# Package 'disdat'

February 7, 2023

**Type** Package

**Title** Data for Comparing Species Distribution Modeling Methods

**Version** 1.0-1

**Date** 2023-02-07

**Maintainer** Roozbeh Valavi <valavi.r@gmail.com>

**License** GPL (>= 3)

**Encoding** UTF-8

**LazyLoad** yes

**Depends** R (>= 3.5.0)

**Suggests** knitr, rmarkdown, terra, sf, tmap, mapview

**VignetteBuilder** knitr

**NeedsCompilation** no

**Description** Easy access to species distribution data for 6 regions in the world, for a to-
tal of 226 anonymised species. These data are described and made avail-
able by Elith et al (2020) <doi:10.17161/bi.v15i2.13384> to compare species distribution mod-
elling methods.

**Author** Robert J. Hijmans [aut] (<https://orcid.org/0000-0001-5872-2872>),
Roozbeh Valavi [cre, aut],
Jane Elith [aut]

**Repository** CRAN

**Date/Publication** 2023-02-07 22:22:34 UTC

# R topics documented:

---

| disdat-package | *Data for species distribution modeling* |
|----------------|-------------------------------------------|

---

### Description

This package allows for easy use of a collection of datasets that can be used to compare species distribution models. There are data for 6 regions in the world, for a total of 226 anonymised species including birds, vascular plants, reptiles and bats. Each data set has presence-only (and optionally background) training data to build models, and presence/absence data to evaluate models.

The data were compiled and used by a species distribution modeling working group sponsored by the National Center for Ecological Analysis and Synthesis (NCEAS), at UC Santa Barbara, USA. Full details of the dataset are provided in the first publication listed below, from the NCEAS data group.

### Details

The data are fully described in the first publication listed below, and also supplied with metadata on Open Science Framework (OSF). On the OSF site, rasters (gridded data) of all environmental data are also available for download.

### Author(s)

Package by Robert J. Hijmans, Roozbeh Valavi, and Jane Elith. Data collation and processing by the NCEAS data group (see first reference below, and the manual package for specific datasets).

### References

The main reference for the these data is:

- Elith, J., Graham, C.H., Valavi, R., Abegg, M., Bruce, C., Ferrier, S., Ford, A., Guisan, A., Hijmans, R.J., Huettmann, F., Lohmann, L.G., Loiselle, B.A., Moritz, C., Overton, J.McC., Peterson, A.T., Phillips, S., Richardson, K., Williams, S., Wiser, S.K., Wohlgemuth, T. & Zimmermann, N.E., (2020). Presence-only and presence-absence data for comparing species distribution modeling methods. Biodiversity Informatics 15:69-80.

Other papers using these data include:

- Dudík, M. & Phillips, S. J. (2009). Generative and Discriminative Learning with Unknown Labeling Bias. in Advances in Neural Information Processing Systems 21 (eds. Koller, D., Schuurmans, D., Bengio, Y. & Bottou, L.) 401-408. Curran Associates, Inc.

- Dudík, M., Phillips, S. J. & Schapire, R. E. (2006). Correcting sample selection bias in maximum entropy density estimation. in Advances in Neural Information Processing Systems 18 (eds. Weiss, Y., Schölkopf, B. & Platt, J. C.) 323-330 (MIT Press).

- Elith, J. & Leathwick, J. R. (2007). Predicting species distributions from museum and herbarium records using multiresponse models fitted with multivariate adaptive regression splines. Diversity and Distributions 13, 165-175.

- Elith, J., Graham, C.H., Anderson, R.P., Dudík, M., Ferrier, S., Guisan, A., Hijmans, R.J., Huettmann, F., Leathwick, J.R., Lehmann, A., Li, J., Lohmann, L.G., Loiselle, B.A., Manion, G., Moritz, C., Nakamura, M., Nakazawa, Y., Overton, J.McC., Peterson, A.T., Phillips, S.J., Richardson, K.S., Scachetti-Pereira, R., Schapire, R.E., Soberón, J., Williams, S., Wisz, M.S., Zimmermann, N.E. (2006). Novel methods improve prediction of species' distributions from occurrence data. Ecography 29, 129–151

- Graham, C.H., Elith, J., Hijmans, R.J., Guisan, A., Peterson, A.T., Loiselle, B.A. (2008). The influence of spatial errors in species occurrence data used in distribution models. Journal of Applied Ecology 45, 239–247.

