

Package ‘rEDM’

August 30, 2026

Type Package

Title Empirical Dynamic Modeling ('EDM')

Version 2.0.2

Date 2026-08-21

Maintainer Joseph Park <JosephPark@IEEE.org>

Description An implementation of 'EDM' algorithms based on research software developed at the Sugihara Lab ('UCSD/SIO'). Primary methods include 'Simplex' projection from Sugihara & May (1990) <doi:10.1038/344734a0>, Sequential locally-weighted global linear maps 'S-map': Sugihara (1994) <doi:10.1098/rsta.1994.0106>, Convergent cross mapping described in Sugihara et al. (2012) <doi:10.1126/science.1227079>, and, 'Multiview embedding' from Ye & Sugihara (2016) <doi:10.1126/science.aag0863>.

License BSD_2_clause + file LICENSE

Encoding UTF-8

LazyData true

LazyLoad yes

Imports stats, utils, graphics, parallel, RANN

Suggests testthat (>= 3.0.0), knitr, rmarkdown, formatR

Config/testthat/edition 3

RoxygenNote 8.0.0

VignetteBuilder knitr

NeedsCompilation no

Author Joseph Park [aut, cre] (ORCID: <<https://orcid.org/0000-0001-5411-1409>>),
George Sugihara [aut, ccp] (ORCID:
<<https://orcid.org/0000-0002-2863-6946>>),
Ethan Deyle [aut] (ORCID: <<https://orcid.org/0000-0001-8704-8434>>),
The Regents of the University of California [cph]

Repository CRAN

Date/Publication 2026-08-30 05:10:18 UTC

Contents

block_3sp	2
CCM	3
circle	5
ComputeError	5
Embed	6
EmbedDimension	7
EvergladesFlow	8
Lorenz5D	8
Multiview	9
paramecium_didinium	11
PlotObsPred	11
PlotSmap	12
PredictExclusionRadius	12
PredictInterval	13
PredictNonlinear	15
rEDM	16
sardine_anchovy_sst	17
Simplex	18
SMap	20
SurrogateData	22
TentMap	23
TentMapNoise	23
Thrips	24
Index	25

block_3sp	<i>Time series for a three-species coupled model.</i>
-----------	---

Description

Time series generated from a discrete-time coupled Lotka-Volterra model exhibiting chaotic dynamics.

Usage

block_3sp

Format

- A data frame with 198 rows and 10 columns:
- time time index (# of generations)
 - x_t abundance of simulated species x at time t
 - x_t-1 abundance of simulated species x at time t-1
 - x_t-2 abundance of simulated species x at time t-2

`y_t` abundance of simulated species y at time t
`y_t-1` abundance of simulated species y at time t-1
`y_t-2` abundance of simulated species y at time t-2
`z_t` abundance of simulated species z at time t
`z_t-1` abundance of simulated species z at time t-1
`z_t-2` abundance of simulated species z at time t-2

CCM

Convergent cross mapping

Description

The state-space of a multivariate dynamical system (not a purely stochastic one) encodes coherent phase-space variable trajectories. If enough information is available, one can infer the presence or absence of cross-variable interactions associated with causal links between variables. CCM measures the extent to which states of variable Y can reliably estimate states of variable X. This can happen if X is causally influencing Y.

If cross-variable state predictability converges as more state-space information is provided, this indicates a causal link. CCM performs this cross-variable mapping using Simplex, with convergence assessed across a range of observational library sizes as described in *Sugihara et al. 2012*.

Usage

```
CCM(dataFrame = NULL, columns, target, E, Tp = 0, knn = 0, tau = -1,
     exclusionRadius = 0, libSizes, sample = 30, seed = NULL,
     embedded = FALSE, validLib = logical(0), includeData = FALSE,
     numProcess = 4, backend = "RANN", pathIn = "./", dataFile = "",
     pathOut = "./", predictFile = "", parameterList = FALSE, showPlot = FALSE)
```

Arguments

<code>dataFrame</code>	A data.frame of input data. The first column must be a time index or time values unless <code>noTime = TRUE</code> . The columns must be named.
<code>columns</code>	Column name(s) to build the embedding: character vector or space-separated string of columns to create the library. If individual column names contain whitespace place names in a vector, or, append ',' to the name.
<code>target</code>	Target column name to predict.
<code>E</code>	Embedding dimension.
<code>Tp</code>	Forecast interval (prediction horizon).
<code>knn</code>	Number of nearest neighbours. 0 selects the method default.
<code>tau</code>	Embedding delay (negative selects past lags).
<code>exclusionRadius</code>	Temporal (Theiler) exclusion radius around each prediction point.

