Documentation for package ‘COST’ version 0.1.0

- DESCRIPTION file.

**Help Pages**

<table>
<thead>
<tr>
<th>Function</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data.COST</td>
<td>Data Generation</td>
</tr>
<tr>
<td>example.forecast</td>
<td>example for one-step ahead forecast</td>
</tr>
<tr>
<td>example.prediction</td>
<td>example for new location prediction</td>
</tr>
<tr>
<td>Forecasts.CF</td>
<td>one-step ahead forecast by separate time series analysis</td>
</tr>
<tr>
<td>Forecasts.COST.G</td>
<td>one-step ahead forecast by Gaussian copula</td>
</tr>
<tr>
<td>Forecasts.COST.t</td>
<td>one-step ahead forecast by t copula</td>
</tr>
<tr>
<td>Forecasts.GP</td>
<td>one-step ahead forecast by Gaussian process fitting</td>
</tr>
<tr>
<td>logL.CF</td>
<td>negative log-likelihood for separate time series analysis</td>
</tr>
<tr>
<td>logL.COST.G</td>
<td>negative log-likelihood for Gaussian copula</td>
</tr>
<tr>
<td>logL.COST.t</td>
<td>negative log-likelihood for t copula</td>
</tr>
<tr>
<td>logL.GP</td>
<td>negative log-likelihood of Gaussian process</td>
</tr>
<tr>
<td>Predictions.COST.G</td>
<td>new location prediction by Gaussian copula</td>
</tr>
<tr>
<td>Predictions.COST.t</td>
<td>new location prediction by t copula</td>
</tr>
<tr>
<td>Predictions.GP</td>
<td>new location prediction by Gaussian process method</td>
</tr>
<tr>
<td>rank.multivariate</td>
<td>multivariate rank of a vector</td>
</tr>
<tr>
<td>Wind_6month</td>
<td>Wind speed data from 10 sites</td>
</tr>
</tbody>
</table>
Package: COST
Type: Package
Title: Copula-Based Semiparametric Models for Spatio-Temporal Data
Version: 0.1.0
Author: Yanlin Tang, Huixia Judy Wang
Maintainer: Yanlin Tang <tangyl@tongji.edu.cn>
Description: Parameter estimation, one-step ahead forecast and new location prediction methods for spatio-temporal data
Depends: copula,mvtnorm
License: GPL
Encoding: UTF-8
LazyData: true
RoxygenNote: 5.0.1
ByteCompile: yes
NeedsCompilation: no
Packaged: 2018-5-25 09:58:07 UTC+8; Shanghai
Built: R 3.2.1; ; 2018-08-10 16:24:34 UTC; unix
Data Generation

Description
Generating data from COST DGP, assuming Markov process of order one

Usage
Data.COST(n, n.total, seed1, coord, par.t)

Arguments
n number of time points for parameter estimation
n.total number of total time points, with a burning sequence
seed1 random seed to generate a data set, for reproducibility
coord coordinates of the locations
par.t the true copula parameters

Value
Y.all data from all locations and time points, may include data at time point n+1, or
data from new locations
mean.true true conditional mean of observed locations at time point n+1

Author(s)
Yanlin Tang, Huixia Judy Wang

References
Yanlin Tang, Huixia Judy Wang, Ying Sun, Amanda Hering. Copula-based Semiparametric
Models for Spatio-Temporal Data.

Examples
library(COST)
N = 2000
n.total = 5001
seed1 = 22222
coord = cbind(rep(c(1, 3, 5)/6, each=3), rep(c(1, 3, 5)/6, 3))
par.t = c(0, 1, 1, 0.5, 1.5, 100)
dat = Data.COST(n, n.total, seed1, coord, par.t)
#it returns a data set with dimension 2001*9
example for one-step ahead forecast

Description

description for one-step ahead forecast, where the data are generated from COST DGP, and parameter estimation and one-step ahead are performed for t copula, Gaussian copula, separate time series analysis, and Gaussian process method. Assuming that data are observed at d=9 locations, and n+1 time points, where the last time point is for validation.