- Guisan, A., Graham, C. H., Elith, J., Huettmann, F. & NCEAS Species Distribution Modelling Group (2007). Sensitivity of predictive species distribution models to change in grain size: insights from a multi-models experiment across five continents. Diversity and Distributions 13, 332-340.

- Guisan, A., Zimmermann, N. E., Elith, J., Graham, C. H., Phillips, S. P., & Peterson, A. T. (2007). What matters for predicting the occurences of trees: techniques, data, or species' characteristics? Ecological Monographs 77, 615-530.

- Hijmans, R. J. (2012). Cross-validation of species distribution models: removing spatial sorting bias and calibration with a null model. Ecology 93, 679-688.

- Phillips, S. J. & Dudík, M. (2008). Modeling of species distributions with Maxent: new extensions and a comprehensive evaluation. Ecography 31, 161-175.

- Phillips, S. J. & Elith, J. (2010). POC-plots: calibrating species distribution models with presence-only data. Ecology 91, 2476-2484.

- Phillips, S.J., Dudík, M., Elith, J., Graham, C.H., Lehmann, A., Leathwick, J., Ferrier, S. (2009). Sample selection bias and presence-only distribution models: implications for background and pseudo-absence data. Ecological Applications 19, 181–197.

- Phillips, S. J., Anderson, R. P., Dudík, M., Schapire, R. E. & Blair, M. E. (2017). Opening the black box: an open-source release of Maxent. Ecography 40, 887-893.

- Valavi, R., Elith, J., Lahoz-Monfort, J.J. & Guillera-Arroita, G. (2023). Flexible species distribution modelling methods perform well on spatially separated testing data. Global Ecology and Biogeography, geb.13639.

- Valavi, R., Elith, J., Lahoz-Monfort, J.J. & Guillera-Arroita, G. (2021). Modelling species presence-only data with random forests. Ecography, 44, 1731–1742.

- Valavi, R., Guillera-Arroita, G., Lahoz-Monfort, J.J. & Elith, J. (2022). Predictive performance of presence-only species distribution models: a benchmark study with reproducible code. Ecological Monographs, 92(1). 10.1002/ecm.1486

- Wisz, M.S., Hijmans, R.J., Li, J., Peterson, A.T., Graham, C.H., Guisan, A., & NCEAS Species Distribution Modelling Group (2008). Effects of sample size on the performance of species distribution models. Diversity and Distributions 14, 763–773.

| AWT | *Australian Wet Tropics species distribution data* |
|---|---|

**Description**

Species occurrence data for 40 species (20 vascular plants, 20 birds) in the Australian Wet Tropics (AWT) and associated environmental data. Full details of the dataset are provided in the reference below. There are four data sets with training (po and bg) and test (pa, env) data:

po (training data) includes site names, species names, coordinates, occurrence ("1" for all, since all are presence records), group (plant or bird), and site values for 13 environmental variables (below).

bg (training data) has 10000 sites selected at random across the study region. It is structured identically to po, with "0" for occurrence (not implying absence, but denoting a background record in a way suited to most modelling methods) and NA for group.

env (testing data) includes group, site names, coordinates, and site values for 13 environmental variables (below). These are for sites from different surveys for plants (102 sites) and birds (340 sites), and can be returned as separate datasets by disEnv, or in one long format dataset by disData. These data are suited to make predictions to.

pa (testing data) includes group, site names, coordinates, and presence-absence records, one column per species (in the wide format returned by disPa). They can also be returned in long format using disData. The sites are identical to the sites in env. These data are suited to evaluating the predictions made with env.

Raster (gridded) data for all environmental variables are available - see the reference below for details.

The coordinate reference system of the x and y coordinates is UTM, zone 55, spheroid GRS 1980, datum GDA94 (EPSG:28355).

The vignette provided with this package provides an example of how to fit and evaluate a model with these data.

