Package 'ltsk'

December 5, 2019

Type Package

Title Local Time Space Kriging

Version 1.0.8

Date 2019-12-06

Author Naresh Kumar, Dong Liang, Jin Chen, Jun Chen

Maintainer Dong Liang <dliang@umces.edu>

Description Implements local spatial and local spatiotemporal Kriging based on local spatial and local spatiotemporal variograms, respectively. The method is documented in Kumar et al (2013) <https://www.nature.com/articles/jes201352>.

License GPL-2

Depends parallel,R (>= 2.10)

Imports fields,gstat,sp

NeedsCompilation yes

Repository CRAN

Date/Publication 2019-12-05 21:20:02 UTC

R topics documented:

ltsk-package	*Local Time Space Kriging*

Description

ltsk library is a collection of programs for implementing local spatial and local spatiotemporal Kriging. Unlike global Kriging, ltsk subsets the sample around a given location and time where predicted is needed; estimates variogram using the subset of sample data. Product-sum model is implemented and automatically estimated using the data points within the local neighbourhood. A unique advantage of ltsk is that it addresses non-stationarity, which is difficult to handle in large spatiotemporal dataset.

Details

Package:	ltsk
Type:	Package
Version:	1.0
Date:	2014-12-31
License:	GPL2.0

Author(s)

Naresh Kumar (NKumar@med.miami.edu) Dong Liang (dliang@umces.edu) Jun chen (wdidwlia@gmail.com) Jin Chen (jc.chenjin@gmail.com)

References

Haas, Timothy C. "Local prediction of a spatio-temporal process with an application to wet sulfate deposition." Journal of the American Statistical Association 90.432 (1995): 1189-1199.

Iaco, S. De & Myers, D. E. & Posa, D., 2001. "Space-time analysis using a general product-sum model," Statistics & Probability Letters, Elsevier, vol. 52(1), pages 21-28, March.

Kumar, N., et al. (2013). "Satellite-based PM concentrations and their application to COPD in Cleveland, OH." Journal of Exposure Science and Environmental Epidemiology 23(6): 637-646.

Liang, D. and N. Kumar (2013). "Time-space Kriging to address the spatiotemporal misalignment in the large datasets." Atmospheric Environment 72: 60-69.

---

cltsk	*Function calls* ltsk *using cumulatively expanding time space thresholds. This function is useful when predictions are needed using data points at different spatiotemporal intervals. For example, if predictions are needed at a given location for the past 30 days at an interval of 3 days. Instead of using* ltsk *10 times,* cltsk *can compute all 10 values simultaneously.*

Description

Function calls ltsk using cumulatively expanding time space thresholds.

Usage

```
cltsk(query, obs, th, nbins, xcoord = "x", ycoord = "y", tcoord = "t",
zcoord = "z", vth = NULL, vlen = NULL, llim = c(3, 3),
verbose = T, Large = 2000, future=T,cl = NULL)
```

Arguments

query	data frame containing query point (X,Y,T i.e. XY coordinates and time) where predictions are needed
obs	data frame containing sample data with XY coordinates, time and observed (measured) values
th	a priori chosen distance and time thresholds for neighbor search
nbins	a vector, number of distance and time bins for cumulative neighbor search and kriging.
xcoord	a character constant, the field name for x coordinate in both query and obs
ycoord	a character constant, the field name for y coordinate in both query and obs
tcoord	a character constant, the field name for time coordinate in both query and obs
zcoord	a character constant, the field name for data in obs
vth	thresholds for local spatiotemporal variogram (default 75% of the max lag difference)
vlen	numbers of bins for local spatiotemporal variogram(default, space 15, temporal for each day)
llim	lower limits for number of regions and intervals with observed data to calculate Kriging (default 3 spatial regions, 3 temporal intervals)
verbose	logical, whether print details information
Large	a numeric constant, upper limit of neighbor points, beyond which subsampling is performance
future	logical, whether including observed points in future relative to query points.
cl	a **parallel** cluster object (default number of cores in the local PC minue one), 0 means single core.

Details

Function performs automatic variogram estimation for each query location using the observed data within th thresholds. The estimated variogram is used for ordinary kriging, but using data in expanding local neighborhoods for ordinary kriging. For example, if predictions are needed at a given location for the past 30 days at an interval of 3 days,data within 3 days are used first, followed by 6 days and so on until data within 30 days. The same applies for distance thresholds.

