Package 'SpTe2M'

Type Package

Title Nonparametric Modeling and Monitoring of Spatio-Temporal Data

Version 1.0.3

Date 2023-9-29

Description Spatio-temporal data have become increasingly popular in many research fields. Such data often have complex structures that are difficult to describe and estimate. This package provides reliable tools for modeling complicated spatio-temporal data. It also includes tools of online process monitoring to detect possible change-points in a spatio-temporal process over time. More specifically, the package implements the spatio-temporal mean estimation procedure described in Yang and Qiu (2018) <doi:10.1002/sim.7622>, the spatio-temporal covariance estimation procedure discussed in Yang and Qiu (2019) <doi:10.1002/sim.8315>, the three-step method for the joint estimation of spatio-temporal mean and covariance functions suggested by Yang and Qiu (2022) <doi:10.1007/s10463-021-00787-2>, the spatio-temporal disease surveillance method discussed in Qiu and Yang (2021) <doi:10.1002/sim.9150> that can accommodate the covariate effect, the spatial-LASSO-based process monitoring method proposed by Qiu and Yang (2023) <doi:10.1080/00224065.2022.2081104>, and the online spatio-temporal disease surveillance method described in Yang and Qiu (2020) <doi:10.1080/24725854.2019.1696496>.

License GPL (>= 3)

NeedsCompilation yes

RoxygenNote 7.2.0

Depends R (>= 3.5.0)

Imports glmnet, MASS, ggplot2, maps, mapproj, knitr, rmarkdown

VignetteBuilder knitr, rmarkdown

Author Kai Yang [aut, cre],
Peihua Qiu [ctb]

Maintainer Kai Yang <kayang@mcw.edu>

Repository CRAN

Date/Publication 2023-09-30 01:12:49 UTC

Contents

SpTe2M-package   *Nonparametric Modeling and Monitoring of Spatio-Temporal Data*

Description

Spatio-temporal data have become increasingly popular in many research fields. Such data often have complex structures that are difficult to describe and estimate. This package provides reliable tools for modeling complicated spatio-temporal data. It also includes tools of online process monitoring to detect possible change-points in a spatio-temporal process over time. More specifically, it implements the nonparametric spatio-temporal data modeling methods described in Yang and Qiu (2018, 2019, and 2022), as well as the online spatio-temporal process monitoring methods discussed in Qiu and Yang (2021 and 2023) and Yang and Qiu (2020).

Author(s)

Kai Yang <kayang@mcw.edu> and Peihua Qiu Maintainer: Kai Yang <kayang@mcw.edu>

References

Qiu, P. and Yang, K. (2021). Effective Disease Surveillance by Using Covariate Information. *Statistics in Medicine*, **40**, 5725-5745.

Qiu, P. and Yang, K. (2023). Spatio-Temporal Process Monitoring Using Exponentially Weighted Spatial LASSO. *Journal of Quality Technology*, **55**, 163-180.

Yang, K. and Qiu, P. (2018). Spatio-Temporal Incidence Rate Data Analysis by Nonparametric Regression. *Statistics in Medicine*, **37**, 2094-2107.

Yang, K. and Qiu, P. (2019). Nonparametric Estimation of the Spatio-Temporal Covariance Structure. *Statistics in Medicine*, **38**, 4555-4565.

Yang, K. and Qiu, P. (2020). Online Sequential Monitoring of Spatio-Temporal Disease Incidence Rates. *IISE Transactions*, **52**, 1218-1233.

Yang, K. and Qiu, P. (2022). A Three-Step Local Smoothing Approach for Estimating the Mean and Covariance Functions of Spatio-Temporal Data. *Annals of the Institute of Statistical Mathematics*, **74**, 49-68.

---

cv_mspe	*Cross-validation mean squared prediction error*

Description

The spatio-temporal covariance function is estimated by the weighted moment estimation method in Yang and Qiu (2019). The function `cv_mspe` is developed to select the bandwidths (`gt`,`gs`) used in the estimation of the spatio-temporal covariance function.