Usage

element.forecast(n,n.total,seed1)

Arguments

n: number of time points for parameter estimation
n.total: number of total time points, with a burning sequence
seed1: random seed to generate a data set, for reproducibility

Value

pars.t.COST: parameter of the t copula
COST.t.fore.ECP: a vector of length d, with value 1 or 0, 1 means the verifying value from the corresponding location lies in the 95% forecast interval, 0 means not
COST.t.fore.ML: a vector of length d, each element is the length of forecast interval of the corresponding location
COST.t.fore.rank: multivariate rank of the verifying vector by t copula
pars.G.COST: parameter of the Gaussian copula
COST.G.fore.ECP: same as COST.t.fore.ECP
COST.G.fore.ML: same as COST.t.fore.ML
COST.G.fore.rank: multivariate rank of the verifying vector by Gaussian copula
pars.CF: parameter of the separate time series analysis
CF.fore.ECP: same as COST.t.fore.ECP
CF.fore.ML: same as COST.t.fore.ML
CF.fore.rank: multivariate rank of the verifying vector by separate time series analysis
pars.GP: parameter of the Gaussian process
GP.fore.ECP: same as COST.t.fore.ECP
GP.fore.ML: same as COST.t.fore.ML
GP.fore.rank: multivariate rank of the verifying vector by Gaussian process method

Author(s)

Yanlin Tang and Huixia Judy Wang

References


Examples

library(COST)
#settings
seed1 = 2222222
n.total = 501 #number of total time points, including the burning sequence
n = 200 #number of time points we observed
example.forecast(n,n.total,seed1)
#OUTPUTS
# #OUTPUTS
# $pars.t.COST #estimated parameter vector for t copula
# [1] 1.5707963 1.0326174 1.0727730 0.5106655 1.5637617 100.0000000
# $COST.t.fore.ECP #whether the forecast interval includes the true value at n+1
# [1] 1 1 1 0 1 1
# $COST.t.fore.ML #length of the forecast interval
# [1] 0.4396424 1.6924811 2.6800425 1.6136172 4.0542908 4.9896388 4.1697370 5.4200409 10.8829349
# $COST.t.fore.rank #multivariate rank
# [1] 54
# $pars.G.COST #estimated parameter vector for Gaussian copula
# [1] 0.01600678 0.98119456 1.07373350 0.50945930 1.56609592
#
# $COST.G.fore.ECP #whether the forecast interval includes the true value at n+1
# [1] 1 1 1 0 1 1 1 1 1 1
#
# $COST.G.fore.ML #length of the forecast interval
# [1] 0.4331987 1.6923304 2.6484572 1.6065395 4.0489496 4.9044345 4.1553041 5.3904896 10.8712031
#
# $COST.G.fore.rank #multivariate rank
# [1] 51
#
# $pars.CF #estimated parameter vector for CF
# [1] 100.0000000 0.7309756 0.6820088 0.6937469 0.6196984 0.6763829 0.7001407 0.5463127 0.5906863 0.6751806
# [10] 0.6751806
#
# $CF.fore.ECP #whether the forecast interval includes the true value at n+1
# [1] 1 1 1 1 1 1 1 1 1 1
#
# $CF.fore.ML #length of the forecast interval
#
# $CF.fore.rank #multivariate rank
# [1] 16
#
# $pars.GP #estimated parameter vector for Gaussian process
# [1] 0.7070787 1.0718393 1.2045607 0.5159298 1.6009586
#
# $GP.fore.ECP #whether the forecast interval includes the true value at n+1
# [1] 1 1 1 1 1 1 1 1 1 1
#
# $GP.fore.ML #length of the forecast interval
#
# $GP.fore.rank #multivariate rank
# [1] 171
example for new location prediction

Description
example for new location prediction, where the data are generated from COST DGP, and parameter estimation and new location prediction are performed for t copula, Gaussian copula, and Gaussian process method. Data are generated at 13 locations and n time points, and assume that 9 locations are observed, and 4 new locations need prediction at time n, conditional on 9 locations at time points n-1 and n.

