# Package 'gdm'

March 17, 2021

**Type** Package

**Title** Generalized Dissimilarity Modeling

**Version** 1.4.2.2

**Date** 2021-03-17

**Author** Matthew C. Fitzpatrick, Karel Mokany, Glenn Manion, Matthew Lisk, Simon Ferrier, Diego Nieto-Lugilde

**Maintainer** Matthew C. Fitzpatrick <mfitzpatrick@umces.edu>

**Description**
A toolkit with functions to fit, plot, summarize, and apply Generalized Dissimilarity Models.

**License** GPL-2

**Depends** R (>= 2.15.2), raster, foreach, doParallel, parallel

**LinkingTo** Rcpp

**Imports** Rcpp (>= 0.10.4), reshape2, vegan, methods

**Suggests** R.rsp

**VignetteBuilder** R.rsp

**NeedsCompilation** yes

**Repository** CRAN

**Date/Publication** 2021-03-17 19:20:02 UTC

## R topics documented:

---

gdm-package                     *Overview of the functions in the gdm package*

---

#### Description

Generalized Dissimilarity Modeling is a statistical technique for modelling spatial variation in bio-
diversity between pairs of geographical locations. The **gdm** package currently provides basic func-
tions to fit, summarize, and plot Generalized Dissimilarity Models and to make predictions (in both
space and time) and map biological patterns by transforming environmental predictor variables.
Future updates will incorporate support for genomic data.

#### Details

The functions in the **gdm** package provide the tools necessary for fitting GDMs, including functions
to prepare biodiversity and environmental data. Major functionality includes:

- Formatting various types of biodiversity and environmental data to **gdm**'s site-pair format
  used in model fitting

- Fitting GDMs using geographic and environmental distances between sites

- Plotting fitted functions & extracting I-spline values

- Predicting pairwise dissimiliarites between sites or times and transforming envirnmental pre-
  dictors to biological importance and mapping these patterns.

To see the preferable citation of the package, type `citation("gdm")`.

#### I. Formatting input data

GDM fits biological distances to pairwise site geographical and environmental distances. Most
users will need to first format their data to **gdm**'s site-pair table format:

| | |
|---|---|
| formatsitepair | To convert biodiversity and environmental data to site-pair format |
| ———————— | ———————————————————————————————————— |

#### II. Model fitting and summary

| | |
|---|---|
| gdm | To fit a GDM model |
| summary | To summarize a GDM model |
| ———————— | ———————————————————————————————————— |

### III. Model prediction and transformation of environmental data

| | |
|---|---|
| predict | To predict biological dissimilarities between sites in space or between time periods |
| gdm.transform | To transform each environmental predictor to biological importance |

### IV. Plotting model output and fitted functions

| | |
|---|---|
| plot | To plot model fit and I-splines |
| isplineExtract | To extract I-spline values to allow for custom plotting |

### Author(s)

The **gdm** development team is Matt Fitzpatrick, Karel Mokany, Simon Ferrier, and Diego Nieto-Lugilde, using code originally developed by Glenn Manion. Where others have contributed to individual functions, credits are provided in function help pages.

The maintainer of the R version of **gdm** is Matt Fitzpatrick <mfitzpatrick@umces.edu>.

---

| formatsitepair | *Converts Common Biological (Response) and Environmental (Predictor) Data Formats to a GDM Site-Pair Table* |
|---|---|

---

### Description

This function takes input biological (response) and environmental, geographic, and other predictor variables and builds a site-pair table required for fitting a Generalized Dissimilarity Model using the gdm function. NOTE: x-y coordinates of sites MUST be present in either the biological or the environmental data.

The input biological data can be in one of the following four formats. Note that the general term "species" is used, but any classification of biological entities (e.g. functional types, haplotypes, etc) can be used as long as an appropriate distance metric is also supplied (see "dist" argument):

1. site-by-species matrix

2. x, y, species list

3. site-by-site biological distance (dissimilarity) matrix

4. an existing site-pair table (see Details)

Predictor data can be provided in three formats:

- a site-by-predictor matrix with a column for each predictor variable and a row for each site

- a raster stack, with one raster for each predictor variable

- one or more site-by-site distance matrices using the "distPreds" argument (see below).

## Usage

```
formatsitepair(bioData, bioFormat, dist="bray", abundance=FALSE, siteColumn=NULL, XColumn,
YColumn, sppColumn=NULL, abundColumn=NULL, sppFilter=0, predData, distPreds=NULL,
weightType="equal", custWeights=NULL, sampleSites=1)
```

## Arguments

| | |
|---|---|
| bioData | The input biological (response) data table, in one of the four formats defined above (see Details). |
| bioFormat | An integer code specifying the format of bioData. Acceptable values are 1, 2, 3, or 4 (see Details). |
| dist | Default = "bray". A character code indicating the metric to quantify pairwise site distances / dissimilarities. Calls the [vegdist](#) function from the **vegan** package to calculate dissimilarity and therefore accepts any method available from [vegdist](#). |
| abundance | Default = FALSE. Indicates whether the biological data are abundance data (TRUE) or presence-absence (0, 1) data (FALSE). |
| siteColumn | The name of the column in either the biological or environmental data table containing site codes/names. If a site column is provided in both the biological and environmental data, the site column name must be the same in both tables. |
| XColumn | The name of the column containing x-coordinates of sample sites. X-coordinates can be provided in either the biological or environmental data tables, but MUST be in at least one of them. If an x-coordinate column is provided in both the biological and environmental data, the column name must be identical. |
| YColumn | The name of the column containing y-coordinates of sample sites. Y-coordinates can be provided in either the biological or environmental data tables, but MUST be in at least one of them. If a y-coordinate column is provided in both the biological and environmental data, the column name must be identical. |
| sppColumn | Only used if bioFormat = 2 (x, y, species list). The name of the column containing unique name / identifier for each species. |
| abundColumn | If abundance = TRUE, this parameter identifies the column containing the measure of abundance at each site. Only used if bioFormat = 2 (i.e., x, y, species list), though in the case of abundance data, the format would be: x, y, species, abundance. |
| sppFilter | Default = 0. To account for limited sampling effort at some sites, sppFilter removes all sites at which the number of recorded species (i.e., observed species richness) is less than the specified value. For example, if sppFilter = 5, all sites with fewer than 5 recorded species will be removed. |

