Package 'rcage'

June 14, 2021

Type Package

Title Regionalization of Multiscale Spatial Processes

Version 1.1

Date 2021-06-09

Author Jonathan R. Bradley [aut],
Christopher K. Wikle [aut],
Scott H. Holan [aut],
Shannon T. Holloway [aut, cre]

Maintainer Shannon T. Holloway <sthollow@ncsu.edu>

Description Regionalization of multiscale spatial processes based on a criterion for spatial aggregation error. The multiscale representation is a truncated Karhunen-Loaeve expansion using Obled-Creutin eigenfunctions. The method is incorporated within a Bayesian framework using an MCMC implementation of a latent spatial model. Bradley, J. R., Wikle, C. K., and Holan, S. H. (2017) <doi:10.1111/rssb.12179>.

License GPL-2

Depends R (>= 2.10), ff

Imports ClustGeo, MASS, fields, ggplot2, methods, pracma, raster, rgeos, sp, gridExtra, rlang, Matrix, sf, LaplacesDemon

NeedsCompilation no

Encoding UTF-8

RoxygenNote 7.1.1

Collate 'knots.R' 'GBFObj.R' 'semiDefinite.R' 'miPrior.R' 'MIpriorInv.R' 'checkInputs.R' 'checkDB.R' 'checkNW.R' 'checkSpatialData.R' 'generatingBasis.R' 'getArea.R' 'wMatrix.R' 'choleskyInvW.R' 'obledCruetinBasis.R' 'OCbasis.R' 'QPriorObj.R' 'QPriorObj_MI.R' 'QPriorObj_Wishart.R' 'biSquare.R' 'cage.R' 'countyExampleData.R' 'yOpt.R' 'hMatrix.R' 'dCCAGE.R' 'gibbs.R' 'verifyQPrior.R' 'verifyffdir.R' 'verifyCovariates.R' 'verifyNumericVector.R' 'inputPrep.R' 'summaryInfo.R' 'map.R' 'minCAGE.R' 'minCAGESH.R' 'oceanExampleData.R' 'optimalDCAGE.R' 'optRegion.R' 'plot.rcage.R' 'radial.R' 'region.R' 'wendland.R' 'windExampleData.R' 'yCageSource.R' 'yFinest.R'

Repository CRAN

Date/Publication 2021-06-14 07:40:06 UTC

R topics documented:

`bisquare` *Bisquare Basis Functions*

Description

Provides an implementation of the bisquare basis functions defined as

$$\Psi_j(s) = \{1 - (||s - c_j||/w)^2\}^2 \mathrm{I}(||s - c_j|| \leq w).$$

Usage

```
bisquare(crd, knots, w = NULL, ..., longlat = TRUE)
```

Arguments

`crd`	A matrix object. The (x,y) coordinates of the reference points {nCrd x 2}.
`knots`	A matrix object. The (x,y) coordinates of the knots {r x 2}.
`w`	A numeric object. The positive scaling factor (bandwidth).
`...`	ignored. Included only to require naming of inputs that follow.
`longlat`	A logical object. If FALSE, Euclidean distance is calculated; if TRUE, Great Circle distance is calculated. See ?sp::spDists for more information.

Details

Distances between reference coordinates and knots are obtained using sp::spDists().

Value

A matrix of bisquare functions evaluated at all combinations of crd and knots { nCrd x r }.

References

Cressie, N. and Johannesson, G. (2008). Fixed rank kriging for very large spatial data sets. Journal of the Royal Statistical Society, Series B, 70, 209–226. <doi:10.1111/j.1467-9868.2007.00633.x>.

Examples

```
data(countyExample)

bisquare(crd = sp::coordinates(county), knots = knots)
```

---

countyExample *U.S. Legal/Statistical Area Data*

Description

The U.S. legal/statistical areas, their median household income, and a matrix of knots provided only to facilitate illustrative examples.

Usage

```
data(countyExample)
```

Format

countyExample contains 3 objects:

county A SpatialPolygonsDataFrame containing 3109 areas. For each areal unit, the following data are defined

- **GEO_ID** identifier for legal/statistical area
- **STATE** state containing the legal/statistical area
- **NAME** name of the legal/statistical area
- **LSAD** legal/statistical area description
- **SHAPE_AREA** area of the county in square meters.
- **SHAPE_LEN** perimeter of the county in meters.

ACS A 3109x2 matrix containing the public-use American Community Survey (ACS) 2013 5-year estimated county-level median household income and its variance.

- **LogMedianIncome** log of the estimated median household income.
- **VarianceLMI** approximate variance of the log median household income (using the Delta method).

knots A 75x2 matrix of knots spanning the county data.

- **x** x coordinate
- **y** y coordinate

References

U.S. Census Bureau. 2013-2017 American Community Survey 5-year Estimates; generated by Jonathan Bradley; using American FactFinder.

