Package ‘ProbForecastGOP’

February 19, 2015

Type Package

Title Probabilistic weather forecast using the GOP method

Version 1.3.2

Date 2010-05-31

Author Veronica J. Berrocal <veroberrocal@gmail.com>, Yulia Gel, Adrian E. Raftery, Tilmann Gneiting

Maintainer Veronica J. Berrocal <veroberrocal@gmail.com>

Depends R (>= 1.8.0), RandomFields, fields

Description The ProbForecastGOP package contains a main function, called ProbForecastGOP and other functions, to produce probabilistic weather forecasts of weather fields using the Geostatistical Output Perturbation (GOP) method of Gel, Raftery, and Gneiting (JASA, 2004).

License GPL (>= 2)

LazyLoad yes

Repository CRAN

Date/Publication 2012-10-29 08:57:28

NeedsCompilation no

R topics documented:

`ProbForecastGOP-package`

Probabilistic weather forecast using the GOP method

Description

The ProbForecastGOP package contains a main function, called ProbForecastGOP and other functions, to produce probabilistic weather forecasts of weather fields using the Geostatistical Output Perturbation (GOP) method of Gel, Raftery, and Gneiting (JASA, 2004).

Details

Package:	ProbForecastGOP
Type:	Package
Version:	1.3.2
Date:	2010-05-31
License:	GPL (>= 2)
LazyLoad:	yes

For an overview of how to use the package, including the most important functions, please refer to the PDF file describing the package.

Author(s)

Veronica J. Berrocal <veroberrocal@gmail.com>, Yulia Gel, Adrian E. Raftery, Tilmann Gneiting

Maintainer: Veronica J. Berrocal <veroberrocal@gmail.com>

References

Gel, Y., Raftery, A. E., Gneiting, T. (2004). Calibrated probabilistic mesoscale weather field forecasting: The Geostatistical Output Perturbation (GOP) method (with discussion). *Journal of the American Statistical Association*, **Vol. 99 (467)**, 575–583.

Gel, Y., Raftery, A. E., Gneiting, T., Berrocal, V. J. (2004). Rejoinder. *Journal of the American Statistical Association*, **Vol. 99 (467)**, 588–590.

Examples

```
library(fields)
library(RandomFields)
```

```
data(slp)
day <-slp$date.obs
id <- slp$id.stat
coord1 <- slp$lon.stat
coord2 <- slp$lat.stat
obs <- slp$obs
forecast <-slp$forecast

data(gridlong)
coord1.grid <- gridlong$gridded.lon

data(gridlat)
coord2.grid <- gridlat$gridded.lat

data(forecast.grid)
forecast.grid <- forecast.grid$gridded.forecast

## Specified cutpoints, default values for all the other fields.
## Only empirical variogram computation
empirical.variog <- ProbForecastGOP(day=day,obs=obs,forecast=forecast,
id=id,coord1=coord1,coord2=coord2,cut.points=seq(0,1000,by=2),
max.dist=NULL,nbins=NULL,variog.model="exponential",max.dist.fit=NULL,
init.val=NULL,fix.nugget=FALSE,coord1.grid=coord1.grid,
coord2.grid=coord2.grid,forecast.grid=forecast.grid,n.sim=10,
out="VARIOG",n.displ=4,qt.displ=c(5,50,95))


## Unspecified cutpoints.
## Fit of a parametric variogram to an empirical variogram.
fitted.variog <- ProbForecastGOP(day=day,obs=obs,forecast=forecast,id=id,
coord1=coord1,coord2=coord2,cut.points=NULL,max.dist=NULL,nbins=NULL,
variog.model="exponential",max.dist.fit=NULL,init.val=NULL,fix.nugget=FALSE,
coord1.grid=coord1.grid,coord2.grid=coord2.grid,forecast.grid=forecast.grid,
n.sim=10,out="FIT",n.displ=4,qt.displ=c(5,50,95))


## Unspecified cutpoints.
## Whole routine.
simulation.fields <-
ProbForecastGOP(day=day,obs=obs,forecast=forecast,id=id,coord1=coord1,
coord2=coord2,cut.points=NULL,max.dist=NULL,nbins=NULL,variog.model=NULL,
max.dist.fit=NULL,init.val=NULL,fix.nugget=FALSE,coord1.grid=coord1.grid,
coord2.grid=coord2.grid,forecast.grid=forecast.grid,n.sim=4,out="SIM",
n.displ=4,qt.displ=c(5,50,95))
```

`avg.variog` *Empirical variogram of a random variable averaged over time*

Description

Calculates the empirical variogram of a random variable averaged over time.

Usage

```
avg.variog(day, coord1, coord2, id, variable, cut.points=NULL, max.dist=NULL, nbins=300)
```

Arguments

`day`	numeric vector containing the day of observation.
`coord1`	vector containing the longitudes of the metereological stations.
`coord2`	vector containing the latitudes of the metereological stations.
`id`	vector with the id of the metereological stations.
`variable`	numeric vector containing the variable for which the empirical varigram is to be computed.
`cut.points`	numeric vector containing the cutpoints used for variogram binning.
`max.dist`	a numerical value giving the upper bound for the distance considered in the variogram computation.
`nbins`	a numerical value giving the number of bins for variogram binning. If both `cut.points` and `nbins` are entered, the entry for `nbins` will be ignored and the vector with the cutpoints will instead be used for variogram binning.

Details

The empirical variogram of the given random variable is calculated by determining, for each day, the distance among all pairs of stations that have been observed in the same day and by calculating for each day the sum of all the squared differences in the given random variable within each bin. These sums are then averaged over time, with weights for each bin given by the sum over time of the number of pairs of stations within the bin.

The formula used is:

$$\gamma(h) = \sum_d \frac{1}{2N_{(h,d)}} (\sum_i (Y(x_i + h, d) - Y(x_i, d))^2)$$

where $\gamma(h)$ is the empirical variogram at distance h, $N_{(h,d)}$ is the number of pairs of stations that have been recorded at day d and whose distance is equal to h, and $Y(x_i + h, d)$ and $Y(x_i, d)$ are, respectively, the values of the given variable observed on day d at stations located at $x_i + h$ and x_i. Variogram binning is ignored in this formula.

- Defaults -

If the vector with the cutpoints is not specified, the cutpoints are determined so that there are `nbins` bins with approximately the same number of pairs per bin.

If both the vector with the cutpoints and the number of bins, `nbins`, are unspecified, the function by default determines the cutpoints so that there are 300 bins with approximately the same number of pairs per bin. If both the vector with the cutpoints and the number of bins are provided, the entry for the number of bins is ignored and the vector with the cutpoints is used for variogram binning.

The default value for the maximum distance considered in the variogram computation is the 90-th percentile of the distances between the stations.

Value

The function returns a list with components given by:

`mar.var`	Marginal variance of the variable for which the empirical variogram is computed.
`bin.midpoints`	Numeric vector with midpoints of the bins used in the empirical variogram computation.
`number.pairs`	Numeric vector with the number of pairs per bin.
`empir.variog`	Numeric vector with the empirical variogram values.

Note

Depending on the data, the function might require substantial computing time.

Author(s)

Berrocal, V. J. <veroberrocal@gmail.com>, Gel, Y., Raftery, A. E., Gneiting, T.

References

Gel, Y., Raftery, A. E., Gneiting, T. (2004). Calibrated probabilistic mesoscale weather field forecasting: The Geostatistical Output Perturbation (GOP) method (with discussion). *Journal of the American Statistical Association*, **Vol. 99 (467)**, 575–583.

Cressie, N. A. C. (1993). *Statistics for Spatial Data* (revised ed.). Wiley: New York.

See Also

`avg.variog.dir` for directional empirical variogram of a random variable averaged over time, `Emp.variog` and `EmpDir.variog` for, respectively, empirical and directional empirical variogram of forecast errors averaged over time, and `Variog.fit` for estimation of parameters in a parametric variogram model.

Examples

```
## Loading data
data(slp)
day <- slp$date.obs
id <- slp$id.stat
coord1 <- slp$lon.stat
coord2 <- slp$lat.stat
obs <- slp$obs
forecast <- slp$forecast

## Computing variogram of observed temperature
## No specified cutpoints, no specified maximum distance
## Default number of bins
variogram <- avg.variog(day=day, coord1=coord1,coord2=coord2,id=id,variable=obs,cut.points=NULL,max.dist=NULL,n
## Plotting variogram
plot(variogram$bin.midpoints,variogram$empir.variog,xlab="Distance", ylab="Semi-variance",main="Empirical vario
```

avg.variog.dir	*Directional empirical variogram of a random variable averaged over time*

Description

Calculates directional empirical variogram of a random variable averaged over time.