Usage

```
cv_mspe(y, st, gt = NULL, gs = NULL)
```

Arguments

y	A vector of length N containing data of the observed response $y(t,s)$, where N is the total number of observations over space and time.
st	An $N \times 3$ matrix specifying the spatial locations (i.e., (s_u, s_v)) and times (i.e., t) for all the observations in y. The three columns of st correspond to s_u, s_v and t, respectively.
gt	A sequence of temporal kernel bandwidth gt provided by users; default is NULL, and cv_mspe will choose its own sequence if gt=NULL.
gs	A sequence of spatial kernel bandwidth gs provided by users; default is NULL, and cv_mspe will choose its own sequence if gs=NULL.

Value

bandwidth	A matrix containing all the bandwidths (gt, gs) provided by users.
mspe	The mean squared prediction errors for all the bandwidths provided by users.
bandwidth.opt	The bandwidths (gt, gs) that minimizes the mean squared prediction error.
mspe.opt	The minimal mean squared prediction error.

Author(s)

Kai Yang <kayang@mcw.edu> and Peihua Qiu

References

Yang, K. and Qiu, P. (2019). Nonparametric Estimation of the Spatio-Temporal Covariance Structure. *Statistics in Medicine*, **38**, 4555-4565.

Examples

```
library(SpTe2M)
data(sim_dat)
y <- sim_dat$y; st <- sim_dat$st
gt <- seq(0.3,0.4,0.1); gs <- seq(0.3,0.4,0.1)
ids <- 1:500; y.sub <- y[ids]; st.sub <- st[ids,]
mspe <- cv_mspe(y.sub,st.sub,gt,gs)
```

---

ili_dat   *Florida influenza-like illness data*

Description

Daily influenza-like illness (ILI) incidence rates at 67 Florida counties during years 2012-2014. The ILI incidence rates were collected by the Electronic Surveillance System for the Early Notification of Community-based Epidemics (ESSENCE) that was developed by the Florida Department of Health. Researchers can have an access to the ESSENCE database after a proper online registration. Moreover, some weather conditions during years 2012-2014 can be obtained from the official website of the National Oceanic and Atmospheric Administration of the United States. The ILI dataset used here contains 8 variables, including County, Date, Lat, Long, Time, Rate (ILI incidence rate), Temp (temperature) and RH (relative humidity), from the two databases mentioned above, where Long and Lat refer to the longitude and latitude of the geometric centers of each Florida county, respectively.

Usage

```
data(ili_dat)
```

Format

A dataframe containing $N = 73,432$ observations of 8 variables.

Author(s)

Kai Yang <kayang@mcw.edu> and Peihua Qiu

---

mod_cv   *Modifed cross-validation for bandwidth selection*

Description

The spatio-temporal mean function can be estimated by the local linear kernel smoothing procedure (cf., Yang and Qiu 2018). The function mod_cv provides a reliable tool for selecting bandwidths (ht, hs) used in the local linear kernel smoothing procedure in cases when data are spatio-temporally correlated.

Usage

```
mod_cv(y, st, ht = NULL, hs = NULL, eps = 0.1)
```

Arguments

y	A vector of the spatio-temporal response $y(t,s)$.
st	A three-column matrix specifying the spatial locations and times for all the spatio-temporal observations in y.
ht	A sequence of temporal kernel bandwidth ht provided by users; default is NULL, and mod_cv chooses its own sequence if ht=NULL.
hs	A sequence of temporal kernel bandwidth hs provided by users; default is NULL, and mod_cv chooses its own sequence if hs=NULL.
eps	The value of this parametric is between 0 and 1. Default is 0.1. The following bimodal kernel function (cf., Yang and Qiu 2018) is used when calculting the modified cross-validation score:

$$K_\epsilon(x) = \frac{4}{4 - 3\epsilon - \epsilon^3} \begin{cases} \frac{3}{4}(1 - x^2)\mathrm{I}(|x| \le 1), & \text{if } |x| \ge \epsilon, \\ \frac{3(1-\epsilon^2)}{4\epsilon}|x|, & \text{otherwise.} \end{cases}$$

The argument eps represents the parameter ϵ in the above bimodal kernel, which controls the closeness of the bimodal kernel to the Epanechnikov kernel $K_e(x) = 0.75(1 - x^2)\mathrm{I}(|x| \le 1)$. The smaller the value, the closer the two kernels.

