Package 'RiskMap'

January 20, 2025

Type Package

Title Geo-Statistical Modeling of Spatially Referenced Data

Version 0.1.0

Description Provides functions for geo-statistical analysis of both continuous and count data using maximum likelihood methods. The models implemented in the package use stationary Gaussian processes with Matern correlation function to carry out spatial prediction in a geographical area of interest. The underpinning theory of the methods implemented in the package are found in Diggle and Giorgi (2019, ISBN: 978-1-138-06102-7).

License MIT + file LICENSE

Encoding UTF-8

LazyData true

RoxygenNote 7.3.1

URL https://claudiofronterre.github.io/RiskMap/

Imports sf, stats, methods, ggplot2, Matrix, maxLik, terra, xtable

Depends R (>= 3.5.0)

NeedsCompilation no

Author Emanuele Giorgi [aut, cre] (<https://orcid.org/0000-0003-0640-181X>), Claudio Fronterre [ctb] (<https://orcid.org/0000-0001-6723-9727>)

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

Repository CRAN

Date/Publication 2024-06-25 14:40:05 UTC

Contents

`anopheles` *Anopheles mosquitoes in Southern Cameroon*

Description

These data contain 116 georeferenced locations on the counts of Anopheles gambiae and Anopheles coluzzii in Southern Cameroon.

- web_x x-coordinate of the spatial locations.
- web_y y-coordinate of the spatial locations.
- Locality: name of the place of the sampled location.
- An.coluzzii: counts of Anopheles coluzzi.
- An.gambiae: counts of Anopheles gambiae.
- Total: total counts of Anopheles coluzzi and Anopheles gambiae.
- elevation: elevation in meters of the sampled location.

The coordinate reference system is 3857.

Usage

```
data(anopheles)
```

Format

A data frame with 116 rows and 7 variables

Source

Tene Fossog, B., Ayala, D., Acevedo, P., Kengne, P., Ngomo Abeso Mebuy, I., Makanga, B., et al. (2015) Habitat segregation and ecological character displacement in cryptic African malaria mosquitoes. Evolutionary Applications, 8 (4), 326-345.

---

coef.RiskMap *Extract Parameter Estimates from a "RiskMap" Model Fit*

Description

This `coef` method for the "RiskMap" class extracts the maximum likelihood estimates from model fits obtained from the `glgpm` function.

Usage

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

Arguments

`object`	An object of class "RiskMap" obtained as a result of a call to `glgpm`.
`...`	other parameters.

Details

The function processes the `RiskMap` object to extract and name the estimated parameters appropriately, transforming them if necessary.

Value

A list containing the maximum likelihood estimates:

`beta`	A vector of coefficient estimates.
`sigma2`	The estimate for the variance parameter σ^2.
`phi`	The estimate for the spatial range parameter ϕ.
`tau2`	The estimate for the nugget effect parameter τ^2, if applicable.
`sigma2_me`	The estimate for the measurement error variance σ^2_{me}, if applicable.
`sigma2_re`	A vector of variance estimates for the random effects, if applicable.

Note

This function handles both Gaussian and non-Gaussian families, and accounts for fixed and random effects in the model.

Author(s)

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

Claudio Fronterre <c.fronterr@lancaster.ac.uk>

See Also

glgpm

---

create_grid *Create Grid of Points Within Shapefile*

Description

Generates a grid of points within a given shapefile. The grid points are created based on a specified spatial resolution.

Usage

```
create_grid(shp, spat_res, grid_crs = NULL)
```

Arguments

shp	An object of class 'sf' representing the shapefile within which the grid of points will be created.
spat_res	Numeric value specifying the spatial resolution in kilometers for the grid.
grid_crs	Coordinate reference system for the grid. If NULL, the CRS of 'shp' is used. The shapefile 'shp' will be transformed to this CRS if specified.

Details

This function creates a grid of points within the boundaries of the provided shapefile ('shp'). The grid points are generated using the specified spatial resolution ('spat_res'). If a coordinate reference system ('grid_crs') is provided, the shapefile is transformed to this CRS before creating the grid.

Value

An 'sf' object containing the generated grid points within the shapefile.

Author(s)

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

Claudio Fronterre <c.fronterr@lancaster.ac.uk>

See Also

st_make_grid, st_intersects, st_transform, st_crs

Examples

```
library(sf)

# Example shapefile data
nc <- st_read(system.file("shape/nc.shp", package="sf"))

# Create grid with 10 km spatial resolution
grid <- create_grid(nc, spat_res = 10)

# Plot the grid
plot(st_geometry(nc))
plot(grid, add = TRUE, col = 'red')
```

---

dist_summaries *Summaries of the distances*

Description

Computes the distances between the locations in the data-set and returns summary statistics of these.