**Environmental variables:**

| Code | Description | Units | Type |
|---|---|---|---|
| bc01 | Annual mean temperature | degrees C | Continuous |
| bc04 | Temperature seasonality | dimensionless | Continuous |
| bc05 | Max. temperature of warmest period | degrees C | Continuous |
| bc06 | Min. temperature of coldest period | degrees C | Continuous |
| bc12 | Annual precipitation | mm | Continuous |
| bc15 | Precipitation seasonality | dimensionless | Continuous |
| bc17 | Precipitation of driest quarter | mm | Continuous |
| bc20 | Annual mean radiation | MJ/m2/day | Continuous |
| bc31 | Moisture index seasonality | dimensionless | Continuous |
| bc33 | Mean moisture index of lowest quarter (MI) | dimensionless | Continuous |
| slope | Slope | percent | Continuous |
| topo | Topographic position | 0 is a gully and 100 a ridge, 50 mid-slope | Continuous |
| tri | Terrain ruggedness index | Sum of variation in a 1 km moving window | Continuous |

**Source**

Environmental predictors prepared by Karen Richardson, Caroline Bruce and Catherine Graham. Species data supplied by Andrew Ford, Stephen Williams and Karen Richardson.

See the reference below for further details on source, accuracy, cleaning, and particular characteristics of these datasets.

**References**

Elith, J., Graham, C.H., Valavi, R., Abegg, M., Bruce, C., Ferrier, S., Ford, A., Guisan, A., Hijmans, R.J., Huettmann, F., Lohmann, L.G., Loiselle, B.A., Moritz, C., Overton, J.McC., Peterson, A.T., Phillips, S., Richardson, K., Williams, S., Wiser, S.K., Wohlgemuth, T. & Zimmermann, N.E., (2020). Presence-only and presence-absence data for comparing species distribution modeling methods. Biodiversity Informatics 15:69-80.

**Examples**

```
awt_po <- disPo("AWT")
awt_bg <- disBg("AWT")

awt_pa_plant <- disPa("AWT", "plant")
awt_env_plant <- disEnv("AWT", "plant")
awt_pa_bird <- disPa("AWT", "bird")
awt_env_bird <- disEnv("AWT", "bird")


# Or all in one list
awt <- disData("AWT")
sapply(awt, head)

disCRS("AWT")
```

---

CAN                                    *Canadian bird species distribution data*

---

**Description**

Species occurrence data for 20 bird species from Ontario, a province in Canada (CAN), and associated environmental data. Full details of the dataset are provided in the reference below. There are four data sets with training (po and bg) and test (pa, env) data:

po (training data) includes site names, species names, coordinates, occurrence ("1" for all, since all are presence records), group (bird), and site values for 11 environmental variables (below).

bg (training data) has 10000 sites selected at random across the study region. It is structured identically to CANtrain_po, with "0" for occurrence (not implying absence, but denoting background in a way suited to most modelling methods) and "NA" for group.

env (testing data) includes group, site names, coordinates, and site values for 11 environmental variables (below), at 14571 sites. This file is suited to making predictions.

pa (testing data) includes group, site names, coordinates, and presence-absence records, one column per species. The sites are identical to the sites in env. This file is suited to evaluating the predictions made to env.

Raster (gridded) data for all environmental variables are available - see the reference below for details.

The reference system of the x and y coordinates is unprojected with Clarke 1866 ellipsoid . Latitude and longitude are in geographical coordinates using unknown datum based upon the Clarke 1866 ellipsoid (EPSG:4008).

The vignette provided with this package provides an example of how to fit and evaluate a model with these data.

**Environmental variables:**

| Code | Description |
|------|-------------|
| alt | Digital elevation |
| asp2 | Aspect |
| ontprec | Annual Precipitation |
| ontprec4 | April precipitation |
| ontprecsd | Precipitation Seasonality |
| ontslp | Slope |
| onttemp | Annual mean temperature |
| onttempsd | Temperature standard deviation |
| onttmin4 | April minimum temperature |
| ontveg | Vegetation, from Ontario Land Cover Database (OLC) vegetation map, derived from a mosaic of Landsat images |
| watdist | Distance from Hudson Bay |

**Source**

Environmental predictors prepared by Falk Huettmann, Jane Elith and Catherine Graham. Species data: PO from the Ontario Nest Records database, Royal Ontario Museum (ROM) and supplied by M. Peck to Falk Huettmann; PA from Breeding Bird Atlas for Ontario, provided by M. Cadman to Falk Huettmann.

See the reference below for further details on source, accuracy, cleaning, and particular characteristics of these datasets.

**References**

Elith, J., Graham, C.H., Valavi, R., Abegg, M., Bruce, C., Ferrier, S., Ford, A., Guisan, A., Hijmans, R.J., Huettmann, F., Lohmann, L.G., Loiselle, B.A., Moritz, C., Overton, J.McC., Peterson, A.T., Phillips, S., Richardson, K., Williams, S., Wiser, S.K., Wohlgemuth, T. & Zimmermann, N.E., (2020). Presence-only and presence-absence data for comparing species distribution modeling methods. Biodiversity Informatics 15:69-80.

