.diggle@lancaster.ac.uk>

## See Also

control.mcmc.MCML.

---

# linear.model.Bayes *Bayesian estimation for the geostatistical linear Gaussian model*

---

## Description

This function performs Bayesian estimation for the geostatistical linear Gaussian model.

## Usage

```
linear.model.Bayes(
  formula,
  coords,
  data,
  kappa,
  control.mcmc,
  control.prior,
  low.rank = FALSE,
  knots = NULL,
  messages = TRUE
)
```

## Arguments

| | |
|---|---|
| formula | an object of class "formula" (or one that can be coerced to that class): a symbolic description of the model to be fitted. |
| coords | an object of class formula indicating the geographic coordinates. |
| data | a data frame containing the variables in the model. |
| kappa | shape parameter of the Matern covariance function. |
| control.mcmc | output from control.mcmc.Bayes. |
| control.prior | output from control.prior. |
| low.rank | logical; if low.rank=TRUE a low-rank approximation is fitted. |
| knots | if low.rank=TRUE, knots is a matrix of spatial knots used in the low-rank approximation. Default is knots=NULL. |
| messages | logical; if messages=TRUE then status messages are printed on the screen (or output device) while the function is running. Default is messages=TRUE. |

## Details

This function performs Bayesian estimation for the geostatistical linear Gaussian model, specified as

$$Y = d'\beta + S(x) + Z,$$

where $Y$ is the measured outcome, $d$ is a vector of coavariates, $\beta$ is a vector of regression coefficients, $S(x)$ is a stationary Gaussian spatial process and $Z$ are independent zero-mean Gaussian variables with variance tau2. More specifically, $S(x)$ has an isotropic Matern covariance function with variance sigma2, scale parameter phi and shape parameter kappa. The shape parameter kappa is treated as fixed.

**Priors definition.** Priors can be defined through the function control.prior. The hierarchical structure of the priors is the following. Let $\theta$ be the vector of the covariance parameters $(\sigma^2, \phi, \tau^2)$; then each component of $\theta$ can have independent priors freely defined by the user. However, uniform and log-normal priors are also available as default priors for each of the covariance parameters. To remove the nugget effect $Z$, no prior should be defined for tau2. Conditionally on sigma2, the vector of regression coefficients beta has a multivariate Gaussian prior with mean beta.mean and covariance matrix sigma2*beta.covar, while in the low-rank approximation the covariance matrix is simply beta.covar.

**Updating the covariance parameters using a Metropolis-Hastings algorithm.** In the MCMC algorithm implemented in linear.model.Bayes, the transformed parameters

$$(\theta_1, \theta_2, \theta_3) = (\log(\sigma^2)/2, \log(\sigma^2/\phi^{2\kappa}), \log(\tau^2))$$

are independently updated using a Metropolis Hastings algorithm. At the $i$-th iteration, a new value is proposed for each from a univariate Gaussian distrubion with variance, say $h_i^2$, tuned according the following adaptive scheme

$$h_i = h_{i-1} + c_1 i^{-c_2}(\alpha_i - 0.45),$$

where $\alpha_i$ is the acceptance rate at the $i$-th iteration (0.45 is the optimal acceptance rate for a univariate Gaussian distribution) whilst $c_1 > 0$ and $0 < c_2 < 1$ are pre-defined constants. The starting values $h_1$ for each of the parameters $\theta_1$, $\theta_2$ and $\theta_3$ can be set using the function control.mcmc.Bayes through the arguments h.theta1, h.theta2 and h.theta3. To define values for $c_1$ and $c_2$, see the documentation of control.mcmc.Bayes.

**Low-rank approximation.** In the case of very large spatial data-sets, a low-rank approximation of the Gaussian spatial process $S(x)$ might be computationally beneficial. Let $(x_1, \ldots, x_m)$ and $(t_1, \ldots, t_m)$ denote the set of sampling locations and a grid of spatial knots covering the area of interest, respectively. Then $S(x)$ is approximated as $\sum_{i=1}^m K(\|x - t_i\|; \phi, \kappa) U_i$, where $U_i$ are zero-mean mutually independent Gaussian variables with variance sigma2 and $K(.; \phi, \kappa)$ is the isotropic Matern kernel (see matern.kernel). Since the resulting approximation is no longer a stationary process (but only approximately), sigma2 may take very different values from the actual variance of the Gaussian process to approximate. The function adjust.sigma2 can then be used to (approximately) explore the range for sigma2. For example if the variance of the Gaussian process is 0.5, then an approximate value for sigma2 is 0.5/const.sigma2, where const.sigma2 is the value obtained with adjust.sigma2.

## Value

An object of class "Bayes.PrevMap". The function summary.Bayes.PrevMap is used to print a summary of the fitted model. The object is a list with the following components:

estimate: matrix of the posterior samples for each of the model parameters.

S: matrix of the posterior samplesfor each component of the random effect. This is only returned for the low-rank approximation.

y: response variable.

D: matrix of covariarates.

coords: matrix of the observed sampling locations.

kappa: vaues of the shape parameter of the Matern function.

knots: matrix of spatial knots used in the low-rank approximation.

const.sigma2: vector of the values of the multiplicative factor used to adjust the sigma2 in the low-rank approximation.

h1: vector of values taken by the tuning parameter h.theta1 at each iteration.

h2: vector of values taken by the tuning parameter h.theta2 at each iteration.

h3: vector of values taken by the tuning parameter h.theta3 at each iteration.

call: the matched call.

## Author(s)

Emanuele Giorgi <e.giorgi@lancaster.ac.uk>

Peter J. Diggle <p.diggle@lancaster.ac.uk>

## References

Diggle, P.J., Giorgi, E. (2019). *Model-based Geostatistics for Global Public Health.* CRC/Chapman & Hall.

Giorgi, E., Diggle, P.J. (2017). *PrevMap: an R package for prevalence mapping.* Journal of Statistical Software. 78(8), 1-29. doi: 10.18637/jss.v078.i08

Higdon, D. (1998). *A process-convolution approach to modeling temperatures in the North Atlantic Ocean.* Environmental and Ecological Statistics 5, 173-190.

### See Also

control.prior, control.mcmc.Bayes, shape.matern, summary.Bayes.PrevMap, autocor.plot, trace.plot, dens.plot, matern, matern.kernel, adjust.sigma2.

---

## linear.model.MLE — *Maximum Likelihood estimation for the geostatistical linear Gaussian model*

---

### Description

This function performs maximum likelihood estimation for the geostatistical linear Gaussian Model.

### Usage

```
linear.model.MLE(
  formula,
  coords = NULL,
  data,
  ID.coords = NULL,
  kappa,
  fixed.rel.nugget = NULL,
  start.cov.pars,
  method = "BFGS",
  low.rank = FALSE,
  knots = NULL,
  messages = TRUE,
  profile.llik = FALSE,
  SPDE = FALSE,
  mesh = NULL,
  SPDE.analytic.hessian = FALSE
)
```

### Arguments

| | |
|---|---|
| `formula` | an object of class "formula" (or one that can be coerced to that class): a symbolic description of the model to be fitted. |
| `coords` | an object of class formula indicating the geographic coordinates. |
| `data` | a data frame containing the variables in the model. |
| `ID.coords` | vector of ID values for the unique set of spatial coordinates obtained from create.ID.coords. These must be provided in order to define a geostatistical model where locations have multiple observations. Default is `ID.coords=NULL`. See the **Details** section for more information. |

kappa shape parameter of the Matern covariance function.

fixed.rel.nugget
: fixed value for the relative variance of the nugget effect; default is `fixed.rel.nugget=NULL` if this should be included in the estimation.

start.cov.pars
: if `ID.coords=NULL`, a vector of length two with elements corresponding to the starting values of `phi` and the relative variance of the nugget effect `nu2`, respectively, that are used in the optimization algorithm; if `ID.coords` is provided, a third starting value for the relative variance of the individual unexplained variation `nu2.star = omega2/sigma2` must be provided. If `nu2` is fixed through `fixed.rel.nugget`, then start.cov.pars represents the starting value for `phi` only, if `ID.coords=NULL`, or for `phi` and `nu2.star`, otherwise.

method
: method of optimization. If `method="BFGS"` then the `maxBFGS` function is used; otherwise `method="nlminb"` to use the `nlminb` function. Default is `method="BFGS"`.

low.rank
: logical; if `low.rank=TRUE` a low-rank approximation of the Gaussian spatial process is used when fitting the model. Default is `low.rank=FALSE`.

knots
: if `low.rank=TRUE`, `knots` is a matrix of spatial knots that are used in the low-rank approximation. Default is `knots=NULL`.

messages
: logical; if `messages=TRUE` then status messages are printed on the screen (or output device) while the function is running. Default is `messages=TRUE`.

profile.llik
: logical; if `profile.llik=TRUE` the maximization of the profile likelihood is carried out. If `profile.llik=FALSE` the full-likelihood is used. Default is `profile.llik=FALSE`.

SPDE
: logical; if `SPDE=TRUE` the SPDE approximation for the Gaussian spatial model is used. Default is `SPDE=FALSE`.

mesh
: an object obtained as result of a call to the function `inla.mesh.2d`.

SPDE.analytic.hessian
: logical; if `SPDE.analytic.hessian=TRUE` computation of the hessian matrix using the SPDE approximation is carried out using analytical expressions, otherwise a numerical approximation is used. Defauls is `SPDE.analytic.hessian=FALSE`.

## Details

This function estimates the parameters of a geostatistical linear Gaussian model, specified as

$$Y = d'\beta + S(x) + Z,$$

where $Y$ is the measured outcome, $d$ is a vector of coavariates, $\beta$ is a vector of regression coefficients, $S(x)$ is a stationary Gaussian spatial process and $Z$ are independent zero-mean Gaussian variables with variance `tau2`. More specifically, $S(x)$ has an isotropic Matern covariance function with variance `sigma2`, scale parameter `phi` and shape parameter `kappa`. In the estimation, the shape parameter `kappa` is treated as fixed. The relative variance of the nugget effect, `nu2=tau2/sigma2`, can be fixed though the argument `fixed.rel.nugget`; if `fixed.rel.nugget=NULL`, then the variance of the nugget effect is also included in the estimation.

**Locations with multiple observations.** If multiple observations are available at any of the sampled locations the above model is modified as follows. Let $Y_{ij}$ denote the random variable associated to the measured outcome for the j-th individual at location $x_i$. The linear geostatistical model assumes the form

$$Y_{ij} = d'_{ij}\beta + S(x_i) + Zi + U_{ij},$$

where $S(x_i)$ and $Z_i$ are specified as mentioned above, and $U_{ij}$ are i.i.d. zer0-mean Gaussian variable with variance $\omega^2$. his model can be fitted by specifing a vector of ID for the unique set locations thourgh the argument `ID.coords` (see also `create.ID.coords`).