`MIpriorInv` *Inverse of the MI Prior Distribution*

Description

Calculates the inverse of the MI prior distribution, which is motivated by specifying the covariance so that it is "close" to the covariance from an Intrinsic Conditional Auto-Regressive model on the domain of the finest resolution. See reference for details.

Usage

```
MIpriorInv(psi, dB, ..., dMax = NULL)
```

Arguments

`psi`	A matrix object. The estimated OC basis {nB x r}, where nB is the number of areal units in the finest resolution spatial object (dB)
`dB`	A SpatialPolygons or SpatialPoints object. The finest resolution {nB}
`...`	Ignored.
`dMax`	Numeric maximum distance between points to be considered adjacent. Ignored if dB is SpatialPolygons. If dB is SpatialPoints and dMax is not specified, it is taken to be the 0.1 quantile of the distances.

Details

For clarity – function returns

$$K^{-1} = R_B^{-1} A^{+} Q_B' (I - A) Q_B R_B^{-1},$$

where A^{+} is the first order adjacency matrix and Q_B, R_B is the QR decomposition of the basis matrix.

Value

The inverse of the MI prior as an rxr matrix.

References

Bradley, J. R., Wikle, C. K., and Holan, S. H. (2017). Regionalization of Multiscale Spatial Processes using a Criterion for Spatial Aggregation Error. Journal of the Royal Statistical Society - Series B, 79, 815–832. <doi:10.1111/rssb.12179>

Examples

```
data(countyExample)

nc <- county[county@data[,"STATE"] == 37, ]
psi <- matrix(data = rbinom(n = 1000, size = 1, prob = 0.5),
              nrow = 100L, ncol = 10L)
MIpriorInv(psi = psi, dB = nc)
```

`OCbasis` *Obled-Cruetin Basis Function*

Description

Performs a reweighting of radial basis functions ensuring orthonormality.

Usage

```
OCbasis(
  ...,
  spatialData,
  gbf,
  knots,
  dB = NULL,
  w = NULL,
  nw = NULL,
  nCore = 1L,
  longlat = TRUE
)
```

Arguments

`...`	Ignored. Included only to require named inputs.
`spatialData`	A SpatialPoints object, SpatialPolygons object, or a list of said objects. The source support data. If provided as a list, input dB must be integer or SpatialPolygons.
`gbf`	A function or character object. The function to use to calculate the radial basis functions of the expansion of the Obled-Creutin eigenfunction. The bi-square, wendland, and radial functions are available through this implementation as 'bisquare', 'wendland' and 'gaussian', respectively. All others must be defined by user. See details for further information.
`knots`	A matrix or integer. If a matrix, the knots of the radial basis functions. If an integer, the number of knots to generate using fields::cover.design().

dB	NULL, integer, or a SpatialPolygons object defining the spatial region to be sampled when using Monte Carlo estimates. If spatialData is a list of spatial objects, dB must be in an integer specifying the element of spatialData to use as the sampling region or a SpatialPolygons object.
w	A numeric object. The scaling factor for radial basis functions. See details for further information.
nw	An integer object or NULL. The number of MC replicates to generate for estimating the O-C eigenfunctions. If <=0 or NULL, spatialData must be or include SpatialPoints data.
nCore	An integer object or NULL. The number of cores if parallel methods are to be used in the Monte Carlo step.
longlat	A logical object. TRUE if spatialData is longitude/latitude data.

Details

Input 'gbf' allows users to specify a radial basis function beyond the internally implemented bi-square, wendland, and gaussian functions. If user provides a function, the function must use the following formal arguments:

- crd - the coordinates at which the basis functions are to be evaluated;
- knots - the knots of the basis functions;
- w - the scaling factor for the basis function; and
- ... - an ellipsis to avoid argument errors.

The function must return a matrix of dimension {nrow(crd) x nrow(knots)}.

For completeness, the bi-square functions implemented in the package are of the form

$$\Psi_j(s) = \{1 - (||s - c_j||/w)^2\}^2 \mathrm{I}(||s - c_j|| \leq w).$$

Note that input 'w' is equivalent to w in the expression above. In addition, if a user were to define an equivalent function inputs $s \equiv$ 'crd' and $c_j \equiv$ 'knots'.

The Wendland basis functions defined as

$$\Psi_j(s) = \{1 - d_j(s)\}^6 \{35 d_j(s)^2 + 18 d_j(s) + 3\}/3 \mathrm{I}(0 \leq d_j \leq 1),$$

where

$$d_j(s) = ||s - c_j||/w.$$

The Gaussian radial basis functions defined as

$$\Psi_j(s) = \exp\{-\frac{1}{2}(||s - c_j||/w)^2\}.$$

Value

A list containing:

basis	The {nSpatial x r} radial basis function.
OCnorm	The {r x r} Obled Cruetin weighting matrix.
knots	The {nKnots x 2} matrix of knots.
w	The scaling factor used in basis.

