Package 'occTest'

November 18, 2022

Title Characterizing and Filtering Species Occurrence Data

Version 0.1.1

Description Perform multiple tests for potential errors in species occurrence data and filter, and filter data according to users specifications.

License MIT + file LICENSE

Encoding UTF-8

RoxygenNote 7.2.1

VignetteBuilder knitr

Imports raster, data.table, rgeos, rgdal, biogeo, stringr, alphahull, plyr, sp, sf, methods, stats, utils, outliers, CoordinateCleaner, tictoc, rgbif, countrycode, rnaturalearth, svMisc, spocc, tidyverse, rnaturalearthdata, pingr, ggplot2, DT, Matrix, dplyr, graphics, dataPreparation

Suggests GNRS, testthat, knitr, rmarkdown, rworldmap, cleangeo, geosphere, originr, parallelsugar, assertthat

Additional_repositories https://pepbioalerts.github.io/drat

NeedsCompilation no

Author Josep M Serra-Diaz [aut, cre], Cory Merow [aut, ctb] (Cory co-developped the structure for testing blocks), Jeremy Borderieux [aut, ctb] (Jeremy developped plotting and filtering functions), Brian Maitner [aut, ctb] (Brian developped and incorporated centroid and GNRS analysis)

Maintainer Josep M Serra-Diaz <pep.agroparistech@gmail.com>

Depends R (>= 3.5.0)

Repository CRAN

Date/Publication 2022-11-18 12:00:03 UTC

R topics documented:

`.addmainfields2` *Add main fields*

Description

Incorporate fields in the initial data frame

Usage

```
.addmainfields2(dat, species, verbose = FALSE)
```

Arguments

`dat`	A dataframe containing occurrence data for checking.
`species`	character. Name of the species
`verbose`	logical. Print messages? Default FALSE

Details

Inspired by addmainfields but modified (hence number 2 after the function original name)

Value

Original dataframe, dat.

Author(s)

Mark Robertson and Vernon Visser (original function), Josep M Serra Diaz (modifs)

See Also

Other checks: `.check.geospatial.data()`, `.checkdatastr2()`, `.checkfields()`, `.status.tracker.and.escaping()`

.ah2sp *Convert Alpha Hull object into a shapefile*

Description

Convert Alpha Hull object into a shapefile

Usage

```
.ah2sp(
  x,
  increment = 360,
  rnd = 10,
  proj4string = sp::CRS(as.character(NA)),
  tol = 1e-04
)
```

Arguments

`x`	an alpha hull object
`increment`	numeric. Increments
`rnd`	numeric. Decimal rounding
`proj4string`	crs object with the spatial projection.
`tol`	numeric. tolerance

Details

Function written by Andrew Bevan, found on R-sig-Geo, and modified by Pascal Title

Value

a sp polygon object

Author(s)

Pascal Title (original version), Josep M Serra-Diaz (modifications)

See Also

Alpha hulls are created with `ahull`.

`.cc_round_occTest` *Flag records with regular pattern interval*

Description

own version of coordinate cleaner cc_round

Usage

```
.cc_round_occTest(
  x,
  lon = "decimallongitude",
  lat = "decimallatitude",
  ds = "dataset",
  T1 = 7,
  reg_out_thresh = 2,
  reg_dist_min = 0.1,
  reg_dist_max = 2,
  min_unique_ds_size = 4,
  graphs = FALSE,
  test = "both",
  value = "flagged",
  verbose = TRUE
)
```

Arguments

`x`	Data.frame of species occurrences
`lon`	character. Column name in x with decimal longitude values
`lat`	character. Column name in x with decimal latitude values
`ds`	character. Column name in x with dataset name of the record
`T1`	numeric. Defaults to 7
`reg_out_thresh`	numeric. Defaults to 7
`reg_dist_min`	numeric. Defaults to 7
`reg_dist_max`	numeric. Defaults to 7
`min_unique_ds_size`	numeric. Defaults to 7

graphs	logical. Defaults to FALSE.
test	character. Defaults to 'both'
value	character. Defaults to flagged
verbose	logical. Defaults to TRUE.