Value

bandwidth	A matrix containing all the bandwidths (ht, hs) provided by users.
mcv	The modified cross-validation scores for all the bandwidths provided by users.
bandwidth.opt	The selected bandwidths (ht, hs) by the modified cross-validation.
mcv.opt	The modified cross-validation score of the selected bandwidths.

Author(s)

Kai Yang <kayang@mcw.edu> and Peihua Qiu

References

Yang, K. and Qiu, P. (2018). Spatio-Temporal Incidence Rate Data Analysis by Nonparametric Regression. *Statistics in Medicine*, **37**, 2094-2107.

Examples

```
library(SpTe2M)
data(sim_dat)
y <- sim_dat$y; st <- sim_dat$st
ht <- seq(0.10,0.15,0.05); hs <- seq(0.20,0.30,0.10)
ids <- 1:500; y.sub <- y[ids]; st.sub <- st[ids,]
mcv <- mod_cv(y.sub,st.sub,ht,hs,eps=0.1)
```

---

pm25_dat *PM2.5 concentration data*

Description

Daily PM2.5 concentration levels at 183 major cities in China during years 2014-2016. This dataset was collected by the China National Environmental Monitoring Centre (CNEMC). It can be downloaded directly from the CNEMC offical web page. The PM2.5 dataset used here contains 6 variables, including `Year`, `Time`, `Long` (longitude), `Lat` (latitude), `City`, and `PM2.5`.

Usage

```
data(pm25_dat)
```

Format

A dataframe containing $N = 200, 385$ observations of 6 variables.

Author(s)

Kai Yang <kayang@mcw.edu> and Peihua Qiu

---

sim_dat *A simulated spatio-temporal dataset*

Description

This simulated dataset is saved as a list, and it contains the following three elements:

y A vector of length N; it contains the data of the observed response variable y.

x A vector of length N; it contains the data of the covariate x.

st An $N \times 3$ matrix containing the spatial locations and times for all the observations in the dataset.

Usage

```
data(sim_dat)
```

Format

A list containing $N = 10, 000$ observations.

Author(s)

Kai Yang <kayang@mcw.edu> and Peihua Qiu

Examples

```
library(MASS)
set.seed(100)
n <- 100; m <- 100; N <- n*m
t <- rep(seq(0.01,1,0.01),each=m)
su <- sv <- seq(0.1,1,0.1)
su <- rep(su,each=10); sv <- rep(sv,10)
su <- rep(su,n); sv <- rep(sv,n)
st <- matrix(0,N,3)
st[,1] <- su; st[,2] <- sv; st[,3] <- t
mu <- rep(0,N)
for(i in 1:N) {
  mu[i] <- 2+sin(pi*su[i])*sin(pi*sv[i])+sin(2*pi*t[i])
}
dist <- matrix(0,m,m) # distance matrix
for(i in 1:m) {
  for(j in 1:m) {
    dist[i,j] <- sqrt((su[i]-su[j])^2+(sv[i]-sv[j])^2)
  }
}
cov.s <- matrix(0,m,m) # spatial correlation
for(i in 1:m) {
  for(j in 1:m) {
    cov.s[i,j] <- 0.3^2*exp(-30*dist[i,j])
  }
}
noise <- matrix(0,n,m)
noise[1,] <- MASS::mvrnorm(1,mu=rep(0,m),Sigma=cov.s)
for(i in 2:n) {
  noise[i,] <- 0.1*noise[i-1,]+sqrt(1-0.1^2)*
    MASS::mvrnorm(1,mu=rep(0,m),Sigma=cov.s)
}
noise <- c(t(noise)); x <- rnorm(N,0,0.3)
beta <- 0.5; y <- mu+x*beta+noise
sim_dat <- list(); sim_dat$y <- y
sim_dat$x <- x; sim_dat$st <- st
```

---

sptemnt_cusum   *Online spatio-temporal process monitoring by a CUSUM chart*

Description

The function `sptemnt_cusum` implements the sequential online monitoring procedure described in Yang and Qiu (2020).