**Low-rank approximation.** In the case of very large spatial data-sets, a low-rank approximation of the Gaussian spatial process $S(x)$ can be computationally beneficial. Let $(x_1, \ldots, x_m)$ and $(t_1, \ldots, t_m)$ denote the set of sampling locations and a grid of spatial knots covering the area of interest, respectively. Then $S(x)$ is approximated as $\sum_{i=1}^m K(\|x - t_i\|; \phi, \kappa)U_i$, where $U_i$ are zero-mean mutually independent Gaussian variables with variance `sigma2` and $K(.; \phi, \kappa)$ is the isotropic Matern kernel (see `matern.kernel`). Since the resulting approximation is no longer a stationary process, the parameter `sigma2` is adjusted by a factor`constant.sigma2`. See `adjust.sigma2` for more details on the the computation of the adjustment factor `constant.sigma2` in the low-rank approximation.

## Value

An object of class "PrevMap". The function `summary.PrevMap` is used to print a summary of the fitted model. The object is a list with the following components:

`estimate`: estimates of the model parameters; use the function `coef.PrevMap` to obtain estimates of covariance parameters on the original scale.

`covariance`: covariance matrix of the ML estimates.

`log.lik`: maximum value of the log-likelihood.

`y`: response variable.

`D`: matrix of covariates.

`coords`: matrix of the observed sampling locations.

`ID.coords`: set of ID values defined through the argument `ID.coords`.

`method`: method of optimization used.

`kappa`: fixed value of the shape parameter of the Matern function.

`knots`: matrix of the spatial knots used in the low-rank approximation.

`const.sigma2`: adjustment factor for `sigma2` in the low-rank approximation.

`fixed.rel.nugget`: fixed value for the relative variance of the nugget effect.

`mesh`: the mesh used in the SPDE approximation.

`call`: the matched call.

## Author(s)

Emanuele Giorgi <e.giorgi@lancaster.ac.uk>

Peter J. Diggle <p.diggle@lancaster.ac.uk>

### References


Diggle, P.J., Giorgi, E. (2019). *Model-based Geostatistics for Global Public Health.* CRC/Chapman & Hall.

Giorgi, E., Diggle, P.J. (2017). *PrevMap: an R package for prevalence mapping.* Journal of Statistical Software. 78(8), 1-29. doi: 10.18637/jss.v078.i08

Higdon, D. (1998). *A process-convolution approach to modeling temperatures in the North Atlantic Ocean.* Environmental and Ecological Statistics 5, 173-190.


### See Also

`shape.matern`, `summary.PrevMap`, `coef.PrevMap`, `matern`, `matern.kernel`, `maxBFGS`, `nlminb`.

---

`lm.ps.MCML` *Monte Carlo Maximum Likelihood estimation of the geostatistical linear model with preferentially sampled locations*

---

### Description

This function performs Monte Carlo maximum likelihood (MCML) estimation for a geostatistical linear model with preferentially sampled locations. For more details on the model, see below.

### Usage

```
lm.ps.MCML(
  formula.response,
  formula.log.intensity = ~1,
  coords,
  which.is.preferential = NULL,
  data.response,
  data.intensity = NULL,
  par0,
  control.mcmc,
  kappa1,
  kappa2,
  mesh,
  grid.intensity,
  start.par = NULL,
  method = "nlminb",
  messages = TRUE,
  plot.correlogram = TRUE
)
```

## Arguments

formula.response
: an object of class formula (or one that can be coerced to that class): a symbolic description of the sub-model for the response variable.

formula.log.intensity
: an object of class formula (or one that can be coerced to that class): a symbolic description of the log-Gaussian Cox process sub-model.

coords
: an object of class formula indicating the spatial coordinates in the data.

which.is.preferential
: a vector of 0 and 1, where 1 indicates a location in the data from a prefential sampling scheme and 0 from a non-preferential. This option is used to fit a model with a mix of preferentally and non-preferentiall sampled locations. For more, details on the model structure see the 'Details' section.

data.response
: a data frame containing the variables in the sub-model of the response variable.

data.intensity
: a data frame containing the variables in the log-Gaussian Coz process sub-model. This data frame must be provided only when explanatory variables are used in the log-Gaussian Cox process model. Each row in the data frame must correspond to a point in the grid provided through the argument 'grid.intensity'. Deafult is `data.intensity=NULL`, which corresponds to a model with only the intercept.

par0
: an object of class 'coef.PrevMap.ps'. This argument is used to define the parameters of the importance sampling distribution used in the MCML algorithm. The input of this argument must be defined using the set.par.ps function.

control.mcmc
: output from control.mcmc.MCML which defined the control parameters of the Monte Carlo Markv chain algorithm.

kappa1
: fixed value for the shape parameter of the Matern covariance function of the spatial process of the sampling intensity (currently only `kappa1=1` is implemented).

kappa2
: fixed value for the shape parameter of the Matern covariance function of the spatial process of the response variable.

mesh
: an object obtained as result of a call to the function `inla.mesh.2d`.

grid.intensity
: a regular grid covering the geographical region of interest, used to approximate the density function of the log-Gaussian Cox process.

start.par
: starting value of the optimization algorithm. This is an object of class 'coef.PrevMap.ps' and must be defined using the function set.par.ps. Default is `start.cov.pars=NULL`, so that the starting values are set automatically.

method
: method of optimization. If `method="BFGS"` then the maxBFGS function is used; otherwise `method="nlminb"` to use the nlminb function. Default is `method="BFGS"`.

messages
: logical; if `messages=TRUE` then status messages are printed on the screen (or output device) while the function is running. Default is `messages=TRUE`.

plot.correlogram
: logical; if `plot.correlogram=TRUE` the autocorrelation plot of the samples of the random effect is displayed after completion of conditional simulation. Default is `plot.correlogram=TRUE`.

## Details

This function performs parameter estimation for a geostatistical linear model with preferentially sampled locations. Let $S_1$ and $S_2$ denote two independent, stationary and isotropic Gaussian processes. The overall model consists of two sub-models: the log-Gaussian Cox process model for the preferentially sampled locations, say $X$; the model for the response variable, say $Y$. The model assumes that

$$[X, Y, S_1, S_2] = [S_1][S_2][X|S_1][Y|X, S_1, S_2],$$

where [.] denotes 'the distribution of .'. Each of the two submodels has an associated linear predictor. Let $\Lambda(x)$ denote the intensity of the Poisson process $X$, continionally on $S_1$. Then

$$\log\{\Lambda(x)\} = d(x)'\alpha + S_1$$

, where $d(x)$ is vector of explanatory variables with regression coefficient $\alpha$. This linear predictor is defined through the argument `formula.log.intensity`. The density of $[X|S_1]$ is given by

$$\frac{\Lambda(x)}{\int_A \Lambda(u)du}$$

, where $A$ is the region of interest. The integral at the denominator is intractable and is then approximated using a quadrature procedure. The regular grid covering $A$, used for the quadrature, must be provided through the argument `grid.intensity`. Conditionally on $X$, $S_1$ and $S_2$, the response variable model is given by

$$Y = d(x)'\beta + S_2 + \gamma S_1,$$

where $\beta$ is another vector of regression coefficients and $\gamma$ is the preferentiality parameter. If $\gamma = 0$ then we recover the standard geostatistical model. More details on the fitting procedure can be found in Diggle and Giorgi (2016).

**When the data have a mix of preferentially and non-preferentially sampled locations.** In some cases the set of locations may consist of a sub-set which is preferentially sampled, $X$, and a standard non-prefential sample, $X^*$. Let $Y$ and $Y^*$ denote the measurments at locations $X$ and $X^*$. In the current implementation, the model has the following form

$$[X, X^*, Y, Y^*, S_1, S_2, S_2^*] = [S_1][S_2][S_2^*][X|S_1][Y|X, S_1, S_2][X^*][Y^*|X^*, S_2^*],$$

where $S_2$ and $S_2^*$ are two independent Gaussian process but with shared parameters, associated with $Y$ and $Y^*$, respectively. The linear predictor for $Y$ is the same as above. The measurements $Y^*$, instead, have linear predicotr

$$Y^* = d(x)'\beta + S_2^*,$$

where $\beta^*$ is vector of regression coefficients, different from $\beta$. The linear predictor for $Y$ and $Y^*$ is specified though `formula.response`. For example, `response ~ x | x + z` defines a linear predictor for $Y$ with one explanatory variable `x` and a linear predictor for $Y^*$ with two explanatory variables `x` and `z`. An example on the application of this model is given in Diggle and Giorgi (2016).

## Value

An object of class "PrevMap.ps". The function `summary.PrevMap.ps` is used to print a summary of the fitted model. The object is a list with the following components:

`estimate`: estimates of the model parameters; use the function `coef.PrevMap.ps` to obtain estimates of covariance parameters on the original scale.

covariance: covariance matrix of the MCML estimates.

log.lik: maximum value of the approximated log-likelihood.

y: observed values of the response variable. If which.is.preferential has been provided, then y is a list with components y$preferential, for the data with prefentially sampled locations, and y$non.preferential, for the remiaining.

D.response: matrix of covariates used to model the mean component of the response variable. If which.is.preferential has been provided, then D.response is a list with components D.response$preferential, for the data with prefentially sampled locations, and D.response$non.preferential, for the remiaining.

D.intensity: matrix of covariates used to model the mean component of log-intensity of the log-Gaussian Cox process.

grid.intensity: grid of locations used to approximate the intractable integral of the log-Gaussian Cox process model.

coords: matrix of the observed sampling locations. If which.is.preferential has been provided, then coords is a list with components y$preferential, for the data with prefentially sampled locations, and y$non.preferential, for the remiaining.

method: method of optimization used.

ID.coords: set of ID values defined through the argument ID.coords.

kappa.response: fixed value of the shape parameter of the Matern covariance function used to model the spatial process associated with the response variable.

mesh: the mesh used in the SPDE approximation.

samples: matrix of the random effects samples from the importance sampling distribution used to approximate the likelihood function.

call: the matched call.

## Author(s)

Emanuele Giorgi <e.giorgi@lancaster.ac.uk>

## References

Diggle, P.J., Giorgi, E. (2019). *Model-based Geostatistics for Global Public Health.* CRC/Chapman & Hall.

Giorgi, E., Diggle, P.J. (2017). *PrevMap: an R package for prevalence mapping.* Journal of Statistical Software. 78(8), 1-29. doi: 10.18637/jss.v078.i08

Diggle, P.J., Giorgi, E. (2017). *Preferential sampling of exposures levels.* In: Handbook of Environmental and Ecological Statistics. Chapman & Hall.

Diggle, P.J., Menezes, R. and Su, T.-L. (2010). *Geostatistical analysis under preferential sampling (with Discussion).* Applied Statistics, 59, 191-232.

Lindgren, F., Havard, R., Lindstrom, J. (2011). *An explicit link between Gaussian fields and Gaussian Markov random fields: the stochastic partial differential equation approach (with discussion).* Journal of the Royal Statistical Society, Series B, 73, 423–498.