| | |
|---|---|
| predData | The environmental predictor data. Accepts either a site-by-predictor table or a raster stack. |
| distPreds | An optional list of distance matrices to be used as predictors in combination with predData. For example, a site-by-site dissimilarity matrix for one biological group (e.g., trees) can be used as a predictor for another group (e.g., ferns). Each distance matrix must have as the first column the names of the sites (therefore the matrix will not be square). The name of the column containing the site names should have the same name as that provided for siteColumn argument. Site IDs are required here to ensure correct ordering of sites in the construction of the site-pair table. Note that the formatsitepair function will not accept only distances matrices as predictors (i.e., at least one predictor variable is required). If you wish to fit GDM using only distance matrices, provide one fake predictor (e.g., with all sites have the same value), plus site and coordinate columns if needed. The s1 and s2 columns for this variable can then be removed before fitting the GDM. |
| weightType | Default = "equal". Defines the weighting for sites. Can be either: (1) "equal" (weights for all sites set = 1), (2) "richness" (each site weighted according to number of species recorded), or (3) "custom" (user defined). If weight-Type="custom", the user must provide a vector of site weights equal to the number of rows in the full site-pair table (i.e., before species filtering (sppFilter argument) or sub-sampling is taken into account (sampleSites argument)). |
| custWeights | A two column matrix or data frame of user-defined site weights. The first column should be the site ID and should be named the same as that provided for siteColumn argument. The second column should be numeric weight values and should be named "weights". The weight values represent the importance of each site in model fitting, and the values in the output site-pair table is an average of the two sites in each site-pair. Required when weightType = "custom". Ignored otherwise. |
| sampleSites | Default = 1. A number $0 < x <= 1$ indicating the fraction of sites to be used to construct the site-pair table. This argument can be used to reduce the number of sites to overcome memory limitations when fitting models with very large numbers of sites. |

### Details

bioData and bioFormat: The function accepts biological data in the following formats:

bioData = site-by-species matrix; bioFormat = 1: assumes that the response data are provided with a site ID column (specified by siteCol) and, optionally, two columns for the x & y coordinates of the sites. All remaining columns contain the biological data, with a column for each biological entity (most commonly species). In the case that a raster stack is provided for the environmental data (predData), x-y coordinates MUST be provided in bioData to allow extraction of the environmental data at site locations. The x-y coordinates will be intersected with the raster stack and, if the number of unique cells intersected by the points is less than the number of unique site IDs (i.e. multiple sites fall within a single cell), the function will use the raster cell as the site ID and aggregate sites accordingly. Therefore, model fitting will be sensitive to raster cell size. If the environmental data are in tabular format, they should have the same number of sites (i.e., same number of rows) as bioData. The x-y coordinate and site ID columns must have the same names in bioData and predData.

bioData = x, y, species list (optionally a fourth column with abundance can be provided); bioFormat = 2: assumes a table of 3 or 4 columns, the first two being the x & y coordinates of species records, the third (sppCol) being the name / identifier of the species observed at that location, and optionally a fourth column indicating a measure of abundance. If an abundance column is not provided, presence-only data are assumed. In the case that a raster stack is provided for the environmental data (predData), the x-y coordinates will be intersected with the raster stack and, if the number of unique cells intersected by the points is less than the number of unique site IDs (i.e. multiple sites fall within a single cell), the function will use the raster cell as the site ID and aggregate sites accordingly. Therefore, model fitting will be sensitive to raster cell size.

bioData = site-by-site distance (dissimilarity) matrix; bioFormat = 3: is used when a site-by-site distance (dissimilarity) matrix has already been created for the biological response (e.g., Fst for genetic data). The distance matrix must have as the first column the names of the sites (therefore the matrix will not be square). The column of site names should have the same name as the siteColumn argument. Only the lower half (triangle) of the matrix is needed to create the site-pair output table, but this function automatically removes the upper half if present. This is the only bioFormat in which the environmental data CANNOT be provided as a raster object.

bioData = site-pair table; bioFormat = 4: with an already created site-pair table, this option allows the user to add one or more distance matrices (see distPreds above) to the existing site-pair table and/or sub-sample the site-pair table (see sample above). If the site-pair table was not created using the formatsitepair function, the user will need to ensure the order of the sites matches that in other tables being provided to the function.

NOTES: (1) The function assumes that the x-y coordinates and the raster stack (if used) are in the same coordinate system. No checking is performed to confirm this is the case. (2) The function assumes that the association between the provided site and x-y coordinate columns are singular and unique. Therefore, the function will fail should a given site have multiple coordinates associated with it, as well as multiple sites being given the exact same coordinates.

**Value**

A site-pair formatted table containing the response (biological distance or dissimilarity), predictors, and weights as required for fitting Generalized Dissimilarity Models.