<code>libSizes</code>	Library sizes: string of 3 whitespace separated integer values specifying the initial library size, the final library size, and the library size increment. Can also be a list of strictly ncreasing library sizes.
<code>sample</code>	Number of random subsamples per library size.
<code>seed</code>	Random seed for reproducibility (not bit-comparable to pyEDM/NumPy).
<code>embedded</code>	If TRUE, columns already form the embedding.
<code>validLib</code>	Logical vector marking admissible library rows (or length 0 for all).
<code>includeData</code>	Add per-library-size sample-variance columns to the result.
<code>numProcess</code>	Number of worker processes for the parameter sweep or task grid.
<code>backend</code>	Nearest-neighbour backend: "RANN" (default) or "brute".
<code>pathIn</code>	File path for input dataFile.
<code>dataFile</code>	Input dataFile, .csv format. The first column must be a time index or time values unless noTime is TRUE. The first row must be column names.
<code>pathOut</code>	Output file path for predictFile
<code>predictFile</code>	Output file name, .csv format.
<code>parameterList</code>	Append named list of parameters/values to return.
<code>showPlot</code>	If TRUE, draw a base-graphics plot of the result.

Value

A data.frame with 3 columns. The first column is LibSize specifying the subsampled library size. Columns 2 and 3 report Pearson correlation coefficients for the prediction of X from Y, and Y from X.

If `includeData = TRUE` and `parameterList = TRUE` a named list "parameters" is added.

References

Sugihara G., May R., Ye H., Hsieh C., Deyle E., Fogarty M., Munch S., 2012. Detecting Causality in Complex Ecosystems. *Science* 338:496-500.

Examples

```
data(sardine_anchovy_sst)
df = CCM( dataFrame = sardine_anchovy_sst, E = 3, Tp = 0, columns = "anchovy",
target = "np_sst", libSizes = "10 70 10", sample = 100 )
```

circle	<i>2-D timeseries of a circle.</i>
--------	------------------------------------

Description

Time series of of circle in 2-D (sin and cos).

Usage

circle

Format

A data frame with 200 rows and 3 columns:

Time time index.

x sin component.

y cos component.

ComputeError	<i>Prediction skill statistics</i>
--------------	------------------------------------

Description

Pearson rho, MAE, CAE and RMSE between observations and predictions.

Usage

ComputeError(obs, pred, digits = 6)

Arguments

- obs Numeric vector of observations.
- pred Numeric vector of predictions aligned with obs.
- digits Rounding applied to each statistic.

Value

A named list with elements rho, MAE, CAE and RMSE.

Examples

ComputeError(1:10 + 0, c(1:9 + 0, 8))

Embed	<i>Time-delay embedding</i>
-------	-----------------------------

Description

Takens time-delay embedding of one or more columns, returned as a named data.frame matching pyEDM's Embed().

Usage

```
Embed(dataFrame = NULL, E = 0, tau = -1, columns = "",
      includeTime = FALSE, pathIn = "./", dataFile = "")
```

Arguments

dataFrame	A data.frame of input data. Column 1 is the time vector.
E	Embedding dimension.
tau	Embedding delay in time-column rows (negative selects past lags).
columns	Column name(s) to embed: a whitespace-separated string, or a character vector of column names.
includeTime	If TRUE, prepend column 1 of dataFrame (the time vector), keeping its original name, as the first output column.
pathIn	Directory for dataFile (used when reading input from disk).
dataFile	CSV file name to read as input; when set it takes precedence over dataFrame.

Details

Each columns item produces E lagged vectors. For column c and step $i = 0 \dots E-1$ the lag is $i * |\tau|$, and the column is named $c(t-lag)$ when $\tau < 0$ or $c(t+lag)$ when $\tau > 0$. For example columns X, Y with $E = 2$, $\tau = -1$ yield $X(t-0) X(t-1) Y(t-0) Y(t-1)$; with $\tau = -2$ they yield $X(t-0) X(t-2) Y(t-0) Y(t-2)$.

If dataFile is supplied the CSV at file.path(pathIn, dataFile) is read (preserving column names) and used as the input, taking precedence over dataFrame.

All rows are retained; the first $|\tau| * (E-1)$ partial rows contain NA.

Value

A data.frame of the embedding (with the time column first when includeTime = TRUE).

Examples

```
Embed(data.frame(x = 1:10), E = 3, tau = -1, columns = "x")
```

EmbedDimension	<i>Prediction skill versus embedding dimension</i>
----------------	--

Description

`EmbedDimension` uses `Simplex` to evaluate prediction accuracy as a function of embedding dimension.