Usage
example.prediction(n,n.total,seed1)

Arguments
n number of time points for parameter estimation
n.total number of total time points, with a burning sequence
seed1 random seed to generate a data set, for reproducibility

Value
pars.t.COST parameter of the t copula
COST.t.pre.ECP a vector of length K=4 (number of new locations), with value 1 or 0, 1 means the verifying value from the corresponding location lies in the 95% prediction interval, 0 means not
COST.t.pre.ML a vector of length K=4, each element is the length of prediction interval of the corresponding location
COST.t.pre.med.error prediction error based on conditional median
pars.G.COST parameter of the Gaussian copula
COST.G.pre.ECP same as COST.t.pre.ECP
COST.G.pre.ML same as COST.t.pre.ML
COST.G.pre.med.error same as COST.t.pre.med.error
pars.GP parameter of the Gaussian process
GP.pre.ECP same as COST.t.pre.ECP
GP.pre.ML same as COST.t.pre.ML
GP.pre.med.error same as COST.t.pre.med.error

Author(s)
Yanlin Tang and Huixia Judy Wang

References

Examples
library(COST)
# settings
n.total = 501 # number of total time points, including the burning sequence
n = 200 # number of time points we observed
seed1 = 22222
example.prediction(n,n.total,seed1)
# OUTPUTS
# $pars.t.COST estimated parameter vector for t copula
# [1] 1.4119487 1.1772323 1.0235709 0.5007877 1.4677992 62.9905729
# $COST.t.pre.ECP whether the prediction interval includes the true value at new location, time point n
# [1] 1 1 1 1
# $COST.t.pre.ML length of the prediction interval
# [1] 1.475864 2.814303 1.803352 2.879387
# $COST.t.pre.med.error point prediction error, using conditional median
# pars.G.COST #estimated parameter vector for Gaussian copula
# [1] 0.0000000 0.8588502 1.1054749 0.4979816 1.4695421

# COST.G.pre.ECP #whether the prediction interval includes the true value at new location, time point n
# [1] 1 1 1 1

# COST.G.pre.ML #length of the prediction interval
# [1] 1.536104 2.928396 1.875486 2.969367

# COST.G.pre.med.error #point prediction error, using conditional median
# [1] 0.1736040 0.3155241 0.2543634 0.2796725

# pars.GP #estimated parameter vector for Gaussian process
# [1] 0.0000000 1.0631387 1.0196981 0.4550216 1.5378387

# GP.pre.ECP #whether the prediction interval includes the true value at new location, time point n
# [1] 1 1 1 1

# GP.pre.ML #length of the prediction interval
# [1] 1.364275 2.017914 2.588556 3.351089

# GP.pre.med.error #point prediction error, using conditional median
# [1] 0.2341244 0.3873859 0.2585801 0.4093439
one-step ahead forecast by separate time series analysis

Description

One-step ahead forecast, analyzing the time series at each location separately with a t copula, including: (i) point forecast, either conditional median or mean; (ii) 95% forecast intervals, which can also be adjusted by the users; (iii) m (m=500 by default) random draws from the conditional distribution for each location, can be used for multivariate rank after combining all the locations together.

Usage

Forecasts.CF(par,Y,seed1,m)

Arguments

par parameters in the copula function
Y observed data
seed1 random seed used to generate random draws from the conditional distribution, for reproducibility
m number of random draws to approximate the conditional distribution

Value

y.qq 0.025-, 0.975- and 0.5-th conditional quantiles of the conditional distribution for each location
mean.est conditional mean estimate for each location
y.draw.random m random draws from the conditional distribution

Author(s)

Yanlin Tang and Huixia Judy Wang

References

one-step ahead forecast by Gaussian copula

Description

one-step ahead forecast by Gaussian copula, including: (i) point forecast, either conditional median or mean; (ii) 95% forecast intervals, which can also be adjusted by the users; (iii) m (m=500 by default) random draws from the conditional distribution, can be used for multivariate rank

Usage

Forecasts.COST.G(par,Y,s.ob,seed1,m,isotropic)