Examples

```
# create 5x5 square

poly <- raster::rasterToPolygons(raster::raster(nrows = 5, ncols = 5,
                                                xmn = -1.25, xmx = 1.25,
                                                ymn = -1.25, ymx = 1.25,
                                                res = 0.5,
                                                crs = "+proj=longlat +datum=WGS84"))

df <- data.frame("x" = stats::rnorm(n = 25))

dt <- sp::SpatialPolygonsDataFrame(poly, df)

knots <- expand.grid(c(-0.75,0.0,0.75),c(-0.75,0.0,0.75))

OCbasis(spatialData = dt,
        gbf = 'bisquare',
        knots = knots,
        nw = 200L,
        nCore = 1L)

OCbasis(spatialData = dt,
        gbf = 'gaussian',
        knots = knots,
        nw = 200L,
        nCore = 1L)
```

`oceanExample` *Ocean Color Observations*

Description

This dataset is a sample from SeaWiFS Ocean Color Satellite Observations taken over the coastal Gulf of Alaska from 1998 to 2001 and ROMS Ocean Model. Data are provided only to facilitate illustrative examples.

Usage

```
data(oceanExample)
```

Format

oceanExample contains 2 objects:

ocean A SpatialPointsDataFrame containing 5255 points containing the following 6 columns:

- **X1** intercept
- **X2** ROMS ocean model output for chlorophyll

X3 ROMS ocean model output for sea surface temperature

X4 ROMS ocean model output for Sea Surface Height (SSH)

CHA SeaWiFs satellite ocean color data.

sigma2 sample variance

oceanKnots A 170x2 matrix of knots in long/lat for the region spanned by data set ocean

References

Leeds, W.B., Wikle, C.K., and Fiechter, J. (2014). Emulator-assisted reduced-rank ecological data assimilation for multivariate dynamical spatio-temporal processes. Statistical Methodology, 17, 126–138.

Wikle, C. K., Milliff, R. F., Herbei, R., and Leeds, W. B. (2013). Modern statistical methods in oceanography: A hierarchical perspective. Statistical Science, 28, 466-486.

`optRegion` *Regionalization of Multi-scale Spatial Processes*

Description

Implements a regionalization algorithm for multi-scale spatial processes based on a criterion for spatial aggregation error. The multi-scale representation is a truncated Karhunen-Loaeve expansion using Obled-Creutin eigenfunctions. The method is incorporated within a Bayesian framework using a Markov chain Monte Carlo implementation of a latent spatial model.

Usage

```
optRegion(
  spatialData,
  response,
  sigmavar,
  ...,
  cage = TRUE,
  longlat = TRUE,
  dB = NULL,
  gbf = "bisquare",
  w = NULL,
  knots = 75L,
  nw = 0L,
  nGibbs = 10000,
  nBurn = 1000,
  nThin = 1L,
  x = NULL,
  sigma2_beta = 1e-04,
  sigma2_xi = 1,
  Qprior = "MI",
  lambda = 1,
```

```
  wishartScale = 100,
  dMax = NULL,
  cMethod = "kmeans",
  gL = NULL,
  gU = NULL,
  alpha = 0.5,
  ffdir = NULL,
  nCore = 1L,
  parallelLog = NULL,
  plot = TRUE,
  dataScale = NULL,
  palette = "plasma"
)
```

Arguments

spatialData	A SpatialPolygonsDataFrame object, a SpatialPointsDataFrame object as defined by package sp, or a list of said objects. The source support.
response	A numeric vector, character, or list of said objects. The value of interest. If numeric, the vector must contain the response for all areal units of 'spatialData'. Specifically, if 'spatialData' is a list, the vector must include the value of interest for all data of element [[1]] followed by that for all data of element [[2]], etc. If a character, the column header of the data slot that holds the value of interest; if a list is provided in 'spatialData' each Spatial object must have the specified column header. If provided as a list, the elements must correspond to the elements of the 'spatialData' list.
sigmavar	A numeric vector/matrix, character, or list of said objects. The survey variance. If numeric, the vector is the diagonal elements of the variance matrix and it must contain the variance for all areal units of 'spatialData'. Specifically, if 'spatialData' is a list, the vector must include the variance for all data of element [[1]] followed by that for all data of element [[2]], etc. Similarly, if a matrix. If a character, the column header of the data slot that holds the diagonal elements of the variance; if a list is provided in 'spatialData' each Spatial object must have the specified column header. If provided as a list, the elements must correspond to the elements of the 'spatialData' list.
...	ignored. Used to require named input for remaining formals.
cage	A logical. If TRUE, CAGE is minimized to determine the optimal clustering. If FALSE, DCAGE is used.
longlat	A logical. TRUE indicates that 'spatialData' is in long/lat coordinate system.
dB	NULL, integer or a SpatialPolygons object defining the finest resolution spatial object to be used for inference. If NULL, and only one SpatialPolygon object is provided in 'spatialData', the finest resolution is taken to be the 'spatialData' object. If 'spatialData' contains a list of objects, this input must be provided.
gbf	NULL, function or function name. The function to use to calculate the radial basis functions of the expansion of the Obled-Creutin eigenfunction. The bisquare (default), the Wendland, and the Gaussian radial functions are available

through this implementation ('bisquare', 'wendland', 'gaussian'). All others must be defined by the user. See details for further information.