Usage

```
avg.variog.dir(day, coord1, coord2, id, variable, tol.angle1=45,
tol.angle2=135, cut.points=NULL, max.dist=NULL, nbins=300, type)
```

Arguments

`day`	numeric vector containing the day of observation.
`coord1`	vector containing the longitudes of the metereological stations.
`coord2`	vector containing the latitudes of the metereological stations.
`id`	vector with the id of the metereological stations.
`variable`	numeric vector containing the variable for which the empirical variogram is to be computed.
`tol.angle1`	number giving a lower bound for the tolerance angle (measured in degrees).
`tol.angle2`	number giving an upper bound for the tolerance angle (measured in degrees).
`cut.points`	numeric vector containing the cutpoints used for variogram binning.
`max.dist`	a numerical value giving the upper bound for the distance considered in the variogram computation.
`nbins`	a numerical value giving the number of bins to use for variogram binning. If both `cut.points` and `nbins` are entered, the entry for `nbins` will be ignored and the vector with the cutpoints will instead be used for variogram binning.
`type`	character string indicating the direction to use for variogram computations. Possible values are either 'E' (for East-West) or 'N' (for North-South).

Details

The directional empirical variogram of the given random variable is calculated by determining, for each day, the "directional" distance among all pairs of stations that have been observed in the same day and by calculating for each day the sum of all the squared differences in the given random variable within each bin. These sums are then averaged over time, with weights for each bin given by the sum over time of the number of pairs of stations within the bin.

The formula used is:

$$\gamma(h) = \sum_d \frac{1}{2N_{(h,d)}} (\sum_i (Y(x_i + h, d) - Y(x_i, d))^2)$$

where $\gamma(h)$ is the empirical variogram at distance h, $N_{(h,d)}$ is the number of pairs of stations that have been recorded at day d and whose distance is equal to h, and $Y(x_i + h, d)$ and $Y(x_i, d)$ are, respectively, the values of the given variable observed on day d at stations located at $x_i + h$ and x_i. Variogram binning is ignored in this formula.

The "directional" distance between two locations is defined to be equal to the distance between the two locations if the angle between the two locations is within the allowed range, while it is set equal to infinity if the angle between the two locations is outside the allowed range.

- Defaults -

By default, `tol.angle1` and `tol.angle2` are set to 45 and 135 degrees, respectively. If the vector with the cutpoints is not specified, the cutpoints are determined so that there are `nbins` number of bins with approximately the same number of pairs per bin.

If both the vector with the cutpoints and the number of bins, `nbins`, are not provided, the function by default determines the cutpoints so that there are a total of 300 bins with approximately the same number of pairs per bin. If both the vector with the cutpoints and the number of bins are provided, the entry for the number of bins is ignored and the vector with the cutpoints is used for variogram binning.

The default value for the maximum distance considered in the variogram computation is the 90-th percentile of the distances between the stations.

Value

The function returns a list with components given by:

`bin.midpoints`	Numeric vector with midpoints of the bins used in the directional empirical variogram computation.
`number.pairs`	Numeric vector with the number of pairs per bin.
`dir.variog`	Numeric vector with the directional empirical variogram values.

Note

The function might require some time to return an output.

Author(s)

Berrocal, V. J. <`veroberrocal@gmail.com`>, Gel, Y., Raftery, A. E., Gneiting, T.

References

Gel, Y., Raftery, A. E., Gneiting, T. (2004). Calibrated probabilistic mesoscale weather field forecasting: The Geostatistical Output Perturbation (GOP) method (with discussion). *Journal of the American Statistical Association*, **Vol. 99 (467)**, 575–583.

Gel, Y., Raftery, A. E., Gneiting, T., Berrocal, V. J. (2004). Rejoinder. *Journal of the American Statistical Association*, **Vol. 99 (467)**, 588–590.

Cressie, N. A. C. (1993). *Statistics for Spatial Data* (revised ed.). Wiley: New York.

See Also

avg.variog for empirical variogram of a random variable averaged over time, Emp.variog and EmpDir.variog for, respectively, empirical and directional empirical variogram of forecast errors averaged over time, and Variog.fit for estimation of parameters in a parametric variogram model.

Examples

```
## Loading data
data(slp)
day <- slp$date.obs
id <- slp$id.stat
coord1 <- slp$lon.stat
coord2 <- slp$lat.stat
obs <- slp$obs
forecast <- slp$forecast

## Computing directional variogram of observed temperature
## No specified cutpoints, no specified maximum distance
## No specified tolerance angles and default number of bins
dir.variog <- avg.variog.dir(day,coord1,coord2,id,variable=obs, tol.angle1=NULL,tol.angle2=NULL,cut.points=NULL
## Plotting directional variogram
plot(dir.variog$bin.midpoints,dir.variog$dir.variog,xlab="Distance", ylab="Semi-variance",main="Empirical Direc
```

Emp.variog *Empirical variogram of forecast errors averaged over time*

Description

Calculates the empirical variogram of forecast errors, averaged over time.

Usage

```
Emp.variog(day, obs, forecast, id, coord1, coord2, cut.points=NULL, max.dist=NULL, nbins=300)
```

Arguments

day	numeric vector containing the day of observation.
obs	numeric vector containing the observed weather quantity.
forecast	numeric vector containing the forecasted weather quantity.
id	vector with the id of the metereological stations.
coord1	vector containing the longitudes of the metereological stations.
coord2	vector containing the latitudes of the metereological stations.
cut.points	numeric vector containing the cutpoints used for variogram binning.

`max.dist`	a numerical value giving the upper bound for the distance considered in the variogram computation.
`nbins`	a numerical value giving the number of bins for variogram binning. If both `cut.points` and `nbins` are entered, the entry for `nbins` will be ignored and the vector with the cutpoints will instead be used for variogram binning.

Details

The function includes bias-correction; it regresses the forecasts on the observed weather quantity and computes the residuals. The empirical variogram of the residuals is then calculated by determining, for each day, the distance among all pairs of stations that have been observed in the same day and by calculating for each day the sum of all the squared differences in the residuals within each bin. These sums are then averaged over time, with weights for each bin given by the sum over time of the number of pairs of stations within the bin.

The formula used is:

$$\gamma(h) = \sum_d \frac{1}{2N_{(h,d)}} (\sum_i (Y(x_i + h, d) - Y(x_i, d))^2)$$

where $\gamma(h)$ is the empirical variogram at distance h, $N_{(h,d)}$ is the number of pairs of stations that have been recorded at day d and whose distance is equal to h, and $Y(x_i + h, d)$ and $Y(x_i, d)$ are, respectively, the values of the residuals on day d at stations located at $x_i + h$ and x_i. Variogram binning is ignored in this formula.

- Defaults

If the vector with the cutpoints is not specified, the cutpoints are determined so that there are `nbins` bins with approximately the same number of pairs per bin.

If both the vector with the cutpoints and the number of bins, `nbins`, are unspecified, the function by default determines the cutpoints so that there are 300 bins with approximately the same number of pairs per bin. If both the vector with the cutpoints and the number of bins are provided, the entry for the number of bins is ignored and the vector with the cutpoints is used for variogram binning.

The default value for the maximum distance considered in the variogram computation is the 90-th percentile of the distances between the stations.

Value

The function returns a list with components given by:

`mar.var`	Marginal variance of the forecast errors.
`bin.midpoints`	Numeric vector with midpoints of the bins used in the empirical variogram computation.
`number.pairs`	Numeric vector with the number of pairs per bin.
`empir.variog`	Numeric vector with the empirical variogram values.

Note

Depending on the data, the function might require substantial computing time. As a consequence, if the interest is in producing probabilistic weather forecasts and generating ensemble members, it is advised to save the output in a file and then use the `Variog.fit` and `Field.sim` functions.

Author(s)

Berrocal, V. J. (<veroberrocal@gmail.com>), Raftery, A. E., Gneiting, T., Gel, Y.