Value

a clean data.frame

Author(s)

A Zizka (original author) Josep M Serra-Diaz (adapted from CoordinateCleaner)

See Also

CoordinateCleaner-package

.check.geospatial.data

Checks on the projection of the spatial data

Description

Verify that all data are in the same projection

Usage

```
.check.geospatial.data(list.geospatial.objects, verbose = FALSE)
```

Arguments

list.geospatial.objects	A list of geospatial objects.Default list includes: 'countries.shapefile','r.env','r.dem','ras.hii','points.proj4s
verbose	logical. Print messages?

Value

None. Used to generate warning messages.

Author(s)

Josep M Serra Diaz

See Also

Other checks: .addmainfields2(), .checkdatastr2(), .checkfields(), .status.tracker.and.escaping()

Examples

```
{
example<-"goes here"
}
```

`.checkdatastr2` *Check data structure*

Description

Verify that all main data fields are correctly structured

Usage

```
.checkdatastr2(dat, xf, yf, verbose = FALSE)
```

Arguments

`dat`	A dataframe containing occurrence data for checking.
`xf`	character. Name of the field where the x coordinate is stored (typically longitude). Default is x.field
`yf`	character. Name of the field where the y coordinate is stored (typically latitude). Default is y.field
`verbose`	logical. Print messages? Defaults to FALSE

Details

Inspired by checkdatastr but modified (hence number 2 after the function original name)

Value

Original dataframe, dat. Used primarily to generate warning messages.

Author(s)

Mark Robertson and Vernon Visser (original function), Josep M Serra Diaz (modifs)

See Also

Other checks: `.addmainfields2()`, `.check.geospatial.data()`, `.checkfields()`, `.status.tracker.and.escaping()`

`.checkfields` *Checking main fields*

Description

Checking main fields

Usage

```
.checkfields(dat, xf, yf, ef, tf, lf, cf, idf, verbose = FALSE)
```

Arguments

`dat`	A dataframe containing occurrence data for checking.
`xf`	character. Name of the field where the x coordinate is stored (typically longitude). Default is x.field
`yf`	character. Name of the field where the y coordinate is stored (typically latitude). Default is y.field
`ef`	character. Name of the field where the elevation of data collection is stored in the original dataset. Default is e.field.
`tf`	character. Name of the field where the date of data collection is stored in the original dataset. Default is t.field.
`lf`	character. Name of the field where the toponim/location of data collection is stored in the original dataset. Default is l.field.
`cf`	character. Name of the field where the registered country of data collection is stored in the original dataset. Default is c.field.
`idf`	character. Name of the field of the id of the observation
`verbose`	logical. Print messages? Default to FALSE

Details

checking main fields (inspired by addmainfields .

Value

Original dataframe, dat. Used primarily to generate warning messages.

Author(s)

Mark Robertson and Vernon Visser (original function), Josep M Serra Diaz (modifs)

See Also

Other checks: `.addmainfields2()`, `.check.geospatial.data()`, `.checkdatastr2()`, `.status.tracker.and.escaping`

`.checkPolygonsGEOS2` *Check polygon geometry*

Description

Check polygon geometry

Usage

```
.checkPolygonsGEOS2(obj, properly = TRUE, force = TRUE, useSTRtree = FALSE)
```

Arguments

`obj`	an alpha hull object
`properly`	logic.
`force`	logic.
`useSTRtree`	logic.

Details

inspired provided by maptools package and from P Title in rangeBuilder

Value

a sp polygon object

Author(s)

Pascal Title (original version), Josep M Serra-Diaz (modifications)

See Also

Alpha hulls are created with `ahull`.
see maptools and RangeBuilder package

`.coords2country` *Extracts country ISO3 based on locations*

Description

Extracts country ISO3 based on locations

Usage

```
.coords2country(
  xydat,
  .countries.shapefile = NULL,
  .points.proj4string = NULL,
  ctryNameField = NULL,
  verbose = FALSE
)
```

Arguments

`xydat`	A dataframe with x and y coordinates
`.countries.shapefile`	SpatialPolygonsDataFrame of world countries and their associated ISO3 codes
`.points.proj4string`	Proj4string for the occurrence data
`ctryNameField`	character. Column name in .countries shapefile where ISO3 are indicated
`verbose`	logical. Print messages?