Usage

```
EmbedDimension(dataFrame = NULL, columns, target, lib, pred, maxE = 10, Tp = 1,
               tau = -1, exclusionRadius = 0, embedded = FALSE,
               validLib = logical(0), noTime = FALSE, ignoreNan = TRUE,
               numProcess = 4, pathIn = "./", dataFile = "", pathOut = "./",
               predictFile = "", backend = "RANN",
               parameterList = FALSE, showPlot = FALSE)
```

Arguments

<code>dataFrame</code>	A data.frame of input data. The first column must be a time index or time values unless <code>noTime = TRUE</code> . The columns must be named.
<code>columns</code>	Column name(s) to build the embedding: character vector or space-separated string.
<code>target</code>	Target column name to predict.
<code>lib</code>	Library (training) index range as (start end) pairs.
<code>pred</code>	Prediction index range as (start end) pairs.
<code>maxE</code>	Largest embedding dimension to evaluate.
<code>Tp</code>	Forecast interval (prediction horizon).
<code>tau</code>	Embedding delay (negative selects past lags).
<code>exclusionRadius</code>	Temporal (Theiler) exclusion radius around each prediction point.
<code>embedded</code>	If TRUE, columns already form the embedding.
<code>validLib</code>	Logical vector marking admissible library rows (or length 0 for all).
<code>noTime</code>	If TRUE, synthesise a 1..N time index instead of using column 1.
<code>ignoreNan</code>	Remove rows with NaN in the embedding from the library and prediction sets.
<code>numProcess</code>	Number of worker processes for the parameter sweep or task grid.
<code>pathIn</code>	File path for input dataFile.
<code>dataFile</code>	Input dataFile, .csv format.
<code>pathOut</code>	Output file path for predictFile
<code>predictFile</code>	Output file name, .csv format.
<code>backend</code>	Nearest-neighbour backend: "RANN" (default) or "brute".
<code>parameterList</code>	Append named list of parameters/values to return.
<code>showPlot</code>	If TRUE, draw a base-graphics plot of the result.

Value

A data.frame with columns E, rho.

Examples

```
data(TentMap)
E.rho = EmbedDimension( dataFrame = TentMap, lib = "1 100", pred = "201 500",
  columns = "TentMap", target = "TentMap", showPlot = FALSE )
```

EvergladesFlow	<i>Water flow to NE Everglades</i>
----------------	------------------------------------

Description

Cumulative weekly water flow into northeast Everglades from water control structures S12C, S12D and S333 from 1980 through 2005.

Usage

EvergladesFlow

Format

A data frame with 1379 rows and 2 columns:
Date Date.
S12CD_S333_CFS Cumulative weekly flow (CFS).

Lorenz5D	<i>5-D Lorenz'96</i>
----------	----------------------

Description

5-D Lorenz'96 timeseries with F = 8.

Usage

Lorenz5D

Format

Data frame with 1000 rows and 6 columns
Time Time.
V1 variable 1.
V2 variable 2.
V3 variable 3.
V4 variable 4.
V5 variable 5.

References

Lorenz, Edward (1996). Predictability - A problem partly solved, Seminar on Predictability, Vol. I, ECMWF.

Multiview

Multiview embedding forecast

Description

Multiview embedding forecast (Ye & Sugihara 2016): rank D-column views and average the best to find optimal combinations of variables that best represent the dynamics.

Usage

```
Multiview(dataFrame = NULL, columns, target, lib = numeric(0), pred = numeric(0),
          D = 0, E = 1, Tp = 1, knn = 0, tau = -1, multiview = 0,
          exclusionRadius = 0, trainLib = TRUE, excludeTarget = FALSE,
          ignoreNan = TRUE, numProcess = 4, backend = "RANN",
          pathIn = "./", dataFile = "", pathOut = "./", predictFile = "",
          parameterList = FALSE, showPlot = FALSE)
```

Arguments

dataFrame	A data.frame of input data. The first column is time unless noTime = TRUE.
columns	Column name(s) to build the embedding: character vector or space-separated string. If individual column names contain whitespace place names in a vector, or, append ',' to the name.
target	Target column name to predict.
lib	Library (training) index range as (start end) pairs.
pred	Prediction index range as (start end) pairs.
D	View dimension for Multiview (number of embedding columns combined).
E	Embedding dimension.
Tp	Forecast interval (prediction horizon).
knn	Number of nearest neighbours. 0 selects the method default.
tau	Embedding delay (negative selects past lags).
multiview	Number of top-ranked views to average (default floor(sqrt(nCombo))).
exclusionRadius	Temporal (Theiler) exclusion radius around each prediction point.
trainLib	Rank views on the library (in-sample) rather than the prediction set.
excludeTarget	Exclude the target variable from candidate views.
ignoreNan	Remove rows with NaN in the embedding from the library and prediction sets.
numProcess	Number of worker processes for the parameter sweep or task grid.

backend	Nearest-neighbour backend: "RANN" (default) or "brute".
pathIn	File path for input dataFile.
dataFile	Input dataFile, .csv format.
pathOut	Output file path for predictFile
predictFile	Output file name, .csv format.
parameterList	Append named list of parameters/values to return.
showPlot	If TRUE, draw a base-graphics plot of the result.

Details

Multiview embedding is a method to identify variables in a multivariate dynamical system that are most likely to contribute to the observed dynamics. It is a multistep algorithm with these general steps:

1. Compute D-dimensional variable combination forecasts.
2. Rank forecasts.
3. Compute predictions of top combinations.
4. Compute multiview averaged prediction.