Usage

```
dist_summaries(data, convert_to_utm = TRUE, scale_to_km = FALSE)
```

Arguments

data	an object of class sf containing the variable for which the variogram is to be computed and the coordinates
convert_to_utm	a logical value, indicating if the conversion to UTM shuold be performed (convert_to_utm = TRUE) or the coordinate reference system of the data must be used without any conversion (convert_to_utm = FALSE). By default convert_to_utm = TRUE. Note: if convert_to_utm = TRUE the conversion to UTM is performed using the epsg provided by propose_utm.
scale_to_km	a logical value, indicating if the distances used in the variogram must be scaled to kilometers (scale_to_km = TRUE) or left in meters (scale_to_km = FALSE). By default scale_to_km = FALSE

Value

a list containing the following components

`min` the minimum distance

`max` the maximum distance

`mean` the mean distance

`median` the minimum distance

`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 July 2000. The variables are as follows:

- x x-coordinate of the spatial locations.
- y y-coordinate of the spatial locations.
- lead number of tested people for the presence nodules.

The coordinate reference system of the data is 32629.

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.

glgpm *Estimation of Generalized Linear Gaussian Process Models*

Description

Fits generalized linear Gaussian process models to spatial data, incorporating spatial Gaussian processes with a Matern correlation function. Supports Gaussian, binomial, and Poisson response families.

Usage

```
glgpm(
  formula,
  data,
  family,
  distr_offset = NULL,
  cov_offset = NULL,
  crs = NULL,
  convert_to_crs = NULL,
  scale_to_km = TRUE,
  control_mcmc = set_control_sim(),
  par0 = NULL,
  S_samples = NULL,
  return_samples = TRUE,
  messages = TRUE,
  fix_var_me = NULL,
 start_pars = list(beta = NULL, sigma2 = NULL, tau2 = NULL, phi = NULL, sigma2_me =
    NULL, sigma2_re = NULL)
)
```

Arguments

formula	A formula object specifying the model to be fitted. The formula should include fixed effects, random effects (specified using `re()`), and spatial effects (specified using `gp()`).
data	A data frame or sf object containing the variables in the model.
family	A character string specifying the distribution of the response variable. Must be one of "gaussian", "binomial", or "poisson".
distr_offset	Optional offset for binomial or Poisson distributions. If not provided, defaults to 1 for binomial.
cov_offset	Optional numeric vector for covariate offset.
crs	Optional integer specifying the Coordinate Reference System (CRS) if data is not an sf object. Defaults to 4326 (long/lat).
convert_to_crs	Optional integer specifying a CRS to convert the spatial coordinates.
scale_to_km	Logical indicating whether to scale coordinates to kilometers. Defaults to TRUE.

control_mcmc	Control parameters for MCMC sampling. Must be an object of class "mcmc.RiskMap" as returned by set_control_sim.
par0	Optional list of initial parameter values for the MCMC algorithm.
S_samples	Optional matrix of pre-specified sample paths for the spatial random effect.
return_samples	Logical indicating whether to return MCMC samples when fitting a Binomial or Poisson model. Defaults to FALSE.
messages	Logical indicating whether to print progress messages. Defaults to TRUE.
fix_var_me	Optional fixed value for the measurement error variance.
start_pars	Optional list of starting values for model parameters: beta (regression coefficients), sigma2 (spatial process variance), tau2 (nugget effect variance), phi (spatial correlation scale), sigma2_me (measurement error variance), and sigma2_re (random effects variances).

Details

Generalized linear Gaussian process models extend generalized linear models (GLMs) by incorporating spatial Gaussian processes to account for spatial correlation in the data. This function fits GLGPMs using maximum likelihood methods, allowing for Gaussian, binomial, and Poisson response families. In the case of the Binomial and Poisson families, a Monte Carlo maximum likelihood algorithm is used.

The spatial Gaussian process is modeled with a Matern correlation function, which is flexible and commonly used in geostatistical modeling. The function supports both spatial covariates and unstructured random effects, providing a comprehensive framework to analyze spatially correlated data across different response distributions.

Additionally, the function allows for the inclusion of unstructured random effects, specified through the re() term in the model formula. These random effects can capture unexplained variability at specific locations beyond the fixed and spatial covariate effects, enhancing the model's flexibility in capturing complex spatial patterns.

The convert_to_crs argument can be used to reproject the spatial coordinates to a different CRS. The scale_to_km argument scales the coordinates to kilometers if set to TRUE.

The control_mcmc argument specifies the control parameters for MCMC sampling. This argument must be an object returned by set_control_sim.

The start_pars argument allows for specifying starting values for the model parameters. If not provided, default starting values are used.

Value

An object of class "RiskMap" containing the fitted model and relevant information:

y	Response variable.
D	Covariate matrix.
coords	Unique spatial coordinates.
ID_coords	Index of coordinates.
re	Random effects.

ID_re	Index of random effects.
fix_tau2	Fixed nugget effect variance.
fix_var_me	Fixed measurement error variance.
formula	Model formula.
family	Response family.
crs	Coordinate Reference System.
scale_to_km	Indicator if coordinates are scaled to kilometers.
data_sf	Original data as an sf object.
kappa	Spatial correlation parameter.
units_m	Distribution offset for binomial/Poisson.
cov_offset	Covariate offset.
call	Matched call.

Author(s)

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

Claudio Fronterre <c.fronterr@lancaster.ac.uk>

See Also

set_control_sim, summary.RiskMap, to_table

---

glgpm_sim *Simulation from Generalized Linear Gaussian Process Models*

Description

Simulates data from a fitted Generalized Linear Gaussian Process Model (GLGPM) or a specified model formula and data.