Value

Factor with ISO3 codes for countries

Author(s)

Josep M Serra Diaz

See Also

Other Geo: `.getSRTM()`

`.getSRTM` *Download SRTM elevation raster*

Description

Download SRTM elevation raster

Usage

```
.getSRTM(xydat, download = TRUE, path = tempdir(), verbose = FALSE)
```

Arguments

`xydat`	A dataframe with x and y coordinates
`download`	Default to TRUE. Whether the data should be downloaded
`path`	where the downloads should go. Default to the current directory
`verbose`	if you want to print messages of progress or warnings

Details

Basedd on getData from raster

Value

raster

Note

borrowed from raster package but adapted to work directly within the occTest workflow

See Also

getData

Other Geo: `.coords2country()`

`.status.tracker.and.escaping` *Workflow status tracker*

Description

Track status and write useful output

Usage

```
.status.tracker.and.escaping(
  dataset.to.continue,
  wfo,
  wso,
  xf,
  yf,
  od,
  rsd,
  obf,
  sp,
  verbose = FALSE,
  as,
  ws,
  ts
)
```

Arguments

`dataset.to.continue`	A dataframe containing occurrence data for checking.
`wfo`	write full output
`wso`	write simple output
`xf`	The dataframe field containing the x values (e.g. "longitude")
`yf`	The dataframe field containing the y values (e.g. "latitude")
`od`	The output directory to use
`rsd`	logical. Return spatial data?
`obf`	Output base filename
`sp`	character. Name of the species
`verbose`	logical. Print messages? Defaults to FALSE
`as`	list. analysis settings
`ws`	list. writoutsettings
`ts`	tlist. able setting

Value

Original dataframe, dat. Used primarily to generate warning messages.

Author(s)

Josep M Serra Diaz

See Also

Other checks: `.addmainfields2()`, `.check.geospatial.data()`, `.checkdatastr2()`, `.checkfields()`

`ctryToIso3` *Convert country names to ISO3 codes*

Description

Froma character it uses different methods to derive country ISO3 digit codes

Usage

```
ctryToIso3(x, method = "countrycode")
```

Arguments

`x`	character. country name
`method`	character. Package name used to derive IS03 codes. Options are 'countrycode' (default) or 'GNRS'.

Details

right now not implemented with fuzzy matching, but is case insensitive. Methods implemented, 'countrycode' and 'GNRS'

Author(s)

Josep M Serra-Diaz (pep.serradiaz@agroparistech.fr) (adpatation)
countrycode-package Vincent Arel-Bundock vincent.arel-bundock@umontreal.ca ,
GNRS Brad Boyle, Brian Maitner

See Also

countrycode and GNRS

Other spStatus: `nativeStatusCtry()`

defaultSettings *load default settings for occTest*

Description

Loads a list of lists with the different default parameters for analysis, outputs and grading needed in occTest

Usage

```
defaultSettings()
```

Details

it can be use internally or it can be used by a user to subsequently modify parameters. No input parameters are required

Value

list of lists with all different parameters to use in occProfile function

Author(s)

Josep M Serra-Diaz (pep.serradiaz@agroparistech.fr)

Examples

```
#load default settings
settings <- defaultSettings()
```

geoEnvAccuracy *Coordinate accuracy*

Description

Detect records with low accuracy in space and time

Usage

```
geoEnvAccuracy(
  df,
  xf,
  yf,
  af,
  dsf,
  ef,
```

```
  tf,
  method = "all",
  r.env,
  accept.threshold.cell = 0.5,
  accept.threshold.env = 0.5,
  bearing.classes = 10,
  distance.classes = 5,
  env.quantiles = c(0.3, 0.7),
  elev.threshold = 100,
  raster.elevation = NULL,
  verbose = FALSE,
  do = TRUE,
  doParallel = FALSE,
  mc.cores = 2
)
```

Arguments

df	data.frame of species occurrences
xf	character. column name in df containing the x coordinates
yf	character. column name in df containing the y coordinates
af	character. column name in df containing the coordinate uncertainty value (in the same)
dsf	character. column name in df containing the dataset to which the record belongs to (e.g. Forest Inventory of Spain)
ef	character. column name in df containing the registered elevation for the record.
tf	character. column name in df containing the dataset with the date/time where the species is recorded
method	character. Vector of methods to be used. See details. Default 'all'
r.env	raster. Raster with environmental data
accept.threshold.cell	numeric. Acceptance threshold for how much percentage of the Area of uncertainty in the cell we want to accept. Default to 0.5
accept.threshold.env	numeric. Default 0.5
bearing.classes	numeric. Default to 10.
distance.classes	integer. Default to 5.
env.quantiles	numeric. Default to c(0.3,0.7)
elev.threshold	numeric. Default to 100
raster.elevation	numeric. Default to 100
verbose	logical. Print messages? Default FALSE
do	logical. Should tests be performed? Default TRUE
doParallel	logical. Should computation use parallel functions? Default FALSE
mc.cores	numeric. How many cores to use? (used when doParallel = TRUE). Default 2