`w` A numeric. The scaling factor for radial basis functions. See details for further information.

`knots` A matrix or integer. If a matrix, the x,y coordinates of the knots of the radial basis functions. If an integer, the number of knots to generate using fields::cover.design().

`nw` An integer or NULL. The number of Monte Carlo replicates to generate for estimating the components of the Obled-Cruetin basis when points data is not available or is insufficient for calculating these components. If <=0 or NULL, 'spatialData' must be or include SpatialPoints data.

`nGibbs` An integer. The total number of Gibbs samples generated. See 'nBurn' for more information.

`nBurn` An integer. The number of samples to skip before accepting every nThin sample of the Gibbs sampling algorithm. The samples kept are those defined by seq(from = nBurn+1, to = nGibbs, by = nThin).

`nThin` An integer. Keep every nThin sample of the Gibbs sampling algorithm once the burn specification has been satisfied. See 'nBurn' for more information.

`x` NULL, character, or numeric. The covariates for the large scale variability model. If NULL, an intercept only model is assumed. If numeric (matrix), the covariates; the object must contain the covariates for all spatialData. Specifically, if 'spatialData' is a list, x must include the covariates for all data of element [[1]] followed by that for all data of element [[2]], etc. If a character, the column header(s) of the data slot that holds the covariates; if a list is provided in 'spatialData' each Spatial object must have the specified column header. If provided as a list, the elements must correspond to the elements of the 'spatialData' list.

`sigma2_beta` A numeric. The variance of the large scale variability model.

`sigma2_xi` A numeric. The variance of the fine scale variability. Defaults to inverse Gamma.

`Qprior` A character. The Q prior. Must be one of 'MI' or 'wish' indicating the MI prior or the Inverse Wishart distribution, respectively.

`lambda` A numeric vector. The r eigenvalues of the prior distribution on the {r x r} covariance matrix Q.

`wishartScale` A numeric of NULL. If Qprior is "wish", the value of the diagonal elements of the scale matrix provided to MCMCpack::riwish. Default value is 100. Input is ignored if Qprior = 'MI'.

`dMax` A numeric. If finest resolution data is SpatialPoints and Qprior is MI, the maximum distance at which points are considered adjacent. If not specified, it is taken to be the upper boundary of the lowest decile.

`cMethod` A character. Must be one of {'kmeans', 'hier'}. Indicates if stats::kmeans() or ClustGeo::hclustgeo() is used to obtain clusters.

`gL` An integer. The smallest number of areal units to consider for regionalization.

`gU` An integer. The largest number of areal units to consider for regionalization.

alpha	A numeric. If using cMethod = 'hier', a real value between 0 and 1. This mixing parameter gives the relative importance of D0 compared to D1. D1 is taken to be the distance matrix defined by sp::spDists. See ?ClustGeo::hclustgeo for further details.
ffdir	A character string. The directory in which ff objects are to be saved. See details for further information.
nCore	An integer. If using Monte Carlo, the algorithm can be spread across nCore cores to expedite calculations. If nCore <= 1L, no parallel methods are used.
parallelLog	A character object. A file name for logging parallel executions.
plot	A logical. TRUE indicates that final plot will be generated.
dataScale	A function. If the response of interest was scaled, the function to undo the scaling. Used in plotting only.
palette	A character. The palette preference for plotting. The palette is coded to be the viridis palette with possible values 'viridis', 'magma', 'plasma', 'inferno', 'cividis'.

Details

Input 'nw' is the number of Monte Carlo samples to be used in estimating matrix W and the basis psi(A) when points data are not available or provide insufficient coverage to obtain closed form expressions. It is recommended that 'nw' be large; values below 1000 are automatically adjusted to the default value of 20000. It is only appropriate to specify nw = NULL when the source support data ('spatialData') contains SpatialPoints data. If 'spatialData' does not contain SpatialPoints data but nw is not provided as input, it is set to 20000.

If input 'spatialData' does not represent the finest resolution upon which inference is to be made, input 'dB' must be specified. The area spanned by 'dB' must fully contain all of the source support data, i.e., no elements of 'spatialData' can lie outside of the boundary defined by 'dB'. If 'spatialData' is multi-resolution 'dB' must be provided as the integer element of 'spatialData' to be used as the finest resolution or as a SpatialPolygons object defining the finest resolution.

Input 'gbf' allows users to specify a radial basis function beyond the internally implemented bi-square, Wendland, and Gaussian functions. If a function is provided, the following formal arguments are required: * crd an {nCrd x 2} matrix containing the coordinates at which the basis functions are to be evaluated; * knots an {nKnots x 2} matrix containing the x,y coordinates of the knots; and * w the scaling factor/bandwidth for the basis function. The function must return a matrix of dimension {nCrd x nKnots}.