Usage

```
glgpm_sim(
  n_sim,
  model_fit = NULL,
  formula = NULL,
  data = NULL,
  family = NULL,
  distr_offset = NULL,
  cov_offset = NULL,
  crs = NULL,
  convert_to_crs = NULL,
  scale_to_km = TRUE,
  control_mcmc = NULL,
```

```
 sim_pars = list(beta = NULL, sigma2 = NULL, tau2 = NULL, phi = NULL, sigma2_me = NULL,
    sigma2_re = NULL),
 messages = TRUE
)
```

Arguments

n_sim	Number of simulations to perform.
model_fit	Fitted GLGPM model object of class 'RiskMap'. If provided, overrides 'formula', 'data', 'family', 'crs', 'convert_to_crs', 'scale_to_km', and 'control_mcmc' arguments.
formula	Model formula indicating the variables of the model to be simulated.
data	Data frame or 'sf' object containing the variables in the model formula.
family	Distribution family for the response variable. Must be one of 'gaussian', 'binomial', or 'poisson'.
distr_offset	Offset for the distributional part of the GLGPM. Required for 'binomial' and 'poisson' families.
cov_offset	Offset for the covariate part of the GLGPM.
crs	Coordinate reference system (CRS) code for spatial data.
convert_to_crs	CRS code to convert spatial data if different from 'crs'.
scale_to_km	Logical; if TRUE, distances between locations are computed in kilometers; if FALSE, in meters.
control_mcmc	Control parameters for MCMC simulation if applicable.
sim_pars	List of simulation parameters including 'beta', 'sigma2', 'tau2', 'phi', 'sigma2_me', and 'sigma2_re'.
messages	Logical; if TRUE, display progress and informative messages.

Details

Generalized Linear Gaussian Process Models (GLGPMs) extend generalized linear models (GLMs) by incorporating spatial Gaussian processes to model spatial correlation. This function simulates data from GLGPMs using Markov Chain Monte Carlo (MCMC) methods. It supports Gaussian, binomial, and Poisson response families, utilizing a Matern correlation function to model spatial dependence.

The simulation process involves generating spatially correlated random effects and simulating responses based on the fitted or specified model parameters. For 'gaussian' family, the function simulates response values by adding measurement error.

Additionally, GLGPMs can incorporate unstructured random effects specified through the `re()` term in the model formula, allowing for capturing additional variability beyond fixed and spatial covariate effects.

Value

A list containing simulated data, simulated spatial random effects (if applicable), and other simulation parameters.

Author(s)

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

Claudio Fronterre <c.fronterr@lancaster.ac.uk>

---

gp *Gaussian Process Model Specification*

Description

Specifies the terms, smoothness, and nugget effect for a Gaussian Process (GP) model.

Usage

```
gp(..., kappa = 0.5, nugget = 0)
```

Arguments

`...`	Variables representing the spatial coordinates or covariates for the GP model.
`kappa`	The smoothness parameter κ. Default is 0.5.
`nugget`	The nugget effect, which represents the variance of the measurement error. Default is 0. A positive numeric value must be provided if not using the default.

Details

The function constructs a list that includes the specified terms (spatial coordinates or covariates), the smoothness parameter κ, and the nugget effect. This list can be used as a specification for a Gaussian Process model.

Value

A list of class `gp.spec` containing the following elements:

`term`	A character vector of the specified terms.
`kappa`	The smoothness parameter κ.
`nugget`	The nugget effect.
`dim`	The number of specified terms.
`label`	A character string representing the full call for the GP model.

Note

The nugget effect must be a positive real number if specified.

Author(s)

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

Claudio Fronterre <c.fronterr@lancaster.ac.uk>

italy_sim *Simulated data-set on the Italian peninsula*

Description

This is a simulated data-set over Italy for a continuous outcome. The data-set contains 10 repeated observations for each of the 200 geo-referenced locations. The variables are as follows:

- x1 ordinate of the spatial locations.
- x2 abscissa of the spatial locations.
- y simulated continuous outcome.
- region the name of the region within which a given observation falls.
- province the name of the province within which a given observation falls.
- pop_dens the population density at the location of the observation.
- ID_loc an ID identifying the location to which the observation belong.

The coordinate reference system of the data is 32634.

Usage

```
data(italy_sim)
```

Format

A data frame with 2000 rows and 7 variables

Laplace_sampling_MCMC *Laplace Sampling Markov Chain Monte Carlo (MCMC) for Generalized Linear Gaussian Process Models*

Description

Performs MCMC sampling using Laplace approximation for Generalized Linear Gaussian Process Models (GLGPMs).