**Examples**

```
## table data, species and environmental
load(system.file("./data/gdm.RData", package="gdm"))
sppData <- gdmExpData[, c(1,2,13,14)]
envTab <- gdmExpData[, c(2:ncol(gdmExpData))]

## environmental raster data
## commented out to reduce example run time
#rastFile <- system.file("./extdata/stackedVars.grd", package="gdm")
#envRast <- stack(rastFile)

#########bioFormat = 1
## bioData = site-species table without coordinates
testData1a <- reshape2::dcast(sppData, site~species, fun.aggregate = length)
## site-species table with coordinates
```

```
coords <- unique(sppData[, 2:ncol(sppData)])
testData1b <- merge(testData1a, coords, by="site")
## site-species table,  env data table
exFormat1a <- formatsitepair(testData1a,
                             bioFormat = 1,
                             siteColumn="site",
                             XColumn="Long",
                             YColumn="Lat",
                             predData=envTab)

## site-species table, env raster
## not run
#exFormat1b <- formatsitepair(testData1b,
#                                 bioFormat = 1,
#                                 siteColumn="site",
#                                 XColumn="Long",
#                                 YColumn="Lat",
#                                 predData=envRast)

#########bioFormat = 2
## bioData = site xy spp list, env data table
exFormat2a <- formatsitepair(sppData,
                             bioFormat = 2,
                             XColumn="Long",
                             YColumn="Lat",
                             sppColumn="species",
                             siteColumn="site",
                             predData=envTab)
## site xy spp list, env raster
## commented out to reduce example run time
#exFormat2b <- formatsitepair(sppData,
#                                 bioFormat = 2,
#                                 XColumn="Long",
#                                 YColumn="Lat",
#                                 sppColumn="species",
#                                 siteColumn="site",
#                                 predData=envRast)

#########bioFormat = 3
## bioData = dissim matrix model
site <- unique(sppData$site)
gdmDissim <- cbind(site, gdmDissim)
exFormat3 <- formatsitepair(gdmDissim,
                            bioFormat = 3,
                            XColumn="Long",
                            YColumn="Lat",
                            predData=envTab,
                            siteColumn="site")

#########bioFormat = 4
## adds a distance matrix predictor to an existing site-pair table
## in this case, predData needs to be
## filled, but is not actually used
```

```
gdmDissim <- cbind(site, gdmDissim)
exFormat4 <- formatsitepair(exFormat2a,
                            bioFormat = 4,
                            predData=envTab,
                            siteColumn="site",
                            distPreds=list(as.matrix(gdmDissim)))
```

---

gdm                    *Fit a Generalized Dissimilarity Model to Tabular Site-Pair Data*

---

### Description

For an overview of the functions in the gdm package have a look here: gdm-package.

The gdm function is used to fit a generalized dissimilarity model to tabular site-pair data formatted as follows using the formatsitepair function: distance, weights, s1.xCoord, s1.yCoord, s2.xCoord, s2.yCoord, s1.Pred1, s1.Pred2, ..., s1.PredN, s2.Pred1, s2.Pred2, ..., s2.PredN. The first column (Response) must be any ratio based dissimilarity (distance) measure between Site 1 and Site 2. The second column defines any weighting to be applied during fitting of the model. If equal weighting is required, then all entries in this column should be set to 1.0 (default). The third and fourth columns, s1.X and s1.Y, represent the spatial coordinates of the first site in the site pair (s1). The fifth and sixth columns, s2.X and s2.Y, represent the coordinates of the second site (s2). Note that the first six columns are REQUIRED, even if you do not intend to use geographic distance as a predictor (in which case these columns can be loaded with dummy data if the actual coordinates are unknown - though that would be weird, no?). The next N*2 columns contain values for N predictors for Site 1, followed by values for the same N predictors for Site 2.

The following is an example of a GDM input table header with three environmental predictors (Temp, Rain, Bedrock):

distance, weights, s1.xCoord, s1.yCoord, s2.xCoord, s2.yCoord, s1.Temp, s1.Rain, s1.Bedrock, s2.Temp, s2.Rain, s2.Bedrock

### Usage

```
gdm(data, geo=FALSE, splines=NULL, knots=NULL)
```

### Arguments

data            A data frame containing the site pairs to be used to fit the GDM (typically obtained using the formatsitepair function). The observed response data must be located in the first column. The weights to be applied to each site pair must be located in the second column. If geo is TRUE, then the s1.xCoord, s1.yCoord and s2.xCoord, s2.yCoord columns will be used to calculate the geographic distance between site pairs for inclusion as the geographic predictor term in the model. If geo is FALSE (default), then the s1.xCoord, s1.yCoord, s2.xCoord and s2.yCoord data columns must still be included, but are ignored in fitting the model. Columns containing the predictor data for Site 1, and the predictor data for Site 2, follow.

| | |
|---|---|
| geo | Set to TRUE if geographic distance between sites is to be included as a model term. Set to FALSE if geographic distance is to be omitted from the model. Default is FALSE. |
| splines | An optional vector of the number of I-spline basis functions to be used for each predictor in fitting the model. If supplied, it must have the same length as the number of predictors (including geographic distance if geo is TRUE). If this vector is not provided (splines=NULL), then a default of 3 basis functions is used for all predictors. |
| knots | An optional vector of knots in *units of the predictor variables* to be used in the fitting process. If knots are supplied and splines=NULL, then the knots argument must have the same length as the number of predictors * 3. If both knots and the number of splines are supplied, then the length of the knots argument must be the same as the sum of the values in the splines vector. Note that the default values for knots when the default three I-spline basis functions are 0 (minimum), 50 (median), and 100 (maximum) quantiles. |