For completeness, the bi-square function implemented in the package is of the form

$$\Psi_j(s) \equiv \begin{cases} \{1-(||\text{crd}-\text{knots}_j||/w)^2\}^2 & \text{if } ||\text{crd}-\text{knots}_j|| \leq w, \\ 0 & \text{otherwise} \end{cases}.$$

The Wendland function is

$$\Psi_j(s) \equiv \begin{cases} \{1-d_j(s)\}^6\{35d_j(s)^2+18d_j(s)+3\}/3 & 0 \leq d_j \leq 1, \\ 0 & \text{otherwise} \end{cases},$$

where

$$d_j(s) = ||s-c_j||/w.$$

The Gaussian function is

$$\Psi_j(s) = \exp\{-\frac{1}{2}(||s - c_j||/w)^2\}.$$

The default prior distribution on the covariance matrix is the MI prior function

$$K^{-1} = R_B^{-1}\mathcal{A}\{Q_B^{'}(I - A)Q_B\}R_B^{-1},$$

as defined in the supplemental section of the original manuscript.

This package implements methods of the `ff` package in scenarios when memory size is of concern. The `ff` package stores variables on the disk rather than in RAM. Though it is highly optimized, this choice does increase computation time. Intermediate variables used only internally are stored in a temp directory as defined by the `ff` package. To trigger this implementation, 'ffdir' must be set. Additionally note that some elements available in parallel are not implemented in combination with ffdir.

The repeated calls to a clustering algorithm make this a time consuming calculation. Parallel methods have been employed to improve computation times when possible.

Value

A list.

`call`	The original call structure
`psi`	A list containing the generating basis function and the Oble-Cruetin normalization matrix
`minCAGE`	The minimum CAGE/DCAGE.
`CAGETrack`	For the optimal clustering, the estimated CAGE/DCAGE for each cluster.
`cluster`	The results returned by kmeans() or hclustgeo() for the optimal clustering.
`yOpt`	The estimated value of interest at each Gibbs sample clustered according to the optimal clustering.
`yFinest`	The estimated value of interest at each Gibbs sample clustered according to dB.
`criterion`	"CAGE" or "DCAGE"

References

Bradley, J. R., Wikle, C. K., and Holan, S. H. (2017). Regionalization of Multiscale Spatial Processes using a Criterion for Spatial Aggregation Error. Journal of the Royal Statistical Society - Series B, 79, 815–832.

Examples

```
# create 5x5 square

poly <- raster::rasterToPolygons(raster::raster(nrows = 5, ncols = 5,
                                                xmn = -1.25, xmx = 1.25,
                                                ymn = -1.25, ymx = 1.25,
                                                res = 0.5,
```

```
                                                    crs = "+proj=longlat +datum=WGS84"))

# covariate assumed to be a norm(0,1)
df <- data.frame("x" = stats::rnorm(n = 25))

# convert spatial data to data.frame
dt <- sp::SpatialPolygonsDataFrame(poly, df)

knots <- cbind(c(-0.75, 0.0, 0.75, -0.75, 0.0, 0.75, -0.75, 0.0, 0.75),
               c(-0.75,-0.75, -0.75, 0.0, 0.0, 0.0, 0.75, 0.75, 0.75))

res <- optRegion(spatialData = dt,
                 response = "x",
                 sigmavar = rep(1, 25),
                 gL = 5,
                 gU = 7,
                 nGibbs = 50L,
                 nBurn = 10L,
                 nThin = 1L,
                 nw = 2000L,
                 knots = knots)
```

`plot.rcage` *Generate Plots Summarizing Main Results*

Description

Function generates plots showing the finest resolution, the support-level predictions, the support-level root prediction error, and the support-level CAGE/DCAGE.

Usage

```
## S3 method for class 'rcage'
plot(x, ..., dB, dataScale = NULL, palette = "plasma")
```

Arguments

`x`	An rcage object. The value returned by optRegion() or region().
`...`	ignored.
`dB`	A SpatialPolygons object or SpatialPoints object. The areal units of the finest resolution.
`dataScale`	A function or NULL. If not NULL, the function to be applied to the outcome interest to adjust the scaling in plots.
`palette`	A character. The palette preference for plotting. The palette is assumed to be viridis palette with possible values 'viridis', 'magma', 'plasma', 'inferno', 'cividis'

Value

No value object returned; called to generate plots.