Pati, D., Reich, B. J., and Dunson, D. B. (2011). *Bayesian geostatistical modelling with informative sampling locations.* Biometrika, 98, 35-48.

loaloa *Loa loa prevalence data from 197 village surveys*

## Description

This data-set relates to a study of the prevalence of Loa loa (eyeworm) in a series of surveys undertaken in 197 villages in west Africa (Cameroon and southern Nigeria). The variables are as follows:

- ROW row id: 1 to 197.
- VILLCODE village id.
- LONGITUDE Longitude in degrees.
- LATITUDE Latitude in degrees.
- NO_EXAM Number of people tested.
- NO_INF Number of positive test results.
- ELEVATION Height above sea-level in metres.
- MEAN9901 Mean of all NDVI values recorded at village location, 1999-2001
- MAX9901 Maximum of all NDVI values recorded at village location, 1999-2001
- MIN9901 Minimum of all NDVI values recorded at village location, 1999-2001
- MIN9901 Minimum of all NDVI values recorded at village location, 1999-2001
- STDEV9901 standard deviation of all NDVI values recorded at village location, 1999-2001

## Usage

```
data(loaloa)
```

## Format

A data frame with 197 rows and 11 variables

## References

Diggle, P.J., Thomson, M.C., Christensen, O.F., Rowlingson, B., Obsomer, V., Gardon, J., Wanji, S., Takougang, I., Enyong, P., Kamgno, J., Remme, H., Boussinesq, M. and Molyneux, D.H. (2007). Spatial modelling and prediction of Loa loa risk: decision making under uncertainty. Annals of Tropical Medicine and Parasitology, 101, 499-509.

## loglik.ci *Profile likelihood confidence intervals*

### Description

Computes confidence intervals based on the interpolated profile likelihood computed for a single covariance parameter.

### Usage

```
loglik.ci(object, coverage = 0.95, plot.spline.profile = TRUE)
```

### Arguments

| | |
|---|---|
| `object` | object of class "profile.PrevMap" obtained from `loglik.linear.model`. |
| `coverage` | a value between 0 and 1 indicating the coverage of the confidence interval based on the interpolated profile likelihood. Default is `coverage=0.95`. |
| `plot.spline.profile` | logical; if `TRUE` an interpolating spline of the profile-likelihood of for a univariate parameter is plotted. Default is `FALSE`. |

### Value

A list with elements `lower` and `upper` for the upper and lower limits of the confidence interval, respectively.

### Author(s)

Emanuele Giorgi <`e.giorgi@lancaster.ac.uk`>

Peter J. Diggle <`p.diggle@lancaster.ac.uk`>

## loglik.linear.model *Profile log-likelihood or fixed parameters likelihood evaluation for the covariance parameters in the geostatistical linear model*

### Description

Computes profile log-likelihood, or evaluatesx likelihood keeping the other paramaters fixed, for the scale parameter `phi` of the Matern function and the relative variance of the nugget effect `nu2` in the linear Gaussian model.

### Usage

```
loglik.linear.model(
  object,
  control.profile,
  plot.profile = TRUE,
  messages = TRUE
)
```

### Arguments

| | |
|---|---|
| `object` | an object of class 'PrevMap', which is the fitted linear model obtained with the function `linear.model.MLE`. |
| `control.profile` | control parameters obtained with `control.profile`. |
| `plot.profile` | logical; if `TRUE` a plot of the computed profile likelihood is displayed. |
| `messages` | logical; if `messages=TRUE` then status messages are printed on the screen (or output device) while the function is running. Default is `messages=TRUE`. |

### Value

an object of class "profile.PrevMap" which is a list with the following values

`eval.points.phi`: vector of the values used for `phi` in the evaluation of the likelihood.

`eval.points.rel.nugget`: vector of the values used for `nu2` in the evaluation of the likelihood.

`profile.phi`: vector of the values of the likelihood function evaluated at `eval.points.phi`.

`profile.rel.nugget`: vector of the values of the likelihood function evaluated at `eval.points.rel.nugget`.

`profile.phi.rel.nugget`: matrix of the values of the likelihood function evaluated at `eval.points.phi` and `eval.points.rel.nugget`.

`fixed.par`: logical value; `TRUE` is the evaluation if the likelihood is carried out by fixing the other parameters, and `FALSE` if the computation of the profile-likelihood was performed instead.

### Author(s)

Emanuele Giorgi <e.giorgi@lancaster.ac.uk>

Peter J. Diggle <p.diggle@lancaster.ac.uk>

---

## `matern.kernel` *Matern kernel*

---

### Description

This function computes values of the Matern kernel for given distances and parameters.

### Usage

```
matern.kernel(u, rho, kappa)
```

### Arguments

u — a vector, matrix or array with values of the distances between pairs of data locations.

rho — value of the (re-parametrized) scale parameter; this corresponds to the re-parametrization rho = 2*sqrt(kappa)*phi.

kappa — value of the shape parameter.

### Details

The Matern kernel is defined as:

$$K(u;\phi,\kappa) = \frac{\Gamma(\kappa+1)^{1/2}\kappa^{(\kappa+1)/4}u^{(\kappa-1)/2}}{\pi^{1/2}\Gamma((\kappa+1)/2)\Gamma(\kappa)^{1/2}(2\kappa^{1/2}\phi)^{(\kappa+1)/2}}\mathcal{K}_{\kappa}(u/\phi), u > 0,$$

where $\phi$ and $\kappa$ are the scale and shape parameters, respectively, and $\mathcal{K}_{\kappa}(.)$ is the modified Bessel function of the third kind of order $\kappa$. The family is valid for $\phi > 0$ and $\kappa > 0$.

### Value

A vector matrix or array, according to the argument u, with the values of the Matern kernel function for the given distances.

### Author(s)

Emanuele Giorgi <e.giorgi@lancaster.ac.uk>

Peter J. Diggle <p.diggle@lancaster.ac.uk>

---

## plot.pred.PrevMap — *Plot of a predicted surface*

---

### Description

plot.pred.PrevMap displays predictions obtained from spatial.pred.linear.MLE, spatial.pred.linear.Bayes,spati spatial.pred.binomial.Bayes and spatial.pred.poisson.MCML.

### Usage

```
## S3 method for class 'pred.PrevMap'
plot(x, type = NULL, summary = "predictions", ...)
```

## Arguments

| | |
|---|---|
| `x` | an object of class "PrevMap". |
| `type` | a character indicating the type of prediction to display: 'prevalence','odds', 'logit' or 'probit' for binomial models; "log" or "exponential" for Poisson models. Default is `NULL`. |
| `summary` | character indicating which summary to display: 'predictions','quantiles', 'standard.errors' or 'exceedance.prob'; default is 'predictions'. If `summary="exceedance.prob"`, the argument `type` is ignored. |
| `...` | further arguments passed to `plot` of the 'raster' package. |

## Author(s)

Emanuele Giorgi <`e.giorgi@lancaster.ac.uk`>

Peter J. Diggle <`p.diggle@lancaster.ac.uk`>

---

`plot.pred.PrevMap.ps` *Plot of a predicted surface of geostatistical linear fits with preferentially sampled locations*

---

## Description

`plot.pred.PrevMap.ps` displays predictions obtained from `lm.ps.MCML`.

## Usage

```
## S3 method for class 'pred.PrevMap.ps'
plot(x, target = NULL, summary = "predictions", ...)
```

## Arguments

| | |
|---|---|
| `x` | an object of class "PrevMap". |
| `target` | a integer value indicating the predictive target: `target=1` to visualize summaries of the surface associated with the response variable; `target=2` to visualize summaries of the surface associated with the sampling intensity. If only one target has been predicted, this argument is ignored. |
| `summary` | character indicating which summary to display: 'predictions','quantiles' or 'standard.errors'. Default is `summary='predictions'`. If `summary="exceedance.prob"`, the argument `type` is ignored. |
| `...` | further arguments passed to `plot` of the 'raster' package. |

## Author(s)

Emanuele Giorgi <`e.giorgi@lancaster.ac.uk`>

plot.PrevMap.diagnostic *Plot of the variogram-based diagnostics*

## Description

Displays the results from a call to `variog.diagnostic.lm` and `variog.diagnostic.glgm`.

## Usage

```
## S3 method for class 'PrevMap.diagnostic'
plot(x, ...)
```

## Arguments

| | |
|---|---|
| x | an object of class "PrevMap.diagnostic". |
| ... | further arguments passed to `plot` of the 'raster' package. |

## Author(s)

Emanuele Giorgi <e.giorgi@lancaster.ac.uk>

Peter J. Diggle <p.diggle@lancaster.ac.uk>

## See Also

`variog.diagnostic.lm`, `variog.diagnostic.glgm`

plot.profile.PrevMap *Plot of the profile log-likelihood for the covariance parameters of the Matern function*

## Description

This function displays a plot of the profile log-likelihood that is computed by the function `loglik.linear.model`.

## Usage

```
## S3 method for class 'profile.PrevMap'
plot(x, log.scale = FALSE, plot.spline.profile = FALSE, ...)
```

### Arguments

| | |
|---|---|
| x | object of class "profile.PrevMap" obtained as output from `loglik.linear.model`. |
| log.scale | logical; if `log.scale=TRUE`, the profile likleihood is plotted on the log-scale of the parameter values. |
| plot.spline.profile | logical; if `TRUE` an interpolating spline of the profile-likelihood of for a univariate parameter is plotted. Default is `FALSE`. |
| ... | further arugments passed to `plot` if the profile log-likelihood is for only one parameter, or to `contour` for the bi-variate profile-likelihood. |

### Value

A plot is returned. No value is returned.

### Author(s)

Emanuele Giorgi <e.giorgi@lancaster.ac.uk>

Peter J. Diggle <p.diggle@lancaster.ac.uk>

---

## plot.shape.matern *Plot of the profile likelihood for the shape parameter of the Matern covariance function*

---

### Description

This function plots the profile likelihood for the shape parameter of the Matern covariance function using the output from `shape.matern` function.

### Usage

```
## S3 method for class 'shape.matern'
plot(x, plot.spline = TRUE, ...)
```

### Arguments

| | |
|---|---|
| x | an object of class 'shape.matern' obtained as result of a call to `shape.matern` |
| plot.spline | logical; if `TRUE` an interpolating spline of the profile likelihood is added to the plot. |
| ... | further arguments passed to `plot`. |

### Value

The function does not return any value.

### Author(s)

Emanuele Giorgi <e.giorgi@lancaster.ac.uk>

Peter J. Diggle <p.diggle@lancaster.ac.uk>

### See Also

shape.matern

---

## point.map — *Point map*

---

### Description

This function produces a plot with points indicating the data locations. Arguments can control the points sizes, patterns and colors. These can be set to be proportional to data values, ranks or quantiles. Alternatively, points can be added to the current plot.