## Value

gdm returns a gdm model object. The function summary.gdm can be used to obtain or print a synopsis of the results. A gdm model object is a list containing at least the following components:

| | |
|---|---|
| dataname | The name of the table used as the data argument to the model. |
| geo | Whether geographic distance was used as a predictor in the model. |
| gdmdeviance | The deviance of the fitted GDM model. |
| nulldeviance | The deviance of the null model. |
| explained | The percentage of null deviance explained by the fitted GDM model. |
| intercept | The fitted value for the intercept term in the model. |
| predictors | A list of the names of the predictors that were used to fit the model. |
| coefficients | A list of the coefficients for each spline for each of the predictors considered in model fitting. |
| knots | A vector of the knots derived from the x data (or user defined), for each predictor. |
| splines | A vector of the number of I-spline basis functions used for each predictor. |
| creationdate | The date and time of model creation. |
| observed | The observed response for each site pair (from data column 1). |
| predicted | The predicted response for each site pair, from the fitted model (after applying the link function). |
| ecological | The linear predictor (ecological distance) for each site pair, from the fitted model (before applying the link function). |

## References

Ferrier S, Manion G, Elith J, Richardson, K (2007) Using generalized dissimilarity modelling to analyse and predict patterns of beta diversity in regional biodiversity assessment. *Diversity & Distributions* 13, 252-264.

**See Also**

formatsitepair, summary.gdm, plot.gdm, predict.gdm, gdm.transform

**Examples**

```
##fit table environmental data
##sets up site-pair table, environmental tabular data
load(system.file("./data/gdm.RData", package="gdm"))
sppData <- gdmExpData[c(1,2,13,14)]
envTab <- gdmExpData[c(2:ncol(gdmExpData))]
sitePairTab <- formatsitepair(sppData, 2, XColumn="Long", YColumn="Lat", sppColumn="species",
siteColumn="site", predData=envTab)

##fit table GDM
gdmTabMod <- gdm(sitePairTab, geo=TRUE)
summary(gdmTabMod)

##fit raster environmental data
##sets up site-pair table
rastFile <- system.file("./extdata/stackedVars.grd", package="gdm")
envRast <- stack(rastFile)

##environmental raster data
sitePairRast <- formatsitepair(sppData, 2, XColumn="Long", YColumn="Lat", sppColumn="species",
siteColumn="site", predData=envRast)
##sometimes raster data returns NA in the site-pair table, these rows will have to be removed
##before fitting gdm
sitePairRast <- na.omit(sitePairRast)

##fit raster GDM
gdmRastMod <- gdm(sitePairRast, geo=TRUE)
summary(gdmRastMod)
```

---

gdm.transform | *Transform Environmental Data Using a Generalized Dissimilarity Model*

---

**Description**

This function transforms geographic and environmental predictors using (1) the fitted functions from a model object returned from gdm and (2) a data frame or raster stack containing predictor data for a set of sites.

**Usage**

```
gdm.transform(model, data)
```

## Arguments

| | |
|---|---|
| `model` | A gdm model object resulting from a call to gdm. |
| `data` | Either (i) a data frame containing values for each predictor variable in the model, formatted as follows: X, Y, var1, var2, var3, ..., varN or (ii) a raster stack with one layer per predictor variable used in the model, excluding X and Y (rasters for x- and y-coordinates are built automatically from the input rasters if the model was fit with geo=T). The order of the columns (data frame) or raster layers (raster stack) MUST be the same as the order of the predictors in the site-pair table used in model fitting. There is currently no checking to ensure that the order of the variables to be transformed are the same as those in the site-pair table used in model fitting. If geographic distance was not used as a predictor in model fitting, the x- and y-columns need to be removed from the data to be transformed. Output is provided in the same format as the input data. |

## Value

gdm.transform returns either a data frame with the same number of rows as the input data frame or a raster stack, depending on the format of the input data. If the model uses geographic distance as a predictor the output object will contain columns or layers for the transformed X and Y values for each site. The transformed environmental data will be in the remaining columns or layers.

## References

Ferrier S, Manion G, Elith J, Richardson, K (2007) Using generalized dissimilarity modelling to analyse and predict patterns of beta diversity in regional biodiversity assessment. *Diversity & Distributions* 13, 252-264.

Fitzpatrick MC, Keller SR (2015) Ecological genomics meets community-level modeling of biodiversity: Mapping the genomic landscape of current and future environmental adaptation. *Ecology Letters* 18: 1-16

## Examples

```
load(system.file("./data/gdm.RData", package="gdm"))

# grab the columns with xy, site ID, and species data
sppTab <- gdmExpData[, c("species", "site", "Lat", "Long")]

##fit gdm using rasters
rastFile <- system.file("./extdata/stackedVars.grd", package="gdm")
envRast <- stack(rastFile)
sitePairRast <- formatsitepair(sppTab, 2, XColumn="Long", YColumn="Lat", sppColumn="species",
siteColumn="site", predData=envRast)
##remove NA values
sitePairRast <- na.omit(sitePairRast)

##fit raster GDM
gdmRastMod <- gdm(sitePairRast, geo=TRUE)

##raster input, raster output
```

```
transRasts <- gdm.transform(gdmRastMod, envRast)

# map biological patterns
rastDat <- sampleRandom(transRasts, 10000)
pcaSamp <- prcomp(rastDat)

# note the use of the 'index' argument
pcaRast <- predict(transRasts, pcaSamp, index=1:3)

# scale rasters
pcaRast[[1]] <- (pcaRast[[1]]-pcaRast[[1]]@data@min) /
  (pcaRast[[1]]@data@max-pcaRast[[1]]@data@min)*255
pcaRast[[2]] <- (pcaRast[[2]]-pcaRast[[2]]@data@min) /
  (pcaRast[[2]]@data@max-pcaRast[[2]]@data@min)*255
pcaRast[[3]] <- (pcaRast[[3]]-pcaRast[[3]]@data@min) /
  (pcaRast[[3]]@data@max-pcaRast[[3]]@data@min)*255

plotRGB(pcaRast, r=1, g=2, b=3)
```