Usage

```
Laplace_sampling_MCMC(
  y,
  units_m,
  mu,
  Sigma,
  ID_coords,
  ID_re = NULL,
```

```
  sigma2_re = NULL,
  family,
  control_mcmc,
  Sigma_pd = NULL,
  mean_pd = NULL,
  messages = TRUE
)
```

Arguments

y	Response variable vector.
units_m	Units of measurement for the response variable.
mu	Mean vector of the response variable.
Sigma	Covariance matrix of the spatial process.
ID_coords	Indices mapping response to locations.
ID_re	Indices mapping response to unstructured random effects.
sigma2_re	Variance of the unstructured random effects.
family	Distribution family for the response variable. Must be one of 'gaussian', 'binomial', or 'poisson'.
control_mcmc	List with control parameters for the MCMC algorithm: **n_sim** Number of MCMC iterations. **burnin** Number of burn-in iterations. **thin** Thinning parameter for saving samples. **h** Step size for proposal distribution. Defaults to 1.65/(n_tot^(1/6)). **c1.h, c2.h** Parameters for adaptive step size tuning.
Sigma_pd	Precision matrix (optional) for Laplace approximation.
mean_pd	Mean vector (optional) for Laplace approximation.
messages	Logical; if TRUE, print progress messages.

Details

This function implements a Laplace sampling MCMC approach for GLGPMs. It maximizes the integrand using 'maxim.integrand' function for Laplace approximation if 'Sigma_pd' and 'mean_pd' are not provided.

The MCMC procedure involves adaptive step size adjustment based on the acceptance probability ('acc_prob') and uses a Gaussian proposal distribution centered on the current mean ('mean_curr') with variance 'h'.

Value

An object of class "mcmc.RiskMap" containing:

samples$S Samples of the spatial process.

samples$<re_names[i >] Samples of each unstructured random effect, named according to columns of ID_re if provided.

tuning_par Vector of step size (h) values used during MCMC iterations.

acceptance_prob Vector of acceptance probabilities across MCMC iterations.

Author(s)

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

Claudio Fronterre <c.fronterr@lancaster.ac.uk>

`liberia` *River-blindness in Liberia*

Description

This data-set contains counts of reported onchocerciasis (or riverblindess) cases from 91 villages sampled across Liberia. The variables are as follows:

- lat latitude of the of sampled villages.
- long longitude of the sampled villages.
- ntest number of tested people for the presence nodules.
- npos number of people that tested positive for the presence of nodules.
- elevation the elevation in meters of the sampled village.
- log_elevation the log-transformed elevation in meters of the sampled village.

Usage

```
data(liberia)
```

Format

A data frame with 90 rows and 6 variables

Source

Zouré, H. G. M., Noma, M., Tekle, Afework, H., Amazigo, U. V., Diggle, P. J., Giorgi, E., and Remme, J. H. F. (2014). The Geographic Distribution of Onchocerciasis in the 20 Participating Countries of the African Programme for Onchocerciasis Control: (2) Pre-Control Endemicity Levels and Estimated Number Infected. Parasites & Vectors, 7, 326.

`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.

`malkenya` *Malaria Transmission in the Western Kenyan Highlands*

Description

The dataset contains information on 82014 individuals enrolled in concurrent school and community cross-sectional surveys, conducted in 46 school clusters in the western Kenyan highlands. Malaria was assessed by rapid diagnostic test (RDT).

The variables are as follows:

- Cluster: unique ID for each of the 46 school clusters.
- Long: longitude coordinate of the household location.
- Lat: latitude coordinate of the household location.
- RDT: binary variable indicating the outcome of the RDT: 1, if positive, and 0, if negative.
- Gender: factor variable indicating the gender of the sampled individual.
- Age: age in years of the sampled individual.
- NetUse: binary variable indicating whether the sampled individual slept under a bed net the previous night: 1, if yes, 0, if no.
- MosqCntl: binary variable indicating whether the household has used some kind of mosquito control, such as sprays and coils: 1, if yes, 0, if no.
- IRS: binary variables in indicating whether there has been indoor residual spraying (IRS) in the house in the last 12 months: 1, if yes, 0, if no.
- Travel: binary variable indicating whether the sampled individual has traveled outside the village in the last three months: 1, if yes, 0, if no.
- SES: ordinal variable indicating the socio-economic status (SES) of the household. The variables is an integer score from 1(=poor) to 5(=rich).
- District: factor variable indicating the village of the sampled individual, "Kisii Central" or "Rachuonyo".
- Survey: factor variable indicating the survey in which the participant was enrolled, "community" or "school".
- elevation: elevation, in meters, of the recorded household location

Usage

```
data(malkenya)
```

Format

A data frame with 82014 rows and 13 variables

Source

Stevenson, J.C., Stresman, G.H., Gitonga, C.W., Gillig, J., Owaga, C., et al. (2013). Reliability of School Surveys in Estimating Geographic Variation in Malaria Transmission in the Western Kenyan Highlands. PLOS ONE 8(10): e77641. doi: 10.1371/journal.pone.0077641

`malnutrition` *Malnutrition in Ghana*

Description

This geostatistical dataset was extracted from the Demographic and Health Survey 2014 conducted in Ghana.

- lng Longitude of the sampling cluster.
- lat Latitude of the sampling cluster.
- age age in months of the child.
- sex sex of the child.
- HAZ height-for-age Z-score.
- WAZ weight-for-age Z-score
- urb binary indicator: urban area=1; rural area=0.
- etn ethnic group.
- edu level of education of the mother, which takes integer values from 1="Poorly educated" to 3="Highly educated".
- wealth wealth score of the household, which takes integer values from 1="Poor" to 3="Rich".

The coordinate reference system is 3857.