References

Gel, Y., Raftery, A. E., Gneiting, T. (2004). Calibrated probabilistic mesoscale weather field forecasting: The Geostatistical Output Perturbation (GOP) method (with discussion). *Journal of the American Statistical Association*, **Vol. 99 (467)**, 575–583.

Cressie, N. A. C. (1993). *Statistics for Spatial Data* (revised ed.). Wiley: New York.

See Also

`EmpDir.variog` for directional empirical variogram of forecast errors averaged over time, `avg.variog` and `avg.variog.dir` for, respectively, empirical and directional empirical variogram of a random variable averaged over time, and `Variog.fit` for estimation of parameters in a parametric variogram model.

Examples

```
## Loading data
data(slp)
day <- slp$date.obs
id <- slp$id.stat
coord1 <- slp$lon.stat
coord2 <- slp$lat.stat
obs <- slp$obs
forecast <- slp$forecast

## Computing variogram
## No specified cutpoints, no specified maximum distance
## Default number of bins
variogram <- Emp.variog(day=day,obs=obs,forecast=forecast,id=id,
coord1=coord1,coord2=coord2,cut.points=NULL,max.dist=NULL,nbins=NULL)
## Plotting variogram
plot(variogram$bin.midpoints,variogram$empir.variog,xlab="Distance",
ylab="Semi-variance",main="Empirical variogram")

## Computing variogram
## Specified cutpoints, specified maximum distance
## Unspecified number of bins
variogram <-
Emp.variog(day=day,obs=obs,forecast=forecast,id=id,coord1=coord1,
coord2=coord2,cut.points=seq(0,1000,by=5),max.dist=800,nbins=NULL)
## Plotting variogram
plot(variogram$bin.midpoints,variogram$empir.variog,xlab="Distance",
ylab="Semi-variance",main="Empirical variogram")
```

EmpDir.variog *Directional empirical variogram of forecast errors averaged over time*

Description

Calculates directional empirical variogram of forecast errors, averaged over time.

Usage

```
EmpDir.variog(day, obs, forecast, id, coord1, coord2, tol.angle1=45, tol.angle2=135, cut.points=NULL,
```

Arguments

`day`	numeric vector containing the day of observation.
`obs`	numeric vector containing the observed weather quantity.
`forecast`	numeric vector containing the forecasted weather quantity.
`id`	vector with the id of the metereological stations.
`coord1`	vector containing the longitudes of the metereological stations.
`coord2`	vector containing the latitudes of the metereological stations.
`tol.angle1`	number giving a lower bound for the tolerance angle (measured in degrees).
`tol.angle2`	number giving an upper bound for the tolerance angle (measured in degrees).
`cut.points`	numeric vector containing the cutpoints used for variogram binning.
`max.dist`	a numerical value giving the upper bound for the distance considered in the variogram computation.
`nbins`	a numerical value giving the number of bins to use for variogram binning. If both `cut.points` and `nbins` are entered, the entry for `nbins` will be ignored and the vector with the cutpoints will instead be used for variogram binning.
`type`	character string indicating the direction to use for variogram computations. Possible values are either 'E' (for East-West) or 'N' (for North-South).

Details

The function includes bias-correction; it regresses the forecasts on the observed weather quantity and computes the residuals. The directional empirical variogram of the residuals is then calculated by determining, for each day, the "directional" distance among all pairs of stations that have been observed in the same day and by calculating for each day the sum of all the squared differences in the residuals within each bin. These sums are then averaged over time, with weights for each bin given by the sum over time of the number of pairs of stations within the bin.

The formula used is:

$$\gamma(h) = \sum_d \frac{1}{2N_{(h,d)}} (\sum_i (Y(x_i + h, d) - Y(x_i, d))^2)$$

where $\gamma(h)$ is the empirical variogram at distance h, $N_{(h,d)}$ is the number of pairs of stations that have been recorded at day d and whose distance is equal to h, and $Y(x_i + h, d)$ and $Y(x_i, d)$ are, respectively, the values of the residuals on day d at stations $x_i + h$ and x_i. Variogram binning is ignored in this formula.

The "directional" distance between two locations is defined to be equal to the distance between the two locations if the angle between the two locations is within the allowed range, while it is set equal to infinity if the angle between the two locations is outside the allowed range.

- Defaults -

By default, `tol.angle1` and `tol.angle2` are set to 45 and 135 degrees, respectively. If the vector with the cutpoints is not specified, the cutpoints are determined so that there are `nbins` number of bins with approximately the same number of pairs per bin.

If both the vector with the cutpoints and the number of bins, `nbins`, are not provided, the function by default determines the cutpoints so that there are a total of 300 bins with approximately the same number of pairs per bin. If both the vector with the cutpoints and the number of bins are provided, the entry for the number of bins is ignored and the vector with the cutpoints is used for variogram binning.

The default value for the maximum distance considered in the variogram computation is the 90-th percentile of the distances between the stations.

Value

The function returns a list with components given by:

- `bin.midpoints` Numeric vector with midpoints of the bins used in the directional empirical variogram computation.
- `number.pairs` Numeric vector with the number of pairs per bin.
- `dir.variog` Numeric vector with the directional empirical variogram values.

Note

The function might require some time to return an output.

Author(s)

Berrocal, V. J. (<veroberrocal@gmail.com>), Raftery, A. E., Gneiting, T., Gel, Y.

References

Gel, Y., Raftery, A. E., Gneiting, T. (2004). Calibrated probabilistic mesoscale weather field forecasting: The Geostatistical Output Perturbation (GOP) method (with discussion). *Journal of the American Statistical Association*, **Vol. 99 (467)**, 575–583.

Gel, Y., Raftery, A. E., Gneiting, T., Berrocal, V. J. (2004). Rejoinder. *Journal of the American Statistical Association*, **Vol. 99 (467)**, 588–590.

Cressie, N. A. C. (1993). *Statistics for Spatial Data* (revised ed.). Wiley: New York.

See Also

Emp.variog for empirical variogram averaged over time, avg.variog and avg.variog.dir for, respectively, empirical and directional empirical variogram of a random variable averaged over time, and Variog.fit for estimation of parameters in a parametric variogram model.

Examples

```
## Loading data
data(slp)
day <- slp$date.obs
id <- slp$id.stat
coord1 <- slp$lon.stat
coord2 <- slp$lat.stat
obs <- slp$obs
forecast <- slp$forecast

## Computing directional variogram
## No specified cutpoints, no specified maximum distance
## No specified tolerance angles and default number of bins
dir.variog <- EmpDir.variog(day,obs,forecast,id,coord1,coord2,
tol.angle1=NULL,tol.angle2=NULL,cut.points=NULL,max.dist=NULL,
nbins=NULL,type='E')
## Plotting directional variogram
plot(dir.variog$bin.midpoints,dir.variog$dir.variog,xlab="Distance",
ylab="Semi-variance",main="Empirical Directional variogram")


## Computing directional variogram
## Specified cutpoints, specified maximum distance
## Specified tolerance angles and unspecified number of bins
dir.variog <-
EmpDir.variog(day,obs,forecast,id,coord1,coord2,tol.angle1=30,
tol.angle2=150,cut.points=seq(0,1000,by=5),max.dist=800,nbins=NULL,
type='N')
## Plotting directional variogram
plot(dir.variog$bin.midpoints,dir.variog$dir.variog,xlab="Distance",ylab="Semi-variance",main="Empirical Direc
```

Field.sim *Simulation of weather random field*

Description

Simulates and displays realizations of forecast weather fields using a geostatistical model.

Usage

```
Field.sim(obs, forecast, coord1.grid, coord2.grid, forecast.grid,variog.model="exponential", param.e
qt.displ=c(10,50,90))
```

Arguments

`obs`	numeric vector containing the observed weather quantity.
`forecast`	numeric vector containing the forecasted weather quantity.
`coord1.grid`	numeric vector containing the longitudes of the grid points for the forecast.
`coord2.grid`	numeric vector containing the latitudes of the grid points for the forecast.
`forecast.grid`	numeric vector containing the forecast grid.
`variog.model`	character string with the name of the variogram model to be used for the simulations. Implemented models are `exponential`, `spherical`, `gauss`, `matern`, and `gencauchy`.
`param.est`	numeric vector containing values for the parameters to use in the parametric variogram model. If the variog.model specified is `exponential`, `spherical` or `gauss`, then the parameters required are, in order, the nugget effect (a non-negative number), the variance and the range (both positive numbers). If the variog.model specified is `gencauchy`, the parameters required are, in order, the nugget effect (a non-negative number), the variance, the range (both positive numbers), the smoothness parameter a (a number in $(0, 2]$), and the long-range parameter b (a positive number). If the variog.model specified is `matern` the parameters required are, in order, the nugget effect (a non-negative number), the variance, the range, and the smoothness parameter a (all positive numbers). For more details on the equation of the variogram models listed above, look below at the section "Details".
`n.sim`	number of realizations to be simulated.
`n.displ`	number of realizations to be displayed on screen.
`qt.displ`	numeric vector containing the quantiles to be displayed.