Usage

```
sptemnt_cusum(
  y,
```

```
  st,
  type,
  ARL0 = 200,
  gamma = 0.1,
  B = 1000,
  bs = 5,
  T = 1,
  ht = NULL,
  hs = NULL,
  gt = NULL,
  gs = NULL
)
```

Arguments

y	A vector of N spatio-temporal observations.
st	An $N \times 3$ matrix specifying the spatial locations and times for all the spatio-temporal observations in y.
type	A vector of N characters specifying the types of the observations. Here, type could be IC1, IC2 or Mnt, where type='IC1' denotes the in-control (IC) observations used to perform the block bootstrap procedure to determine the control limit of the CUSUM chart, type='IC2' denotes the IC observations used to estimate the spatio-temporal mean and covariance functions by spte_meanest and spte_covest, and type='Mnt' denotes the observations used for online process monitoring (cf., Yang and Qiu 2020). If there are only data points with either type='IC1' or type='IC2', then these data points will be used to estimate the model and conduct the bootstrap procedure as well. This function will return an error if there are no observations with type='IC1' or type='IC2'.
ARL0	The pre-specified IC average run length. Default is 200.
gamma	The pre-specified allowance constant in the CUSUM chart. Default is 0.1.
B	The bootstrap sizes used in the block bootstrap procedure for determining the control limit. Default value is 1,000.
bs	The block size of the block bootstrap procedure. Default value is 5.
T	The period of the spatio-temporal mean and covariance. Default value is 1.
ht	The temporal kernel bandwidth ht; default is NULL and it will be chosen by the modified cross-validation via mod_cv if ht=NULL.
hs	The spatial kernel bandwidth hs; default is NULL, and it will be chosen by the function mod_cv if hs=NULL.
gt	The temporal kernel bandwidth gt; default is NULL and it will be chosen by minimizing the mean squared prediction error via cv_mspe if gt=NULL.
gs	The spatial kernel bandwidth gs; default is NULL, and it will be chosen by the function cv_mspe if gs=NULL.

Value

ARL0	Same as the one in the arguments.

gamma	Same as the one in the arguments.
cstat	The charting statistics which can be used to make a plot for the control chart.
cl	The control limit that is determined by the block bootstrap.
signal_time	The signal time (i.e., the first time point when the charting statistic `cstat` exceeds the control limit `cl`).

Author(s)

Kai Yang <kayang@mcw.edu> and Peihua Qiu

References

Yang, K. and Qiu, P. (2020). Online Sequential Monitoring of Spatio-Temporal Disease Incidence Rates. *IISE Transactions*, **52**, 1218-1233.

Examples

```
library(SpTe2M)
data(ili_dat)
n <- 365; m <- 67
y <- ili_dat$Rate; st <- ili_dat[,3:5]
type <- rep(c('IC1','IC2','Mnt'),c(m*(n+1),(m*n),(m*n)))
ids <- c(1:(5*m),((n+1)*m+1):(m*(n+6)),((2*n+1)*m+1):(m*(2*n+6)))
y.sub <- y[ids]; st.sub <- st[ids,]; type.sub <- type[ids]
ili.cusum <- sptemnt_cusum(y.sub,st.sub,type.sub,ht=0.05,hs=6.5,gt=0.25,gs=1.5)
```

---

sptemnt_ewmac	*Spatio-temporal process monitoring using covariate information*

Description

The function `sptemnt_ewmac` is developed to solve the spatio-temporal process montoring problems in cases when the information in covariates needs to be used. Please refer to Qiu and Yang (2021) for more details of the method.