### Usage

```
point.map(data, var.name, coords, ...)
```

### Arguments

| | |
|---|---|
| `data` | an object of class "data.frame" containing the data. |
| `var.name` | a `formula` object indicating the variable to display. |
| `coords` | a `formula` object indicating the geographical coordinates. |
| `...` | additional arguments to be passed to `points.geodata`. |

---

## poisson.log.MCML — *Monte Carlo Maximum Likelihood estimation for the Poisson model*

---

### Description

This function performs Monte Carlo maximum likelihood (MCML) estimation for the geostatistical Poisson model with log link function.

### Usage

```
poisson.log.MCML(
  formula,
  units.m = NULL,
  coords,
  data,
  ID.coords = NULL,
  par0,
  control.mcmc,
  kappa,
  fixed.rel.nugget = NULL,
  start.cov.pars,
  method = "BFGS",
  low.rank = FALSE,
  knots = NULL,
  messages = TRUE,
  plot.correlogram = TRUE
)
```

### Arguments

| | |
|---|---|
| formula | an object of class formula (or one that can be coerced to that class): a symbolic description of the model to be fitted. |
| units.m | an object of class formula indicating the multiplicative offset for the mean of the Poisson model; if not specified this is then internally set as 1. |
| coords | an object of class formula indicating the geographic coordinates. |
| data | a data frame containing the variables in the model. |
| ID.coords | vector of ID values for the unique set of spatial coordinates obtained from create.ID.coords. These must be provided if, for example, spatial random effects are defined at location-level but some of the covariates are at individual level. **Warning**: the spatial coordinates must all be distinct otherwise see jitterDupCoords. Default is NULL. |
| par0 | parameters of the importance sampling distribution: these should be given in the following order c(beta,sigma2,phi,tau2), where beta are the regression coefficients, sigma2 is the variance of the Gaussian process, phi is the scale parameter of the spatial correlation and tau2 is the variance of the nugget effect (if included in the model). |
| control.mcmc | output from control.mcmc.MCML. |
| kappa | fixed value for the shape parameter of the Matern covariance function. |
| fixed.rel.nugget | fixed value for the relative variance of the nugget effect; fixed.rel.nugget=NULL if this should be included in the estimation. Default is fixed.rel.nugget=NULL. |
| start.cov.pars | a vector of length two with elements corresponding to the starting values of phi and the relative variance of the nugget effect nu2, respectively, that are used in the optimization algorithm. If nu2 is fixed through fixed.rel.nugget, then start.cov.pars represents the starting value for phi only. |

| | |
|---|---|
| `method` | method of optimization. If `method="BFGS"` then the `maxBFGS` function is used; otherwise `method="nlminb"` to use the `nlminb` function. Default is `method="BFGS"`. |
| `low.rank` | logical; if `low.rank=TRUE` a low-rank approximation of the Gaussian spatial process is used when fitting the model. Default is `low.rank=FALSE`. |
| `knots` | if `low.rank=TRUE`, `knots` is a matrix of spatial knots that are used in the low-rank approximation. Default is `knots=NULL`. |
| `messages` | logical; if `messages=TRUE` then status messages are printed on the screen (or output device) while the function is running. Default is `messages=TRUE`. |
| `plot.correlogram` | logical; if `plot.correlogram=TRUE` the autocorrelation plot of the samples of the random effect is displayed after completion of conditional simulation. Default is `plot.correlogram=TRUE`. |

## Details

This function performs parameter estimation for a geostatistical Poisson model with log link function. Conditionally on a zero-mean stationary Gaussian process $S(x)$ and mutually independent zero-mean Gaussian variables $Z$ with variance `tau2`, the observations `y` are generated from a Poisson distribution with mean $m\lambda$, where $m$ is an offset defined through the argument `units.m`. A canonical log link is used, thus the linear predictor assumes the form

$$\log(\lambda) = d'\beta + S(x) + Z,$$

where $d$ is a vector of covariates with associated regression coefficients $\beta$. The Gaussian process $S(x)$ has isotropic Matern covariance function (see `matern`) with variance `sigma2`, scale parameter `phi` and shape parameter `kappa`. In the `poisson.log.MCML` function, the shape parameter is treated as fixed. The relative variance of the nugget effect, `nu2=tau2/sigma2`, can also be fixed through the argument `fixed.rel.nugget`; if `fixed.rel.nugget=NULL`, then the relative variance of the nugget effect is also included in the estimation.

**Monte Carlo Maximum likelihood.** The Monte Carlo maximum likelihood method uses conditional simulation from the distribution of the random effect $T(x) = d(x)'\beta + S(x) + Z$ given the data `y`, in order to approximate the high-dimensiional intractable integral given by the likelihood function. The resulting approximation of the likelihood is then maximized by a numerical optimization algorithm which uses analytic epression for computation of the gradient vector and Hessian matrix. The functions used for numerical optimization are `maxBFGS` (`method="BFGS"`), from the **maxLik** package, and `nlminb` (`method="nlminb"`).

**Low-rank approximation.** In the case of very large spatial data-sets, a low-rank approximation of the Gaussian spatial process $S(x)$ might be computationally beneficial. Let $(x_1, \ldots, x_m)$ and $(t_1, \ldots, t_m)$ denote the set of sampling locations and a grid of spatial knots covering the area of interest, respectively. Then $S(x)$ is approximated as $\sum_{i=1}^{m} K(\|x - t_i\|; \phi, \kappa)U_i$, where $U_i$ are zero-mean mutually independent Gaussian variables with variance `sigma2` and $K(.; \phi, \kappa)$ is the isotropic Matern kernel (see `matern.kernel`). Since the resulting approximation is no longer a stationary process (but only approximately), the parameter `sigma2` is then multiplied by a factor `constant.sigma2` so as to obtain a value that is closer to the actual variance of $S(x)$.

## Value

An object of class "PrevMap". The function `summary.PrevMap` is used to print a summary of the fitted model. The object is a list with the following components:

estimate: estimates of the model parameters; use the function coef.PrevMap to obtain estimates of covariance parameters on the original scale.

covariance: covariance matrix of the MCML estimates.

log.lik: maximum value of the log-likelihood.

y: observations.

units.m: offset.

D: matrix of covariates.

ID.coords: set of ID values defined through the argument ID.coords.

coords: matrix of the observed sampling locations.

method: method of optimization used.

kappa: fixed value of the shape parameter of the Matern function.

knots: matrix of the spatial knots used in the low-rank approximation.

const.sigma2: adjustment factor for sigma2 in the low-rank approximation.

h: vector of the values of the tuning parameter at each iteration of the Langevin-Hastings MCMC algorithm; see Laplace.sampling, or Laplace.sampling.lr if a low-rank approximation is used.

samples: matrix of the random effects samples from the importance sampling distribution used to approximate the likelihood function.

fixed.rel.nugget: fixed value for the relative variance of the nugget effect.

call: the matched call.

## Author(s)

Emanuele Giorgi <e.giorgi@lancaster.ac.uk>

Peter J. Diggle <p.diggle@lancaster.ac.uk>

## References

Diggle, P.J., Giorgi, E. (2019). *Model-based Geostatistics for Global Public Health.* CRC/Chapman & Hall.

Giorgi, E., Diggle, P.J. (2017). *PrevMap: an R package for prevalence mapping.* Journal of Statistical Software. 78(8), 1-29. doi: 10.18637/jss.v078.i08

Christensen, O. F. (2004). *Monte carlo maximum likelihood in model-based geostatistics.* Journal of Computational and Graphical Statistics 13, 702-718.

Higdon, D. (1998). *A process-convolution approach to modeling temperatures in the North Atlantic Ocean.* Environmental and Ecological Statistics 5, 173-190.

## See Also

Laplace.sampling, Laplace.sampling.lr, summary.PrevMap, coef.PrevMap, matern, matern.kernel, control.mcmc.MCML.

`set.par.ps` *Define the model coefficients of a geostatistical linear model with preferentially sampled locations*

## Description

`set.par.ps` defines the model coefficients of a geostatistical linear model with preferentially sampled locations. The output of this function can be used to: 1) define the parameters of the importance sampling distribution in `lm.ps.MCML`; 2) the starting values of the optimization algorithm in `lm.ps.MCML`.

## Usage

```
set.par.ps(p = 1, q = 1, intensity, response, preferentiality.par)
```

## Arguments

| | |
|---|---|
| `p` | number of covariates used in the response variable model, including the intercept. Default is p=1. |
| `q` | number of covariates used in the log-Guassian Cox process model, including the intercept. Default is q=1. |
| `intensity` | a vector of parameters of the log-Gaussian Cox process model. These must be provided in the following order: regression coefficients of the explanatory variables; variance and scale of the spatial correlation for the isotropic Gaussian process. In the case of a model with a mix of preferentially and non-preferentially sampled locations, the order of the regression coefficients should be the following: regression coefficients for the linear predictor with preferential sampling; regression coefficients for the linear predictor with non-preferential samples. |
| `response` | a vector of parameters of the response variable model. These must be provided in the following order: regression coefficients of the explanatory variables; variance and scale of the spatial correlation for the isotropic Gaussian process; and variance of the nugget effect. |
| `preferentiality.par` | value of the preferentiality paramter. |

## Value

a list of coefficients of class `coef.PrevMap.ps`.

## Author(s)

Emanuele Giorgi <e.giorgi@lancaster.ac.uk>

shape.matern *Profile likelihood for the shape parameter of the Matern covariance function*

## Description

This function plots the profile likelihood for the shape parameter of the Matern covariance function used in the linear Gaussian model. It also computes confidence intervals of coverage `coverage` by interpolating the profile likelihood with a spline and using the asymptotic distribution of a chi-squared with one degree of freedom.

## Usage

```
shape.matern(
  formula,
  coords,
  data,
  set.kappa,
  fixed.rel.nugget = NULL,
  start.par,
  coverage = NULL,
  plot.profile = TRUE,
  messages = TRUE
)
```

## Arguments

- `formula` an object of class `formula` (or one that can be coerced to that class): a symbolic description of the model to be fitted.
- `coords` an object of class `formula` indicating the geographic coordinates.
- `data` a data frame containing the variables in the model.
- `set.kappa` a vector indicating the set values for evluation of the profile likelihood.
- `fixed.rel.nugget` a value for the relative variance `nu2` of the nugget effect, that is then treated as fixed. Default is `NULL`.
- `start.par` starting values for the scale parameter `phi` and the relative variance of the nugget effect `nu2`; if `fixed.rel.nugget` is provided, then a starting value for `phi` only should be provided.
- `coverage` a value between 0 and 1 indicating the coverage of the confidence interval based on the interpolated profile liklelihood for the shape parameter. Default is `coverage=NULL` and no confidence interval is then computed.
- `plot.profile` logical; if `TRUE` the computed profile-likelihood is plotted together with the interpolating spline.
- `messages` logical; if `messages=TRUE` then status messages are printed on the screen (or output device) while the function is running. Default is `messages=TRUE`.

## Value

The function returns an object of class 'shape.matern' that is a list with the following components

set.kappa set of values of the shape parameter used to evaluate the profile-likelihood.

val.kappa values of the profile likelihood.

If a value for coverage is specified, the list also contains lower, upper and kappa.hat that correspond to the lower and upper limits of the confidence interval, and the maximum likelihood estimate for the shape parameter, respectively.