**Examples**

```
can_po <- disPo("CAN")
can_bg <- disBg("CAN")
```

```
can_pa <- disPa("CAN")
can_env <- disEnv("CAN")


# Or all in one list
x <- disData("CAN")
sapply(x, head)

disCRS("CAN")
```

---

disCRS                          *Coordinate reference system*

---

### Description

Get the coordinate reference system for the data of a region.

### Usage

```
disCRS(region, format="proj4")
```

### Arguments

| region | character. One of "AWT", "CAN", "NSW", "NZ", "SA", "SWI" |
|--------|----------------------------------------------------------|
| format | character. Either "proj4" or "EPSG" |

### Value

character vector

### Examples

```
disCRS("AWT")
disCRS("NSW")
```

---

disData                         *Get disdat datasets*

---

### Description

disPo returns the presence-only (po) data for a region

disBg returns the background (bg) data for a region

disPa returns the presence-absence (pa) data for a region and group

disEnv returns the environmental (env) data for sites matching those in the pa data, for a region and group

disData returns a list with all data for a region.

disBorder returns a polygon for one of the regions.

## Usage

```
disData(region)

disPo(region)

disBg(region)

disPa(region, group)

disEnv(region, group)

disBorder(region, pkg="sf")
```

## Arguments

| | |
|---|---|
| region | character. One of "AWT", "CAN", "NSW", "NZ", "SA", "SWI" |
| group | character. If region is "NSW", one of "ba", "db", "nb", "ot", "ou", "rt", "ru", "sr". region is "AWT" "bird", "plant". The other regions each have only one group, so group should not be specified |
| pkg | character. Either "sf" or "terra" to get polygons as defined by that package |

## Details

disData returns a list with env, pa, bg and po data in that order. For regions with more than one group, the testing data (env and pa) will come from different surveys, and the model testing should be targeted to the relevant group. The first column of the env and pa data.frames is "group", which can be used to extract the correct data.

## Value

data.frame (disPo, disBg, disPa and disEnv) or list with four data.frames (disData)

## Examples

```
awt_po <- disPo("AWT")

awt_bg <- disBg("AWT")

awt_pa_plants <- disPa("AWT", "plant")

awt_env_plants <- disEnv("AWT", "plant")


x <- disData("NSW")

names(x)

sapply(x, head)
```

```
z <- disBorder("NSW")

plot(z)
```

---

disMapBook                    *Generating maps of disdat species*

---

### Description

A helper function for automatically generating maps for the species data in PDF format.

### Usage

```
disMapBook(region, output_pdf, verbose = TRUE)
```

### Arguments

| | |
|---|---|
| region | A character vector. The name of the region(s) to generate plots. |
| output_pdf | Output pdf file to be saved. |
| verbose | Logical. control amount of screen reporting. |

### See Also

[disPo](#), [disPa](#) and [disBorder](#)

### Examples

```
disMapBook(c("AWT", "NSW"), "~/Desktop/sp_mapbook.pdf")
```

---

disPredictors                 *Predictor variables*

---

### Description

Get the names of the predictor variables for a region.

### Usage

```
disPredictors(region)
```

### Arguments

| | |
|---|---|
| region | character. One of "AWT", "CAN", "NSW", "NZ", "SA", "SWI" |

## Value

character vector

## Examples

```
disPredictors("NSW")
```

---

NSW                          *New South Wales species distribution data*

---

## Description

Species occurrence data for 54 species from 8 biological groups in New South Wales (NSW, a state in Australia) and associated environmental data. Full details of the dataset are provided in the reference below. There are four data sets with training (po and bg) and test (pa, env) data:

po (training data) includes site names, species names, coordinates, occurrence ("1" for all, since all are presence records), group [ba = bats (7 species); db = diurnal birds (8 species); nb = nocturnal birds (2 species); ot = open-forest trees (8 species); ou = open-forest understorey plants (8 species); rt = rainforest trees (7 species); ru = rainforest understorey plants (6 species); sr = small reptiles (8 species)], and site values for 13 environmental variables (below).

bg (training data) has 10000 sites selected at random across the study region. It is structured identically to po, with "0" for occurrence (not implying absence, but denoting a background record in a way suited to most modelling methods) and NA for group.

env (testing data) includes group, site names, coordinates, and site values for 13 environmental variables (below). These are for sites from different surveys for each biological group (from 570 to 2075 sites per group), and can be returned as separate datasets by disEnv, or in one long format dataset by disData. This set of files is suited to making predictions.

pa (testing data) includes group, site names, coordinates, and presence-absence records, one column per species (in the wide format returned by disPa). They can also be returned in long format using disData. The sites are identical to the sites in env. These data are suited to evaluating the predictions made with env.