If $E > 1$, all variables are embedded to dimension E . If `trainLib` is TRUE initial forecasts and ranking are done in-sample (`lib=pred`) and predictions using the top ranked combinations use the specified `lib` and `pred`. If `trainLib` is FALSE initial forecasts and ranking use the specified `lib` and `pred`, the step of computing predictions of the top combinations is skipped.

Value

Named list with `data.frames` `[[View, Predictions]]`.

`data.frame` View columns:

Col_1	column index
...	column index
Col_D	column index
rho	Pearson correlation
MAE	mean absolute error
RMSE	root mean square error
name_1	column name
...	column name
name_D	column name

If `parameterList = TRUE` a named list "parameters" is added.

References

Ye H., and G. Sugihara, 2016. Information leverage in interconnected ecosystems: Overcoming the curse of dimensionality. *Science* 353:922-925.

Examples

```
data(block_3sp)
L = Multiview( dataFrame = block_3sp, lib = "1 100", pred = "101 190",
E = 2, columns = "x_t y_t z_t", target = "x_t" )
```

paramecium_didinium	<i>Time series for the Paramecium-Didinium laboratory experiment</i>
---------------------	--

Description

Time series of Paramecium and Didinium abundances (#/mL) from an experiment by Veilleux (1979)

Usage

```
paramecium_didinium
```

PlotObsPred	<i>Plot observed versus predicted values</i>
-------------	--

Description

Base-graphics plot of Observations and Predictions against time.

Usage

```
PlotObsPred(df, dataName = "", E = NULL, Tp = NULL)
```

Arguments

- df A data.frame with a time column plus Observations and Predictions.
- dataName Optional label used in the plot title.
- E Embedding dimension.
- Tp Forecast interval (prediction horizon).

Value

Invisibly NULL.

PlotSmap	<i>Plot S-map predictions and coefficients</i>
----------	--

Description

Base-graphics multi-panel plot of S-map predictions and coefficient series.

Usage

```
PlotSmap(SmapList, dataName = "", E = NULL, Tp = NULL)
```

Arguments

SmapList	A list returned by SMap (predictions and coefficients).
dataName	Optional label used in the plot title.
E	Embedding dimension.
Tp	Forecast interval (prediction horizon).

Value

Invisibly NULL.

PredictExclusionRadius	<i>Prediction skill versus exclusion radius</i>
------------------------	---

Description

Simplex prediction skill (ρ) swept over the temporal exclusion radius.

Usage

```
PredictExclusionRadius(dataFrame = NULL, columns, target, lib, pred,
  exclusionRadius = NULL, E = 1, Tp = 1, tau = -1,
  embedded = FALSE, validLib = logical(0),
  noTime = FALSE, ignoreNan = TRUE, numProcess = 4,
  backend = "RANN", pathIn = "./", dataFile = "",
  pathOut = "./", predictFile = "",
  parameterList = FALSE, showPlot = FALSE)
```

Arguments

dataFrame	A data.frame of input data. The first column must be a time index or time values unless noTime = TRUE.
columns	Column name(s) to build the embedding: character vector or space-separated string.
target	Target column name to predict.
lib	Library (training) index range as (start end) pairs.
pred	Prediction index range as (start end) pairs.
exclusionRadius	Temporal (Theiler) exclusion radius around each prediction point.
E	Embedding dimension.
Tp	Forecast interval (prediction horizon).
tau	Embedding delay (negative selects past lags).
embedded	If TRUE, columns already form the embedding.
validLib	Logical vector marking admissible library rows (or length 0 for all).
noTime	If TRUE, synthesise a 1..N time index instead of using column 1.
ignoreNan	Remove rows with NaN in the embedding from the library and prediction sets.
numProcess	Number of worker processes for the parameter sweep or task grid.
backend	Nearest-neighbour backend: "RANN" (default) or "brute".
pathIn	File path for input dataFile.
dataFile	Input dataFile, .csv format.
pathOut	Output file path for predictFile
predictFile	Output file name, .csv format.
parameterList	Append named list of parameters/values to return.
showPlot	If TRUE, draw a base-graphics plot of the result.

Value

A data.frame with columns ExclusionRadius and rho.

PredictInterval	<i>Prediction skill versus forecast interval</i>
-----------------	--

Description

`PredictInterval` uses `Simplex` to evaluate prediction accuracy as a function of forecast interval `Tp`.