## Author(s)

Emanuele Giorgi <e.giorgi@lancaster.ac.uk>

Peter J. Diggle <p.diggle@lancaster.ac.uk>

---

spat.corr.diagnostic *Diagnostics for residual spatial correlation*

---

## Description

This function performs two variogram-based tests for residual spatial correlation in real-valued and count (Binomial and Poisson) data.

## Usage

```
spat.corr.diagnostic(
  formula,
  units.m = NULL,
  coords,
  data,
  likelihood,
  ID.coords = NULL,
  n.sim = 200,
  nAGQ = 1,
  uvec = NULL,
  plot.results = TRUE,
  lse.variogram = FALSE,
  kappa = 0.5,
  which.test = "both"
)
```

## Arguments

| | |
|---|---|
| formula | an object of class formula (or one that can be coerced to that class): a symbolic description of the model to be fitted. |
| units.m | vector of binomial denominators, or offset if the Poisson model is used. |
| coords | an object of class formula indicating the geographic coordinates. |

| | |
|---|---|
| data | an object of class "data.frame" containing the data. |
| likelihood | a character that can be set to "Gaussian","Binomial" or "Poisson" |
| ID.coords | vector of ID values for the unique set of spatial coordinates obtained from `create.ID.coords`. These must be provided if, for example, spatial random effects are defined at household level but some of the covariates are at individual level. **Warning**: the household coordinates must all be distinct otherwise see `jitterDupCoords`. Default is `NULL`. |
| n.sim | number of simulations used to perform the selected test(s) for spatial correlation. |
| nAGQ | integer scalar (passed to `glmer`) - the number of points per axis for evaluating the adaptive Gauss-Hermite approximation to the log-likelihood. Defaults to 1, corresponding to the Laplace approximation. Values greater than 1 produce greater accuracy in the evaluation of the log-likelihood at the expense of speed. A value of zero uses a faster but less exact form of parameter estimation for GLMMs by optimizing the random effects and the fixed-effects coefficients in the penalized iteratively reweighted least squares step. |
| uvec | a vector with values used to define the variogram binning. If `uvec=NULL`, then `uvec` is then set to `seq(MIN_DIST,(MAX_DIST-MIN_DIST)/2,length=15)` where `MIN_DIST` and `MAX_DIST` are the minimum and maximum observed distances. |
| plot.results | if `plot.results=TRUE`, a plot is returned showing the results for the selected test(s) for spatial correlation. By default `plot.results=TRUE`. |
| lse.variogram | if `lse.variogram=TRUE`, a weighted least square fit of a Matern function (with fixed `kappa`) to the empirical variogram is performed. If `plot.results=TRUE` and `lse.variogram=TRUE`, the fitted weighted least square fit is displayed as a dashed line in the returned plot. |
| kappa | smothness parameter of the Matern function for the Gaussian process to approximate. The deafault is `kappa=0.5`. |
| which.test | a character specifying which test for residual spatial correlation is to be performed: "variogram", "test statistic" or "both". The default is `which.test="both"`. See 'Details'. |

## Details

The function first fits a generalized linear mixed model using the for an outcome $Y_i$ which, conditionally on i.i.d. random effects $Z_i$, are mutually independent GLMs with linear predictor

$$g^{-1}(\eta_i) = d_i'\beta + Z_i$$

where $d_i$ is a vector of covariates which are specified through `formula`. Finally, the $Z_i$ are assumed to be zero-mean Gaussian variables with variance $\sigma^2$

**Variogram-based graphical diagnostic**

This graphical diagnostic is performed by setting `which.test="both"` or `which.test="variogram"`. The output are 95 (see below `lower.lim` and `upper.lim`) that are generated under the assumption of spatial indepdence through the following steps

1. Fit a generalized linear mixed model as indicated by the equation above.
2. Obtain the mode, say $\hat{Z}_i$, of the $Z_i$ conditioned on the data $Y_i$.

3. Compute the empirical variogram using $\hat{Z}_i$
4. Permute the locations specified in `coords`, `n.sim` time while holding the $\hat{Z}_i$ fixed.
5. For each of the permuted data-sets compute the empirical variogram based on the $\hat{Z}_i$.
6. From the `n.sim` variograms obtained in the previous step, compute the 95

If the observed variogram (`obs.variogram` below), based on the un-permuted $\hat{Z}_i$, falls within the 95 residual spatial correlation; if, instead, that partly falls outside the 95

**Test for spatial independence**

This diagnostic test is performed if `which.test="both"` or `which.test="test statistic"`. Let $\hat{v}(B)$ denote the empirical variogram based on $\hat{Z}_i$ for the distance bin $B$. The test statistic used for testing residual spatial correlation is

$$T = \sum_B N(B)\{v(B) - \hat{\sigma}^2\}$$

where $N(B)$ is the number of pairs of data-points falling within the distance bin $B$ (`n.bins` below) and $\hat{\sigma}^2$ is the estimate of $\sigma^2$.

To obtain the distribution of the test statistic $T$ under the null hypothesis of spatial independence, we use the simulated empirical variograms as obtained in step 5 of the iterative procedure described in "Variogram-based graphical diagnostic." The p-value for the test of spatial independence is then computed by taking the proportion of simulated values for $T$ under the null the hypothesis that are larger than the value of $T$ based on the original (un-permuted) $\hat{Z}_i$

## Value

An object of class "PrevMap.diagnostic" which is a list containing the following components:

`obs.variogram`: a vector of length `length(uvec)-1` containing the values of the variogram for each of the distance bins defined through `uvec`.

`distance.bins`: a vector of length `length(uvec)-1` containing the average distance within each of the distance bins defined through `uvec`.

`n.bins`: a vector of length `length(uvec)-1` containing the number of pairs of data-points falling within each distance bin.

`lower.lim`: (available only if `which.test="both"` or `which.test="variogram"`) a vector of length `length(uvec)-1` containing the lower limits of the 95 generated under the assumption of absence of spatial correlation at each fo the distance bins defined through `uvec`.

`upper.lim`: (available only if `which.test="both"` or `which.test="variogram"`) a vector of length `length(uvec)-1` containing the upper limits of the 95 generated under the assumption of absence of spatial correlation at each fo the distance bins defined through `uvec`.

`mode.rand.effects`: the predictive mode of the random effects from the fitted non-spatial generalized linear mixed model.

`p.value`: (available only if `which.test="both"` or `which.test="test statistic"`) p-value of the test for residual spatial correlation.

`lse.variogram`: (available only if `lse.variogram=TRUE`) a vector of length `length(uvec)-1` containing the values of the estimated Matern variogram via a weighted least square fit.

spatial.pred.binomial.Bayes

*Bayesian spatial prediction for the binomial logistic and binary probit models*

## Description

This function performs Bayesian spatial prediction for the binomial logistic and binary probit models.

## Usage

```
spatial.pred.binomial.Bayes(
  object,
  grid.pred,
  predictors = NULL,
  type = "marginal",
  scale.predictions = "prevalence",
  quantiles = c(0.025, 0.975),
  standard.errors = FALSE,
  thresholds = NULL,
  scale.thresholds = NULL,
  messages = TRUE
)
```

## Arguments

`object` an object of class "Bayes.PrevMap" obtained as result of a call to `binomial.logistic.Bayes` or `binary.probit.Bayes`.

`grid.pred` a matrix of prediction locations.

`predictors` a data frame of the values of the explanatory variables at each of the locations in `grid.pred`; each column correspond to a variable and each row to a location. **Warning:** the names of the columns in the data frame must match those in the data used to fit the model. Default is `predictors=NULL` for models with only an intercept.

`type` a character indicating the type of spatial predictions: `type="marginal"` for marginal predictions or `type="joint"` for joint predictions. Default is `type="marginal"`. In the case of a low-rank approximation only joint predictions are available.

`scale.predictions` a character vector of maximum length 3, indicating the required scale on which spatial prediction is carried out: "logit", "prevalence", "odds" and "probit". Default is `scale.predictions="prevalence"`.

`quantiles` a vector of quantiles used to summarise the spatial predictions.

`standard.errors` logical; if `standard.errors=TRUE`, then standard errors for each `scale.predictions` are returned. Default is `standard.errors=FALSE`.

`thresholds` a vector of exceedance thresholds; default is `NULL`.

`scale.thresholds` a character value ("logit", "prevalence", "odds" or "probit") indicating the scale on which exceedance thresholds are provided.

`messages` logical; if `messages=TRUE` then status messages are printed on the screen (or output device) while the function is running. Default is `messages=TRUE`.

## Value

A "pred.PrevMap" object list with the following components: `logit`; `prevalence`; `odds`; `probit`;`exceedance.prob`, corresponding to a matrix of the exceedance probabilities where each column corresponds to a specified value in `thresholds`; `samples`, corresponding to a matrix of the posterior samples at each prediction locations for the linear predictor; `grid.pred` prediction locations. Each of the three components `logit`, `prevalence`, `odds` and `probit` is also a list with the following components:

`predictions`: a vector of the predictive mean for the associated quantity (logit, odds or prevalence).

`standard.errors`: a vector of prediction standard errors (if `standard.errors=TRUE`).

`quantiles`: a matrix of quantiles of the resulting predictions with each column corresponding to a quantile specified through the argument `quantiles`.

## Author(s)

Emanuele Giorgi <e.giorgi@lancaster.ac.uk>

Peter J. Diggle <p.diggle@lancaster.ac.uk>

---

`spatial.pred.binomial.MCML` *Spatial predictions for the binomial logistic model using plug-in of MCML estimates*

---

## Description

This function performs spatial prediction, fixing the model parameters at the Monte Carlo maximum likelihood estimates of a geostatistical binomial logistic model.

## Usage

```
spatial.pred.binomial.MCML(
  object,
  grid.pred,
  predictors = NULL,
  control.mcmc,
  type = "marginal",
  scale.predictions = c("logit", "prevalence", "odds"),
  quantiles = c(0.025, 0.975),
  standard.errors = FALSE,
  thresholds = NULL,
```

```
  scale.thresholds = NULL,
  plot.correlogram = FALSE,
  messages = TRUE
)
```

## Arguments

`object` an object of class "PrevMap" obtained as result of a call to `binomial.logistic.MCML`.

`grid.pred` a matrix of prediction locations.

`predictors` a data frame of the values of the explanatory variables at each of the locations in `grid.pred`; each column correspond to a variable and each row to a location. **Warning:** the names of the columns in the data frame must match those in the data used to fit the model. Default is `predictors=NULL` for models with only an intercept.

`control.mcmc` output from `control.mcmc.MCML`.

`type` a character indicating the type of spatial predictions: `type="marginal"` for marginal predictions or `type="joint"` for joint predictions. Default is `type="marginal"`. In the case of a low-rank approximation only joint predictions are available.

`scale.predictions` a character vector of maximum length 3, indicating the required scale on which spatial prediction is carried out: "logit", "prevalence" and "odds". Default is `scale.predictions=c("logit","prevalence","odds")`.