Usage

```
sptemnt_ewmac(
  y,
  x,
  st,
  type,
  ARL0 = 200,
  ARL0.z = 200,
  lambda = 0.1,
  B = 1000,
  bs = 5,
```

```
    T = 1,
    ht = NULL,
    hs = NULL,
    gt = NULL,
    gs = NULL
)
```

Arguments

y	A vector of N spatio-temporal observations.
x	An $N \times p$ matrix containing the data of p covariates.
st	An $N \times 3$ matrix specifying the spatial locations and times for all the spatio-temporal observations in y.
type	A vector of N characters specifying the types of the observations. Here, type could be IC1, IC2 or Mnt, where type='IC1' denotes the in-control (IC) observations used to perform the block bootstrap procedure to determine the control limit of the CUSUM chart, type='IC2' denotes the IC observations used to estimate the spatio-temporal mean and covariance functions by spte_meanest and spte_covest, and type='Mnt' denotes the observations used for online process monitoring (cf., Yang and Qiu 2021). If there are only data points with either type='IC1' or type='IC2', then these data points will be used to estimate the model and conduct the bootstrap procedure as well. This function will return an error if there are no observations with type='IC1' or type='IC2'.
ARL0	The pre-specified IC average run length. Default is 200.
ARL0.z	The pre-specified IC average run length for the covariate chart. Default is 200. Usually, set ARL0.z=ARL0.
lambda	The pre-specified weighting parameter in the EWMAC chart. Default is 0.1.
B	The bootstrap sizes used in the block bootstrap procedure for determining the control limit. Default value is 1,000.
bs	The block size of the block bootstrap procedure. Default value is 5.
T	The period of the spatio-temporal mean and covariance. Default value is 1.
ht	The temporal kernel bandwidth ht; default is NULL and it will be chosen by the modified cross-validation via mod_cv if ht=NULL.
hs	The spatial kernel bandwidth hs; default is NULL, and it will be chosen by the function mod_cv if hs=NULL.
gt	The temporal kernel bandwidth gt; default is NULL and it will be chosen by minimizing the mean squared prediction error via cv_mspe if gt=NULL.
gs	The spatial kernel bandwidth gs; default is NULL, and it will be chosen by the function cv_mspe if gs=NULL.

Value

ARL0	Same as the one in the arguments.
lambda	Same as the one in the arguments.

cstat	The charting statistics which can be used to make a plot for the control chart.
cl	The control limit that is determined by the block bootstrap.
signal_time	The signal time (i.e., the first time point when the charting statistic cstat exceeds the control limit cl).

Author(s)

Kai Yang <kayang@mcw.edu> and Peihua Qiu

References

Qiu, P. and Yang, K. (2021). Effective Disease Surveillance by Using Covariate Information. *Statistics in Medicine*, **40**, 5725-5745.

Examples

```
library(SpTe2M)
data(ili_dat)
n <- 365; m <- 67
y <- ili_dat$Rate; x <- as.matrix(ili_dat[,7:8]); st <- ili_dat[,3:5]
type <- rep(c('IC1','IC2','Mnt'),c(m*(n+1),(m*n),(m*n)))
ids <- c(1:(5*m),((n+1)*m+1):(m*(n+6)),((2*n+1)*m+1):(m*(2*n+6)))
y.sub <- y[ids]; x.sub <- x[ids,]; st.sub <- st[ids,]; type.sub <- type[ids]
ili.ewmac <- sptemnt_ewmac(y.sub,x.sub,st.sub,type.sub,ht=0.05,hs=6.5,gt=0.25,gs=1.5)
```

---

sptemnt_ewsl	*Spatio-temporal process monitoring using exponentially weighted spatial LASSO*

Description

Implementation of the online spatio-temporal process monitoring procedure described in Qiu and Yang (2023), in which spatial locations with the detected shifts are guaranteed to be small clustered spatal regions by the exponentially weighted spatial LASSO.