Usage

```
PredictInterval(dataFrame = NULL, columns, target, lib, pred, maxTp = 10, E = NULL,
  tau = -1, exclusionRadius = 0, embedded = FALSE,
  validLib = logical(0), noTime = FALSE, ignoreNan = TRUE,
  numProcess = 4, backend = "RANN", pathIn = "./",
  dataFile = "", pathOut = "./", predictFile = "",
  parameterList = FALSE, showPlot = FALSE)
```

Arguments

dataFrame	A data.frame of input data. The first column must be a time index or time values unless noTime = TRUE. The columns must be named.
columns	Column name(s) to build the embedding: character vector or space-separated string.
target	Target column name to predict.
lib	Library (training) index range as (start end) pairs.
pred	Prediction index range as (start end) pairs.
maxTp	Largest forecast interval to evaluate.
E	Embedding dimension. Required (no default); must be a positive integer unless embedded = TRUE, in which case it is inferred as the number of columns.
tau	Embedding delay (negative selects past lags).
exclusionRadius	Temporal (Theiler) exclusion radius around each prediction point.
embedded	If TRUE, columns already form the embedding.
validLib	Logical vector marking admissible library rows (or length 0 for all).
noTime	If TRUE, synthesise a 1..N time index instead of using column 1.
ignoreNan	Remove rows with NaN in the embedding from the library and prediction sets.
numProcess	Number of worker processes for the parameter sweep or task grid.
backend	Nearest-neighbour backend: "RANN" (default) or "brute".
pathIn	File path for input dataFile.
dataFile	Input dataFile, .csv format.
pathOut	Output file path for predictFile
predictFile	Output file name, .csv format.
parameterList	Append named list of parameters/values to return.
showPlot	If TRUE, draw a base-graphics plot of the result.

Value

A data.frame with columns Tp, rho.

Examples

```
data(TentMap)
Tp.rho = PredictInterval( dataFrame = TentMap, lib = "1 100",
  pred = "201 500", E = 2, columns = "TentMap", target = "TentMap",
  showPlot = FALSE )
```

PredictNonlinear	<i>Prediction skill versus S-map localisation</i>
------------------	---

Description

PredictNonlinear uses **SMap** to evaluate prediction accuracy as a function of the state-space localization parameter θ .

Usage

```
PredictNonlinear(dataFrame = NULL, columns, target, lib, pred, theta = NULL, E = NULL,
  Tp = 1, knn = 0, tau = -1, exclusionRadius = 0,
  embedded = FALSE, validLib = logical(0), noTime = FALSE,
  ignoreNan = TRUE, numProcess = 4, backend = "RANN",
  pathIn = "./", dataFile = "", pathOut = "./", predictFile = "",
  parameterList = FALSE, showPlot = FALSE, verbose = FALSE)
```

Arguments

dataFrame	A data.frame of input data. The first column must be a time index or time values unless noTime = TRUE. The columns must be named.
columns	Column name(s) to build the embedding: character vector or space-separated string.
target	Target column name to predict.
lib	Library (training) index range as (start end) pairs.
pred	Prediction index range as (start end) pairs.
theta	S-map localisation parameter ($\emptyset$ = global linear map).
E	Embedding dimension. Required (no default); must be a positive integer unless embedded = TRUE, in which case it is inferred as the number of columns.
Tp	Forecast interval (prediction horizon).
knn	Number of nearest neighbours. 0 selects the method default.
tau	Embedding delay (negative selects past lags).
exclusionRadius	Temporal (Theiler) exclusion radius around each prediction point.
embedded	If TRUE, columns already form the embedding.
validLib	Logical vector marking admissible library rows (or length 0 for all).
noTime	If TRUE, synthesise a 1..N time index instead of using column 1.

ignoreNaN	Remove rows with NaN in the embedding from the library and prediction sets.
numProcess	Number of worker processes for the parameter sweep or task grid.
backend	Nearest-neighbour backend: "RANN" (default) or "brute".
pathIn	File path for input dataFile.
dataFile	Input dataFile, .csv format.
pathOut	Output file path for predictFile
predictFile	Output file name, .csv format.
parameterList	Append named list of parameters/values to return.
showPlot	If TRUE, draw a base-graphics plot of the result.
verbose	Not used.

Value

A data.frame with columns Theta and rho.

Examples

```
data(TentMapNoise)
theta.rho = PredictNonlinear( dataFrame = TentMapNoise, E = 2,
  lib = "1 100", pred = "201 500", columns = "TentMap",
  target = "TentMap", showPlot = FALSE )
```

rEDM

Empirical dynamic modeling

Description

rEDM provides tools for Empirical Dynamic Modeling time series analyses. It is based on reconstructing multivariate state space representations from uni- or multivariate time series, then projecting state changes using various metrics applied to nearest neighbors.