Usage

```
data(malnutrition)
```

Format

A data frame with 2671 rows and 10 variables

Source

Demographic and Health Survey, dhsprogram.com

matern.grad.phi *First Derivative with Respect to ϕ*

Description

Computes the first derivative of the Matern correlation function with respect to ϕ.

Usage

```
matern.grad.phi(U, phi, kappa)
```

Arguments

U	A vector of distances between pairs of data locations.
phi	The scale parameter ϕ.
kappa	The smoothness parameter κ.

Value

A matrix with the values of the first derivative of the Matern function with respect to ϕ for the given distances.

Author(s)

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

Claudio Fronterre <c.fronterr@lancaster.ac.uk>

matern.hessian.phi *Second Derivative with Respect to ϕ*

Description

Computes the second derivative of the Matern correlation function with respect to ϕ.

Usage

```
matern.hessian.phi(U, phi, kappa)
```

Arguments

U	A vector of distances between pairs of data locations.
phi	The scale parameter ϕ.
kappa	The smoothness parameter κ.

Value

A matrix with the values of the second derivative of the Matern function with respect to ϕ for the given distances.

Author(s)

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

Claudio Fronterre <c.fronterr@lancaster.ac.uk>

`matern_cor` *Matern Correlation Function*

Description

Computes the Matern correlation function.

Usage

```
matern_cor(u, phi, kappa, return_sym_matrix = FALSE)
```

Arguments

`u`	A vector of distances between pairs of data locations.
`phi`	The scale parameter ϕ.
`kappa`	The smoothness parameter κ.
`return_sym_matrix`	A logical value indicating whether to return a symmetric correlation matrix. Defaults to `FALSE`.

Details

The Matern correlation function is defined as

Value

A vector of the same length as `u` with the values of the Matern correlation function for the given distances, if `return_sym_matrix=FALSE`. If `return_sym_matrix=TRUE`, a symmetric correlation matrix is returned.

Author(s)

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

Claudio Fronterre <c.fronterr@lancaster.ac.uk>

$$\rho(u;\phi;\kappa) = (2^{\kappa-1})^{-1}(u/\phi)^{\kappa}K_{\kappa}(u/\phi)$$

where ϕ and κ are the scale and smoothness parameters, and $K_{\kappa}(\cdot)$ denotes the modified Bessel function of the third kind of order κ. The parameters ϕ and κ must be positive.

`maxim.integrand` *Maximization of the Integrand for Generalized Linear Gaussian Process Models*

Description

Maximizes the integrand function for Generalized Linear Gaussian Process Models (GLGPMs), which involves the evaluation of likelihood functions with spatially correlated random effects.

Usage

```
maxim.integrand(
  y,
  units_m,
  mu,
  Sigma,
  ID_coords,
  ID_re = NULL,
  family,
  sigma2_re = NULL,
  hessian = FALSE,
  gradient = FALSE
)
```

Arguments

`y`	Response variable vector.
`units_m`	Units of measurement for the response variable.
`mu`	Mean vector of the response variable.
`Sigma`	Covariance matrix of the spatial process.
`ID_coords`	Indices mapping response to locations.
`ID_re`	Indices mapping response to unstructured random effects.
`family`	Distribution family for the response variable. Must be one of 'gaussian', 'binomial', or 'poisson'.
`sigma2_re`	Variance of the unstructured random effects.
`hessian`	Logical; if TRUE, compute the Hessian matrix.
`gradient`	Logical; if TRUE, compute the gradient vector.

Details

This function maximizes the integrand for GLGPMs using the Nelder-Mead optimization algorithm. It computes the likelihood function incorporating spatial covariance and unstructured random effects, if provided.

The integrand includes terms for the spatial process (Sigma), unstructured random effects (sigma2_re), and the likelihood function (llik) based on the specified distribution family ('gaussian', 'binomial', or 'poisson').

Value

A list containing the mode estimate, and optionally, the Hessian matrix and gradient vector.

Author(s)

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

Claudio Fronterre <c.fronterr@lancaster.ac.uk>

---

plot.RiskMap_pred_target_grid

Plot Method for RiskMap_pred_target_grid Objects

Description

Generates a plot of the predicted values or summaries over the regular spatial grid from an object of class 'RiskMap_pred_target_grid'.

Usage

```
## S3 method for class 'RiskMap_pred_target_grid'
plot(x, which_target = "linear_target", which_summary = "mean", ...)
```

Arguments

x	An object of class 'RiskMap_pred_target_grid'.
which_target	Character string specifying which target prediction to plot.
which_summary	Character string specifying which summary statistic to plot (e.g., "mean", "sd").
...	Additional arguments passed to the plot function of the terra package.

Details

This function requires the 'terra' package for spatial data manipulation and plotting. It plots the values or summaries over a regular spatial grid, allowing for visual examination of spatial patterns.

Value

A ggplot object representing the specified prediction target or summary statistic over the spatial grid.

Author(s)

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

Claudio Fronterre <c.fronterr@lancaster.ac.uk>

See Also

pred_target_grid

`plot.RiskMap_pred_target_shp` *Plot Method for RiskMap_pred_target_shp Objects*

Description

Generates a plot of predictive target values or summaries over a shapefile.