Usage

```
sptemnt_ewsl(
  y,
  st,
  type,
  ARL0 = 200,
  lambda = 0.1,
  B = 1000,
  bs = 5,
  T = 1,
  ht = NULL,
  hs = NULL,
```

```
    gt = NULL,
    gs = NULL
)
```

Arguments

y	A vector of N spatio-temporal observations.
st	An $N \times 3$ matrix specifying the spatial locations and times for all the spatio-temporal observations in y.
type	A vector of N characters specifying the types of the observations. Here, type could be IC1, IC2 or Mnt, where type='IC1' denotes the in-control (IC) observations used to perform the block bootstrap procedure to determine the control limit of the CUSUM chart, type='IC2' denotes the IC observations used to estimate the spatio-temporal mean and covariance functions by spte_meanest and spte_covest, and type='Mnt' denotes the observations used for online process monitoring (cf., Yang and Qiu 2021). If there are only data points with either type='IC1' or type='IC2', then these data points will be used to estimate the model and conduct the bootstrap procedure as well. This function will return an error if there are no observations with type='IC1' or type='IC2'.
ARL0	The pre-specified IC average run length. Default is 200.
lambda	The pre-specified weighting parameter in the EWMAC chart. Default is 0.1.
B	The bootstrap sizes used in the block bootstrap procedure for determining the control limit. Default value is 1,000.
bs	The block size of the block bootstrap procedure. Default value is 5.
T	The period of the spatio-temporal mean and covariance. Default value is 1.
ht	The temporal kernel bandwidth ht; default is NULL and it will be chosen by the modified cross-validation via mod_cv if ht=NULL.
hs	The spatial kernel bandwidth hs; default is NULL, and it will be chosen by the function mod_cv if hs=NULL.
gt	The temporal kernel bandwidth gt; default is NULL and it will be chosen by minimizing the mean squared prediction error via cv_mspe if gt=NULL.
gs	The spatial kernel bandwidth gs; default is NULL, and it will be chosen by the function cv_mspe if gs=NULL.

Value

ARL0	Same as the one in the arguments.
lambda	Same as the one in the arguments.
cstat	The charting statistics which can be used to make a plot for the control chart.
cl	The control limit that is determined by the block bootstrap.
signal_time	The signal time (i.e., the first time point when the charting statistic cstat exceeds the control limit cl).

Author(s)

Kai Yang <kayang@mcw.edu> and Peihua Qiu

References

Qiu, P. and Yang, K. (2023). Spatio-Temporal Process Monitoring Using Exponentially Weighted Spatial LASSO. *Journal of Quality Technology*, **55**, 163-180.

Examples

```
library(SpTe2M)
data(ili_dat)
n <- 365; m <- 67
y <- ili_dat$Rate; st <- ili_dat[,3:5]
type <- rep(c('IC1','IC2','Mnt'),c(m*(n+1),(m*n),(m*n)))
ids <- c(1:(5*m),((n+1)*m+1):(m*(n+6)),((2*n+1)*m+1):(m*(2*n+6)))
y.sub <- y[ids]; st.sub <- st[ids,]; type.sub <- type[ids]
ili.ewsl <- sptemnt_ewsl(y.sub,st.sub,type.sub,ht=0.05,hs=6.5,gt=0.25,gs=1.5)
```

---

spte_covest	*Estimate the spatio-temporal covariance function*

Description

The function `spte_covest` is developed to estimate the spatio-temporal covariance $V(t, t'; s, s') =$ $\mathrm{Cov}(y(t, s), y(t', s'))$ by the weighted moment estimation procedure (cf., Yang and Qiu 2019). It should be noted that the estimated covariance from `spte_covest` may not be positive semidefinite and thus it may not be a legitimate covariance function. In such cases, the projection-based modification needs to be used to make it positive semidefinite (cf., Yang and Qiu 2019).