---

gdm.varImp                  *Quantify model significance and variable importance/significance in*
                            *gdm using matrix permutation.*

---

### Description

This function uses matrix permutation to perform model and variable significance testing and to
estimate variable importance in a generalized dissimilarity model. The function can be run in par-
allel on multicore machines to reduce computation time (recommended until we learn to program
in C++).

### Usage

```
gdm.varImp(spTable, geo, splines = NULL, knots = NULL, fullModelOnly = FALSE,
nPerm = 50, parallel = FALSE, cores = 2, sampleSites = 1, sampleSitePairs = 1,
outFile = NULL)
```

### Arguments

spTable          A site-pair table, same as used to fit a gdm

geo              Similar to the gdm geo argument. The only difference is that the geo argument
                 does not have a default in this function.

splines          Same as the gdm splines argument.

knots            Same as the gdm knots argument.

fullModelOnly    Set to TRUE to test only the full variable set. Set to FALSE to estimate model
                 significance and variable importance and significance using matrix permutation
                 and backward elimination. Default is FALSE.

| | |
|---|---|
| nPerm | Number of permutations to use to estimate p-values. Default is 50. |
| parallel | Whether or not to run the matrix permutations and model fitting in parallel. Parallel processing is highly recommended when either (i) the nPerms argument is large (>100) or (ii) a large number of site-pairs (and or variables) are being used in model fitting (note computation demand can be reduced using subsampling - see next arguments). The default is FALSE. |
| cores | When the parallel argument is set to TRUE, the number of cores to be registered for parallel processing. Must be <= the number of cores in the machine running the function. |
| sampleSites | The fraction (0-1, though a value of 0 would be silly, wouldn't it?) of *sites to retain* from the full site-pair table. If less than 1, this argument will completely remove a fraction of sites such that they are not used in the permutation routines. |
| sampleSitePairs | |
| | The fraction (0-1) of *site-pairs (i.e., rows) to retain* from the full site-pair table - in other words, all sites will be used in the permutation routines (assuming sampleSites = 1), but not all *site-pair combinations*. In the case where both the sampleSites and the sampleSitePairs argument have values less than 1, sites first will be removed using the sampleSites argument, followed by removal of site-pairs using the sampleSitePairs argument. Note that the number of site-pairs removed is based on the fraction of the resulting site-pair table after sites have been removed, not on the size of the full site-pair table. |
| outFile | An optional character string to write the object returned by the function to disk as an .RData object (".RData"" is not required as part of the file name). The .RData object will contain a single list with the name of "outObject". The default is NULL, meaning that no file will be written. |

## Details

To test model significance, first a "full model" is fit using un-permuted environmental data. Next, the environmental data are permuted nPerm times (by randomizing the order of the rows) and a GDM is fit to each permuted table. Model significance is determined by comparing the deviance explained by GDM fit to the un-permuted table to the distribution of deviance explained values from GDM fit to the nPerm permuted tables. To assess variable significance, this process is repeated for each predictor individually (i.e., only the data for the variable being tested is permuted rather than the entire environmental table). Variable importance is quantified as the percent change in deviance explained between a model fit with and without that variable (technically speaking, with the variable permuted and un-permuted). If fullModelOnly=FALSE, this process continues by then permutating the site-pair table nPerm times, but removing one variable at a time and reassessing variable importance and significance. At each step, the least important variable is dropped (backward elimination) and the process continues until all non-significant predictors are removed.

## Value

A list of four tables. The first table summarizes full model deviance, percent deviance explained by the full model, the p-value of the full model, and the number of permutations used to calculate the statistics for each fitted model (i.e., the full model and each model with variables removed in succession during the backward elimination procedure if fullModelOnly=F). The remaining three

tables summarize (1) variable importance, (2) variable significance, and (3) the number of permutations used to calculate the statistics for that model, which is provided because some GDMs may fail to fit for some permutations / variable combinations and you might want to know how many permutations were used when clacuating statistics. Or maybe you don't, you decide.

Variable importance is measured as the percent change in deviance explained by the full model and the deviance explained by a model fit with that variable permuted. Significance is estimated using the bootstrapped p-value when the variable has been permuted. For most cases, the number of permutations will equal the nPerm argument. However, the value may be less should any of the permutations fail to fit.

If fullModelOnly=T, the tables will have values only in the first column and NAs elsewhere.

NOTE: In some cases, GDM may fail to fit if there is a weak relationship between the response and predictors (e.g., when an important variable is removed). Such cases are indicated by -9999 values in the variable importance, variable significance, and number of permutations tables.

### Author(s)

Karel Mokany, Matthew Lisk, and Matt Fitzpatrick

### References

Ferrier S, Manion G, Elith J, Richardson, K (2007) Using generalized dissimilarity modelling to analyse and predict patterns of beta diversity in regional biodiversity assessment. *Diversity & Distributions* 13, 252-264.