Details

Geoenvironmental accuracy function will implement differnt methods to assess occurrence accuracy in environmnental and geographic space.
Current implmented methods are: 'lattice' : tests for lattice arrangement in occurrence datasets. Borrowed from cd_round .
'elevDiff' : assess the elevation difference between a given raster (or automatically downloaded fro SRTM), and the elevation recorded. If differences >elev.threshold then the record is considered as a low accuracy threshold
'noDate' : assess whether there is a date or timestamp information in the record.
'noDateFormatKnown' : assess whether the information in the timestamp agrees with different formatting of Dates.
'outDateRange' : (not implemented) assess whether the record is within a user specified time frame.
'percDiffCell' : assess whether the record may be falling in a different raster cell given an information of coordinate accuracy.
'envDeviation' : assess whether the climate in a given record may be outside of the interval 30th-70th (default values) for a given variable due to coordinate uncertainty.

Value

data.frame

Author(s)

Josep M Serra-Diaz (pep.serradiaz@agroparistech.fr), A Zizka (CoordinateCleaner package)

See Also

cd_round

Other analysis: `.nearestcell3()`, `centroidDetection()`, `countryStatusRangeAnalysis()`, `duplicatesexcludeAnalysis()`, `humanDetection()`

Examples

```
#see examples in vignetteXtra-occTest
```

---

get_occTest_settings *Get occTest Settings*

Description

Get the settings used to create a occTest or occFilter object

Usage

```
get_occTest_settings(x)
```

Arguments

x	An occTest or occFilter object returned by occTest or occFilter

Value

list of lists with all different parameters to use in occTest

Author(s)

Jeremy Borderieux (jeremy.borderieux@agroparistech.fr)

See Also

occTest; occFilter

Examples

```
### THIS IS A CUT DOWN  EXAMPLE
### visit vignetteXtra-occTest for more info

#load output from occTest
occTest_output <- readRDS (system.file('ext/out.rds',package = 'occTest'))
get_occTest_settings(occTest_output)
```

---

landSeaFilter *Filter occurrence by habitat (terrestrial/non-Terrestrial)*

Description

Filter the occurrence recoreds according to whether they should be in land masses or not

Usage

```
landSeaFilter(df, xf, yf, habType = NULL, verbose = TRUE, habPol = NULL)
```

Arguments

df	Data.frame of species occurrences
xf	the field in the dataframe containing the x cordinates
yf	the field in the dataframe containing the y cordinates
habType	character. Define the species habitat. Only "terrestrial" and "sea" implemented.
verbose	logical. Print messages? Default TRUE
habPol	sf polygon boject. Shows land masses

Value

list

Author(s)

Josep M Serra-Diaz (pep.serradiaz@agroparistech.fr)

Examples

```
xyDF <- data.frame (x=c(0,42),y=c(0,1),Reason=NA)
landSeaFilter(xyDF,xf='x',yf='y')
```

---

minimalSettings *Load minimal settings for occTest*

Description

Loads a list of lists with the different default parameters for analysi. It avoids using some functions of the pkg under development.

Usage

```
minimalSettings()
```

Details

it can be use internally or it can be used by a user to subsequently modify parameters

Value

list of lists with all different parameters to use in occTest function

Author(s)

Josep M Serra-Diaz (pep.serradiaz@agroparistech.fr)

Examples

```
#load default settings
settings <- minimalSettings()
```

`occFilter` *Filter occurrence records from occTest outputs*

Description

Select occurrence records based on aggregated values of different tests

Usage

```
occFilter(
  df,
  by = "testBlock",
  errorAcceptance = "relaxed",
  errorThreshold = NULL,
  custom = NULL
)
```

Arguments

`df`	data.frame. Output of occTest
`by`	character. Applying thresholds to either blocks of test ('testBlock') or single test types ('testType')
`errorAcceptance`	character. Philosophy for filtering based on threshold. Option are majority, relaxed, strict. Default are 'relaxed'
`errorThreshold`	double. Value from 0 to 1, specifying the threshold of wrong tests (potentally erroneous records) to filter. It overrides the parameters in thresholds. We recommend building that table based on the functio buildCustomThresholds.
`custom`	data.frame or equivalent, custom rules created adding a "errorThreshold" (ranging from 0, strict, to 1, relaxed) column to to the result of readRDS(system.file('ext/fieldMetadata.rds',pac