Value

1. krig Kriging estimates at each space and time neighborhood
2. legend The legend for space and time neighborhood

Author(s)

Naresh Kumar (nkumar@med.miami.edu) Dong Liang (dliang@umces.edu)

References

Iaco, S. De & Myers, D. E. & Posa, D., 2001. "Space-time analysis using a general product-sum model," Statistics & Probability Letters, Elsevier, vol. 52(1), pages 21-28, March.

Kumar, N., et al. (2013). "Satellite-based PM concentrations and their application to COPD in Cleveland, OH." Journal of Exposure Science and Environmental Epidemiology 23(6): 637-646.

Liang, D. and N. Kumar (2013). "Time-space Kriging to address the spatiotemporal misalignment in the large datasets." Atmospheric Environment 72: 60-69.

Examples

```
## load the data
data(ex)
data(epa_cl)
## apply log transformation
obs[,'pr_pm25'] = log(obs[,'pr_pm25'])
## run kriging
system.time(out <- cltsk(ex2.query[1:2,],obs,c(0.10,10),
  zcoord='pr_pm25',nbins=c(4,5),verbose=FALSE,cl=0))
table(out$flag)
```

---

dnb   *Search Neighbours in Time and Space Within Specified Ranges*

Description

A brute force neighbor search implementation to identify observed data points within a given distance around location and time interval.

Usage

```
dnb(query, obs, th, future=TRUE)
```

Arguments

query	a vector; the x, y coordinates and the time stamp of the query point
obs	a matrix; the x, y coordinates and time stamps of the spatiotemporal locations
th	a vector; the distance threshold and time lag
future	logical, whether include observed spatiotemporal points future in time relative to the query spatiotemporal location.

Details

Implementation involves first calculating the time lags between query point and observed data (with locational coordinates and time); for observed locations within time lag of query, the function calculates the Euclidean distances between query location and all potential neighbors and select those within specified distance threshold.

The future argument can be used to exclude data in the future in neighbor search. This is useful in an extrapolation application.

Value

A vector, row numbers in the observed data matrix, that are within the given distance threshold and time lag of the query location.

Note

For large dataset, use ANN (for spatial kriging) and Range Tree for spatiotemporal Kriging.

Author(s)

Dong Liang (dliang@umces.edu)

See Also

get.knn in **FNN**

Examples

```
data(epa_cl)
coords <- c('x','y','t')
ii <- dnb(query[1,coords],obs[,coords],c(0.1,10))
```

---

lk   *Local Ordinary Kriging*

Description

Function implements ANN based neighbor search, automatic variogram estimation and ordinary Kriging at query locations.

Usage

```
lk(query, obs, th, xcoord = "x", ycoord = "y", zcoord = "z",  vlen = 15, cl=NULL)
```

Arguments

query	a data frame containing query spatial locations
obs	a data frame containing spatial locations and observed data
th	a vector, distance threshold for neighbor search
xcoord	a character constant, field name for x coordinate in both query and obs
ycoord	a character constant, field name for y coordinate in both query and obs
zcoord	a character constant, field name for data in obs
vlen	number of bins to compute local spatial variogram
cl	a cluster object for multi-core computing (default number of cores in the PC minue one), 0 means single core

Details

This function implements a local spatial Kriging method for spatial prediction using large data sets. At each location where prediction is required (termed as query point), a conservative distance threshold is used to select neighboring locations from the observed (or sample) data, which has XY coordinates.

A local variogram is used to estimate ordinary Kriging within this neighborhood.

Implementation of the local Kriging requires specifying a distance threshold (th). It is an upper bound within which spatial auto-correlation is believed to exist. The distance threshold may be obtained from variogram estimates or chosen based on prior knowledge of the underlying spatial process. For the given threshold, ANN tree is used to efficiently identify neighbors.

This function does not perform local kriging when the neighborhood contains less than *five* distinct spatial locations with observed data in the neighborhood. More data points are recommended for robust varioram estimation.

Four variogram models: Gaussian, exponential, spherical and Matern are automatically fit to the empirical variogram. The range parameter is estimated from the first distance lag where the empirical variogram exceeds 80

The estimated range parameter is used to reduce the neighborhood. Only initial neighbors that are spatially correlated with the process at the query point are used in prediction. In case the estimated range is small, indicating a locally weak spatial process, a sample of 10

Field names for geographic coordinates must match between query and observed data frames.