Fitzpatrick, MC, Sanders NJ, Ferrier S, Longino JT, Weiser MD, and RR Dunn. 2011. Forecasting the Future of Biodiversity: a Test of Single- and Multi-Species Models for Ants in North America. *Ecography* 34: 836-47.

### Examples

```
##fit table environmental data
##sets up site-pair table, environmental tabular data
load(system.file("./data/gdm.RData", package="gdm"))
sppData <- gdmExpData[c(1,2,13,14)]
envTab <- gdmExpData[c(2:ncol(gdmExpData))]
sitePairTab <- formatsitepair(sppData, 2, XColumn="Long", YColumn="Lat", sppColumn="species",
siteColumn="site", predData=envTab)

## not run
#modTest <- gdm.varImp(sitePairTab, geo=T, nPerm=50, parallel=T, cores=10)
#barplot(sort(modTest[[2]][,1], decreasing=T))
```

---

gdmDissim                            *A sample biological dissimilarity matrix for use in GDM examples*

---

### Description

A full dissimilarity matrix.

**Format**

A data frame with 180 rows and 180 columns.

**Source**

Created using the example species data from southwest Australia.

---

gdmExpData                    *Species and environmental data used in GDM examples*

---

**Description**

A dataset containing the species, site, and environmental data used to fit GDM in examples.

**Format**

A data frame with 42,142 rows and 14 variables

**Details**

-species: Species identifier.

-site: Site identification.

-awcA: plant-available water capacity of soil

-phTotal: soil pH

-sandA: percent sand content of soil

-shcA: saturated hydraulic conductivity of soil

-solumDepth: depth of solum layer (unknown units)

-bio5: Maximum Temperature of Warmest Month (C x 10)

-bio6: Minimum Temperature of Coldest Month (C x 10)

-bio15: Precipitation Seasonality (Coefficient of Variation)

-bio18: Precipitation of Warmest Quarter (mm)

-bio19: Precipitation of Coldest Quarter (mm)

-Lat: Latitude Coordinate of site

-Long: Longitude Coordinate of site

**Source**

Fitzpatrick MC, Sanders NJ, Normand S, Svenning J-C, Ferrier S, Gove AD, Dunn RR (2013). Environmental and historical imprints on beta diversity: insights from variation in rates of species turnover along gradients. Proceedings of the Royal Society: Series B 280, art. 1768

---

**isplineExtract**                                    *Extract I-spline Values From a gdm Object.*

---

### Description

Extracts the I-spline values from a gdm object. There is one I-spline for each predictor that has at least one non-zero coefficient in the fitted model.

### Usage

```
isplineExtract(model)
```

### Arguments

model                    A gdm object from gdm.

### Value

A list with two items. The first item contains the x-values (actual values of the predictors) of the I-splines and the second item contains the y-values (partial ecological distances) of the fitted I-splines.

### References

Ferrier S, Manion G, Elith J, Richardson, K (2007) Using generalized dissimilarity modelling to analyse and predict patterns of beta diversity in regional biodiversity assessment. *Diversity & Distributions* 13, 252-264.

Fitzpatrick MC, Sanders NJ, Normand S, Svenning J-C, Ferrier S, Gove AD, Dunn RR (2013). Environmental and historical imprints on beta diversity: insights from variation in rates of species turnover along gradients. Proceedings of the Royal Society: Series B 280, art. 1768

### Examples

```
##sets up site-pair table
load(system.file("./data/gdm.RData", package="gdm"))
sppData <- gdmExpData[, c(1,2,14,13)]
envTab <- gdmExpData[, c(2:ncol(gdmExpData))]
sitePairTab <- formatsitepair(sppData, 2, XColumn="Long", YColumn="Lat", sppColumn="species",
siteColumn="site", predData=envTab)

##create GDM
gdmMod <- gdm(sitePairTab, geo=TRUE)

##extracts splines
exSplines <- isplineExtract(gdmMod)

##plot spline(s)
#spline for winter precip (bio19)
```

```
plot(exSplines[[1]][,"bio19"], exSplines[[2]][,"bio19"], type="l",
lwd=3, xlab="Winter precipitation (mm)", ylab="Partial Ecological Distance")
```

---

plot.gdm                    *Plot model fit and I-splines from a generalized dissimilarity model*

---

### Description

plot is used to plot the I-splines and fit of a generalized dissimilarity model created using the gdm function.

### Usage

```
## S3 method for class 'gdm'
## S3 method for class 'gdm'
plot(x, plot.layout = c(2, 2), plot.color = "blue", plot.linewidth = 2,
include.rug = FALSE, rug.sitepair = NULL, ...)
```

### Arguments

| | |
|---|---|
| x | A gdm model object returned from gdm. |
| plot.layout | This argument specifies the row and column layout for the plots, including: (1) a single page plot of observed response data against the raw linear predictor (ecological distance) from the model, and (2) a single page plot of the observed response against the predicted response from the model, i.e. after applying the link function, 1.0 - exp(-y), to the linear predictor, and (3) the I-splines fitted to the individual predictors. Default is 2 rows by 2 columns. To produce one predictor plot per page set plot.layout to c(1,1). The first two model plots are always produced on a single page each and therefore the layout parameter affects only the layout of the I-spline plots for those predictors that featured in the model fitting process (i.e., predictors with all-zero I-spline coefficients are not plotted). |
| plot.color | Color of the data points that are plotted for the overall plots. |
| plot.linewidth | The line width for the regression line over-plotted in the two overall plots to optimize the display of the line over the data points. |
| include.rug | Whether or not to include a rug plot of the predictor values used to fit the gdm in the I-spline plots. When set to TRUE, a site-pair table must be supplied for the rug.sitepair argument. Default is FALSE. |
| rug.sitepair | A site-pair table used to add a rug plot of the predictor values used to fit the gdm in the I-spline plots. This should be the same site-pair table used to fit the gdm model being plotted. The function does not check whether the supplied site-pair table matches that used in model fitting. |
| ... | Ignored. |

### Value

plot returns NULL. Use summary.gdm to obtain a synopsis of the model object.