`quantiles` a vector of quantiles used to summarise the spatial predictions.

`standard.errors` logical; if `standard.errors=TRUE`, then standard errors for each `scale.predictions` are returned. Default is `standard.errors=FALSE`.

`thresholds` a vector of exceedance thresholds; default is `thresholds=NULL`.

`scale.thresholds` a character value indicating the scale on which exceedance thresholds are provided; `"logit"`, `"prevalence"` or `"odds"`. Default is `scale.thresholds=NULL`.

`plot.correlogram` logical; if `plot.correlogram=TRUE` the autocorrelation plot of the conditional simulations is displayed.

`messages` logical; if `messages=TRUE` then status messages are printed on the screen (or output device) while the function is running. Default is `messages=TRUE`.

## Value

A "pred.PrevMap" object list with the following components: `logit`; `prevalence`; `odds`; `exceedance.prob`, corresponding to a matrix of the exceedance probabilities where each column corresponds to a specified value in `thresholds`; `samples`, corresponding to a matrix of the predictive samples at each prediction locations for the linear predictor of the binomial logistic model (if `scale.predictions="logit"` and neither the SPDE nor the low-rank approximations have been used, this component is `NULL`); `grid.pred` prediction locations. Each of the three components `logit`, `prevalence` and `odds` is also a list with the following components:

`predictions`: a vector of the predictive mean for the associated quantity (logit, odds or prevalence).

standard.errors: a vector of prediction standard errors (if standard.errors=TRUE).

quantiles: a matrix of quantiles of the resulting predictions with each column corresponding to a quantile specified through the argument quantiles.

## Author(s)

Emanuele Giorgi <e.giorgi@lancaster.ac.uk>

Peter J. Diggle <p.diggle@lancaster.ac.uk>

---

spatial.pred.linear.Bayes

*Bayesian spatial predictions for the geostatistical Linear Gaussian model*

---

## Description

This function performs Bayesian prediction for a geostatistical linear Gaussian model.

## Usage

```
spatial.pred.linear.Bayes(
  object,
  grid.pred,
  predictors = NULL,
  type = "marginal",
  scale.predictions = c("logit", "prevalence", "odds"),
  quantiles = c(0.025, 0.975),
  standard.errors = FALSE,
  thresholds = NULL,
  scale.thresholds = NULL,
  messages = TRUE
)
```

## Arguments

| | |
|---|---|
| object | an object of class "Bayes.PrevMap" obtained as result of a call to linear.model.Bayes. |
| grid.pred | a matrix of prediction locations. |
| predictors | a data frame of the values of the explanatory variables at each of the locations in grid.pred; each column correspond to a variable and each row to a location. **Warning:** the names of the columns in the data frame must match those in the data used to fit the model. Default is predictors=NULL for models with only an intercept. |
| type | a character indicating the type of spatial predictions: type="marginal" for marginal predictions or type="joint" for joint predictions. Default is type="marginal". In the case of a low-rank approximation only joint predictions are available. |

scale.predictions
: a character vector of maximum length 3, indicating the required scale on which spatial prediction is carried out: "logit", "prevalence" and "odds". Default is `scale.predictions=c("logit","prevalence","odds")`.

quantiles
: a vector of quantiles used to summarise the spatial predictions.

standard.errors
: logical; if `standard.errors=TRUE`, then standard errors for each `scale.predictions` are returned. Default is `standard.errors=FALSE`.

thresholds
: a vector of exceedance thresholds; default is `thresholds=NULL`.

scale.thresholds
: a character value indicating the scale on which exceedance thresholds are provided: `"logit"`, `"prevalence"` or `"odds"`. Default is `scale.thresholds=NULL`.

messages
: logical; if `messages=TRUE` then status messages are printed on the screen (or output device) while the function is running. Default is `messages=TRUE`.

## Value

A "pred.PrevMap" object list with the following components: `logit`; `prevalence`; `odds`; `exceedance.prob`, corresponding to a matrix of the exceedance probabilities where each column corresponds to a specified value in `thresholds`; `grid.pred` prediction locations. Each of the three components `logit`, `prevalence` and `odds` is also a list with the following components:

`predictions`: a vector of the predictive mean for the associated quantity (logit, odds or prevalence).

`standard.errors`: a vector of prediction standard errors (if `standard.errors=TRUE`).

`quantiles`: a matrix of quantiles of the resulting predictions with each column corresponding to a quantile specified through the argument `quantiles`.

## Author(s)

Emanuele Giorgi <e.giorgi@lancaster.ac.uk>

Peter J. Diggle <p.diggle@lancaster.ac.uk>

---

# spatial.pred.linear.MLE

*Spatial predictions for the geostatistical Linear Gaussian model using plug-in of ML estimates*

---

## Description

This function performs spatial prediction, fixing the model parameters at the maximum likelihood estimates of a linear geostatistical model.

### Usage

```
spatial.pred.linear.MLE(
  object,
  grid.pred,
  predictors = NULL,
  predictors.samples = NULL,
  type = "marginal",
  scale.predictions = c("logit", "prevalence", "odds"),
  quantiles = c(0.025, 0.975),
  n.sim.prev = 0,
  standard.errors = FALSE,
  thresholds = NULL,
  scale.thresholds = NULL,
  messages = TRUE,
  include.nugget = FALSE
)
```

### Arguments

`object` an object of class "PrevMap" obtained as result of a call to `linear.model.MLE`.

`grid.pred` a matrix of prediction locations.

`predictors` a data frame of the values of the explanatory variables at each of the locations in `grid.pred`; each column correspond to a variable and each row to a location. **Warning:** the names of the columns in the data frame must match those in the data used to fit the model. Default is `predictors=NULL` for models with only an intercept.

`predictors.samples` a list of data frame objects. This argument is used to average over repeated simulations of the predictor variables in order to obtain an "average" map over the distribution of the explanatory variables in the model. Each component of the list is a simulation. The number of simulations passed through `predictors.samples` must be the same as `n.sim.prev`. NOTE: This argument can currently only be used only for a linear regression model that does not use any approximation of the spatial Gaussian process.

`type` a character indicating the type of spatial predictions: `type="marginal"` for marginal predictions or `type="joint"` for joint predictions. Default is `type="marginal"`. In the case of a low-rank approximation only marginal predictions are available.

`scale.predictions` a character vector of maximum length 3, indicating the required scale on which spatial prediction is carried out: "logit", "prevalence" and "odds". Default is `scale.predictions=c("logit","prevalence","odds")`.

`quantiles` a vector of quantiles used to summarise the spatial predictions.

`n.sim.prev` number of simulation for non-linear predictive targets. Default is `n.sim.prev=0`.

`standard.errors` logical; if `standard.errors=TRUE`, then standard errors for each `scale.predictions` are returned. Default is `standard.errors=FALSE`.

`thresholds` a vector of exceedance thresholds; default is `thresholds=NULL`.

`scale.thresholds` a character value indicating the scale on which exceedance thresholds are provided; `"logit"`, `"prevalence"` or `"odds"`. Default is `scale.thresholds=NULL`.

`messages` logical; if `messages=TRUE` then status messages are printed on the screen (or output device) while the function is running. Default is `messages=TRUE`.

`include.nugget` logical; if `include.nugget=TRUE` then the nugget effect is included in the predictions. This option is available only for fitted linear models with locations having multiple observations. Default is `include.nugget=FALSE`.

### Value

A "pred.PrevMap" object list with the following components: `logit`; `prevalence`; `odds`; `exceedance.prob`, corresponding to a matrix of the exceedance probabilities where each column corresponds to a specified value in `thresholds`; `grid.pred` prediction locations; `samples`, corresponding to the predictive samples of the linear predictor (only if `any(scale.predictions=="prevalence")`). Each of the three components `logit`, `prevalence` and `odds` is also a list with the following components:

`predictions`: a vector of the predictive mean for the associated quantity (logit, odds or prevalence).

`standard.errors`: a vector of prediction standard errors (if `standard.errors=TRUE`).

`quantiles`: a matrix of quantiles of the resulting predictions with each column corresponding to a quantile specified through the argument `quantiles`.

`samples`: If `n.sim.prev` > `0`, the function returns `n.sim.prev` samples of the linear predictor at each of the prediction locations.

### Author(s)

Emanuele Giorgi <`e.giorgi@lancaster.ac.uk`>

Peter J. Diggle <`p.diggle@lancaster.ac.uk`>

---

## `spatial.pred.lm.ps` *Spatial predictions for the geostatistical Linear Gaussian model using plug-in of ML estimates*

---

### Description

This function performs spatial prediction, fixing the model parameters at the maximum likelihood estimates of a linear geostatistical model.

### Usage

```
spatial.pred.lm.ps(
  object,
  grid.pred = NULL,
  predictors = NULL,
  predictors.intensity = NULL,
```

```
  control.mcmc = NULL,
  target = 3,
  type = "marginal",
  quantiles = NULL,
  standard.errors = FALSE,
  messages = TRUE,
  return.samples = FALSE
)
```

## Arguments

`object` an object of class "PrevMap" obtained as result of a call to `linear.model.MLE`.

`grid.pred` a matrix of prediction locations. Default is `grid.pred=NULL`, in which case the grid used to approximate the intractable integral in the log-Gaussian Cox process model is used for prediction.

`predictors` a data frame of the values of the explanatory variables at each of the locations in `grid.pred`, for the response variable model; each column correspond to a variable and each row to a location. **Warning:** the names of the columns in the data frame must match those in the data used to fit the model. Default is `predictors=NULL` for models with only an intercept.

`predictors.intensity` a data frame of the values of the explanatory variables at each of the locations in `grid.pred`, for the log-Gaussian Cox process model; each column correspond to a variable and each row to a location. **Warning:** the names of the columns in the data frame must match those in the data used to fit the model. Default is `predictors=NULL` for models with only an intercept.

`control.mcmc` output from `control.mcmc.MCML` which defined the control parameters of the Monte Carlo Markv chain algorithm.

`target` an integeter indicating the predictive target: `target=1` if the predictive target is the linear predictor of the response; `target=2` is the predictive target is the sampling intensity of the preferentially sampled data; `target=3` if both of the above are the predictive targets. Default is `target=3`.

`type` a character indicating the type of spatial predictions for `target=1`: `type="marginal"` for marginal predictions or `type="joint"` for joint predictions. Default is `type="marginal"`. Note that predictions for the sampling intensity (`target=2`) are always joint.

`quantiles` a vector of quantiles used to summarise the spatial predictions.

`standard.errors` logical; if `standard.errors=TRUE`, then standard errors for each `scale.predictions` are returned. Default is `standard.errors=FALSE`.

`messages` logical; if `messages=TRUE` then status messages are printed on the screen (or output device) while the function is running. Default is `messages=TRUE`.

`return.samples` logical; if `return.samples=TRUE` a matrix of the predictive samples for the prediction target (as specified in `target`) are returned in the output.