Value

kriging prediction and standard deviation and quality flags.

krig	predictive mean
sigma	predictive standard error
Hs	estimated range
nugget	estimated nugget
psill	estimated partial sill
model	variogram model

where the variogram model are

1	Spherical
2	Gaussian
3	Exponential
4	Matern

Author(s)

Jun chen (wdidwlia@gmail.com) Jin Chen (jc.chenjin@gmail.com) Naresh Kumar (NKumar@med.miami.edu) Dong Liang (dliang@umces.edu)

See Also

autoKrige in **automap**

Examples

```
data(epa_cl)
data(ex)
out <- lk(ex1.grid[1:10,],ex1.data,0.5,cl=0)
```

---

ltsk   *Ordinary Local Time and Space Kriging*

Description

Function implements ordinary time and space kriging for large data sets, with automatic product-sum variogram estimation.

Usage

```
ltsk(query, obs, th, xcoord = "x", ycoord = "y", tcoord = "t",
zcoord = "z", vth = NULL, vlen = NULL, llim = c(3, 3),
verbose = T, Large = 2000, future=T, cl = 0)
```

Arguments

query	a data frame containing query spatiotemporal locations for which predictions are needed
obs	a data frame containing spatiotemporal locations and observed data
th	a vector, distance threshold and time lag to define neighbors of a query point
xcoord	a character constant, the field name for x coordinate in both query and obs
ycoord	a character constant, the field name for y coordinate in both query and obs
tcoord	a character constant, the field name for time coordinate in both query and obs
zcoord	a character constant, the field name for data in obs
vth	thersholds for local spatiotemporal varigoram (default 75% of the max lag difference)
vlen	numbers of bins for local spatiotemporal varigram(default,space 15, temporal for each day)
llim	lower limits for number of regions and intervals with observed data to calculate Kriging (default 3 spatial regions, 3 temporal intervals)
verbose	logical, whether print details information
Large	a numeric constant, upper limit of neighbor points, beyond which subsampling is performance
future	logical, whether including observed points in future relative to query points.
cl	a **parallel** cluster object, NULL means total number of cores, (default 0 means single core)

Details

Function implements automatic variogram estimation (when possible) within a local spatiotemporal neighborhoods, and ordinary krigng based on the produce-sum variogram within that neighborhood. An variogram is estimated for each query point to allow for possible non-stationarity in the data generating field.

If the number of neighbors exceeds a user-specified upper limit (Large), neighbors are sub-sampled in a balanced way to reduce the neighborhood size.

Four variogram models: Gaussian, exponential, spherical and Matern are automatically fit to the empirical space and time variogram in the first lag. The range parameter is estimated from the first distance lag where the empirical variogram exceeds 80% of the maximum. Weighted least square is then used to estimate the nugget and partial sill parameters. Model with minimal residual sum of squares between the empirical and fitted variogram is chosen as the variogram model.

Value

Kriging mean and standard deviation and quality flags.

0	valid prediction
1	not enough temporal neighbors
2	not enough spatial neighbors
3	not enough neighbors
4	variogram could not be fit

Author(s)

Naresh Kumar (NKumar@med.miami.edu) Dong Liang (dliang@umces.edu)

References

Haas, Timothy C. "Local prediction of a spatio-temporal process with an application to wet sulfate deposition." Journal of the American Statistical Association 90.432 (1995): 1189-1199.

Iaco, S. De & Myers, D. E. & Posa, D., 2001. "Space-time analysis using a general product-sum model," Statistics & Probability Letters, Elsevier, vol. 52(1), pages 21-28, March.

Kumar, N., et al. (2013). "Satellite-based PM concentrations and their application to COPD in Cleveland, OH." Journal of Exposure Science and Environmental Epidemiology 23(6): 637-646.

Liang, D. and N. Kumar (2013). "Time-space Kriging to address the spatiotemporal misalignment in the large datasets." Atmospheric Environment 72: 60-69.

Examples

```
## load the data
data(ex)
data(epa_cl)
## apply log transformation
obs[,'pr_pm25'] = log(obs[,'pr_pm25'])
## run kriging
system.time(out <- ltsk(ex2.query[1:2,],obs,c(0.10,10),zcoord='pr_pm25',verbose=FALSE,cl=0))
table(out$flag)
```

---

ltsk-interval   *Internal functions to ltsk*

Description

These functions are working R functions that are called by the ltsk function. They should not be directly used.