Functionality includes:

- Simplex projection (Sugihara and May 1990)
- Sequential Locally Weighted Global Linear Maps (S-map) (Sugihara 1994)
- Multivariate embeddings (Dixon et. al. 1999)
- Convergent cross mapping (Sugihara et. al. 2012)
- Multiview embedding (Ye and Sugihara 2016)

Details**Main Functions:**

- [Simplex](#) - simplex projection
- [SMap](#) - S-map projection
- [CCM](#) - convergent cross mapping
- [Multiview](#) - multiview forecasting

Helper Functions:

- [Embed](#) - time delay embedding
- [ComputeError](#) - forecast skill metrics
- [EmbedDimension](#) - optimal embedding dimension
- [PredictInterval](#) - optimal prediction interval
- [PredictNonlinear](#) - evaluate nonlinearity

Author(s)

Maintainer: Joseph Park

Authors: Joseph Park, Ethan Deyle, George Sugihara

References

- Sugihara G. and May R. 1990. Nonlinear forecasting as a way of distinguishing chaos from measurement error in time series. *Nature*, 344:734-741.
- Sugihara G. 1994. Nonlinear forecasting for the classification of natural time series. *Philosophical Transactions: Physical Sciences and Engineering*, 348 (1688) : 477-495.
- Dixon, P. A., M. Milicich, and G. Sugihara, 1999. Episodic fluctuations in larval supply. *Science* 283:1528-1530.
- Sugihara G., May R., Ye H., Hsieh C., Deyle E., Fogarty M., Munch S., 2012. Detecting Causality in Complex Ecosystems. *Science* 338:496-500.
- Ye H., and G. Sugihara, 2016. Information leverage in interconnected ecosystems: Overcoming the curse of dimensionality. *Science* 353:922-925.

sardine_anchovy_sst *Time series for the California Current Anchovy-Sardine-SST system*

Description

Time series of Pacific sardine landings (CA), Northern anchovy landings (CA), and sea-surface temperature (3-year average) at the SIO pier and Newport pier

Usage

sardine_anchovy_sst

Format

year year of measurement
 anchovy anchovy landings, scaled to mean = 0, sd = 1
 sardine sardine landings, scaled to mean = 0, sd = 1
 sio_sst 3-year running average of sea surface temperature at SIO pier, scaled to mean = 0, sd = 1
 np_sst 3-year running average of sea surface temperature at Newport pier, scaled to mean = 0, sd = 1

Simplex

*Simplex projection***Description**

Simplex performs time series forecasting based on weighted nearest neighbors projection in the time series phase space as described in *Sugihara and May*.

Usage

```
Simplex(dataFrame = NULL, columns, target, lib, pred, E = NULL, Tp = 1, knn = 0,
        tau = -1, exclusionRadius = 0, embedded = FALSE,
        validLib = logical(0), noTime = FALSE, ignoreNan = TRUE,
        backend = "RANN", pathIn = "./", dataFile = "", pathOut = "./",
        predictFile = "", parameterList = FALSE, includeState = FALSE,
        verbose = FALSE, showPlot = FALSE, .tieBreak = TRUE)
```

Arguments

dataFrame	A data.frame of input data. The first column is time unless noTime = TRUE. The columns must be named.
columns	Column name(s) to build the embedding: character vector string of whitespace separated column name(s), or vector of column names used to create the library. If individual column names contain whitespace place names in a vector, or, append ',' to the name.
target	Target column name to predict.
lib	Library (training) index range as (start end) pairs. Multiple row index pairs can be specified with each pair defining the first and last rows of time series observation segments used to create the library.
pred	Prediction index range as (start end) pairs. A single contiguous range is supported.
E	Embedding dimension. Required (no default); must be a positive integer unless embedded = TRUE, in which case it is inferred as the number of columns.
Tp	Forecast interval (prediction horizon, number of time column rows).
knn	Number of nearest neighbours. If knn=0 knn is set to E+1..

tau	Embedding delay (negative selects past lags).
exclusionRadius	Temporal (Theiler) exclusion radius around each prediction point excludes vectors from the search space of nearest neighbors if their relative time index is within exclusionRadius.
embedded	If TRUE columns already form the embedding.
validLib	Logical vector marking admissible library rows (or length 0 for all).
noTime	If TRUE synthesise a 1..N time index instead of using column 1.
ignoreNan	Remove rows with NaN in the embedding from the library and prediction sets.
backend	Nearest-neighbour backend: "RANN" (default) or "brute".
pathIn	File path for input dataFile.
dataFile	Input dataFile, .csv format. The first column must be a time index or time values unless noTime is TRUE. The first row must be column names.
pathOut	Output file path for predictFile
predictFile	Output file name, .csv format.
parameterList	Append named list of parameters/values to return.
includeState	If TRUE, also return an internal list of engine state: knn_neighbors, knn_distances, lib_i, pred_i, targetVec, embedding (1-based row indices; 0 = neighbour sentinel).
verbose	Emit diagnostic messages.
showPlot	If TRUE draw a base-graphics plot of the result.
.tieBreak	If TRUE implement nearest neighbor tie breaking.

Details

If embedded is FALSE, the data column(s) are embedded to dimension E with time lag tau. This embedding forms an E-dimensional phase space for the [Simplex](#) projection. If embedded is TRUE, the data are assumed to contain an E-dimensional embedding with E equal to the number of columns. Predictions are made using leave-one-out cross-validation, i.e. observation vectors are excluded from the prediction simplex.