Raster (gridded) data for all environmental variables are available - see the reference below for details.

The reference system of the x and y coordinates is unprojected. Latitude and longitude are in geographical coordinates using the WGS84 datum (EPSG:4326).

The vignette provided with this package provides an example of how to fit and evaluate a model with these data.

**Environmental variables:**

| Code | Description |
|------|-------------|
| cti | "compound topographic index" - a quantification of the position of a site in the local landscape. It is often referred |
| disturb | disturbance (clearing, logging etc) index. |
| mi | moisture index. Index of site wetness derived from a water balance algorithm using rainfall, evaporation, radiation |
| rainann | mean annual rainfall |

| raindq | mean rainfall of the driest quarter |
| rugged | ruggedness. Coefficient of variation of grid cells within 1km of cell of interest |
| soildepth | mean soil depth predicted from a model relating sampled soil depths to climate, geology and topography |
| soilfert | soil fertility ordinal class, derived from soil maps and modeling of geochemical data |
| solrad | annual mean solar radiation (terrain adjusted) |
| tempann | annual mean temperature |
| tempmin | minimum temperature of the coldest month |
| topo | topographic position. Mean difference in elevation between grid cell of interest and all cells within 1km radius (-v |
| vegsys | broad vegetation type |

## Source

All data were compiled and provided by Simon Ferrier and colleagues.

## References

Elith, J., Graham, C.H., Valavi, R., Abegg, M., Bruce, C., Ferrier, S., Ford, A., Guisan, A., Hijmans, R.J., Huettmann, F., Lohmann, L.G., Loiselle, B.A., Moritz, C., Overton, J.McC., Peterson, A.T., Phillips, S., Richardson, K., Williams, S., Wiser, S.K., Wohlgemuth, T. & Zimmermann, N.E., (2020). Presence-only and presence-absence data for comparing species distribution modeling methods. Biodiversity Informatics 15:69-80.

## Examples

```
nsw_po <- disPo("NSW")
nsw_bg <- disBg("NSW")

nsw_pa_bat <- disPa("NSW", "ba")
nsw_env_bat <- disEnv("NSW", "ba")
nsw_pa_reptile <- disPa("NSW", "sr")
nsw_env_reptile <- disEnv("NSW", "sr")


# Or all in one list
nsw <- disData("NSW")
sapply(nsw, head)

disCRS("NSW")
```

---

NZ                              *New Zealand species distribution data*

---

## Description

Species occurrence data for 52 vascular plant species - mostly trees and shrubs from indigenous forests - in New Zealand (NZ), and associated environmental data. Full details of the dataset are

provided in the reference below. There are four data sets with training (po and bg) and test (pa, env) data:

po (training data) includes site names, species names, coordinates, occurrence ("1" for all, since all are presence records), group (plant), and site values for 13 environmental variables (below).

bg (training data) has 10000 sites selected at random across the study region. It is structured identically to po, with "0" for occurrence (not implying absence, but denoting a background record in a way suited to most modelling methods) and NA for group.

env (testing data) includes group, site names, coordinates, and site values for 13 environmental variables (below), at 19120 sites. These data are suited to making predictions.

pa (testing data) includes group, site names, coordinates, and presence-absence records, one column per species. The sites are identical to the sites in env. This file is suited to evaluating the predictions made to env.

Raster (gridded) data for all environmental variables are available - see the reference below for details.

The coordinate reference system of the x and y coordinates is New Zealand Map Grid (NZMG), Datum: NZGD49 (New Zealand Geodetic Datum 1949), Ellipsoid: International 1924 (EPSG:27200).

The vignette provided with this package provides an example of how to fit and evaluate a model with these data.