---

ltsk.cv   *Local Time and Space Kriging Cross Validation, n-Fold or Leave-one-out*

Description

Cross validation functions for local time space kriging

Usage

```
ltsk.cv(nfold, obs, th, nbins, part=NULL,zcoord = "z",...)
```

Arguments

nfold	integer, apply n-fold cross validation; if larger than number of observed data, apply leave-one-out cross validation
obs	data frame containing spatiotemporal locations and observed data
th	vector of length two; a priori chosen distance threshold and time lag for neighbor search
nbins	vector of length two; a priori chosen bins to divide distance threshold and time lag equally
part	vector of random digits between 1 and nfold; if NULL, it was sampled with replacement from seq(1,nfold) of length nrow(obs)
zcoord	character constant, the field name for data in obs
...	other arguments that will be passed to cltsk

Details

Leave-one-out cross validation visits a data point, and predicts the value at that location by leaving out the observed value, and proceeds with the next data point. N-fold cross validation makes a partitions the data set in N parts. For all observations in a part, predictions are made based on the remaining N-1 parts; this is repeated for each of the N parts.

Value

a matrix of the cross validation residual, each column corresponds to a given distance threshold and time lag; a data frame containing the summary statistics of the cross validation residuals, including number of non-missing kriging, the sum of square prediction errors and the mean square prediction errors. Each individual row is a combination of distance threshold and time lag.

Author(s)

Naresh Kumar (NKumar@med.miami.edu)

Dong Liang (dliang@umces.edu)

References

Iaco, S. De & Myers, D. E. & Posa, D., 2001. "Space-time analysis using a general product-sum model," Statistics & Probability Letters, Elsevier, vol. 52(1), pages 21-28, March.

Kumar, N., et al. (2013). "Satellite-based PM concentrations and their application to COPD in Cleveland, OH." Journal of Exposure Science and Environmental Epidemiology 23(6): 637-646.

Liang, D. and N. Kumar (2013). "Time-space Kriging to address the spatiotemporal misalignment in the large datasets." Atmospheric Environment 72: 60-69.

Examples

```
## load the data
set.seed(123)
data(epa_cl)
```

```
ii= with(obs,which(amonth==5 & aday <13)) ## first week of Januray 2005;
x=obs[sample(ii,400),]
## apply log transformation
x[,'pr_pm25'] = log(x[,'pr_pm25'])
## run kriging
out <- ltsk.cv(nfold=10,obs=x,th=c(0.10,10),nbins=c(2,2),zcoord='pr_pm25',verbose=FALSE,cl=0)
```

---

obs   *example data sets for Cleveland OH*

Description

query and observed data for Cleveland OH

Usage

```
data(epa_cl)
```

---

tsbk   *Ordinary Global Time and Space Block Kriging*

Description

Function for block kriging in time and space based on the product-sum variogram model.

Usage

```
tsbk(query, obs, xcoord = "x", ycoord = "y", tcoord = "t", zcoord = "z",
bcoord='block', gcoord='g',vth = NULL, vlen = NULL,
    llim = c(3, 3), verbose = T, Large = 2000, future = T)
```

Arguments

query	a data frame containing query spatiotemporal locations
obs	a data frame containing spatiotemporal locations and observed data
xcoord	field name for x coordinate in both query and obs
ycoord	field name for y coordinate in both query and obs
tcoord	field name for time coordinate in both query and obs
zcoord	field name for data in obs
bcoord	field name for block in query
gcoord	field name identifying each unique query point
vth	thersholds for local spatiotemporal varigoram (default 75% max lag difference)

vlen	numbers of bins for local spatiotemporal varigram(default,space 15, temporal for each day)
llim	lower limits for number of data points to calculate Kriging (default 3 spatial, 3 temporal neighbors)
verbose	boolean whether print details information
Large	upper limit of neighbor points, beyond which subsampling was done
future	include observed points in future relative to query points.

Details

Function implements global time space block kriging based on a product sum model.

If the number of neighbors exceeds a user-specified upper limit (Large), neighbors are sub-sampled in a balanced way to reduce the neighborhood size.

Four variogram models: Gaussian, exponential, spherical and Matern are automatically fit to the empirical space and time variogram in the first lag. The range parameter is estimated from the first distance lag where the empirical variogram exceeds 80% of the maximum. Weighted least square is then used to estimate the nugget and partial sill parameters. Model with minimal residual sum of squares between the empirical and fitted variogram is chosen as the variogram model.

Field names for geographic coordinates and time stamps must match between query and observed data frames.

Value

a matrix containing the prediction and prediction standard error for each block, and a flag denoting the reason for un-successful prediction:

0	valid prediction
1	not enough temporal neighbors
2	not enough spatial neighbors
3	not enough neighbors
4	variogram could not be fit

Author(s)

Naresh Kumar (NKumar@med.miami.edu) Dong Liang (dliang@umces.edu)

References

Iaco, S. De & Myers, D. E. & Posa, D., 2001. "Space-time analysis using a general product-sum model," Statistics & Probability Letters, Elsevier, vol. 52(1), pages 21-28, March.