Details

The function initially estimates the additive and multiplicative bias by regressing the forecasts on the observed weather quantity. Then it simulates and displays realizations of the weather random field using the specified geostatistical model. Percentiles of the weather random field are calculated using the specified variogram model.

- Parametric variogram models implemented -

Below are the equations of the parametric variogram models that can be implemented.

- `exponential`

$$\gamma(d) = \rho + \sigma^2 \cdot (1 - exp(-\frac{d}{r}))$$

where d is the distance, ρ is the nugget effect, σ^2 is the variance, and r is the range.

- `spherical`

$$\gamma(d) = \rho + \sigma^2 \cdot (\frac{3}{2} \cdot \frac{d}{r} - \frac{1}{2} \cdot \frac{d^3}{r^3})$$

where d is the distance, ρ is the nugget effect, σ^2 is the variance, and r is the range.

- `gauss`

$$\gamma(d) = \rho + \sigma^2 \cdot (1 - exp(-\frac{d^2}{r^2}))$$

where d is the distance, ρ is the nugget effect, σ^2 is the variance, and r is the range.

- `gencauchy`

$$\gamma(d) = \rho + \sigma^2 \cdot (1 - (1 + \frac{d^a}{r^a})^{-\frac{b}{a}})$$

where d is the distance, ρ is the nugget effect, σ^2 is the variance, r is the range, a is the smoothness parameter, and b is the long-range parameter.

- `matern`

$$\gamma(d) = \rho + \sigma^2 \cdot (1 - \frac{2^{1-a}}{\Gamma(a)} \cdot \frac{d^a}{r^a} \cdot K_a(\frac{d}{r}))$$

where d is the distance, ρ is the nugget effect, σ^2 is the variance, r is the range, and a is the smoothness parameter.

- Defaults-

By default, 99 weather random fields are simulated, and 4 of them are displayed.

If no vector of percentiles is provided, the 10th, 50th and 90th percentiles are determined and displayed.

Value

The function returns both a graphical and a numerical output. The numerical output is a list with the following components:

`model`	Character string with the name of the variogram model used.
`nugget`	Value of the nugget effect.
`variance`	Value of the variance.
`range`	Value of the rang.
`additional.par`	Value(s) of the additional parameters required and used in the variogram model. This is returned only if the variogram model used is `matern` or `gencauchy`
`sim.fields`	3-dimensional array where each layer contains the values of the simulated weather field at the gridded locations.
`pct.fields`	3-dimensional array where each layer contains the specified percentile.

The graphical output consists in plots of the simulated weather fields displayed on multiple pages. Before displaying each page, the user will be asked for input.

Note

This function depends on the following packages: **RandomFields** and **fields**.

Author(s)

Berrocal, V. J. (<veroberrocal@gmail.com>), Raftery, A. E., Gneiting, T., Gel, Y.

References

Gel, Y., Raftery, A. E., Gneiting, T. (2004). Calibrated probabilistic mesoscale weather field forecasting: The Geostatistical Output Perturbation (GOP) method (with discussion). *Journal of the American Statistical Association*, **Vol. 99 (467)**, 575–583.

Gneiting, T., Schlather, M. (2004). Stochastic models that separate the fractal dimension and the Hurst effect. *SIAM Review* **46**, 269–282.

Schlather, M. (2001). Simulation and Analysis of Random Fields. *R News* **1(2)**, 18–20.

Nychka, D. (2004). *The fields package.* Available at: http:lib.stat.cmu.edu/R/CRAN/doc/package/fields.pdf.

Stein, M. L. (1999). *Interpolation of Spatial Data - Some Theory for Kriging*. Springer-Verlag: New York.

See Also

plotfields, a plotting utility for weather random fields, GaussRF in the **RandomFields** package, for simulation of Gaussian random fields.

Examples

```
## Loading data
library(fields)
library(RandomFields)
data(slp)
data(gridlong)
data(gridlat)
data(forecast.grid)
day <- slp$date.obs
id <- slp$id.stat
coord1 <- slp$lon.stat
coord2 <- slp$lat.stat
obs <- slp$obs
forecast <- slp$forecast
coord1.grid <- gridlong$gridded.lon
coord2.grid <- gridlat$gridded.lat
forecast.grid <- forecast.grid$gridded.forecast

## Computing the empirical variogram
variogram <- Emp.variog(day=day,obs=obs,forecast=forecast,id=id,
coord1=coord1,coord2=coord2,cut.points=NULL,max.dist=NULL,nbins=NULL)
## Estimating parameters
param.est <- Variog.fit(emp.variog=variogram,variog.model="exponential",
max.dist.fit=NULL,init.val=NULL,fix.nugget=FALSE)
```

```
## Simulating realizations of the weather random field
simul <-
Field.sim(obs=obs,forecast=forecast,coord1.grid=coord1.grid,
coord2.grid=coord2.grid,forecast.grid=forecast.grid,variog.model="exponential",
param.est=c(param.est$nugget,param.est$variance,param.est$range),n.sim=4,
n.displ=4,qt.displ=c(10,50,90))
```

`forecast.grid` *48-hour Temperature Forecast for the North American Pacific Northwest*

Description

This data set gives the 48-hour Temperature Forecast verifying January 12, 2002 at 0 hours GMT for the North American Pacific Northwest as provided by the Aviation (AVN) member of the University of Washington MM5 Mesoscale Ensemble, using a 12-km grid resolution.

Format

gridded forecast a numeric vector with 10098 forecasted temperature.

Source

http://metoc.apl.washington.edu/uwme/

References

Gel, Y., Raftery, A. E., Gneiting, T. (2004). Calibrated mesoscale weather field forecasting: The Geostatistical Output Perturbation (GOP) method (with discussion). *Journal of the American Statistical Association*, **Vol. 99 (467)**, 575–583.

Grimit, E. P., Mass, C. F. (2002). Initial results of a short-range ensemble forecasting system over the Pacific Northwest. *Weather and Forecasting* **17**, 192–205.

`gridlat` *Latitudes of localities in the North American Pacific Northwest*

Description

This data set gives the latitudes of the grid points for the 12-km domain of the University of Washington MM5 Mesoscale Ensemble.

Format

gridded.lat a numeric vector with 10098 latitudes.

Source

http://metoc.apl.washington.edu/uwme/

References

Gel, Y., Raftery, A. E., Gneiting, T. (2004). Calibrated mesoscale weather field forecasting: The Geostatistical Output Perturbation (GOP) method (with discussion). *Journal of the American Statistical Association*, **Vol. 99 (467)**, 575–583.

Grimit, E. P., Mass, C. F. (2002). Initial results of a mesoscale short-range ensemble forecasting system over the Pacific Northwest. *Weather and Forecasting* **17**, 192–205.

---

gridlong *Longitudes of localities in the North American Pacific Northwest*

Description

This data set gives the longitudes of the grid points for the 12-km domain of the University of Washington MM5 Mesoscale Ensemble.

Format

gridded.long a numeric vector with 10098 longitudes.

Source

http://metoc.apl.washington.edu/uwme/

References

Gel, Y., Raftery, A. E., Gneiting, T. (2004). Calibrated mesoscale weather field forecasting: The Geostatistical Output Perturbation (GOP) method, (with discussion). *Jorunal of the American Statistical Association*, **Vol. 99 (467)**, 575–583.

Grimit, E. P., Mass, C. F. (2002). Initial results of a short-range ensemble forecasting system over the Pacific Northwest. *Weather and Forecasting* **17**, 192–205.

`linesmodel` *Computation of parametric variogram model*

Description

Computes the value of the parametric variogram model at given distances.

Usage

```
linesmodel(distance, variog.model="exponential", param)
```

Arguments

`distance` numeric vector of distances.