Examples

```
# create 5x5 square

poly <- raster::rasterToPolygons(raster::raster(nrows = 5, ncols = 5,
                                                xmn = -1.25, xmx = 1.25,
                                                ymn = -1.25, ymx = 1.25,
                                                res = 0.5,
                                                crs = "+proj=longlat +datum=WGS84"))

df <- data.frame("x" = stats::rnorm(n = 25))

dt <- sp::SpatialPolygonsDataFrame(poly, df)

knots <- cbind(c(-0.75, 0.0, 0.75, -0.75, 0.0, 0.75, -0.75, 0.0, 0.75),
               c(-0.75,-0.75, -0.75, 0.0, 0.0, 0.0, 0.75, 0.75, 0.75))

res <- optRegion(spatialData = dt,
                 response = "x",
                 sigmavar = rep(1, 25),
                 gL = 5,
                 gU = 7,
                 nGibbs = 50L,
                 nBurn = 10L,
                 nThin = 1L,
                 nw = 2000L,
                 knots = knots)

 plot(x = res, dB = dt)
```

`radial` *Gaussian Radial Basis Functions*

Description

Provides an implementation of the Gaussian radial basis functions defined as

$$\Psi_j(s) = \exp\{-\frac{1}{2}(||s - c_j||/w)^2\}.$$

Usage

```
radial(crd, knots, w = NULL, ..., longlat = TRUE)
```

Arguments

crd	A matrix object. The (x,y) coordinates of the reference points {nCrd x 2}.
knots	A matrix object. The (x,y) coordinates of the knots {r x 2}.
w	A numeric object. The positive scaling factor (bandwidth).
...	ignored. Included only to require naming of inputs that follow.
longlat	A logical object. If FALSE, Euclidean distance is calculated; if TRUE, Great Circle distance is calculated. See ?sp::spDists for more information.

Details

Distances between reference coordinates and knots are obtained using sp::spDists().

Value

A matrix of Gaussian functions evaluated at all combinations of crd and knots { nCrd x r }.

Examples

```
data(countyExample)

radial(crd = sp::coordinates(county), knots = knots)
```

---

region *Regionalization of Multiscale Spatial Processes*

Description

Regionalization of multiscale spatial processes based on a criterion for spatial aggregation error. The multiscale representation is a truncated Karhunen-Loaeve expansion using Obled-Creutin eigenfunctions. The method is incorporated within a Bayesian framework using a Markov chain Monte Carlo implementation of a latent spatial model. This function differs from optRegion() in that clustering algorithms are not used to determine an optimal clustering, but spatial regions are clustered according to the provided regionalization and CAGE/DCAGE is calculated.

Usage

```
region(
  spatialData,
  response,
  sigmavar,
  dC,
  ...,
  cage = TRUE,
  longlat = TRUE,
```

```
  dB = NULL,
  gbf = "bisquare",
  w = NULL,
  knots = 75L,
  nw = 0L,
  nGibbs = 10000L,
  nBurn = 1000L,
  nThin = 1L,
  x = NULL,
  sigma2_beta = 1e-04,
  sigma2_xi = 1,
  Qprior = "MI",
  lambda = 1,
  wishartScale = 100,
  dMax = NULL,
  ffdir = NULL,
  nCore = 1L,
  parallelLog = NULL,
  plot = TRUE,
  dataScale = NULL,
  palette = "plasma"
)
```

Arguments

spatialData	SpatialXXDataFrame as defined by package sp or a list of said objects. Currently, this implementation is limited to use of a SpatialPolygonsDataFrame or a SpatialPointsDataFrame. For multi-resolution, a list of said SpatialXXDataFrames.
response	A numeric vector, character, or list of such. If numeric, the value of interest; the vector must contain the response for all spatialData. Specifically, if 'spatialData' is a list, vector must include the value of interest for all data of element [[1]] followed by that for all data of element [[2]], etc. If a character, the column header of the data slot that holds the value of interest; if a list is provided in 'spatialData' each Spatial object must have the specified column header. If provided as a list, the elements must correspond to the elements of the spatialData list.
sigmavar	A numeric vector/matrix, character, or list of said objects. The survey variance. If numeric, the vector is the diagonal elements of the variance matrix and it must contain the variance for all areal units of 'spatialData'. Specifically, if 'spatialData' is a list, the vector must include the variance for all data of element [[1]] followed by that for all data of element [[2]], etc. Similarly, if a matrix. If a character, the column header of the data slot that holds the diagonal elements of the variance; if a list is provided in 'spatialData' each Spatial object must have the specified column header. If provided as a list, the elements must correspond to the elements of the 'spatialData' list.
dC	A SpatialPolygons object or a vector defining the desired clustering of the spatial data. If vector, it must be of the length of dB and contain the cluster id for each areal unit of dB.
...	ignored. Used to require named input for remaining formals.