**References**

Ferrier S, Manion G, Elith J, Richardson, K (2007) Using generalized dissimilarity modelling to analyse and predict patterns of beta diversity in regional biodiversity assessment. *Diversity & Distributions* 13, 252-264.

**See Also**

[isplineExtract](#)

**Examples**

```
##sets up site-pair table
load(system.file("./data/gdm.RData", package="gdm"))
sppData <- gdmExpData[c(1,2,13,14)]
envTab <- gdmExpData[c(2:ncol(gdmExpData))]
sitePairTab <- formatsitepair(sppData, 2, XColumn="Long", YColumn="Lat", sppColumn="species",
siteColumn="site", predData=envTab)

##create GDM
gdmMod <- gdm(sitePairTab, geo=TRUE)

##plot GDM
plot(gdmMod, plot.layout=c(3,3))
```

---

| plotUncertainty | *Plot I-splines with error bands using bootstrapping.* |
|---|---|

---

**Description**

This function estimates uncertainty in the fitted I-splines using bootstrapping. The function can run in parallel on multicore machines to reduce computation time (recommended for large number of iterations). I-spline plots with error bands (+/- one standard deviation) are produced showing (1) the variance of I-spline coefficients and (2) a rug plot indicating how sites used in model fitting are distributed along each gradient.

**Usage**

```
plotUncertainty(spTable, sampleSites, bsIters, geo=FALSE, splines=NULL, knots=NULL,
          splineCol="blue", errCol="grey80", plot.linewidth=2.0, plot.layout=c(2,2),
               parallel=FALSE, cores=2)
```

**Arguments**

| | |
|---|---|
| spTable | A site-pair table, same as used to fit a [gdm](#). |
| sampleSites | The fraction (0-1) of sites to retain from the full site-pair table when subsampling. |
| bsIters | The number of bootstrap iterations to perform. |

| | |
|---|---|
| geo | Same as the [gdm] geo argument. |
| splines | Same as the [gdm] splines argument. |
| knots | Same as the [gdm] knots argument. |
| splineCol | The color of the plotted mean spline. The default is "blue". |
| errCol | The color of shading for the error bands (+/- one standard deviation around the mean line). The default is "grey80". |
| plot.linewidth | The line width of the plotted mean spline line. The default is 2. |
| plot.layout | Same as the [plot.gdm] plot.layout argument. |
| parallel | Whether or not to run the uncertainty estimation in parallel. The parallel processing is highly recommended when the bsIters argument is hundreds or more. The default is FALSE (no parallel processing). |
| cores | When the parallel argument is set to TRUE, the number of cores to be registered for the foreach loop. Must be <= the number of cores in the machine running the function. |

## Value

plotUncertainty returns NULL.

## References

Shryock, D. F., C. A. Havrilla, L. A. DeFalco, T. C. Esque, N. A. Custer, and T. E. Wood. 2015. Landscape genomics of *Sphaeralcea ambigua* in the Mojave Desert: a multivariate, spatially-explicit approach to guide ecological restoration. *Conservation Genetics* 16:1303-1317.

## See Also

[plot.gdm], [formatsitepair], [removeSitesFromSitePair]

## Examples

```
##sets up site-pair table
load(system.file("./data/gdm.RData", package="gdm"))
sppData <- gdmExpData[c(1,2,13,14)]
envTab <- gdmExpData[c(2:ncol(gdmExpData))]
sitePairTab <- formatsitepair(sppData, 2, XColumn="Long", YColumn="Lat",
                              sppColumn="species", siteColumn="site", predData=envTab)

##plot GDM uncertainty using one core
#not run
#plotUncertainty(sitePairTab, sampleSites=0.70, bsIters=5, geo=TRUE, plot.layout=c(3,3))

##plot GDM uncertainty in parallel
#not run
#plotUncertainty(sitePairTab, sampleSites=0.70, bsIters=50, geo=TRUE, plot.layout=c(3,3),
#parallel=T, cores=10)
```

---

| predict.gdm | *Predict Biological Dissimilarities Between Sites or Times Using a Generalized Dissimilarity Model* |

---

## Description

This function predicts biological distances between sites or times using a model object returned from gdm. Predictions between site pairs require a data frame containing the values of predictors for pairs of locations, formatted as follows: distance, weights, s1.X, s1.Y, s2.X, s2.Y, s1.Pred1, s1.Pred2, ..., s1.PredN, s2.Pred1, s2.Pred2, ..., s2.PredN.. Predictions of biological change through time require two raster stacks or bricks for environmental conditions at two time periods, each with a layer for each environmental predictor in the fitted model.