`variog.model` character string giving the name of the parametric variogram model. Implemented models are: `exponential`, `spherical`, `gauss`, `gencauchy` and `matern`.

`param` numeric vector containing the values of the variogram parameters.

If the parametric model specified is `exponential`, `spherical` or `gauss`, `param` is a vector of length 3 containing, in order: the nugget effect (non negative number), the variance and the range (both positive numbers).

If the parametric model specified is `gencauchy`, `param` is a vector of length 5 whose entries are, in order: the nugget effect (non negative number), the variance, the range (both positive numbers), the smoothness parameter `a` (a number in $(0, 2]$), and the long-range parameter `b` (a positive number).

If the parametric model specified is `matern`, `param` is a vector of length 4 whose entries are, in order: the nugget effect (a non-negative number), the variance, the range, and the smoothness parameter `a` (all three, positive numbers).

Details

The function calculates the value of the parametric variogram at given distances using the following equations:

- If the parametric model is `exponential`

$$\gamma(d) = \rho + \sigma^2 \cdot (1 - exp(-\frac{d}{r}))$$

where ρ is the nugget effect, σ^2 is the variance, r is the range, and d is the distance.

- If the parametric model is `spherical`

$$\gamma(d) = \rho + \sigma^2 \cdot (\frac{3}{2} \cdot \frac{d}{r} - \frac{1}{2} \cdot \frac{d^3}{r^3})$$

where ρ is the nugget effect, σ^2 is the variance, r is the range, and d is the distance.

- If the parametric model is `gauss`

$$\gamma(d) = \rho + \sigma^2 \cdot (1 - exp(-\frac{d^2}{r^2}))$$

where ρ is the nugget effect, σ^2 is the variance, r is the range, and d is the distance.

- If the parametric model is gencauchy

$$\gamma(d) = \rho + \sigma^2 \cdot (1 - (1 + \frac{d^a}{r^a})^{-\frac{b}{a}})$$

where ρ is the nugget effect, σ^2 is the variance, r is the range, d is the distance, a is the smoothness parameter, and b is the long-range parameter.

- If the parametric model is matern

$$\gamma(d) = \rho + \sigma^2 \cdot (1 - (\frac{2^{1-a}}{\Gamma(a)} \cdot \frac{d^a}{r^a} \cdot K_a(\frac{d}{r}))$$

where ρ is the nugget effect, σ^2 is the variance, r is the range, d is the distance, and a is the smoothness parameter.

Value

The function returns a numeric vector with the values of the parametric variogram model at the bin midpoints.

Author(s)

Berrocal, V. J. (<veroberrocal@gmail.com>), Raftery, A. E., Gneiting, T., Gel, Y.

References

Gel, Y., Raftery, A. E., Gneiting, T. (2004). Calibrated probabilistic mesoscale weather field forecasting: The Geostatistical Output Perturbation (GOP) method (with discussion). *Journal of the American Statistical Association*, **Vol. 99 (467)**, 575–583.

Cressie, N. A. C. (1993). *Statistics for Spatial Data* (revised ed.). Wiley: New York.

Gneiting, T., Schlather, M. (2004). Stochastic models that separate fractal dimension and the Hurst effect. *SIAM Review* **46**, 269–282.

Stein, M. L. (1999). *Interpolation of Spatial Data - Some Theory for Kriging*. Springer-Verlag: New York.

Examples

```
## Loading data
data(slp)
day <- slp$date.obs
id <- slp$id.stat
coord1 <- slp$lon.stat
coord2 <- slp$lat.stat
```

```
obs <- slp$obs
forecast <- slp$forecast

## Computing empirical variogram
variogram <- Emp.variog(day=day,obs=obs,forecast=forecast,id=id,coord1=coord1,
coord2=coord2,cut.points=NULL,max.dist=NULL,nbins=NULL)

## Estimating variogram parameters
## Without specifying initial values for the parameters
param.variog <-
Variog.fit(emp.variog=variogram,variog.model="exponential",max.dist.fit=NULL,
init.val=NULL,fix.nugget=FALSE)

## Plotting the empirical variogram with the estimated parametric variogram superimposed
plot(variogram$bin.midpoints,variogram$empir.variog,xlab="Distance",ylab="Semi-variance")
lines(variogram$bin.midpoints,linesmodel(distance=variogram$bin.midpoints,variog.model="exponential",param=c(p
param.variog$variance,param.variog$range)))
```

---

phase1temp *Temperature in the Pacific Northwest during January-June 2000*

Description

This data set gives the observed and the 48-hour forecast temperature (in degree Celsius) at 0 hours GMT for 1092 stations located in the North American Pacific Northwest during the period January-June 2000, as provided by the Aviation (AVN) member of the University of Washington MM5 Mesoscale Ensemble.

Usage

```
data(phase1temp)
```

Format

A matrix with 56489 rows and 6 columns. The columns are:

date.obs a numeric vector giving the date of observations in the YYYYMMDDHH format.

name.stat a character string giving the identifier of the metereological stations.

lon.stat a numeric vector giving the longitudes of the metereological stations.

lat.stat a numeric vector giving the latitudes of the metereological stations.

forecast a numeric vector giving the 48-hr temperature forecasts as provided by the Aviation (AVN) member of the University of Washington MM5 ensemble.

obs a numeric vector giving the observed temperatures.

References

Gel, Y., Raftery, A. E., Gneiting, T. (2004). Calibrated mesoscale weather field forecasting: The Geostatistical Output Perturbation (GOP) method (with discussion). *Journal of the American Statistical Association*, **Vol. 99 (467),** 575–583.

Grimit, E. P., Mass, C. F. (2002). Initial results of a mesoscale short-range ensemble forecasting system over the Pacific Northwest. *Weather and Forecasting* **17**, 192–205.

`plotfields` *Plot of weather fields*

Description

Plot weather fields.

Usage

```
plotfields(field, x.lim, y.lim, country.outline="US", title)
```

Arguments

`field`	numeric square matrix containing the values of the weather field on a grid.
`x.lim`	numeric vector giving the smallest and the largest longitude to be displayed.
`y.lim`	numeric vector giving the smallest and the largest latitude to be displayed.
`country.outline`	character string indicating which country to outline in the plot. Possible values are "US", "world" or "both". If "US" is specified, a medium resolution outline of the US with the states and bodies of water is added to plot. If "world" is specified, a medium resolution of bodies of land and mass of water delimited by the smallest and largest latitude and longitude specified in `x.lim` and `y.lim` is added to the plot. If `country.outline` is set equal to `both`, medium resolution of both the US and of the bodies of land and water enclosed between the specified latitude and longitude are added to the plot. Default value is "US".
`title`	character string with the title for the plot.

Value

The function returns a graphical display of the weather field on a region delimited by the lower and upper bound for the longitude and the latitude.

Note

This function uses the package **fields**.

Author(s)

Gel, Y., Raftery, A. E., Gneiting, T., Berrocal, V. J. <veronica@stat.washington.edu>.

References

Gel, Y., Raftery, A. E., Gneiting, T. (2004). Calibrated probabilistic mesoscale weather field forecasting: The Geostatistical Output Perturbation (GOP) method (with discussion). *Journal of the American Statistical Association*, **Vol. 99 (467)**, 575–583.

Nychka, D. (2004). *The fields package*. Available at: http:lib.stat.cmu.edu/R/CRAN/doc/package/fields.pdf.

See Also

The package **fields** for display of spatial data, and `US` and `world` for a map of the US and the world.