Details

If errorAcceptance is used, a 'relaxed' philosophy corresponds to 0.7 (70

Value

list of 2 data.frames

Author(s)

Josep M Serra-Diaz (pep.serradiaz@agroparistech.fr), Jeremy Borderieux (jeremy.borderieux@agroparistech.fr)

See Also

showTests

Examples

```
### THIS IS A CUT DOWN  EXAMPLE
### visit vignetteXtra-occTest for more info

#load output from occTest
occTest_output <- readRDS (system.file('ext/out.rds',package = 'occTest'))
filtered_dataset <- occFilter (occTest_output)
#inspect results
names (filtered_dataset)
```

`occSimpFilter` *Runs tests and validates data*

Description

Runs tests and validates data

Usage

```
occSimpFilter(
  spOcc,
  env,
  speciesName = "My species",
  x = "x",
  y = "y",
  date = NULL,
  isoCountry = NULL,
  classification = "majority",
  filterCols = TRUE
)
```

Arguments

`spOcc`	data.frame. Object with the coordinate data.
`env`	raster or rasterStack. Environmental data (e.g. typically climatic).
`speciesName`	character. Name of the species.
`x`	name of the field with the coordinate x. Default 'x'
`y`	name of the field with the coordinate y. Default 'y'
`date`	name of the field with the values of the reported timestamp of the record. Default NULL
`isoCountry`	name of the field with the values of the reported country of the record. Default NULL
`classification`	character. Indicates the thresholds philosophy applied to classify errors in occurrence data. Possible values 'strict','relaxed','custom'

`filterCols` logical. Should only the initial input columns be retained in the output (the filtered dataframe)?

Value

a list of two. First element is a data.frame with profiled occurrence records with their associated profiled labels. Second element is a dataframe with all outputs of the analysis implemented.

Note

The majority of function parameters can be adjusted but we provide default values.
We recommend those default values if the user is to use the geospatial data included in the package.
but this automatic implementation (occTest + occFilter) missses some analysis to increase speed.

Author(s)

Josep M Serra-Diaz (pep.serradiaz@agroparistech.fr)

Examples

```
### THIS IS A CUT DOWN  EXAMPLE
### visit vignetteXtra-occTest for more info
#load environmental raster
library (raster)
library (sf)
library (occTest)
#load occurrence data
occData <- read.csv (system.file('ext/exampleOccData.csv',package = 'occTest'))
#load environmental raster
renv <- raster (system.file('ext/AllEnv.tif',package = 'occTest'))
#load elevation raster
dem <- raster (system.file('ext/DEM.tif',package = 'occTest'))
#load settings
settings <- readRDS (system.file('ext/exSettings.rds',package = 'occTest'))
#run occTest
out = occTest(sp.name='MyFake species',
             sp.table = occData,ntv.ctry = 'ESP',inv.ctry = 'FRA',
             tableSettings = settings$tableSettings,
             writeoutSettings = settings$writeoutSettings,
             analysisSettings = settings$analysisSettings,
             r.env = renv,r.dem=dem)
#filter
occFilter(out)
```

occTest *Occurrence tests*

Description

Perform tests for data quality in species occurrence using several methods

Usage

```
occTest(
  sp.name,
  habitat = NULL,
  sp.table,
  r.env,
  tableSettings = NULL,
  analysisSettings = NULL,
  writeoutSettings = NULL,
  gradingSettings = NULL,
  return.spatial.data = FALSE,
  r.dem = NULL,
  ntv.ctry = NULL,
  inv.ctry = NULL,
  resolveAlienCtry = FALSE,
  resolveNativeCtry = FALSE,
  interactiveMode = FALSE,
  verbose = FALSE,
  doParallel = FALSE,
  mc.cores = 2
)
```

Arguments

sp.name	character. Name of the species.
habitat	NULL
sp.table	data.frame. Object with the coordinate data.
r.env	raster or rasterStack. Environmental data(e.g. typically climatic.
tableSettings	list. Elements corresponding to different settings of the input occurrence table.
analysisSettings	list. Elements corresponding to different settings of the analysis functions .
writeoutSettings	list. Elements corresponding to different settings of the analysis functions .
gradingSettings	list. Not implemented yet. Defaults to NULL.Elements corresponding to different settings of the analysis functions .