## Value

A "pred.PrevMap.ps" object list with the following components: `response` (if `target=1` or `target=3`) and `intensity` (if `target=2` pr `target=3`). `grid.pred` prediction locations. Each of the components `intensity` and `response` is a list with the following components:

`predictions`: a vector of the predictive mean for the corresponding target.

`standard.errors`: a vector of prediction standard errors (if `standard.errors=TRUE`).

`quantiles`: a matrix of quantiles of the resulting predictions with each column corresponding to a quantile specified through the argument `quantiles`.

`samples`: a matrix corresponding to the predictive samples of the predictive target (only if `return.samples=TRUE`), with each row corresponding to a samples and column to a prediction location. In the case of a model with a mix of preferential and non-preferential data, if `target=1` or `target=3`, each of the above components will be a list with two components, namely `preferential` and `non.preferential`, associated with `response`.

## Author(s)

Emanuele Giorgi <e.giorgi@lancaster.ac.uk>

---

`spatial.pred.poisson.MCML` *Spatial predictions for the Poisson model with log link function, using plug-in of MCML estimates*

---

## Description

This function performs spatial prediction, fixing the model parameters at the Monte Carlo maximum likelihood estimates of a geostatistical Poisson model with log link function.

## Usage

```
spatial.pred.poisson.MCML(
  object,
  grid.pred,
  predictors = NULL,
  control.mcmc,
  type = "marginal",
  scale.predictions = c("log", "exponential"),
  quantiles = c(0.025, 0.975),
  standard.errors = FALSE,
  thresholds = NULL,
  scale.thresholds = NULL,
  plot.correlogram = FALSE,
  messages = TRUE
)
```

## Arguments

| | |
|---|---|
| object | an object of class "PrevMap" obtained as result of a call to poisson.log.MCML. |
| grid.pred | a matrix of prediction locations. |
| predictors | a data frame of the values of the explanatory variables at each of the locations in grid.pred; each column correspond to a variable and each row to a location. **Warning:** the names of the columns in the data frame must match those in the data used to fit the model. Default is predictors=NULL for models with only an intercept. |
| control.mcmc | output from control.mcmc.MCML. |
| type | a character indicating the type of spatial predictions: type="marginal" for marginal predictions or type="joint" for joint predictions. Default is type="marginal". In the case of a low-rank approximation only joint predictions are available. |
| scale.predictions | a character vector of maximum length 2, indicating the required scale on which spatial prediction is carried out: "log" and "exponential". Default is scale.predictions=c("log","expo |
| quantiles | a vector of quantiles used to summarise the spatial predictions. |
| standard.errors | logical; if standard.errors=TRUE, then standard errors for each scale.predictions are returned. Default is standard.errors=FALSE. |
| thresholds | a vector of exceedance thresholds; default is thresholds=NULL. |
| scale.thresholds | a character value indicating the scale on which exceedance thresholds are provided; "log" or "exponential". Default is scale.thresholds=NULL. |
| plot.correlogram | logical; if plot.correlogram=TRUE the autocorrelation plot of the conditional simulations is displayed. |
| messages | logical; if messages=TRUE then status messages are printed on the screen (or output device) while the function is running. Default is messages=TRUE. |

## Value

A "pred.PrevMap" object list with the following components: log; exponential; exceedance.prob, corresponding to a matrix of the exceedance probabilities where each column corresponds to a specified value in thresholds; samples, corresponding to a matrix of the predictive samples at each prediction locations for the linear predictor of the Poisson model (if scale.predictions="log" this component is NULL); grid.pred prediction locations. Each of the three components log and exponential is also a list with the following components:

predictions: a vector of the predictive mean for the associated quantity (log or exponential).

standard.errors: a vector of prediction standard errors (if standard.errors=TRUE).

quantiles: a matrix of quantiles of the resulting predictions with each column corresponding to a quantile specified through the argument quantiles.

### Author(s)

Emanuele Giorgi <e.giorgi@lancaster.ac.uk>

Peter J. Diggle <p.diggle@lancaster.ac.uk>

---

## summary.Bayes.PrevMap *Summarizing Bayesian model fits*

---

### Description

summary method for the class "Bayes.PrevMap" that computes the posterior mean, median, mode and high posterior density intervals using samples from Bayesian fits.

### Usage

```
## S3 method for class 'Bayes.PrevMap'
summary(object, hpd.coverage = 0.95, ...)
```

### Arguments

| | |
|---|---|
| object | an object of class "Bayes.PrevMap" obatained as result of a call to binomial.logistic.Bayes or linear.model.Bayes. |
| hpd.coverage | value of the coverage of the high posterior density intervals; default is 0.95. |
| ... | further arguments passed to or from other methods. |

### Value

A list with the following values

linear: logical value that is TRUE if a linear model was fitted and FALSE otherwise.

binary: logical value that is TRUE if a binary model was fitted and FALSE otherwise.

probit: logical value that is TRUE if a binary model with probit link function was fitted and FALSE if with logistic link function.

ck: logical value that is TRUE if a low-rank approximation was fitted and FALSE otherwise.

beta: matrix of the posterior summaries for the regression coefficients.

sigma2: vector of the posterior summaries for sigma2.

phi: vector of the posterior summaries for phi.

tau2: vector of the posterior summaries for tau2.

call: matched call.

kappa: fixed value of the shape paramter of the Matern covariance function.

### Author(s)

Emanuele Giorgi <e.giorgi@lancaster.ac.uk>

Peter J. Diggle <p.diggle@lancaster.ac.uk>

summary.PrevMap *Summarizing likelihood-based model fits*

## Description

summary method for the class "PrevMap" that computes the standard errors and p-values of likelihood-based model fits.

## Usage

```
## S3 method for class 'PrevMap'
summary(object, log.cov.pars = TRUE, ...)
```

## Arguments

| | |
|---|---|
| object | an object of class "PrevMap" obatained as result of a call to binomial.logistic.MCML or linear.model.MLE. |
| log.cov.pars | logical; if log.cov.pars=TRUE the estimates of the covariance parameters are given on the log-scale. Note that standard errors are also adjusted accordingly. Default is log.cov.pars=TRUE. |
| ... | further arguments passed to or from other methods. |

## Value

A list with the following components

linear: logical value; linear=TRUE if a linear model was fitted and linear=FALSE otherwise.

poisson: logical value; poisson=TRUE if a Poisson model was fitted and poisson=FALSE otherwise.

ck: logical value; ck=TRUE if a low-rank approximation was used and ck=FALSE otherwise.

spde: logical value; spde=TRUE if the SPDE approximation was used and spde=FALSE otherwise.

coefficients: matrix of the estimates, standard errors and p-values of the estimates of the regression coefficients.

cov.pars: matrix of the estimates and standard errors of the covariance parameters.

log.lik: value of likelihood function at the maximum likelihood estimates.

kappa: fixed value of the shape paramter of the Matern covariance function.

kappa.t: fixed value of the shape paramter of the Matern covariance function for the temporal covariance matrix, if a spatio-temporal model has been fitted.

fixed.rel.nugget: fixed value for the relative variance of the nugget effect.

call: matched call.

## Author(s)

Emanuele Giorgi <e.giorgi@lancaster.ac.uk>

Peter J. Diggle <p.diggle@lancaster.ac.uk>

summary.PrevMap.ps *Summarizing fits of geostatistical linear models with preferentially sampled locations*

## Description

summary method for the class "PrevMap" that computes the standard errors and p-values of likelihood-based model fits.

## Usage

```
## S3 method for class 'PrevMap.ps'
summary(object, log.cov.pars = TRUE, ...)
```

## Arguments

| | |
|---|---|
| object | an object of class "PrevMap.ps" obatained as result of a call to lm.ps.MCML. |
| log.cov.pars | logical; if log.cov.pars=TRUE the estimates of the covariance parameters are given on the log-scale. Note that standard errors are also adjusted accordingly. Default is log.cov.pars=TRUE. |
| ... | further arguments passed to or from other methods. |

## Value

A list with the following components

coefficients.response: matrix of the estimates, standard errors and p-values of the estimates of the regression coefficients for the response variable.

coefficients.intensity: matrix of the estimates, standard errors and p-values of the estimates of the regression coefficients for the sampling intenisty of the log-Gaussian process.

cov.pars.response: matrix of the estimates and standard errors of the covariance parameters for the Gaussian process associated with the response.

cov.pars.intenisty: matrix of the estimates and standard errors of the covariance parameters for the Gaussian process associated with the log-Gaussian process.

log.lik: value of likelihood function at the maximum likelihood estimates.

kappa.response: fixed value of the shape paramter of the Matern covariance function.

call: matched call.

## Author(s)

Emanuele Giorgi <e.giorgi@lancaster.ac.uk>

`trace.plot` *Trace-plots for posterior samples*

## Description

Displays the trace-plots for the posterior samples of the model parameters and spatial random effects.

## Usage

```
trace.plot(object, param, component.beta = NULL, component.S = NULL)
```

## Arguments

| | |
|---|---|
| `object` | an object of class 'Bayes.PrevMap'. |
| `param` | a character indicating for which component of the model the density plot is required: `param="beta"` for the regression coefficients; `param="sigma2"` for the variance of the spatial random effect; `param="phi"` for the scale parameter of the Matern correlation function; `param="tau2"` for the variance of the nugget effect; `param="S"` for the spatial random effect. |
| `component.beta` | if `param="beta"`, `component.beta` is a numeric value indicating the component of the regression coefficients; default is `NULL`. |
| `component.S` | if `param="S"`, `component.S` can be a numeric value indicating the component of the spatial random effect. Default is `NULL`. |

## Author(s)

Emanuele Giorgi <`e.giorgi@lancaster.ac.uk`>

Peter J. Diggle <`p.diggle@lancaster.ac.uk`>

`trace.plot.MCML` *Trace-plots of the importance sampling distribution samples from the MCML method*

## Description

Trace-plots of the MCMC samples from the importance sampling distribution used in `binomial.logistic.MCML`.

## Usage

```
trace.plot.MCML(object, component = NULL, ...)
```

### Arguments

| | |
|---|---|
| `object` | an object of class "PrevMap" obatained as result of a call to `binomial.logistic.MCML`. |
| `component` | a positive integer indicating the number of the random effect component for which a trace-plot is required. If `component=NULL`, then a component is selected at random. Default is `component=NULL`. |
| `...` | further arguments passed to `plot`. |

### Author(s)

Emanuele Giorgi <e.giorgi@lancaster.ac.uk>

Peter J. Diggle <p.diggle@lancaster.ac.uk>

---

## `trend.plot` *Plot of trends*

---

### Description

This function produces a plot of the variable of interest against each of the two geographical coordinates.

### Usage

```
trend.plot(data, var.name, coords, ...)
```

### Arguments

| | |
|---|---|
| `data` | an object of class "data.frame" containing the data. |
| `var.name` | a `formula` object indicating the variable to display. |
| `coords` | a `formula` object indicating the geographical coordinates. |
| `...` | additional arguments to be passed to `plot`. |

---

## `variog.diagnostic.glgm` *Variogram-based validation for generalized linear geostatistical model fits (Binomial and Poisson)*

---

### Description

This function performs model validation for generalized linear geostatistical models (Binomial and Poisson) using Monte Carlo methods based on the variogram.