cage	A logical. If TRUE, CAGE is estimated. If FALSE, DCAGE is estimated.
longlat	A logicial. TRUE indicates that spatialData is long/lat coordinate system.
dB	NULL, integer, or a SpatialPolygons object defining the finest resolution spatial object to be used for inference. If NULL, 'spatialData' is the finest resolution considered. Note that if 'spatialData' is a list, 'dB' must be specified as either the element of 'spatialData' to be used as the finest resolution or as a SpatialPolygons object.
gbf	NULL, function or function name. The function to use to calculate the radial basis functions of the expansion of the Obled-Creutin eigenfunction. The default bi-square function is available through this implementation. All others must be defined by user. See details for further information.
w	A numeric. The scaling factor for radial basis functions. See details for further information.
knots	A matrix or integer. If a matrix, the knots of the radial basis functions. If an integer, the number of knots to generate using fields::cover.design().
nw	An integer. The number of MC replicates to generate for estimating the O-C eigenfunctions. If <=0, 'spatialData' must be or include SpatialPoints data.
nGibbs	An integer. The total number of Gibbs samples generated.
nBurn	An integer. The first sample accepted in the Gibbs sampling algorithm.
nThin	An integer. Keep every nThin sample of the Gibbs sampling algorithm once the burn specification has been satisfied.
x	NULL, character, or numeric. The covariates for the large scale variability model. If NULL, an intercept only model is assumed. If numeric (matrix), the covariates; the object must contain the covariates for all spatialData. Specifically, if 'spatialData' is a list, vector must include the covariates for all data of element [[1]] followed by that for all data of element [[2]], etc. If a character, the column header(s) of the data slot that holds the covariates; if a list is provided in 'spatialData' each Spatial object must have the specified column header. If provided as a list, the elements must correspond to the elements of the spatialData list.
sigma2_beta	A numeric. The variance of the large scale variability model.
sigma2_xi	A numeric. The variance of the fine scale variability. Defaults to inverse Gamma.
Qprior	A character. The Q prior. Must be one of 'MI' or 'wish' indicating the MI prior or the the wishart distribution respectively.
lambda	A numeric. The r eigenvalues of the prior distribution on the {r x r} covariance matrix Q.
wishartScale	A numeric of NULL. If Qprior is "wishart", the value of the diagonal elements of the scale matrix provided to MCMCpack::riwish. Default value is 100. Input is ignored for all other values of Qprior.
dMax	A numeric. If finest resolution data is SpatialPoints and Qprior is MI, the maximum distance at which points are considered adjacent. If not specified, it is taken to be upper boundary of the lowest decile.
ffdir	A character string. The directory in which ff object is to be saved. See details for further information.

nCore	An integer. If using MC, the algorithm can be spread across nCore cores to expedite calculations. If nCore = 1L, no parallelization methods are used.
parallelLog	A character object. A file name for logging parallel executions.
plot	A logical. TRUE indicates that final plot will be generated.
dataScale	A function. If the response of interest was scaled, the function to undo the scaling. Used in plotting only.
palette	A character. The palette preference for plotting. The palette is assumed to be viridis palette with possible values 'viridis', 'magma', 'plasma', 'inferno', 'cividis'

Details

Input 'spatialData' must be a single spatial object or a list of spatial objects as defined by the sp package. If it is a SpatialPointsDataFrame object and nw = 0/NULL, the package will use the point data to obtain Psi and W. If nw>0, the package will use Monte Carlo to estimate Psi and W.

If input 'spatialData' does not represent the finest resolution upon which inference is made, input 'dB' must be set to specify the finest resolution. The area spanned by 'dB' must fully contain all of the source support data, i.e., no elements of 'spatialData' can lie outside of the boundary defined by 'dB'. If 'spatialData' is multi-resolution 'dB' must be provided as the integer element of 'spatialData' to be used as the finest resolution or as a SpatialPolygons object defining the finest resolution.

Input 'gbf' allows users to specify radial basis function beyond the internally implemented bisquare, Wendland, and Gaussian radial functions. ('bisquare', 'wendland', 'gaussian') If user provides a function, the following formal arguments are required: crd - the coordinates at which the basis functions are to be evaluated; knots - the knots of the basis functions; and w - the scaling factor for the basis function. The function must return a matrix of dimension {length(crd) x nrow(knots)}.

For completeness, the bi-square function implemented in the package is of the form

$$\Psi_j(s) \equiv \begin{cases} \{1 - (||\text{crd} - \text{knots}_j||/w)^2\}^2 & \text{if } ||\text{crd} - \text{knots}_j|| \leq w, \\ 0 & \text{otherwise} \end{cases}.$$

The Wendland function is

$$\Psi_j(s) \equiv \begin{cases} \{1 - d_j(s)\}^6\{35d_j(s)^2 + 18d_j(s) + 3\}/3 \; 0 \leq d_j \leq 1, \\ 0 \quad \text{otherwise} \end{cases},$$

where

$$d_j(s) = ||s - c_j||/w.$$

The Gaussian radial function is

$$\Psi_j(s) = \exp\{-\frac{1}{2}(||s - c_j||/w)^2\}.$$

For clarity, the default MI prior function returns

$$K^{-1} = R_B^{-1}\mathcal{A}\{Q_B^{'}(I - A)Q_B\}R_B^{-1},$$

as defined in the supplemental section of the original manuscript.