Examples

```
## Loading data
library(fields)
library(RandomFields)
data(slp)
data(gridlong)
data(gridlat)
data(forecast.grid)
day <- slp$date.obs
id <- slp$id.stat
coord1 <- slp$lon.stat
coord2 <- slp$lat.stat
obs <- slp$obs
forecast <- slp$forecast
coord1.grid <- gridlong$gridded.long
coord2.grid <- gridlat$gridded.lat
forecast.grid <- forecast.grid$gridded.forecast

## Computing the empirical variogram
variogram <- Emp.variog(day,obs,forecast,id,coord1,coord2,cut.points=NULL,
max.dist=NULL,nbins=NULL)

## Estimating parameters
param.est <- Variog.fit(variogram,"exponential",max.dist.fit=NULL,
init.val=NULL,fix.nugget=FALSE)

## Simulating realizations of the weather random field
simul <- Field.sim(obs, forecast, coord1.grid, coord2.grid, forecast.grid,
variog.model="exponential", param.est=c(param.est$nugget,param.est$variance,
param.est$range), n.sim=4, n.displ=0, qt.displ=c(10,50,90))

##Plotting one of the simulated weather random fields
par(mfrow=c(1,1),mai=c(0.8,0.8,0.8,0.8))
plotfields(simul$sim.fields[,,1],x.lim=c(min(coord1.grid),max(coord1.grid)),
y.lim=c(min(coord2.grid),max(coord2.grid)),country.outline="US",title="Simulated weather field")

## Plotting one of the percentiles of the weather field
par(mfrow=c(1,1),mai=c(0.8,0.8,0.8,0.8))
plotfields(simul$pct.fields[,,1],x.lim=c(min(coord1.grid),max(coord1.grid)),
y.lim=c(min(coord2.grid),max(coord2.grid)),country.outline="US",title="10th percentile")
```

ProbForecastGOP *Probabilistic Weather Forecasts using the GOP method*

Description

This function generates probabilistic forecasts of weather fields using the Geostatistical Output Perturbation method as described by Gel, Raftery and Gneiting (2004).

Usage

```
ProbForecastGOP(day, obs, forecast, id, coord1, coord2, cut.points=NULL,
max.dist=NULL, nbins=300, variog.model="exponential", max.dist.fit=NULL,
init.val=NULL, fix.nugget=FALSE, coord1.grid, coord2.grid, forecast.grid,
n.sim=99, out="SIM", n.displ=4, qt.displ=c(10,50,90))
```

Arguments

`day`	numeric vector containing the day of observation.
`obs`	numeric vector containing the observed weather quantity.
`forecast`	numeric vector containing the forecasted weather quantity.
`id`	vector with the id of the metereological stations.
`coord1`	vector containing the longitudes of the metereological stations.
`coord2`	vector containing the latitudes of the metereological stations.
`cut.points`	numeric vector containing the cutpoints used for variogram binning.
`max.dist`	a numerical value giving the upper bound for the distance considered in the variogram computation.
`nbins`	a numerical value giving the number of bins for variogram binning.
`variog.model`	character string giving the name of the parametric model to fit to the empirical variogram. Implemented models are `exponential`, `spherical`, `gauss`, `gencauchy`, and `matern`.
`max.dist.fit`	number giving the maximum distance considered when fitting the variogram.
`init.val`	numeric vector giving the initial values for the parameters in the variogram model. The number of initial values to be entered depends on the parametric model specified. If the variog.model specified is `exponential`, `spherical` or `gauss`, then the parameters required are, in order, the nugget effect, the variance and the range. If the variog.model specified is `gencauchy`, the parameters required are, in order, the nugget effect, the variance, the range, the smoothness parameter `a` and the long-range parameter `b`. If the variog.model specified is `matern` the parameters required are, in order, the nugget effect, the variance, the range, and the smoothness parameter `a`. For more details on the valid range for the parameters and for the equation of the variogram models listed above, look below at the section "Details".

`fix.nugget` logical field indicating whether the nugget should be considered fixed or not.

If `TRUE` the nugget effect will be assumed to be constant, and a value for the fixed nugget effect can be also provided. If the value provided is different from the one entered in the `init.val` field, then the value of the nugget effect is taken to be the one entered in the `init.val` field.

If `FALSE` the nugget effect will be estimated along with the other parameters in the variogram model.

`coord1.grid` numeric vector containing the longitudes of the grid points for the forecast field.

`coord2.grid` numeric vector containing the latitudes of the grid points for the forecast field.

`forecast.grid` numeric vector containing the forecast grid.

`n.sim` number of realizations to be simulated.

`out` character string indicating the output produced by the function. Possible values are `VARIOG`, `FIT` and `SIM`.

If `VARIOG` is specified, the function computes the empirical variogram of the forecast errors averaged over time and displays a plot of the empirical variogram.

If `FIT` is specified, the function, in addition, fits a parametric model to the empirical variogram, returns estimates of the parameters and display both the empirical and the fitted variogram.

If `SIM` is specified, the function simulates also realizations of the weather random field using the parametric variogram model specified. The function displays the simulated weather fields and some specified percentiles.

`n.displ` number of realizations to be displayed on screen.

`qt.displ` numeric vector containing the quantiles to be displayed.

Details

This function generates probabilistic weather field forecasts by generating ensemble members using the Geostatistical Output Perturbation method presented by Gel, Raftery and Gneiting (2004).

The Geostatistical Output Perturbation method consists of the following steps:

1. determination of the residuals, by regressing the forecasts on the observed weather quantity and computation of their empirical variogram.

2. estimation of the parameters of the variogram model fitted to the empirical variogram.

3. simulation of weather random fields using the estimated parametric variogram model.

- *Computation of the empirical variogram of the residuals*

The empirical variogram of the residuals is calculated by determining, for each day, the distance among all pairs of stations that have been observed in the same day and by calculating for each day the sum of all the squared differences in the residuals within each bin. These sums are then averaged over time, with weights for each bin given by the sum over time of the number of pairs of stations within the bin.

The formula used is:

$$\gamma(h) = \sum_d \frac{1}{2N_{(h,d)}} (\sum_h (Y(i+h,d) - Y(i,d))^2)$$

- Estimation of the parameters of the parametric variogram model fitted to the empirical variogram.

The function estimates the parameters in the variogram model by minimizing the weighted least-square loss function

$$LOSS(\theta) = \sum_k n_k [\left(\frac{\hat{\gamma_k} - \gamma_k(\theta)}{\gamma_k(\theta)} \right)^2]$$

where n_k is the number of pairs contributing to the variogram computation in the kth bin, $\hat{\gamma_k}$ is the value of the empirical variogram in the kth bin and $\gamma_k(\theta)$ is the value of the estimated parametric variogram model at the midpoint of the kth bin.

- Parametric variogram models -

The parametric variogram models implemented are: exponential, spherical, gaussian, generalized Cauchy and Whittle-Matern.

- `exponential`

The equation of the exponential variogram with parameters the nugget effect, ρ, the variance, σ^2, and the range, r, is given by:

$$\gamma(d) = \rho + \sigma^2 \cdot (1 - exp(-\frac{d}{r}))$$

where d is the distance between the two locations and $\gamma(d)$ is the value of the exponential variogram at distance d. The nugget effect, ρ, is a non-negative number, while the variance, σ^2, and the range, r, are positive numbers.

- `spherical`

The equation of the spherical variogram with parameters the nugget effect, ρ, the variance, σ^2, and the range, r, is given by:

$$\gamma(d) = \rho + \sigma^2 \cdot (\frac{3}{2} \cdot \frac{d}{r} - \frac{1}{2} \cdot \frac{d^3}{r^3})$$

where d is the distance between the two locations and $\gamma(d)$ is the value of the spherical variogram at distance d. The nugget effect, ρ, is a non-negative number, while the variance, σ^2, and the range, r, are positive numbers.

- `gauss`

The equation of the gaussian variogram with parameters the nugget effect, ρ, the variance, σ^2, and the range, r, is given by:

$$\gamma(d) = \rho + \sigma^2 \cdot (1 - exp(-\frac{d^2}{r^2}))$$

where d is the distance between the two locations and $\gamma(d)$ is the value of the gaussian variogram at distance d. The nugget effect, ρ, is a non-negative number, while the variance, σ^2, and the range, r, are positive numbers.

- `gencauchy`

The equation of the generalized Cauchy variogram with parameters the nugget effect, ρ, the variance, σ^2, the range, r, the smoothness parameter, a, and the long-range parameter, b, is given by:

$$\gamma(d) = \rho + \sigma^2 \cdot (1 - (1 + \frac{d^a}{r^a})^{-\frac{b}{a}})$$

where d is the distance between the two locations and $\gamma(d)$ is the value of the generalized Cauchy variogram at distance d. The nugget effect, ρ is a non-negative number, the variance, σ^2 and the range, r, are positive numbers, the smoothness parameter, a, is a number in $(0, 2]$, and the long-range parameter, b, is a positive number.