Usage

```
spte_covest(y, st, gt = NULL, gs = NULL, stE1 = NULL, stE2 = NULL)
```

Arguments

y	A vector of length N containing data of the observed response.
st	An $N \times 3$ matrix specifying the spatial locations and times for all the spatio-temporal observations in y.
gt	The temporal kernel bandwidth gt; default is NULL, and it will be chosen by minimizing the mean squared prediction error via cv_mspe if gt=NULL.
gs	The spatial kernel bandwidth gs; default is NULL, and it will be chosen by the function cv_mspe if gs=NULL.
stE1	An $N_1 \times 3$ matrix specifying the spatial locations s and times t. Default value is NULL, and stE1=st if stE1=NULL.
stE2	An $N_2 \times 3$ matrix specifying the spatial locations s' and times t'. Default value is NULL, and stE2=st if stE2=NULL.

Value

stE1	Same as the one in the arguments.
stE2	Same as the one in the arguments.
bandwidth	The bandwidths (gt, gs) used in the weighted moment estimation procedure.
covhat	An $N_1 \times N_2$ covariance matrix estimate.

Author(s)

Kai Yang <kayang@mcw.edu> and Peihua Qiu

References

Yang, K. and Qiu, P. (2019). Nonparametric Estimation of the Spatio-Temporal Covariance Structure. *Statistics in Medicine*, **38**, 4555-4565.

Examples

```
library(SpTe2M)
data(sim_dat)
y <- sim_dat$y; st <- sim_dat$st
ids <- 1:500; y.sub <- y[ids]; st.sub <- st[ids,]
cov.est <- spte_covest(y.sub,st.sub)
```

---

spte_decor	*Decorrelate the spatio-temporal data*

Description

The function spte_decor uses the estimated spatio-temporal mean and covariance to decorrelate the observed spatio-temporal data. After data decorrelation, each decorrelated observation should have asymptotic mean of 0 and asymptotic variance of 1, and the decorrelated data should be asymptotically uncorrelated with each other.

Usage

```
spte_decor(y, st, y0, st0, T = 1, ht = NULL, hs = NULL, gt = NULL, gs = NULL)
```

Arguments

y	A vector of N spatio-temporal observations to decorrelate.
st	A three-column matrix specifying the spatial locations and observation times of the observations to decorrelate.
y0	A vector of N_0 in-control (IC) spatio-temporal observations from which the IC spatio-temporal mean and covariance functions can be estimated via spte_meanest and spte_covest.

st0	A three-column matrix specifying the spatial locations and times for all the spatio-temporal observations in y0.
T	The period of the spatio-temporal mean and covariance. Default value is 1.
ht	The temporal kernel bandwidth ht; default is NULL, and it will be chosen by the modified cross-validation mod_cv if ht=NULL.
hs	The spatial kernel bandwidth hs; default is NULL, and it will be chosen by the function mod_cv if hs=NULL.
gt	The temporal kernel bandwidth gt; default is NULL, and it will be chosen by minimizing the mean squared prediction error via cv_mspe if gt=NULL.
gs	The spatial kernel bandwidth gs; default is NULL, and it will be chosen by the function cv_mspe if gs=NULL.

Value

st	Same as the one in the arguments.
std.res	The decorrelated data.

Author(s)

Kai Yang <kayang@mcw.edu> and Peihua Qiu

References

Yang, K. and Qiu, P. (2020). Online Sequential Monitoring of Spatio-Temporal Disease Incidence Rates. *IISE Transactions*, **52**, 1218-1233.

Examples

```
library(SpTe2M)
data(sim_dat)
y <- sim_dat$y; st <- sim_dat$st
ids <- 1:500; y.sub <- y[ids]; st.sub <- st[ids,]
decor <- spte_decor(y.sub,st.sub,y0=y.sub,st0=st.sub)
```

---

spte_meanest	*Estimate the spatio-temporal mean function*

Description

The function spte_meanest provides a major tool for estimating the spatio-temporal mean function nonparametrically (cf., Yang and Qiu 2018 and 2022).