**Environmental variables:**

| Code | Description | Units | Type |
|------|-------------|-------|------|
| age | 3 classes (0 to 2): <2000, 2000-postglacial (app. 30,000), and pre-glacial | number (category) | Categorical |
| deficit | Mean October vapor pressure deficit at 0900 hours | kPa | Continuous |
| dem | Elevation | meters | Continuous |
| hillshade | Hill shading (as surrogate for slope and aspect) | index of brightness | Continuous |
| mas | Mean annual solar radiation | Mj/m2/day | Continuous |
| mat | Mean annual temperature | degrees C * 10 | Continuous |
| r2pet | Average monthly ratio of rainfall and potential evapotranspiration (ratio) | none | Continuous |
| rain | annual precipitation | mm | Continuous |
| slope | Slope | degrees | Continuous |
| sseas | Solar radiation seasonality | dimensionless | Continuous |
| toxicats | Toxic Cations in classes: 0=low, 1=intermediate, 2=high | number (category) | Categorical |
| tseas | Temperature seasonality | degrees C | Continuous |
| vpd | Mean October vapor pressure deficit at 9 AM | kPa | Continuous |

**Source**

Environmental predictors provided by Jake Overton. Species data supplied by Jake Overton and Susan Wiser, from Allan Herbarium and National Vegetation Survey databank.

See the reference below for further details on source, accuracy, cleaning, and particular characteristics of these datasets.

## References

Elith, J., Graham, C.H., Valavi, R., Abegg, M., Bruce, C., Ferrier, S., Ford, A., Guisan, A., Hijmans, R.J., Huettmann, F., Lohmann, L.G., Loiselle, B.A., Moritz, C., Overton, J.McC., Peterson, A.T., Phillips, S., Richardson, K., Williams, S., Wiser, S.K., Wohlgemuth, T. & Zimmermann, N.E., (2020). Presence-only and presence-absence data for comparing species distribution modeling methods. Biodiversity Informatics 15:69-80.

## Examples

```
nz_po <- disPo("NZ")
nz_bg <- disBg("NZ")

nz_pa <- disPa("NZ")
nz_env <- disEnv("NZ")

x <- disData("NZ")
sapply(x, head)

disCRS("NZ")
```

---

SA                                *South American plant species distribution data*

---

## Description

Species occurrence data for 30 vascular plant species (all from the Bignoniaceae family) from Continental Brazil, Ecuador, Colombia, Bolivia, and Peru, South America (SA), and associated environmental data. Full details of the dataset are provided in the reference below. There are four data sets with training (po and bg) and test (pa, env) data:

po (training data) includes site names, species names, coordinates, occurrence ("1" for all, since all are presence records), group (plant), and site values for 11 environmental variables (below).

bg (training data) has 10000 sites selected at random across the study region. It is structured identically to po, with "0" for occurrence (not implying absence, but denoting background in a way suited to most modelling methods) and NA for group.

env (testing data) includes group, site names, coordinates, and site values for 11 environmental variables (below), at 152 sites. This file is suited to making predictions.

pa (testing data) includes group, site names, coordinates, and presence-absence records, one column per species. The sites are identical to the sites in env. This file is suited to evaluating the predictions made to env.

Raster (gridded) data for all environmental variables are available - see the reference below for details.

The coordinate reference system of the x and y coordinates is longitude, latitude, with the WGS84 datum (EPSG:4326).

The vignette provided with this package provides an example of how to fit and evaluate a model with these data.

**Environmental variables (extracted from WorldClim):**

| Code | Description | Units | Type |
|------|-------------|-------|------|
| sabio1 | Annual mean temperature | degrees C * 10 | Continuous |
| sabio2 | Mean Diurnal Range (Mean of monthly (max temp - min temp)) | degrees C * 10 | Continuous |
| sabio4 | Temperature Seasonality (standard deviation *100) | dimensionless | Continuous |
| sabio5 | Max Temperature of Warmest Month | degrees C * 10 | Continuous |
| sabio6 | Min Temperature of Coldest Month | degrees C * 10 | Continuous |
| sabio7 | Temperature Annual Range | degrees C * 10 | Continuous |
| sabio8 | Mean Temperature of Wettest Quarter | mm | Continuous |
| sabio12 | Annual Precipitation | mm | Continuous |
| sabio15 | Precipitation Seasonality (Coefficient of Variation) | mm | Continuous |
| sabio17 | Precipitation of Driest Quarter | mm | Continuous |
| sabio18 | Precipitation of Warmest Quarter | mm | Continuous |

**Source**

Environmental data prepared by Bette Loiselle, Lucia Lohmann and Catherine Graham. Species supplied by Bette Loiselle and Lucia Lohmann. PO data from the Missouri Botanical Gardens database and Lucia Lohmann; PA data collected by Al Gentry.