- `matern`

The equation of the Whittle-Matern variogram with parameters the nugget effect, ρ, the variance, σ^2, the range, r, and the smoothness parameter, a, is given by:

$$\gamma(d) = \rho + \sigma^2 \cdot (1 - (\frac{2^{1-a}}{\Gamma(a)} \cdot \frac{d^a}{r^a} \cdot K_a(\frac{d}{r}))$$

where d is the distance between the two locations, $\gamma(d)$ is the value of the Whittle-Matern variogram at distance d, Γ is the gamma function and K_a is the Bessel function of the third kind with parameter a. The nugget effect, ρ, is a non-negative number, the variance, σ^2, the range, r, and the smoothness parameter, a, are positive numbers.

- Defaults-

If the vector with the cutpoints is not specified, the cutpoints are determined so that there are `nbins` bins with approximately the same number of pairs per bin. If both the vector with the cutpoints and the number of bins, `nbins`, are unspecified, the function by default determines the cutpoints so that there are 300 bins with approximately the same number of pairs per bin.

The default value for the maximum distance considered in the variogram computation is the 90-th percentile of the distances between all stations.

The default value for the maximum distance is $\frac{1}{2 \cdot \sqrt{2}}$ times the maximum distance considered in the empirical variogram.

Default for the initial values of the parameters is `NULL`, in which case the initial values for the parameters are determined internally and depend on the empirical variogram values. By default, the nugget effect is not considered constant. Thus, the default value for the `fix.nugget` field is `FALSE`.

The default for `out` is `SIM`: by default, 99 weather random fields are simulated, and 4 of them are displayed. If no vector of percentiles, is provided, the 10th, 50th and 90th percentiles are determined and displayed.

Value

The function returns both a numerical and a graphical summary. The graphical and numerical summary provided depend on the specification given in the field `out`.

The graphical summary consists of a plot of the empirical variogram, if `out` is equal to `VARIOG`.

It consists of a plot of the empirical variogram with superimposed the fitted parametric model, if `out` is equal to `FIT`.

If `out` is equal to `SIM`, the graphical summary consists of several plots: a plot of the empirical variogram with superimposed the fitted parametric model, plots of the simulated weather fields, and plots of specified percentiles of the weather random field. The plots are printed on multiple pages, and before displaying each page, the user will be asked for input.

The numerical summary include the following:

`bias.coeff`	additive and multiplicative bias obtained by linear regression of the forecasted weather quantity on the observed weather quantity. Always provided.
`se.bias.coeff`	standard errors of the bias coefficients. Always provided.
`res.var`	variance of the forecast errors. Always provided.
`bin.midpoints`	numeric vector with midpoints of the bins used in the empirical variogram computation. Always provided.
`number.pairs`	numeric vector with the number of pairs per bin. Always provided.
`empir.variog`	numeric vector with the empirical variogram values. Always provided.
`model`	name of the parametric model fitted to the empirical variogram. Provided only if `out` is equal to `FIT` or `SIM`.
`nugget`	Estimate of the nugget effect. Provided only if `out` is equal to `FIT` or `SIM`.
`variance`	Estimate of the variance. Provided only if `out` is equal to `FIT` or `SIM`.
`range`	Estimate of the range. Provided only if `out` is equal to `FIT` or `SIM`.
`additional.par`	Numeric vector with the estimates of the additional parameters required by the parametric variogram model (only if the model specified is either `whittlematern` or `gencauchy`). This is provided only if `out` is equal to `FIT` or `SIM`.
`sim.fields`	3-dimensional array where each layer contains the values of the simulated weather field at the gridded locations. Provided only if `out` is equal to `SIM`.
`pct.fields`	3-dimensional array where each layer contains the specified percentiles of the weather field. Provided only if `out` is equal to `SIM`.

Note

Computation of the empirical variogram might require substantial computing time. Hence, it is advisable to run the function once with `out` equal to `VARIOG`, save part of the output and use it as input for the `Variog.fit` function to estimate a parametric variogram model. Finally, `Field.sim` will simulate realizations of the weather field from the specified geostatistical model.

Author(s)

Berrocal, V. J. <veroberrocal@gmail.com>, Raftery, A. E., Gneiting, T., Gel, Y.

References

Gel, Y., Raftery, A. E., Gneiting, T. (2004). Calibrated probabilistic mesoscale weather field forecasting: The Geostatistical Output Perturbation (GOP) method (with discussion). *Journal of the American Statistical Association*, **Vol. 99 (467)**, 575–583.

Gel, Y., Raftery, A. E., Gneiting, T., Berrocal, V. J. (2004). Rejoinder. *Journal of the American Statistical Association*, **Vol. 99 (467)**, 588–590.

Cressie, N. A. C. (1993). *Statistics for Spatial Data* (revised ed.). Wiley: New York.

Gneiting, T., Schlather, M. (2004). Stochastic models that separate fractal dimension and the Hurst effect. *SIAM Review*, **46**, 269–282.

Stein, M. L. (1999). *Interpolation of Spatial Data - Some Theory for Kriging*. Springer-Verlag: New York.

Schlather, M. (2001). Simulation and analysis of Random Fields. *R News 1(2)*, 18–20.

Nychka, D. (2004). *The fields package*. Available at: http:lib.stat.cmu.edu/R/CRAN/doc/package/fields.pdf.

See Also

`Emp.variog` for computation of empirical variogram of forecast errors averaged over time, `Variog.fit` for estimation of parameters in a geostatistical model using Weighted Least Square, `Field.sim` for realizations of weather fields with the specified covariance structure.

Examples

```
## Loading data
library(fields)
library(RandomFields)
data(slp)
day <-slp$date.obs
id <- slp$id.stat
coord1 <- slp$lon.stat
coord2 <- slp$lat.stat
obs <- slp$obs
forecast <-slp$forecast

data(gridlong)
coord1.grid <- gridlong$gridded.lon

data(gridlat)
coord2.grid <- gridlat$gridded.lat

data(forecast.grid)
forecast.grid <- forecast.grid$gridded.forecast

## Specified cutpoints, default values for all the other fields.
## Only empirical variogram computation
empirical.variog <- ProbForecastGOP(day=day,obs=obs,forecast=forecast,
id=id,coord1=coord1,coord2=coord2,cut.points=seq(0,1000,by=2),
max.dist=NULL,nbins=NULL,variog.model="exponential",max.dist.fit=NULL,
init.val=NULL,fix.nugget=FALSE,coord1.grid=coord1.grid,
coord2.grid=coord2.grid,forecast.grid=forecast.grid,n.sim=10,
out="VARIOG",n.displ=4,qt.displ=c(5,50,95))


## Unspecified cutpoints.
## Fit of a parametric variogram to an empirical variogram.
fitted.variog <- ProbForecastGOP(day=day,obs=obs,forecast=forecast,id=id,
coord1=coord1,coord2=coord2,cut.points=NULL,max.dist=NULL,nbins=NULL,
variog.model="exponential",max.dist.fit=NULL,init.val=NULL,fix.nugget=FALSE,
coord1.grid=coord1.grid,coord2.grid=coord2.grid,forecast.grid=forecast.grid,
```

```
n.sim=10,out="FIT",n.displ=4,qt.displ=c(5,50,95))

## Unspecified cutpoints.
## Whole routine.
simulation.fields <-
ProbForecastGOP(day=day,obs=obs,forecast=forecast,id=id,coord1=coord1,
coord2=coord2,cut.points=NULL,max.dist=NULL,nbins=NULL,variog.model=NULL,
max.dist.fit=NULL,init.val=NULL,fix.nugget=FALSE,coord1.grid=coord1.grid,
coord2.grid=coord2.grid,forecast.grid=forecast.grid,n.sim=4,out="SIM",
n.displ=4,qt.displ=c(5,50,95))
```

---

slp *Sea Level Pressure in the Pacific Northwest for the period Jan. 12, 2000 - Jan. 14, 2000*

Description

This data set gives 48-hour forecast and simulated observed Sea Level Pressure (measured in millibars, mb) for 59 stations located in the Pacific Northwest. The 48-hour forecasts of Sea Level Pressure were provided by the Aviation (AVN) member of the University of Washington MM5 Mesoscale Ensemble. The observed Sea Level Pressure values were simulated for each day using a Multivariate Normal distribution with exponential covariance structure. The parameters used for the exponential covariance structure were: nugget effect, equal to $0.1\ mb^2$, σ^2 equal to $0.5\ mb^2$ and range equal to 50 km.