### Usage

```
variog.diagnostic.glgm(
  object,
  n.sim = 200,
  uvec = NULL,
  plot.results = TRUE,
  which.test = "both"
)
```

### Arguments

| | |
|---|---|
| `object` | an object of class "PrevMap" obtained as an output from `binomial.logistic.MCML` and `poisson.log.MCML`. |
| `n.sim` | integer indicating the number of simulations used for the variogram-based diagnostics. Defeault is `n.sim=1000`. |
| `uvec` | a vector with values used to define the variogram binning. If `uvec=NULL`, then `uvec` is then set to `seq(MIN_DIST,(MAX_DIST-MIN_DIST)/2,length=15)` |
| `plot.results` | if `plot.results=TRUE`, a plot is returned showing the results for the selected test(s) for spatial correlation. By default `plot.results=TRUE`. defined as the distance at which the fitted spatial correlation is no less than 0.05. Default is `range.fact=1` |
| `which.test` | a character specifying which test for residual spatial correlation is to be performed: "variogram", "test statistic" or "both". The default is `which.test="both"`. See 'Details.' |

### Details

The function takes as an input through the argument `object` a fitted generalized linear geostaistical model for an outcome $Y_i$, with linear predictor

$$\eta_i = d_i'\beta + S(x_i) + Z_i$$

where $d_i$ is a vector of covariates which are specified through `formula`, $S(x_i)$ is a spatial Gaussian process and the $Z_i$ are assumed to be zero-mean Gaussian. The model validation is performed on the adopted satationary and isotropic Matern covariance function used for $S(x_i)$. More specifically, the function allows the users to select either of the following validation procedures.

**Variogram-based graphical validation**

This graphical diagnostic is performed by setting `which.test="both"` or `which.test="variogram"`. The output are 95 (see below `lower.lim` and `upper.lim`) that are generated under the assumption that the fitted model did generate the analysed data-set. This validation procedure proceed through the following steps.

1. Obtain the mean, say $\hat{Z}_i$, of the $Z_i$ conditioned on the data $Y_i$ and by setting $S(x_i) = 0$ in the equation above.

2. Compute the empirical variogram using $\hat{Z}_i$

3. Simulate `n.sim` data-sets under the fitted geostatistical model.

4. For each of the simulated data-sets and obtain $\hat{Z}_i$ as in Step 1. Finally, compute the empirical variogram based on the resulting $\hat{Z}_i$.

5. From the `n.sim` variograms obtained in the previous step, compute the 95

If the observed variogram (`obs.variogram` below), based on the $\hat{Z}_i$ from Step 2, falls within the 95 evidence against the fitted spatial correlation model; if, instead, that partly falls outside the 95 correlation in the data.

**Test for suitability of the adopted correlation function**

This diagnostic test is performed if `which.test="both"` or `which.test="test statistic"`. Let $v_E(B)$ and $v_T(B)$ denote the empirical and theoretical variograms based on $\hat{Z}_i$ for the distance bin $B$. The test statistic used for testing residual spatial correlation is

$$T = \sum_B N(B)\{v_E(B) - v_T(B)\}$$

where $N(B)$ is the number of pairs of data-points falling within the distance bin $B$ (`n.bins` below).

To obtain the distribution of the test statistic $T$ under the null hypothesis that the fitted model did generate the analysed data-set, we use the simulated empirical variograms as obtained in step 5 of the iterative procedure described in "Variogram-based graphical validation." The p-value for the test of suitability of the fitted spatial correlation function is then computed by taking the proportion of simulated values for $T$ that are larger than the value of $T$ based on the original $\hat{Z}_i$ in Step 1.

## Value

An object of class "PrevMap.diagnostic" which is a list containing the following components:

`obs.variogram`: a vector of length `length(uvec)-1` containing the values of the variogram for each of the distance bins defined through `uvec`.

`distance.bins`: a vector of length `length(uvec)-1` containing the average distance within each of the distance bins defined through `uvec`.

`n.bins`: a vector of length `length(uvec)-1` containing the number of pairs of data-points falling within each distance bin.

`lower.lim`: (available only if `which.test="both"` or `which.test="variogram"`) a vector of length `length(uvec)-1` containing the lower limits of the 95 generated under the assumption of absence of suitability of the fitted model at each fo the distance bins defined through `uvec`.

`upper.lim`: (available only if `which.test="both"` or `which.test="variogram"`) a vector of length `length(uvec)-1` containing the upper limits of the 95 generated under the assumption of absence of suitability of the fitted model at each fo the distance bins defined through `uvec`.

`mode.rand.effects`: the predictive mode of the random effects from the fitted non-spatial generalized linear mixed model.

`p.value`: (available only if `which.test="both"` or `which.test="test statistic"`) p-value of the test for residual spatial correlation.

`lse.variogram`: (available only if `lse.variogram=TRUE`) a vector of length `length(uvec)-1` containing the values of the estimated Matern variogram via a weighted least square fit.

variog.diagnostic.lm *Variogram-based validation for linear geostatistical model fits*

## Description

This function performs model validation for linear geostatistical model using Monte Carlo methods based on the variogram.

## Usage

```
variog.diagnostic.lm(
  object,
  n.sim = 1000,
  uvec = NULL,
  plot.results = TRUE,
  range.fact = 1,
  which.test = "both",
  param.uncertainty = FALSE
)
```

## Arguments

| | |
|---|---|
| `object` | an object of class "PrevMap" obtained as an output from `linear.model.MLE`. |
| `n.sim` | integer indicating the number of simulations used for the variogram-based diagnostics. Defeault is `n.sim=1000`. |
| `uvec` | a vector with values used to define the variogram binning. If `uvec=NULL`, then `uvec` is then set to `seq(MIN_DIST,(MAX_DIST-MIN_DIST)/2,length=15)` |
| `plot.results` | if `plot.results=TRUE`, a plot is returned showing the results for the selected test(s) for spatial correlation. By default `plot.results=TRUE`. |
| `range.fact` | a value between 0 and 1 used to disregard all distance bins provided through `uvec` that are larger than the (pr)x`range.fact`, where pr is the practical range, defined as the distance at which the fitted spatial correlation is no less than 0.05. Default is `range.fact=1` |
| `which.test` | a character specifying which test for residual spatial correlation is to be performed: "variogram", "test statistic" or "both". The default is `which.test="both"`. See 'Details.' |
| `param.uncertainty` | a logical indicating whether uncertainty in the model parameters should be incorporated in the selected diagnostic tests. Default is `param.uncertainty=FALSE`. See 'Details.' |

## Details

The function takes as an input through the argument `object` a fitted linear geostaistical model for an outcome $Y_i$, which is expressed as

$$Y_i = d_i'\beta + S(x_i) + Z_i$$

where $d_i$ is a vector of covariates which are specified through formula, $S(x_i)$ is a spatial Gaussian process and the $Z_i$ are assumed to be zero-mean Gaussian. The model validation is performed on the adopted satationary and isotropic Matern covariance function used for $S(x_i)$. More specifically, the function allows the users to select either of the following validation procedures.

**Variogram-based graphical validation**

This graphical diagnostic is performed by setting which.test="both" or which.test="variogram". The output are 95 (see below lower.lim and upper.lim) that are generated under the assumption that the fitted model did generate the analysed data-set. This validation procedure proceed through the following steps.

1. Obtain the mean, say $\hat{Z}_i$, of the $Z_i$ conditioned on the data $Y_i$.

2. Compute the empirical variogram using $\hat{Z}_i$

3. Simulate n.sim data-sets under the fitted geostatistical model.

4. For each of the simulated data-sets and obtain $\hat{Z}_i$ as in Step 1. Finally, compute the empirical variogram based on the resulting $\hat{Z}_i$.

5. From the n.sim variograms obtained in the previous step, compute the 95

If the observed variogram (obs.variogram below), based on the $\hat{Z}_i$ from Step 2, falls within the 95 evidence against the fitted spatial correlation model; if, instead, that partly falls outside the 95 correlation in the data.

**Test for suitability of the adopted correlation function**

This diagnostic test is performed if which.test="both" or which.test="test statistic". Let $v_E(B)$ and $v_T(B)$ denote the empirical and theoretical variograms based on $\hat{Z}_i$ for the distance bin $B$. The test statistic used for testing residual spatial correlation is

$$T = \sum_B N(B)\{v_E(B) - v_T(B)\}$$

where $N(B)$ is the number of pairs of data-points falling within the distance bin $B$ (n.bins below).

To obtain the distribution of the test statistic $T$ under the null hypothesis that the fitted model did generate the analysed data-set, we use the simulated empirical variograms as obtained in step 5 of the iterative procedure described in "Variogram-based graphical validation." The p-value for the test of suitability of the fitted spatial correlation function is then computed by taking the proportion of simulated values for $T$ that are larger than the value of $T$ based on the original $\hat{Z}_i$ in Step 1.

## Value

An object of class "PrevMap.diagnostic" which is a list containing the following components:

obs.variogram: a vector of length length(uvec)-1 containing the values of the variogram for each of the distance bins defined through uvec.

distance.bins: a vector of length length(uvec)-1 containing the average distance within each of the distance bins defined through uvec.

n.bins: a vector of length length(uvec)-1 containing the number of pairs of data-points falling within each distance bin.

lower.lim: (available only if which.test="both" or which.test="variogram") a vector of length length(uvec)-1 containing the lower limits of the 95 generated under the assumption of absence of suitability of the fitted model at each fo the distance bins defined through uvec.

upper.lim: (available only if which.test="both" or which.test="variogram") a vector of length length(uvec)-1 containing the upper limits of the 95 generated under the assumption of absence of suitability of the fitted model at each fo the distance bins defined through uvec.

mode.rand.effects: the predictive mode of the random effects from the fitted non-spatial generalized linear mixed model.

p.value: (available only if which.test="both" or which.test="test statistic") p-value of the test for residual spatial correlation.

lse.variogram: (available only if lse.variogram=TRUE) a vector of length length(uvec)-1 containing the values of the estimated Matern variogram via a weighted least square fit.

---

variogram *The empirical variogram*

---

## Description

This function computes sample (empirical) variograms with options for the classical or robust estimators. Output can be returned as a binned variogram, a variogram cloud or a smoothed variogram. Data transformation (Box-Cox) is allowed. "Trends" can be specified and are fitted by ordinary least squares in which case the variograms are computed using the residuals.

## Usage

```
variogram(data, var.name, coords, ...)
```

## Arguments

| | |
|---|---|
| data | an object of class "data.frame" containing the data. |
| var.name | a formula object indicating the variable to display. |
| coords | a formula object indicating the geographical coordinates. |
| ... | additional arguments to be passed to variog. |

## Value

An object of the class "variogram" which is list containing components as detailed in variog.

# Index