This package implements methods of the ff package in scenarios when memory size is of concern. The ff package stores variables on the disk rather than in RAM. Though it is highly optimized, this choice does increase computation time. Intermediate variables used only internally are stored in the temp directory specified in input 'ffdir.' To trigger this implementation, 'ffdir' must be set.

Value

A list.

call	The original call structure
psi	A list containing the generating basis and OC weighting matrix.
CAGETrack	The estimated CAGE/DCAGE for each dC cluster.
cluster	Clustering indices mapping dB to dC.
yOpt	The estimated value of interest at each Gibbs sample clustered according to dC.
yFinest	The estimated value of interest at each Gibbs sample clustered according to dB.
criterion	"CAGE" or "DCAGE"

References

Bradley, J. R., Wikle, C. K., and Holan, S. H. (2017). Regionalization of Multiscale Spatial Processes using a Criterion for Spatial Aggregation Error. Journal of the Royal Statistical Society - Series B, 79, 815–832.

Examples

```
# create 5x5 square

poly <- raster::rasterToPolygons(raster::raster(nrows = 5, ncols = 5,
                                                xmn = -1.25, xmx = 1.25,
                                                ymn = -1.25, ymx = 1.25,
                                                res = 0.5,
                                                crs = "+proj=longlat +datum=WGS84"))

df <- data.frame("x" = stats::rnorm(n = 25))

dt <- sp::SpatialPolygonsDataFrame(poly, df)

dC <- raster::rasterToPolygons(raster::raster(xmn = -1.25, xmx = 1.25,
                                              ymn = -1.25, ymx = 1.25,
                                              res = c(0.5,2.5),
                                              crs = "+proj=longlat +datum=WGS84"))

knots <- cbind(c(-0.75, 0.0, 0.75, -0.75, 0.0, 0.75, -0.75, 0.0, 0.75),
               c(-0.75,-0.75, -0.75, 0.0, 0.0, 0.0, 0.75, 0.75, 0.75))

res <- region(spatialData = dt,
              response = "x",
              sigmavar = rep(1, 25),
              dC = dC,
              nGibbs = 50L,
```

```
nBurn = 10L,
nThin = 1L,
nw = 2000L,
knots = knots)
```

---

wendland *Wendland Basis Functions*

Description

Provides an implementation of the Wendland basis functions defined as

$$\Psi_j(s) = \{1 - d_j(s)\}^6\{35d_j(s)^2 + 18d_j(s) + 3\}/3\mathrm{I}(0 \leq d_j \leq 1),$$

where

$$d_j(s) = ||s - c_j||/w.$$

Usage

```
wendland(crd, knots, w = NULL, ..., longlat = TRUE)
```

Arguments

crd	A matrix object. The (x,y) coordinates of the reference points {nCrd x 2}.
knots	A matrix object. The (x,y) coordinates of the knots {r x 2}.
w	A numeric object. The positive scaling factor (bandwidth).
...	ignored. Included only to require naming of inputs that follow.
longlat	A logical object. If FALSE, Euclidean distance is calculated; if TRUE, Great Circle distance is calculated. See ?sp::spDists for more information.

Details

Distances between reference coordinates and knots are obtained using sp::spDists().

Value

A matrix of Wendland functions evaluated at all combinations of crd and knots { nCrd x r }.

References

Wendland, H. (1998). Error estimates for interpolation by compactly supported radial basis functions of minimal degree. Journal of Approximation Theory, 93,258-272. <doi:10.1006/jath.1997.3137>.

Examples

```
data(countyExample)

wendland(crd = sp::coordinates(county), knots = knots)
```

`windExample` *Wind Point Observations*

Description

This data set is a sample of ocean surface wind data with observations from 2 February 2005 at 12:00 Universal Coordinated Time over the Mediterranean Sea using the QuickSCAT scatterometer. Provided only to facilitate illustrative examples.

Usage

```
data(windExample)
```

Format

windExample contains 3 objects

windPt A SpatialPointsDataFrame containing 6916 points. The data.frame contains only one column, "vel", the observed wind velocity.

windPoly A SpatialPolygonsDataFrame containing 4551 areas. The data.frame contains only one column, "vel", the observed wind velocity.

windGrid A SptialPolygons object providing a 0.5 degree grid that covers the measurement area of the Mediterranean Sea.

References

Milliff, R., Bonazzi, A., Wikle,C., Pinardi,N. and Berliner, L. (2011). Ocean ensemble forecasting: Part I, Ensemble Mediterranean winds from a Bayesian hierarchical model. Quarterly Journal of the Royal Meteorological Society, 137, 858–878.

Wikle, C. K., Milliff, R. F., Herbei, R., and Leeds, W. B., (2013). Modern statistical methods in oceanography: A hierarchical perspective. Statistical Science, 28, 466-486. DOI: 10.1214/13-STS436

Index