Usage

```
## S3 method for class 'RiskMap_pred_target_shp'
plot(x, which_target = "linear_target", which_summary = "mean", ...)
```

Arguments

`x`	An object of class 'RiskMap_pred_target_shp' containing computed targets, summaries, and associated spatial data.
`which_target`	Character indicating the target type to plot (e.g., "linear_target").
`which_summary`	Character indicating the summary type to plot (e.g., "mean", "sd").
`...`	Additional arguments passed to 'scale_fill_distiller' in 'ggplot2'.

Details

This function plots the predictive target values or summaries over a shapefile. It requires the 'ggplot2' package for plotting and 'sf' objects for spatial data.

Value

A `ggplot` object showing the plot of the specified predictive target or summary.

Author(s)

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

Claudio Fronterre <c.fronterr@lancaster.ac.uk>

See Also

`pred_target_shp`, `ggplot`, `geom_sf`, `aes`, `scale_fill_distiller`

plot_s_variogram *Plotting the empirical variogram*

Description

Plots the empirical variogram generated by s_variogram

Usage

```
plot_s_variogram(variog_output, plot_envelope = FALSE, color = "royalblue1")
```

Arguments

variog_output	The output generated by the function s_variogram.
plot_envelope	A logical value indicating if the envelope of spatial independence generated using the permutation test must be displayed (plot_envelope = TRUE) or not (plot_envelope = FALSE). By default plot_envelope = FALSE. Note: if n_permutation = 0 when running the function s_variogram, the function will display an error message because no envelope can be generated.
color	If plot_envelope = TRUE, it sets the colour of the envelope; run vignette("ggplot2-specs") for more details on this argument.

Details

This function plots the empirical variogram, which shows the spatial dependence structure of the data. If plot_envelope is set to TRUE, the plot will also include an envelope indicating the range of values under spatial independence, based on a permutation test.

Value

A ggplot object representing the empirical variogram plot, optionally including the envelope of spatial independence.

See Also

s_variogram

`pred_over_grid` *Prediction of the random effects components and covariates effects over a spatial grid using a fitted generalized linear Gaussian process model*

Description

This function computes predictions over a spatial grid using a fitted model obtained from the `glgpm` function. It provides point predictions and uncertainty estimates for the specified locations for each component of the model separately: the spatial random effects; the unstructured random effects (if included); and the covariates effects.

Usage

```
pred_over_grid(
  object,
  grid_pred,
  predictors = NULL,
  re_predictors = NULL,
  pred_cov_offset = NULL,
  control_sim = set_control_sim(),
  type = "marginal",
  messages = TRUE
)
```

Arguments

`object`	A RiskMap object obtained from the ‘glgpm‘ function.
`grid_pred`	An object of class ’sfc’, representing the spatial grid over which predictions are to be made. Must be in the same coordinate reference system (CRS) as the object passed to ’object’.
`predictors`	Optional. A data frame containing predictor variables used for prediction.
`re_predictors`	Optional. A data frame containing predictors for unstructured random effects, if applicable.
`pred_cov_offset`	Optional. A numeric vector specifying covariate offsets at prediction locations.
`control_sim`	Control parameters for MCMC sampling. Must be an object of class "mcmc.RiskMap" as returned by `set_control_sim`.
`type`	Type of prediction. "marginal" for marginal predictions, "joint" for joint predictions.
`messages`	Logical. If TRUE, display progress messages. Default is TRUE.

Value

An object of class ’RiskMap.pred.re’ containing predicted values, uncertainty estimates, and additional information.

Author(s)

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

Claudio Fronterre <c.fronterr@lancaster.ac.uk>

`pred_target_grid` *Predictive Target Over a Regular Spatial Grid*

Description

Computes predictions over a regular spatial grid using outputs from the `pred_over_grid` function. This function allows for incorporating covariates, offsets, and optional unstructured random effects into the predictive target.

Usage

```
pred_target_grid(
  object,
  include_covariates = TRUE,
  include_nugget = FALSE,
  include_cov_offset = FALSE,
  include_re = FALSE,
  f_target = NULL,
  pd_summary = NULL
)
```

Arguments

`object`	Output from 'pred_over_grid', a RiskMap.pred.re object.
`include_covariates`	Logical. Include covariates in the predictive target.
`include_nugget`	Logical. Include the nugget effect in the predictive target.
`include_cov_offset`	Logical. Include the covariate offset in the predictive target.
`include_re`	Logical. Include unstructured random effects in the predictive target.
`f_target`	Optional. List of functions to apply on the linear predictor samples.
`pd_summary`	Optional. List of summary functions to apply on the predicted values.

Value

An object of class 'RiskMap_pred_target_grid' containing predicted values and summaries over the regular spatial grid.

Author(s)

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

Claudio Fronterre <c.fronterr@lancaster.ac.uk>

See Also

pred_over_grid

---

pred_target_shp *Predictive Target over a Shapefile*

Description

Computes predictions over a shapefile using outputs from the pred_over_grid function. This function allows for incorporating covariates, offsets, and optional unstructured random effects into the predictive target.