Arguments

par parameters in the copula function
Y observed data
s.ob coordinates of observed locations
seed1 random seed used to generate random draws from the conditional distribution, for reproducibility
m number of random draws to approximate the conditional distribution
isotropic indicator, True for isotropic correlation matrix, False for anisotropic correlation matrix, and we usually choose False for flexibility

Value

y.qq 0.025-, 0.975- and 0.5-th conditional quantiles of the conditional distribution for each location
mean.est conditional mean estimate for each location
y.draw.random m random draws from the conditional distribution

Author(s)

Yanlin Tang and Huixia Judy Wang

References

one-step ahead forecast by t copula

Description

one-step ahead forecast by t copula, including: (i) point forecast, either conditional median or mean; (ii) 95% forecast intervals, which can also be adjusted by the users; (iii) m (m=500 by default) random draws from the conditional distribution, can be used for multivariate rank

Usage

Forecasts.COST.t(par,Y,s.ob,seed1,m,isotropic)

Arguments

par parameters in the copula function
Y observed data
s.ob coordinates of observed locations
seed1 random seed used to generate random draws from the conditional distribution, for reproducibility
m number of random draws to approximate the conditional distribution
isotropic indicator, True for isotropic correlation matrix, False for anisotropic correlation matrix, and we usually choose False for flexibility

Value

y.qq 0.025-, 0.975- and 0.5-th conditional quantiles of the conditional distribution for each location
mean.est conditional mean estimate for each location
y.draw.random m random draws from the conditional distribution

Author(s)

Yanlin Tang and Huixia Judy Wang

References

one-step ahead forecast by Gaussian process fitting

Description

one-step ahead forecast by Gaussian process fitting, including: (i) point forecast, either conditional mean; (ii) 95% forecast intervals, which can also be adjusted by the users; (iii) m (m=500 by default) random draws from the conditional distribution, can be used for multivariate rank

Usage

Forecasts.GP(par,Y,s.ob,seed1,m,isotropic)

Arguments

- **par**: parameters in the copula function
- **Y**: observed data
- **s.ob**: coordinates of observed locations
- **seed1**: random seed used to generate random draws from the conditional distribution, for reproducibility
- **m**: number of random draws to approximate the conditional distribution
- **isotropic**: indicator, True for isotropic correlation matrix, False for anisotropic correlation matrix, and we usually choose False for flexibility

Value

- **y.qq**: 0.025-, 0.975- and 0.5-th conditional quantiles of the conditional distribution for each location
- **mean.est**: conditional mean estimate for each location
- **y.draw.random**: m random draws from the conditional distribution

Author(s)

Yanlin Tang and Huixia Judy Wang

References

negtive log-likelihood for separate time series analysis

Description

negtive log-likelihood for separate time series analysis, copula-based semiparametric method from Chen and Fan (2006), assuming t copula for each time series and Markov process of order one, with marginal distribution estimated by espirical CDF, and it is for correlation parameter estimation

Usage

logL.CF(par,Yk,dfs)

Arguments

par  correlation parameter in the t copula function, will be obtained by minimizing the negtive log-likelihood
Yk   observed data from k-th location
dfs  degrees of freedom for the t copula, obtained from COST method with t copula

Value

the negative log-likelihood

Author(s)

Yanlin Tang and Huixia Judy Wang

References

negative log-likelihood for Gaussian copula

Description
gives the negative log-likelihood of the Gaussian copula, with empirical CDF plugin, and it is for parameter estimation in the correlation matrix

Usage
`logL.COST.G(par, Y, s.ob)`

Arguments

par parameters in the copula function, will be obtained by minimizing the negative log-likelihood

Y the data set from observed locations, used for parameter estimation

s.ob coordinates of observed locations

Value
the negative log-likelihood

Author(s)
Yanlin Tang and Huixia Judy Wang

References

[Package COST version 0.1.0 Index]
negtive log-likelihood for t copula

Description
gives the negtive log-likelihood of the t copula, with empirical CDF plugin, and it is for parameter estimation in the correlation matrix

Usage

logL.COST.t(par,Y,s.ob)