`return.spatial.data`
: logical. Should the spatial dataset of `analysisSettings` attached to the metadata?, default is FALSE to save memory

`r.dem`
: raster. Elevation data (in meters).

`ntv.ctry`
: character. vector with ISO3 code of the countries where species is considered native

`inv.ctry`
: character. vector with ISO3 code of the countries where species is considered invasive

`resolveAlienCtry`
: logical. To automatically try to detect countries for which species is considered native

`resolveNativeCtry`
: logical. To automatically try to detect countries for which species is considered alien

`interactiveMode`
: logical. Should prompts be ouput for some decisions taken by the workflow? Deafult FALSE,

`verbose`
: logical. Print workflow information? Default to FALSE

`doParallel`
: logical. Should some operations be run in parellel when possible? Default to FALSE

`mc.cores`
: numeric. If doParallel is TRUE, then how many cores to use? Default to 2

Value

data frame with the tests performed (field $_test), specific comment for the tests ($_comments), the exact value of the test ($_value), and scores summarizing results across tests with an objective ($_score)

Note

The output dataframe allows users to classify or scrub the occurrences the way they want to.
The names of the columns in the output object have the following naming convention:
$AnalysisType$_$SpecificTest$_value: numeric or logical. Shows the quantitative result of the test (sometimes the same as in the result of the _test)
$AnalysisType$_$SpecificTest$_test: logical Shows whether the occurrence passes or not the test, being TRUE a flag for a wrong record and NA indicating that the test was not performed on that record.
$AnalysisType$_$SpecificTest$_comment: character. Shows some comments related to the specific test.
Examples: HumanDetection_HumanInfluence_value gives you the score of current human influence in the record HumanDetection_HumanInfluence_test gives you whether we consider the former value an error/bias (TRUE) or not (FALSE) HumanDetection_HumanInfluence_comment gives you a commen that give further detail on the analysis. In this case that the threshold of 45 was used for the test. HumanDetection_score summarizes all the other HumanDetection tests and outputs a value from 0 to 1. A value of 0.5 would indicate that half of the tests used indicate that is an a Human signal in the record.

Examples

```
### THIS IS A CUT DOWN  EXAMPLE
### visit vignetteXtra-occTest for more info
#load environmental raster
library (raster)
library (sf)
library (occTest)
#load occurrence data
occData <- read.csv (system.file('ext/exampleOccData.csv',package = 'occTest'))
#load environmental raster
renv <- raster (system.file('ext/AllEnv.tif',package = 'occTest'))
#load elevation raster
dem <- raster (system.file('ext/DEM.tif',package = 'occTest'))
#load settings
settings <- readRDS (system.file('ext/exSettings.rds',package = 'occTest'))
#run occTest
out = occTest(sp.name='MyFake species',
             sp.table = occData,ntv.ctry = 'ESP',inv.ctry = 'FRA',
             tableSettings = settings$tableSettings,
             writeoutSettings = settings$writeoutSettings,
             analysisSettings = settings$analysisSettings,
             r.env = renv,r.dem=dem)
```

`plot.occTest` *Display the filtering process*

Description

Display the filtering process

Usage

```
## S3 method for class 'occTest'
plot(x, occFilter_list = NULL, show_plot = FALSE, ...)
```

Arguments

`x`	An occTest object returned by occTest, i.e. the unfiltered data.frame
`occFilter_list`	Optional, an occFilter object; a list returned by occFilter, the result of the filtering of `x`
`show_plot`	Logical, should the plots be plotted ?
`...`	not used

Details

If `occFilter_list` is provided, display how the occurences passed the different tests, otherwise only plot the coordinates filtering step

Value

list of ggplots objects, of varying length, depending on whether the filtering was done by testBlock or testType

Author(s)

Jeremy Borderieux (jeremy.borderieux@agroparistech.fr)

See Also

occFilter , occTest , the ggplot2 package

Examples

```
#load output from occTest
occTest_output <- readRDS (system.file('ext/out.rds',package = 'occTest'))
#filter dataset output from occTest
filtered_occTest <- occFilter (occTest_output)
#plot the outputs
descriptive_plots <- plot (x=occTest_output,occFilter_list=filtered_occTest)
```