## Usage

```
## S3 method for class 'gbm'
## S3 method for class 'gdm'
predict(object, data, time=FALSE, predRasts=NULL, ...)
```

## Arguments

| | |
|---|---|
| object | A gdm model object resulting from a call to gdm. |
| data | Either a data frame containing the values of predictors for pairs of sites, in the same format and structure as used to fit the model using gdm or a raster stack if a prediction of biological change through time is needed. |
| | For a data frame, the first two columns - distance and weights - are required by the function but are not used in the prediction and can therefore be filled with dummy data (e.g. all zeros). If geo is TRUE, then the s1.X, s1.Y and s2.X, s2.Y columns will be used for calculating the geographical distance between each site for inclusion of the geographic predictor term into the GDM model. If geo is FALSE, then the s1.X, s1.Y, s2.X and s2.Y data columns are ignored. However these columns are still REQUIRED and can be filled with dummy data (e.g. all zeroes). The remaining columns are for N predictors for Site 1 and followed by N predictors for Site 2. The order of the columns must match those in the site-pair table used to fit the model. |
| | A raster stack should be provided only when time=T and should contain one layer for each environmental predictor in the same order as the columns in the site-pair table used to fit the model. |
| time | TRUE/FALSE: Is the model prediction for biological change through time? |
| predRasts | A raster stack characterizing environmental conditions for a different time in the past or future, with the same extent, resolution, and layer order as the data object. Required only if time=T. |
| ... | Ignored. |

## Value

predict returns either a response vector with the same length as the number of rows in the input data frame or a raster depicting change through time across the study region.

## See Also

[gdm.transform](gdm.transform)

## Examples

```
##sets up site-pair table
load(system.file("./data/gdm.RData", package="gdm"))
sppData <- gdmExpData[, c(1,2,14,13)]
envTab <- gdmExpData[, c(2:ncol(gdmExpData))]

# remove soils (no rasters for these)
envTab <- envTab[,-c(2:6)]
sitePairTab <- formatsitepair(sppData, 2, XColumn="Long", YColumn="Lat", sppColumn="species",
siteColumn="site", predData=envTab)

##create GDM
gdmMod <- gdm(sitePairTab, geo=TRUE)

##predict GDM
predDiss <- predict(gdmMod, sitePairTab)

##time example
rastFile <- system.file("./extdata/stackedVars.grd", package="gdm")
envRast <- stack(rastFile)

##make some fake climate change data
futRasts <- envRast
##reduce winter precipitation by 25%
futRasts[[3]] <- futRasts[[3]]*0.75

timePred <- predict(gdmMod, envRast, time=TRUE, predRasts=futRasts)
plot(timePred)
```

---

removeSitesFromSitePair

*Remove Random Sites from Site-Pair Table*

---

## Description

Randomly selects a number of sites from a given site-pair table and removes them from the site-pair table. It will remove all instances of the sites randomly selected to be removed in both s1 and s2 positions.

**Usage**

```
removeSitesFromSitePair(spTable, sampleSites)
```

**Arguments**

| | |
|---|---|
| spTable | A site-pair table, same as used to fit a gdm. |
| sampleSites | The fraction (0-1, though a value of 0 would be silly, wouldn't it?) of *sites to retain* from the full site-pair table. If less than 1, this argument will completely remove a fraction of sites such that they are not used in the permutation routines. |

**Value**

A site-pair table, such as one created by formatsitepair, ideally smaller than the one given. In the very rare case where the function determines not to remove any sites, or should the sampleSites argument be 1, then the function will return the full site-pair table.

**Note**

This function removes sites, not just site-pairs (rows) from the site-pair table. This function is called from several of the other functions within the gdm package, including the plotUncertainty and gdm.varImp functions, for the purposes of subsampling the sites in the site-pair table.

**See Also**

formatsitepair

**Examples**

```
##fit table environmental data
##sets up site-pair table, environmental tabular data
load(system.file("./data/gdm.RData", package="gdm"))
sppData <- gdmExpData[c(1,2,13,14)]
envTab <- gdmExpData[c(2:ncol(gdmExpData))]
sitePairTab <- formatsitepair(sppData, 2, XColumn="Long", YColumn="Lat", sppColumn="species",
siteColumn="site", predData=envTab)

removeSitesFromSitePair(sitePairTab, sampleSites=0.7)
```

---

stackedVars                          *Worldclim climate variables used to fit GDM in examples*

---

**Description**

A raster stack of five Worldclim gridded climate variables used as environmental data in the GDM examples. The data were originally downloaded from Worldclim using the getData() function in the raster package and clipped to the area encompassing the example data points.

### Format

A raster stack of five layers.

### Details

The specific variables included in this data are: bio5, bio6, bio15, bio18, and bio19. The clipped area is southwestern Australia.

### Source

http://www.worldclim.org/

### References

Hijmans, R.J., S.E. Cameron, J.L. Parra, P.G. Jones and A. Jarvis, 2005. Very high resolution interpolated climate surfaces for global land areas. International Journal of Climatology 25: 1965-1978.

---

| summary.gdm | *Summarize a Fitted Generalized Dissimilarity Model* |
|---|---|

---

### Description

This function summarizes the model object returned from gdm.

### Usage

```
## S3 method for class 'gdm'
summary(object, ...)
```

### Arguments

| | |
|---|---|
| object | A gdm model object resulting from a call to gdm. |
| ... | Ignored. |

### Value

summary prints its output to the R Console window and returns no value.

### See Also

gdm

**Examples**

```
##sets up site-pair table
load(system.file("./data/gdm.RData", package="gdm"))
sppData <- gdmExpData[, c(1,2,14,13)]
envTab <- gdmExpData[, c(2:ncol(gdmExpData))]
sitePairTab <- formatsitepair(sppData, 2, XColumn="Long", YColumn="Lat", sppColumn="species",
siteColumn="site", predData=envTab)

##create GDM
gdmMod <- gdm(sitePairTab, geo=TRUE)

##summary of GDM
summary(gdmMod)
```

# Index