Arguments

par parameters in the copula function, will be obtained by minimizing the negtive log-likelihood
Y the data set from observed locations, used for parameter estimation
s.ob coordinates of observed locations

Value
the negtive log-likelihood

Author(s)
Yanlin Tang and Huixia Judy Wang

References

[Package COST version 0.1.0 Index]
negative log-likelihood of Gaussian process

Description

negative log-likelihood of Gaussian process, with mean vector and variance vector obtained by the empirical version, and it is for parameter estimation in the correlation matrix

Usage

logL.GP(par, Y, s.ob)

Arguments

par parameters in the copula function, will be obtained by minimizing the negative log-likelihood
Y the data set from observed locations, used for parameter estimation
s.ob coordinates of observed locations

Value

the negative log-likelihood

Author(s)

Yanlin Tang and Huixia Judy Wang

References

new location prediction by Gaussian copula

Description

new location prediction by Gaussian copula, where the copula dimension is extended, and the marginal CDF of the new location is estimated by neighboring information; it gives 0.025-, 0.975- and 0.5-th conditional quantiles of the conditional distribution for each new location, at time n, conditional on observed locations at time n-1 and n; both point and interval predictions are provided

Usage

Predictions.COST.G(par,Y,s.ob,s.new,isotropic)

Arguments

par parameters in the copula function
Y observed data
s.ob coordinates of observed locations
s.new coordinates of new locations
isotropic indicator, True for isotropic correlation matrix, False for anisotropic correlation matrix, and we usually choose False for flexibility

Value

0.025-, 0.975- and 0.5-th conditional quantiles of the conditional distribution for each new location, at time n

Author(s)

Yanlin Tang and Huixia Judy Wang

References

new location prediction by t copula

Description

new location prediction by t copula, where the copula dimension is extended, and the marginal CDF of the new location is estimated by neighboring information; it gives 0.025-, 0.975- and 0.5-th conditional quantiles of the conditional distribution for each new location, at time n, conditional on observed locations at time n-1 and n; both point and interval predictions are provided

Usage

Predictions.COST.t(par,Y,s.ob,s.new,isotropic)

Arguments

par parameters in the copula function
Y observed data
s.ob coordinates of observed locations
s.new coordinates of new locations
isotropic indicator, True for isotropic correlation matrix, False for anisotropic correlation matrix, and we usually choose False for flexibility

Value

0.025-, 0.975- and 0.5-th conditional quantiles of the conditional distribution for each new location, at time n

Author(s)

Yanlin Tang and Huixia Judy Wang

References

new location prediction by Gaussian process method

Description

new location prediction by Gaussian process method, and the marginal mean and variance of the new location is estimated by neighboring information; it gives 0.025-, 0.975- and 0.5-th conditional quantiles of the conditional distribution for each new location, at time n, conditional on observed locations at time n-1 and n; both point and interval predictions are provided

Usage

Predictions.GP(par,Y,s.ob,s.new,isotropic)

Arguments

par parameters in the copula function
Y observed data
s.ob coordinates of observed locations
s.new coordinates of new locations
isotropic indicator, True for isotropic correlation matrix, False for anisotropic correlation matrix, and we usually choose False for flexibility

Value

0.025-, 0.975- and 0.5-th conditional quantiles of the conditional distribution for each new location, at time n

Author(s)

Yanlin Tang and Huixia Judy Wang

References

multivariate rank of a vector

Description

calculating the multivariate rank of a vector among a set of vectors, used to evaluate the performance of conditional distribution, and the rank would be uniform when the conditional distribution is estimated well

Usage

rank.multivariate(y.test, y.random, seed1)

Arguments

y.test the observed (verifying) vector at time n+1
y.random m random draws from the conditional distribution
seed1 random seed to solve tie at random

Value

the multivariate rank of the observed (verifying) vector at time n+1

Author(s)

Yanlin Tang and Huixia Judy Wang

References


[Package COST version 0.1.0 Index]
Wind speed data from 10 sites

Description
The data set is a subset of the data we used in the paper, with 10 sites and 6-month long time series.