---

setTableNames *set table names internally*

Description

helper function to set the names for the fields in the input table (tableSettings). By default it provides rbif like column names (not fully consistent yet tho). Alternatively, the user can specify their own field names for the table

Usage

```
setTableNames(
  x.field = NULL,
  y.field = NULL,
  t.field = NULL,
  l.field = NULL,
  c.field = NULL,
  e.field = NULL,
  a.field = NULL,
  ds.field = NULL,
  taxonobservation.id = NULL
)
```

Arguments

x.field	character. Name of the x coordinate field.
y.field	character. Name of the y coordinate field.
t.field	character. Name of the timestamp field.
l.field	character. Name of the locality field.
c.field	character. Name of the country code field.
e.field	character. Name of the eleveation field.
a.field	character. Name of the accuracy field.
ds.field	character. Name of the dataset identifier field.
taxonobservation.id	character. Name of the taxon observartion id field.

Value

list

Author(s)

Josep M Serra-Diaz (pep.serradiaz@agroparistech.fr)

Examples

```
defaultTableNames <- setTableNames()
#only modifying the names for the coordinates
myTable_withMyNames <- setTableNames (x.field='x_coord',y.field = 'y_coord')
```

---

setTestBlocks *Set the tests to run*

Description

function used to select which groups of tests you want in occTest workflow

Usage

```
setTestBlocks(geo = TRUE, lu = TRUE, env = TRUE, time = TRUE)
```

Arguments

geo	logical. Should this family of tests be performed?
lu	logical. Should this family of tests be performed?
env	logical. Should this family of tests be performed?
time	logical. Should this family of tests be performed?

Details

You can turn off an entire type of tests altogether by modifying this seetings. See occTest::showTests for further information on tests used in the packages

Value

list

Author(s)

Josep M Serra-Diaz (pep.serradiaz@agroparistech.fr)

Examples

```
defaultSettings_analysis  <- setTestBlocks()
#now we turn off the block of tests related to land use
mySettings_analysis  <- setTestBlocks(lu=FALSE)
```

`setTestTypes` *Set the tests to run*

Description

function used to select which types of tests you want in occTest workflow (analysisSet)

Usage

```
setTestTypes(
  countryStatusRange = TRUE,
  centroidDetection = TRUE,
  humanDetection = TRUE,
  landUseType = TRUE,
  institutionLocality = TRUE,
  geoOutliers = TRUE,
  envOutliers = TRUE,
  geoenvLowAccuracy = TRUE
)
```

Arguments

`countryStatusRange`	logical. Should this test type be performed?
`centroidDetection`	logical. Should this test type be performed?
`humanDetection`	logical. Should this test type be performed?
`landUseType`	logical. Should this test type be performed?

`institutionLocality`
: logical. Should this test type be performed?

`geoOutliers`
: logical. Should this test type be performed?

`envOutliers`
: logical. Should this test type be performed?

`geoenvLowAccuracy`
: logical. Should this test type be performed?

Details

See occTest::showTests for further information on tests used in the packages

Value

list with user analysis settings

Author(s)

Josep M Serra-Diaz (pep.serradiaz@agroparistech.fr)

Examples

```
defaultSettings_analysis <- setTestTypes()
#now we do not want to perform centroid geoenironmental accuracy type of tests
mySettings_analysis <- setTestTypes(geoenvLowAccuracy=FALSE)
```

---

showTableNames *Print naming conventions in occTest*

Description

prints a table with the the conventions used for column names

Usage

```
showTableNames()
```

Details

The function prints a guide to column naming conventions used by occTest in their default parameters. These defaults can be changed via setTableNames, but the user may also decide to format their input table according to these naming conventions. It does not require input parameters

Value

prints a data.frame

Author(s)

Josep M Serra-Diaz (pep.serradiaz@agroparistech.fr)

Examples

```
showTableNames ()
```

`showTests` *Show implemented tests and types of tests*

Description

prints a table with the column names

Usage

```
showTests()
```

Details

The function prints a guide to column naming conventions used by occTest in their default parameters. These defaults can be changed via setTableNames, but the user may also decide to format their input table according to these naming conventions.

Value

prints a dataframe

Author(s)

Josep M Serra-Diaz (pep.serradiaz@agroparistech.fr)

Examples

```
showTests()
```

Index