Usage

```
spte_meanest(y, st, ht = NULL, hs = NULL, cor = FALSE, stE = NULL)
```

Arguments

y	A vector of spatio-temporal observations.
st	A three-column matrix specifying the spatial locations and times for all the spatio-temporal observations in y.
ht	The temporal kernel bandwidth ht; default is NULL and it will be chosen by the modified cross-validation mod_cv if ht=NULL.
hs	The spatial kernel bandwidth hs; default is NULL, and it will be chosen by the function mod_cv if hs=NULL.
cor	A logical indicator where cor=FALSE implies that the covariance is not taken into account and the local linear kernel smoothing procedure is used for estimating the mean function (cf., Yang and Qiu 2018) and cor=TRUE implies that the covariance is accommodated and the three-step local smoothing approach is used to estimate the mean function (cf., Yang and Qiu 2022). Default is FALSE.
stE	A three-column matrix specifying the spatial locations and times where we want to calculate the estimate of the mean. Default is NULL, and stE=st if stE=NULL.

Value

bandwidth	The bandwidths (ht, hs) used in the estimation procedure.
stE	Same as the one in the arguments.
muhat	The estimated mean values at the spatial locations and times specified by stE.

Author(s)

Kai Yang <kayang@mcw.edu> and Peihua Qiu

References

Yang, K. and Qiu, P. (2018). Spatio-Temporal Incidence Rate Data Analysis by Nonparametric Regression. Statistics in Medicine, **37**, 2094-2107.

Yang, K. and Qiu, P. (2022). A Three-Step Local Smoothing Approach for Estimating the Mean and Covariance Functions of Spatio-Temporal Data. *Annals of the Institute of Statistical Mathematics*, **74**, 49-68.

Examples

```
library(SpTe2M)
data(sim_dat)
y <- sim_dat$y; st <- sim_dat$st
ids <- 1:500; y.sub <- y[ids]; st.sub <- st[ids,]
cov.est <- spte_meanest(y.sub,st.sub)
```

---

spte_semiparmreg *Fit the semiparametric spatio-temporal model*

Description

The function spte_semiparmreg fits the semiparametric spatio-temporal model to study the relationship between the response y and covariates x by the method discussed in Qiu and Yang (2021), in which an iterative algorithm is used to compute the estimated regression coefficients.

Usage

```
spte_semiparmreg(
  y,
  st,
  x,
  ht = NULL,
  hs = NULL,
  maxIter = 1000,
  tol = 10^(-4),
  stE = NULL
)
```

Arguments

y	A vector of length N containing the data of the spatio-temporal response $y(t, s)$.
st	An $N \times 3$ matrix specifying the spatial locations and times for all the spatio-temporal observations in y.
x	An $N \times p$ matrix containing the data of the p covariates.
ht	The temporal kernel bandwidth ht; default is NULL and it will be chosen by the modified cross-validation mod_cv if ht=NULL.
hs	The spatial kernel bandwidth hs; default is NULL, and it will be chosen by the function mod_cv if hs=NULL.
maxIter	A positive integer specifying the maximum number of iterations allowed. Default value is 1,000.
tol	A positive numeric value specifying the tolerance level for the convergence criterion. Default value is 0.0001.
stE	A three-column matrix specifying the spatial locations and times where we want to calculate the estimate of the mean. Default is NULL, and stE=st if stE=NULL.

Value

bandwidth	The bandwidths (ht, hs) used in the estimation procedure.
stE	Same as the one in the arguments.
muhat	The estimated mean values at spatial locations and times specified by stE.
beta	The vector of the estimated regression coefficient vector.

Author(s)

Kai Yang <kayang@mcw.edu> and Peihua Qiu

References

Qiu, P. and Yang, K. (2021). Effective Disease Surveillance by Using Covariate Information. *Statistics in Medicine*, **40**, 5725-5745.

Examples

```
library(SpTe2M)
data(sim_dat)
y <- sim_dat$y; st <- sim_dat$st; x <- sim_dat$x
ids <- 1:500; y.sub <- y[ids]; st.sub <- st[ids,]; x.sub <- x[ids]
semi.est <- spte_semiparmreg(y.sub,st.sub,x.sub,maxIter=2)
```

Index