Usage
data("Wind_6month")

Format
A data frame with 4320 observations on the following 2 variables.

Y.ob
4320*9 matrix from 9 observed sites

Y.newloc
4320-dim vector, from site Hood River

Coordinates of the locations.

s.ob
9*2 matrix for the coordinates of 9 observed sites

s.new
2-dim vector, coordinate of the site Hood River

Source

References

Examples
library(COST)
dim(Y.ob) #4320*9, data at 9 locations, with length 4320 (hours)
length(Y.newloc) #4320, time series at the new location
t = 1:dim(Y.ob)[1]
par(mfrow=c(3,3))
for (k0 in 1:9)
{
  plot(Y.ob[,k0]~t,type="l",ylab=expression(Y[ti]))
}

#In this sample code, we are doing one-step-ahead forecast, based on a rolling window method
d = nrow(s.ob)
m = 500 #number of ensembles for the multivariate rank
n = 2160 #the length of rolling window, i.e. we use data from t,t+1,...,t+2160-1, to forecast t+2160
seed1 = 22222
pj = 1
Y.1 = Y.ob[pj:(pj+2160-1),]
Y.d.test = Y.ob[pj+2160,]

hour = matrix(0,24,24)
diag(hour) = 1
hours = matrix(rep(hour,90),nrow=24)
hour.mean = hours

Y = matrix(0,2160,9)
for (kk in 1:9)
{
\[Y(:,kk) = Y.1(:,kk)-\text{rep}(\text{hour.mean}(:,kk),90)\]

\[\text{hour.1} = \text{as.vector}(\text{hour.mean}[1,])\]

# hourly mean for time points n-24+1, time point n+1 has the same mean as n-24+1

\[\text{pars.COST} <- \text{optim}(\text{par} = c(0.01,1,2,0.5,1.5,3), \text{logL.COST.t}, Y=Y, s.ob=s.ob, \text{method} = \text{"L-BFGS-B"}, \text{lower} = c(0,0.1,0.2,0.1,1.05,5), \text{upper} = c(\pi/2,10,10,0.9,2,50))\]$par

\[\text{COST.fore} <- \text{Forecasts.COST.t}(\text{pars.COST}, Y, s.ob, seed1, m, \text{isotropic} = \text{FALSE})\]

\[\text{med.est} = \text{COST.fore}$y.qq[,3]+\text{hour.1} \]  
# y.qq[,3] is the median point forecast

\[\text{y.low} = \text{COST.fore}$y.qq[,1]+\text{hour.1} \]

\[\text{y.up} = \text{COST.fore}$y.qq[,2]+\text{hour.1} \]

\[\text{y.draw.random} = \text{COST.fore}$y.draw.random+\text{hour.1} \]

\[\text{COST.t.fore.rank} = \text{rank.multivariate}(Y.d.test, y.draw.random, seed1)\]

\[\text{pars.COST.G} <- \text{optim}(\text{par} = c(0.3,0.5,3,0.4,1.2), \text{logL.COST.G}, Y=Y, s.ob=s.ob, \text{method} = \text{"L-BFGS-B"}, \text{lower} = c(0,0.1,2,0.1,1.05), \text{upper} = c(\pi/3,10,10,0.8,1.5))\]$par

\[\text{COST.G.fore} <- \text{Forecasts.COST.G}(\text{pars.COST.G}, Y, s.ob, seed1, m, \text{isotropic} = \text{FALSE})\]

\[\text{med.G.est} = \text{COST.G.fore}$y.qq[,3]+\text{hour.1} \]

\[\text{y.G.low} = \text{COST.G.fore}$y.qq[,1]+\text{hour.1} \]

\[\text{y.G.up} = \text{COST.G.fore}$y.qq[,2]+\text{hour.1} \]

\[\text{y.draw.random.G} = \text{COST.G.fore}$y.draw.random+\text{hour.1} \]

\[\text{COST.G.fore.rank} = \text{rank.multivariate}(Y.d.test, y.draw.random.G, seed1)\]

[Package COST version 0.1.0 Index]