Usage

```
pred_target_shp(
  object,
  shp,
  shp_target = mean,
  weights = NULL,
  standardize_weights = FALSE,
  col_names = NULL,
  include_covariates = TRUE,
  include_nugget = FALSE,
  include_cov_offset = FALSE,
  include_re = FALSE,
  f_target = NULL,
  pd_summary = NULL
)
```

Arguments

object	Output from 'pred_over_grid', a RiskMap.pred.re object.
shp	Spatial dataset (sf or data.frame) representing the shapefile over which predictions are computed.
shp_target	Function defining the aggregation method for shapefile targets (default is mean).
weights	Optional numeric vector of weights for spatial predictions.
standardize_weights	Logical indicating whether to standardize weights (default is FALSE).
col_names	Column name or index in 'shp' containing region names.
include_covariates	Logical indicating whether to include covariates in predictions (default is TRUE).
include_nugget	Logical indicating whether to include the nugget effect (default is FALSE).
include_cov_offset	Logical indicating whether to include covariate offset in predictions (default is FALSE).

include_re	Logical indicating whether to include random effects in predictions (default is FALSE).
f_target	List of target functions to apply to the linear predictor samples.
pd_summary	List of summary functions (e.g., mean, sd) to summarize target samples.

Details

This function computes predictive targets or summaries over a spatial shapefile using outputs from 'pred_S'. It requires the 'terra' package for spatial data manipulation and should be used with 'sf' or 'data.frame' objects representing the shapefile.

Value

An object of class 'RiskMap_pred_target_shp' containing computed targets, summaries, and associated spatial data.

Author(s)

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

Claudio Fronterre <c.fronterr@lancaster.ac.uk>

See Also

pred_target_grid

---

print.summary.RiskMap *Print Summary of RiskMap Model*

Description

Provides a print method for the summary of "RiskMap" objects, detailing the model type, parameter estimates, and other relevant statistics.

Usage

```
## S3 method for class 'summary.RiskMap'
print(x, ...)
```

Arguments

x	An object of class "summary.RiskMap".
...	other parameters.

Details

This function prints a detailed summary of a fitted "RiskMap" model, including:

- The type of geostatistical model (e.g., Gaussian, Binomial, Poisson).
- Confidence intervals for parameter estimates.
- Regression coefficients with their standard errors and p-values.
- Measurement error variance, if applicable.
- Spatial process parameters, including the Matern covariance parameters.
- Variance of the nugget effect, if applicable.
- Unstructured random effects variances, if applicable.
- Log-likelihood of the model.
- Akaike Information Criterion (AIC) for Gaussian models.

Value

This function is used for its side effect of printing to the console. It does not return a value.

Author(s)

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

Claudio Fronterre <c.fronterr@lancaster.ac.uk>

---

propose_utm *EPSG of the UTM Zone*

Description

Suggests the EPSG code for the UTM zone where the majority of the data falls.

Usage

```
propose_utm(data)
```

Arguments

data	An object of class sf containing the coordinates.

Details

The function determines the UTM zone and hemisphere where the majority of the data points are located and proposes the corresponding EPSG code.

Value

An integer indicating the EPSG code of the UTM zone.

Author(s)

Emanuele Giorgi <e.giorgi@lancaster.ac.uk> Claudio Fronterre <c.fronterr@lancaster.ac.uk>

`re` *Random Effect Model Specification*

Description

Specifies the terms for a random effect model.

Usage

```
re(...)
```

Arguments

`...`	Variables representing the random effects in the model.

Details

The function constructs a list that includes the specified terms for the random effects. This list can be used as a specification for a random effect model.

Value

A list of class `re.spec` containing the following elements:

`term`	A character vector of the specified terms.
`dim`	The number of specified terms.
`label`	A character string representing the full call for the random effect model.

Note

At least one variable must be provided as input.

Author(s)

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

Claudio Fronterre <c.fronterr@lancaster.ac.uk>

`set_control_sim` *Set Control Parameters for Simulation*

Description

This function sets control parameters for running simulations, particularly for MCMC methods. It allows users to specify the number of simulations, burn-in period, thinning interval, and various other parameters necessary for the simulation.

Usage

```
set_control_sim(
  n_sim = 12000,
  burnin = 2000,
  thin = 10,
  h = NULL,
  c1.h = 0.01,
  c2.h = 1e-04,
  linear_model = FALSE
)
```

Arguments

`n_sim`	Integer. The total number of simulations to run. Default is 12000.
`burnin`	Integer. The number of initial simulations to discard (burn-in period, used for the MCMC algorithm). Default is 2000.
`thin`	Integer. The interval at which simulations are recorded (thinning interval, used for the MCMC algorithm). Default is 10.
`h`	Numeric. An optional parameter. Must be non-negative if specified.
`c1.h`	Numeric. A control parameter for the simulation. Must be positive. Default is 0.01.
`c2.h`	Numeric. Another control parameter for the simulation. Must be between 0 and 1. Default is 1e-04.
`linear_model`	Logical. If TRUE, the function sets up parameters for a linear model and only returns `n_sim`. Default is FALSE.

Details

The function validates the input parameters and ensures they are appropriate for the simulation that is used in the `glgpm` fitting function. For non-linear models, it checks that `n_sim` is greater than `burnin`, that `thin` is positive and a divisor of `(n_sim - burnin)`, and that `h`, `c1.h`, and `c2.h` are within their respective valid ranges.