To assess an optimal embedding dimension [EmbedDimension](#) can be applied. Accuracy statistics can be estimated by [ComputeError](#).

Value

A data.frame with columns Time, Observations, Predictions and Pred_Variance.

If includeState = TRUE a named list "internal" is added.

References

Sugihara G. and May R. 1990. Nonlinear forecasting as a way of distinguishing chaos from measurement error in time series. *Nature*, 344:734-741.

Examples

```
data( block_3sp )
smplx = Simplex(block_3sp, "x_t", "x_t", lib = c(1, 100), pred = c(101, 195), E = 3)
ComputeError( smplx $ Predictions, smplx $ Observations )
```

SMap

Sequential locally weighted global linear map (S-map)

Description

[SMap](#) performs time series forecasting based on localised (or global) nearest neighbor projection in the time series phase space as described in *Sugihara 1994*.

Usage

```
SMap(dataFrame = NULL, columns, target, lib, pred, E = NULL, Tp = 1, knn = 0,
      tau = -1, theta = 0, exclusionRadius = 0, embedded = FALSE,
      validLib = logical(0), noTime = FALSE, ignoreNan = TRUE,
      backend = "RANN", pathIn = "./", dataFile = "", pathOut = "./",
      predictFile = "", parameterList = FALSE, includeState = FALSE,
      verbose = FALSE, showPlot = FALSE)
```

Arguments

dataFrame	A data.frame of input data. The first column must be a time index or time values unless noTime = TRUE. The columns must be named.
columns	Column name(s) to build the embedding: character vector or space-separated string, or vector of column names used to create the library. If individual column names contain whitespace place names in a vector, or, append ',' to the name.
target	Target column name to predict.
lib	string or vector with start and stop indices of input data rows used to create the library from observations. Multiple row index pairs can be specified with each pair defining the first and last rows of time series observation segments used to create the library.
pred	string with start and stop indices of input data rows used for predictions. A single contiguous range is supported.
E	Embedding dimension. Required (no default); must be a positive integer unless embedded = TRUE, in which case it is inferred as the number of columns.
Tp	Forecast interval (prediction horizon).
knn	Number of nearest neighbours. 0 selects the method default.
tau	Embedding delay (negative selects past lags).
theta	S-map localisation parameter (0 = global linear map).

exclusionRadius	Temporal (Theiler) exclusion radius around each prediction point. Excludes vectors from the search space of nearest neighbors if their relative time index is within exclusionRadius.
embedded	If TRUE, columns already form the embedding.
validLib	Logical vector marking admissible library rows (or length 0 for all).
noTime	If TRUE, synthesise a 1..N time index instead of using column 1.
ignoreNan	Remove rows with NaN in the embedding from the library and prediction sets.
pathIn	File path for input dataFile.
dataFile	Input dataFile, .csv format. The first column must be a time index or time values unless noTime is TRUE. The first row must be column names.
pathOut	Output file path for predictFile
predictFile	Output file name, .csv format.
backend	Nearest-neighbour backend: "RANN" (default) or "brute".
parameterList	Append named list of parameters/values to return.
includeState	If TRUE, also return an internal list of engine state: knn_neighbors, knn_distances, lib_i, pred_i, targetVec, embedding (1-based row indices; 0 = neighbour sentinel).
verbose	Emit diagnostic messages.
showPlot	If TRUE, draw a base-graphics plot of the result.

Details

If embedded is FALSE, the data column(s) are embedded to dimension E with time lag tau. This embedding forms an n-columns * E-dimensional phase space for the [SMap](#) projection. If embedded is TRUE, the data are assumed to contain an E-dimensional embedding with E equal to the number of columns. See the Note below for proper use of multivariate data (number of columns > 1).

If ignoreNan is TRUE, the library (lib) is internally redefined to exclude nan embedding vectors. If ignoreNan is FALSE no library adjustment is made. The (lib) can be explicitly specified to exclude nan library vectors.

Predictions are made using leave-one-out cross-validation, i.e. observation rows are excluded from the prediction regression.

In contrast to [Simplex](#), [SMap](#) uses all available neighbors and weights them with an exponential decay in phase space distance with exponent theta. theta=0 uses all neighbors corresponding to a global autoregressive model. As theta increases, neighbors closer in vicinity to the observation are considered.

Value

A named list with three data.frames [[predictions, coefficients, singularValues]]. predictions has columns Observations, Predictions. The first column contains time or index values.

coefficients data.frame has time or index values in the first column. Columns 2 through E+2 (E+1 columns) are the SMap coefficients.

singularValues data.frame has time or index values in the first column. Columns 2 through E+2 (E+1 columns) are the SVD singularValues. The first value corresponds to the SVD bias (intercept) term.