Kumar, N., et al. (2013). "Satellite-based PM concentrations and their application to COPD in Cleveland, OH." Journal of Exposure Science and Environmental Epidemiology 23(6): 637-646.

Liang, D. and N. Kumar (2013). "Time-space Kriging to address the spatiotemporal misalignment in the large datasets." Atmospheric Environment 72: 60-69.

See Also

krigeST in **gstat**

Examples

```
## load the data
data(ex)
data(epa_cl)
## apply log transformation
obs[,'pr_pm25'] = log(obs[,'pr_pm25'])
ex2.query$block <- 1 ## a single block
ex2.query$g <- 1:nrow(ex2.query)
## run kriging
## system.time(out <- tsbk(ex2.query[1:2,],obs,zcoord='pr_pm25',Large=400))
```

---

tsk	*Ordinary Global Time and Space Kriging*

Description

Function for ordinary kriging in time and space based on the product-sum variogram model, kriging in a local neighbourhood.

Usage

```
tsk(query, obs, subset = T, nmin = 3, nmax = 20, xcoord = "x",
    ycoord = "y", tcoord = "t", zcoord = "z", vth = NULL, vlen = NULL,
    llim = c(3, 3), verbose = T, Large = 2000, future = T)
```

Arguments

query	a data frame containing query spatiotemporal locations
obs	a data frame containing spatiotemporal locations and observed data
subset	logical; for local kriging; if TRUE only observations within the distances of estimated spatial and temporal sills from the prediction location are used for prediction
nmin	for local kriging: if the number of neighbors after subset is less than nmin, a missing value will be generated
nmax	for local kriging: the number of nearest observations that should be used for a kriging prediction, by default all observations are used.
xcoord	field name for x coordinate in both query and obs
ycoord	field name for y coordinate in both query and obs
tcoord	field name for time coordinate in both query and obs
zcoord	field name for data in obs
vth	thersholds for local spatiotemporal varigoram (default 75% max lag difference)

vlen	numbers of bins for local spatiotemporal varigram(default,space 15, temporal for each day)
llim	lower limits for number of data points to calculate Kriging (default 3 spatial, 3 temporal neighbors)
verbose	boolean whether print details information
Large	upper limit of neighbor points, beyond which subsampling was done
future	include observed points in future relative to query points.

Details

Function implements global time space kriging based on a product sum model and support kriging in a local neighborhood.

If the number of neighbors exceeds a user-specified upper limit (Large), neighbors are sub-sampled in a balanced way to reduce the neighborhood size.

Four variogram models: Gaussian, exponential, spherical and Matern are automatically fit to the empirical space and time variogram in the first lag. The range parameter is estimated from the first distance lag where the empirical variogram exceeds 80% of the maximum. Weighted least square is then used to estimate the nugget and partial sill parameters. Model with minimal residual sum of squares between the empirical and fitted variogram is chosen as the variogram model.

Field names for geographic coordinates and time stamps must match between query and observed data frames.

Value

a list of a matrix krig, containing the prediction and prediction standard error and a flag denoting the reason for un-successful prediction:

0	valid prediction
1	not enough temporal neighbors
2	not enough spatial neighbors
3	not enough neighbors
4	variogram could not be fit

a list of estimated time space variogram, and a list of fitted parameter values of the product sum variogram model.

Author(s)

Naresh Kumar (NKumar@med.miami.edu) Dong Liang (dliang@umces.edu)

References

Iaco, S. De & Myers, D. E. & Posa, D., 2001. "Space-time analysis using a general product-sum model," Statistics & Probability Letters, Elsevier, vol. 52(1), pages 21-28, March.

Kumar, N., et al. (2013). "Satellite-based PM concentrations and their application to COPD in Cleveland, OH." Journal of Exposure Science and Environmental Epidemiology 23(6): 637-646.

Liang, D. and N. Kumar (2013). "Time-space Kriging to address the spatiotemporal misalignment in the large datasets." Atmospheric Environment 72: 60-69.

See Also

`krigeST` in **gstat**

Examples

```
## load the data
data(ex)
data(epa_cl)
## apply log transformation
obs[,'pr_pm25'] = log(obs[,'pr_pm25'])
## run kriging
system.time(out <- tsk(ex2.query[1:2,],obs,zcoord='pr_pm25',Large=400))
out$krig
```

Index