If `linear_model` is TRUE, only `n_sim` and `linear_model` are required, and the function returns a list containing these parameters.

If `linear_model` is FALSE, the function returns a list containing `n_sim`, `burnin`, `thin`, `h`, `c1.h`, `c2.h`, and `linear_model`.

Value

A list of control parameters for the simulation with class attribute "mcmc.RiskMap".

Author(s)

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

Claudio Fronterre <c.fronterr@lancaster.ac.uk>

See Also

Matrix, forceSymmetric

Examples

```
# Example with default parameters
control_params <- set_control_sim()

# Example with custom parameters
control_params <- set_control_sim(n_sim = 15000, burnin = 3000, thin = 20)
```

---

summary.RiskMap *Summarize Model Fits*

Description

Provides a summary method for the "RiskMap" class that computes the standard errors and p-values for likelihood-based model fits.

Usage

```
## S3 method for class 'RiskMap'
summary(object, ..., conf_level = 0.95)
```

Arguments

object	An object of class "RiskMap" obtained as a result of a call to glgpm.
...	other parameters.
conf_level	The confidence level for the intervals (default is 0.95).

Details

This function computes the standard errors and p-values for the parameters of a "RiskMap" model, adjusting for the covariance structure if needed.

Value

A list containing:

`reg_coef`	A matrix with the estimates, standard errors, z-values, p-values, and confidence intervals for the regression coefficients.
`me`	A matrix with the estimates and confidence intervals for the measurement error variance, if applicable.
`sp`	A matrix with the estimates and confidence intervals for the spatial process parameters.
`tau2`	The fixed nugget variance, if applicable.
`ranef`	A matrix with the estimates and confidence intervals for the random effects variances, if applicable.
`conf_level`	The confidence level used for the intervals.
`family`	The family of the model (e.g., "gaussian").
`kappa`	The kappa parameter of the model.
`log.lik`	The log-likelihood of the model fit.
`cov_offset_used`	A logical indicating if a covariance offset was used.
`aic`	The Akaike Information Criterion (AIC) for the model, if applicable.

Note

Handles both Gaussian and non-Gaussian families, and accounts for fixed and random effects in the model.

Author(s)

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

Claudio Fronterre <c.fronterr@lancaster.ac.uk>

See Also

`glgpm`, `coef.RiskMap`

`s_variogram` *Empirical variogram*

Description

Computes the empirical variogram using “bins” of distance provided by the user.

Usage

```
s_variogram(
  data,
  variable,
  bins = NULL,
  n_permutation = 0,
  convert_to_utm = TRUE,
  scale_to_km = FALSE
)
```

Arguments

data	an object of class sf containing the variable for which the variogram is to be computed and the coordinates
variable	a character indicating the name of variable for which the variogram is to be computed.
bins	a vector indicating the 'bins' to be used to define the classes of distance used in the computation of the variogram. By default bins=NULL and bins are then computed as seq(0, d_max/2, length=15) where d_max is the maximum distance observed in the data.
n_permutation	a non-negative integer indicating the number of permutation used to compute the 95 level envelope under the assumption of spatial independence. By default n_permutation=0, and no envelope is generated.
convert_to_utm	a logical value, indicating if the conversion to UTM shuold be performed (convert_to_utm = TRUE) or the coordinate reference system of the data must be used without any conversion (convert_to_utm = FALSE). By default convert_to_utm = TRUE. Note: if convert_to_utm = TRUE the conversion to UTM is performed using the epsg provided by propose_utm.
scale_to_km	a logical value, indicating if the distances used in the variogram must be scaled to kilometers (scale_to_km = TRUE) or left in meters (scale_to_km = FALSE). By default scale_to_km = FALSE

Value

an object of class 'variogram' which is a list containing the following components

variogram a data-frame containing the following columns: mid_points, the middle points of the classes of distance provided by bins; obs_vari the values of the observed variogram; obs_vari the number of pairs. If n_permutation > 0, the data-frame also contains lower_bound and upper_bound corresponding to the lower and upper bounds of the 95 used to assess the departure of the observed variogram from the assumption of spatial independence.

scale_to_km the value passed to scale_to_km

n_permutation the number of permutations

Author(s)

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

Claudio Fronterre <c.fronterr@lancaster.ac.uk>

to_table *Create LaTeX Table from Model Fit*

Description

Converts a "RiskMap" model fit into an `xtable` object, which can then be printed as a LaTeX or HTML table.

Usage

```
to_table(object, ...)
```

Arguments

`object`	An object of class "RiskMap" obtained as a result of a call to `glgpm`.
`...`	Additional arguments to be passed to `xtable`.

Details

This function takes a fitted "RiskMap" model and converts it into an `xtable` object. The resulting table includes:

- Regression coefficients with their estimates, confidence intervals, and p-values.
- Spatial process parameters.
- Random effects variances.
- Measurement error variance, if applicable.

The `xtable` object can be customized further using additional arguments and then printed as a LaTeX or HTML table.

Value

An object of class "xtable" which inherits the `data.frame` class and contains several additional attributes specifying the table formatting options.

Author(s)

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

Claudio Fronterre <c.fronterr@lancaster.ac.uk>

See Also

`glgpm`, `xtable`

Index