If parameterList = TRUE a named list "parameters" is added.

If includeState = TRUE a named list "internal" is added.

Note

[SMap](#) should be called with columns explicitly corresponding to dimensions E. In the univariate case (number of columns = 1) with default embedded = FALSE, the time series will be time-delay embedded to dimension E, SMap coefficients correspond to each dimension.

If a multivariate data set is used (number of columns > 1) it must use embedded = TRUE with E equal to the number of columns. This prevents the function from internally time-delay embedding the multiple columns to dimension E. If the internal time-delay embedding is performed, then state-space columns will not correspond to the intended dimensions in the matrix inversion, coefficient assignment, and prediction. In the multivariate case, the user should first prepare the embedding (using [Embed](#) for time-delay embedding), then pass this embedding to [SMap](#) with appropriately specified columns, E, and embedded = TRUE.

References

Sugihara G. 1994. Nonlinear forecasting for the classification of natural time series. Philosophical Transactions: Physical Sciences and Engineering, 348 (1688):477-495.

SurrogateData	<i>Surrogate time series for significance testing</i>
---------------	---

Description

Generate surrogate series (random_shuffle, ebisuzaki phase randomisation, or seasonal).

Usage

```
SurrogateData(ts, method = c("random_shuffle", "ebisuzaki", "seasonal"),
              num_surr = 100, T_period = 1, alpha = 0)
```

Arguments

- | | |
|----------|---|
| ts | Numeric time-series vector to generate surrogates for. |
| method | Surrogate method: "random_shuffle", "ebisuzaki", or "seasonal". |
| num_surr | Number of surrogate series to generate. |
| T_period | Seasonal period (seasonal method). |
| alpha | Gaussian-noise standard deviation (seasonal method). |

Value

A numeric matrix with `length(ts)` rows and `num_surr` columns.

Examples

```
SurrogateData(sin(1:100), method = "ebisuzaki", num_surr = 5)
```

TentMap	<i>Time series for a tent map with $\mu = 2$.</i>
---------	--

Description

First-differenced time series generated from the tent map recurrence relation with $\mu = 2$.

Usage

```
TentMap
```

Format

Data frame with 999 rows and 2 columns

Time time index.

TentMap tent map values.

TentMapNoise	<i>Time series of tent map plus noise.</i>
--------------	--

Description

First-differenced time series generated from the tent map recurrence relation with $\mu = 2$ and random noise.

Usage

```
TentMapNoise
```

Format

Data frame with 999 rows and 2 columns

Time time index.

TentMap tent map values.

Thrips	<i>Apple-blossom Thrips time series</i>
--------	---

Description

Seasonal outbreaks of Thrips imaginis.

References

Davidson and Andrewartha, Annual trends in a natural population of Thrips imaginis *Thysanoptera*, Journal of Animal Ecology, 17, 193-199, 1948.

Index

- * **datagen**
 - SurrogateData, [22](#)
- * **datasets**
 - block_3sp, [2](#)
 - circle, [5](#)
 - EvergladesFlow, [8](#)
 - Lorenz5D, [8](#)
 - paramecium_didinium, [11](#)
 - sardine_anchovy_sst, [17](#)
 - TentMap, [23](#)
 - TentMapNoise, [23](#)
- * **hplot**
 - PlotObsPred, [11](#)
 - PlotSmap, [12](#)
- * **nonlinear**
 - CCM, [3](#)
 - Embed, [6](#)
 - EmbedDimension, [7](#)
 - Multiview, [9](#)
 - PredictExclusionRadius, [12](#)
 - PredictInterval, [13](#)
 - PredictNonlinear, [15](#)
 - Simplex, [18](#)
 - SMap, [20](#)
- * **package**
 - rEDM, [16](#)
- * **regression**
 - ComputeError, [5](#)

block_3sp, [2](#)

CCM, [3](#), [3](#), [17](#)

circle, [5](#)

ComputeError, [5](#), [17](#), [19](#)

EDM (rEDM), [16](#)

Embed, [6](#), [17](#), [22](#)

EmbedDimension, [7](#), [7](#), [17](#), [19](#)

EvergladesFlow, [8](#)

Lorenz5D, [8](#)

Multiview, [9](#), [17](#)

paramecium_didinium, [11](#)

PlotObsPred, [11](#)

PlotSmap, [12](#)

PredictExclusionRadius, [12](#)

PredictInterval, [13](#), [13](#), [17](#)

PredictNonlinear, [15](#), [15](#), [17](#)

rEDM, [16](#)

rEDM-package (rEDM), [16](#)

sardine_anchovy_sst, [17](#)

Simplex, [7](#), [13](#), [17](#), [18](#), [18](#), [19](#), [21](#)

SMap, [15](#), [17](#), [20](#), [20](#), [21](#), [22](#)

SurrogateData, [22](#)

TentMap, [23](#)

TentMapNoise, [23](#)

Thrips, [24](#)