See the reference below for further details on source, accuracy, cleaning, and particular characteristics of these datasets.

**References**

Elith, J., Graham, C.H., Valavi, R., Abegg, M., Bruce, C., Ferrier, S., Ford, A., Guisan, A., Hijmans, R.J., Huettmann, F., Lohmann, L.G., Loiselle, B.A., Moritz, C., Overton, J.McC., Peterson, A.T., Phillips, S., Richardson, K., Williams, S., Wiser, S.K., Wohlgemuth, T. & Zimmermann, N.E., (2020). Presence-only and presence-absence data for comparing species distribution modeling methods. Biodiversity Informatics 15:69-80.

**Examples**

```
sa_po <- disPo("SA")
sa_bg <- disBg("SA")

sa_pa <- disPa("SA")
sa_env <- disEnv("SA")

x <- disData("SA")
sapply(x, head)

disCRS("SA")
```

---

SWI                            *Swiss species distribution data*

---

## Description

Species occurrence data for 30 tree species in Switzerland (SWI, a country in Europe) and associated environmental data. Full details of the dataset are provided in the reference below. There are four data sets with training (po and bg) and test (pa, env) data:

po (training data) includes site names, species names, coordinates, occurrence ("1" for all, since all are presence records), group (tree), and site values for 13 environmental variables (below).

bg (training data) has 10000 sites selected at random across the study region. It is structured identically to po, with "0" for occurrence (not implying absence, but denoting background in a way suited to most modelling methods) and NA for group.

env (testing data) includes group, site names, coordinates, and site values for 13 environmental variables (below), at 10103 sites. This file is suited to making predictions.

pa (testing data) includes group, site names, coordinates, and presence-absence records, one column per species. The sites are identical to the sites in env. This file is suited to evaluating the predictions made to env.

Raster (gridded) data for all environmental variables are available - see the reference below for details.

The reference system of the x and y coordinates is Transverse, spheroid Bessel (EPSG:21781) (note all SWI data has a constant shift applied).

The vignette provided with this package provides an example of how to fit and evaluate a model with these data.

**Environmental variables:**

| Code | Description | Units | Type |
|------|-------------|-------|------|
| bcc | Broadleaved continuous cover (based on Landsat images) | percentage | Continuous |
| calc | Bedrock is strictly calcareous | 1 (yes) or 0 (no) | Categorical |
| ccc | Coniferous continuous cover (based on Landsat images) | percentage | Continuous |
| ddeg | Growing degree-days above a threshold of 0 degrees C | degrees C * days | Continuous |
| nutri | Soil nutrients index between 0-45 | D mval/cm2 | Continuous |
| pdsum | Number of days with rainfall higher than 1 mm | ndays | Continuous |
| precyy | Average yearly precipitation sum | mm | Continuous |
| sfro | Summer Frost Frequency | days | Continuous |
| slope | Slope | degrees x 10 | Continuous |
| sradyy | Potential yearly global radiation (daily average) | (kJ/m2)/day | Continuous |
| swb | Site water balance | mm | Continuous |
| tavecc | Average temperature of the coldest month | degrees C | Continuous |
| topo | Topographic position | dimensionless | Continuous |

## Source

Environmental predictors supplied by Niklaus E. Zimmermann. Species data supplied by Niklaus E. Zimmermann, Thomas Wohlgemuth and Meinrad Abegg.

See the reference below for further details on source, accuracy, cleaning, and particular characteristics of these datasets.

## References

Elith, J., Graham, C.H., Valavi, R., Abegg, M., Bruce, C., Ferrier, S., Ford, A., Guisan, A., Hijmans, R.J., Huettmann, F., Lohmann, L.G., Loiselle, B.A., Moritz, C., Overton, J.McC., Peterson, A.T., Phillips, S., Richardson, K., Williams, S., Wiser, S.K., Wohlgemuth, T. & Zimmermann, N.E., (2020). Presence-only and presence-absence data for comparing species distribution modeling methods. Biodiversity Informatics 15:69-80.

## Examples

```
swi_po <- disPo("SWI")
swi_bg <- disBg("SWI")

swi_pa <- disPa("SWI")
swi_env <- disEnv("SWI")

x <- disData("SWI")
sapply(x, head)

disCRS("SWI")
```

# Index