Usage

```
data(slp)
```

Format

A matrix with 150 rows and 6 columns. The columns are:

date.obs a numeric vector giving the date of observations in the YYYYMMDDHH format.

id.stat a character string giving the identifier of the metereological stations.

lon.stat a numeric vector giving the longitudes of the metereological stations.

lat.stat a numeric vector giving the latitudes of the metereological stations.

forecast a numeric vector giving the 48-hr Sea Level Pressure forecasts as provided by the Aviation (AVN) member of the University of Washington MM5 ensemble.

obs a numeric vector giving the simulated Sea Level Pressures.

References

Gel, Y., Raftery, A. E., Gneiting, T. (2004). Calibrated mesoscale weather field forecasting: The Geostatistical Output Perturbation (GOP) method (with discussion). *Journal of the American Statistical Association*, **Vol. 99 (467)**, 575–583.

Grimit, E. P., Mass, C. F. (2002). Initial results of a mesoscale short-range ensemble forecasting system over the Pacific Northwest. *Weather and Forecasting* **17**, 192–205.

`Variog.fit` *Fitting a parametric variogram model to an empirical variogram*

Description

This function estimates a parametric model fitted to an empirical variogram. The estimates are calculated using the weighted least squares method.

Usage

```
Variog.fit(emp.variog, variog.model="exponential", max.dist.fit=NULL,
init.val=NULL, fix.nugget=FALSE)
```

Arguments

`emp.variog` an object of the class `emp.variog`, output of the function `Emp.variog` or of the function `avg.variog`.

`variog.model` character string giving the name of the parametric model to fit to the empirical variogram. Implemented models are `exponential`, `spherical`, `gauss`, `gencauchy`, and `matern`.

`max.dist.fit` number giving the maximum distance considered when fitting the variogram.

`init.val` numeric vector giving the initial values for the parameters. The number of initial values to be entered depends on the variogram model specified.

If the variog.model specified is `exponential`, `spherical` or `gauss`, then the parameters required are, in order, the nugget effect, the variance and the range.

If the variog.model specified is `gencauchy`, the parameters required are, in order, the nugget effect, the variance, the range, the smoothness parameter `a`, and the long-range parameter `b`.

If the variog.model specified is `matern` the parameters required are, in order, the nugget effect, the variance, the range, and the smoothness parameter `a`.

For more details on the valid range for the parameters and for the equation of the variogram models listed above, look below at the section "Details".

`fix.nugget` logical field indicating whether the nugget should be considered fixed or not. If `TRUE` the nugget effect will be assumed to be constant, and a value for the fixed nugget effect can be also provided. If the value provided is different from the one entered in the `init.val` field, then the value of the nugget effect is taken to be the one entered in the `init.val` field.

If `FALSE` the nugget effect will be estimated along with the other parameters in the variogram model.

Details

The function estimates the parameters in the variogram model by minimizing the weighted least-square loss function

$$LOSS(\theta) = \sum_k n_k [\left(\frac{\hat{\gamma_k} - \gamma_k(\theta)}{\gamma_k(\theta)} \right)^2]$$

where n_k is the number of pairs contributing to the variogram computation in the kth bin, $\hat{\gamma_k}$ is the value of the empirical variogram in the kth bin and $\gamma_k(\theta)$ is the value of the estimated parametric variogram model at the midpoint of the kth bin.

- Parametric variogram models -

The parametric model implemented for the variogram are: exponential (`exponential`), spherical(`spherical`), gaussian (`gauss`), generalized Cauchy (`gencauchy`) and Whittle-Matern (`matern`).

- `exponential`:

The equation of the exponential variogram with parameters the nugget effect, ρ, the variance, σ^2, and the range, r, is given by:

$$\gamma(d) = \rho + \sigma^2 \cdot (1 - exp(-\frac{d}{r}))$$

where d is the distance between the two locations and $\gamma(d)$ is the value of the exponential variogram at distance d.

Notice that: the nugget effect, ρ, is a non-negative number, while the variance, σ^2, and the range, r, are positive numbers.

- `spherical`

The equation of the spherical variogram with parameters the nugget effect, ρ, the variance, σ^2, and the range, r, is given by:

$$\gamma(d) = \rho + \sigma^2 \cdot (\frac{3}{2} \cdot \frac{d}{r} - \frac{1}{2} \cdot \frac{d^3}{r^3})$$

where d is the distance between the two locations and $\gamma(d)$ is the value of the spherical variogram at distance d.

Notice that: the nugget effect, ρ, is a non-negative number, while the variance, σ^2, and the range, r, are positive numbers.

- `gauss`

The equation of the gaussian variogram with parameters the nugget effect, ρ, the variance, σ^2, and the range, r, is given by:

$$\gamma(d) = \rho + \sigma^2 \cdot (1 - exp(-\frac{d^2}{r^2}))$$

where d is the distance between the two locations and $\gamma(d)$ is the value of the gaussian variogram at distance d.

Notice that: the nugget effect, ρ, is a non-negative number, while the variance, σ^2, and the range, r, are positive numbers.

- `gencauchy`

The equation of the generalized Cauchy variogram with parameters the nugget effect, ρ, the variance, σ^2, the range, r, the smoothness parameter, a, and the long-range parameter, b, is given by:

$$\gamma(d) = \rho + \sigma^2 \cdot (1 - (1 + \frac{d^a}{r^a})^{-\frac{b}{a}})$$

where d is the distance between the two locations and $\gamma(d)$ is the value of the generalized Cauchy variogram at distance d.

Notice that: the nugget effect, ρ, is a non-negative number, the variance, σ^2, and the range, r, are positive numbers, the smoothness parameter, a, is a number in $(0, 2]$, and the long-range parameter, b, is a positive number.

- `matern`

The equation of the Whittle-Matern variogram with parameters the nugget effect, ρ, the variance, σ^2, the range, r, and the smoothness parameter, a, is given by:

$$\gamma(d) = \rho + \sigma^2 \cdot (1 - (\frac{2^{1-a}}{\Gamma(a)} \cdot \frac{d^a}{r^a} \cdot K_a(\frac{d}{r}))$$

where d is the distance between the two locations, $\gamma(d)$ is the value of the Whittle-Matern variogram at distance d, Γ is the gamma function and K_a is the Bessel function of the third kind with parameter a.

Notice that: the nugget effect, ρ, is a non-negative number, the variance, σ^2, the range, r, and the smoothness parameter, a, are positive numbers.

- Defaults -

The default value for the maximum distance is $\frac{1}{2 \cdot \sqrt{2}}$ times the maximum distance considered in the empirical variogram.

Default for the initial values of the parameters is `NULL`, in which case the initial values for the parameters are determined internally and depend on the empirical variogram values.

By default, the nugget effect is not considered constant. Thus, the default value for the `fix.nugget` field is `FALSE`.

Value

The function returns a list with components given by:

`model`	name of the parametric model fitted to the empirical variogram.
`nugget`	Estimate of the nugget effect.
`variance`	Estimate of the variance.
`range`	Estimate of the range.
`additional.par`	Numeric vector with the estimates of the additional parameters required by the parametric variogram model. This is returned only if the parametric model fitted are `gencauchy` or `matern`

Author(s)

Berrocal, V. J. (<veroberrocal@gmail.com>), Raftery, A. E., Gneiting, T., Gel, Y.

References

Gel, Y., Raftery, A. E., Gneiting, T. (2004). Calibrated probabilistic mesoscale weather field forecasting: the Geostatistical Output Perturbation (GOP) method (with discussion). *Journal of the American Statistical Association*, **Vol. 99 (467)**, 575–583.

Cressie, N. A. C. (1993) *Statistics for Spatial Data* (revised ed.). Wiley: New York.

Gneiting, T., Schlather, M. (2004). Stochastic models that separate fractal dimension and the Hurst effect. *SIAM Review* **46**, 269–282.

Stein, M. L. (1999). *Interpolation of Spatial Data - Some Theory for Kriging*. Springer-Verlag: New York.

See Also

`Emp.variog` for computation of the empirical variogram of forecast errors, and `avg.variog` for computation of the empirical variogram of a random variable.

Examples

```
## Loading data
library(fields)
library(RandomFields)
data(slp)
day <- slp$date.obs
id <- slp$id.stat
coord1 <- slp$lon.stat
coord2 <- slp$lat.stat
obs <- slp$obs
forecast <- slp$forecast

## Computing the empirical variogram
variogram <-
Emp.variog(day=day,obs=obs,forecast=forecast,id=id,coord1=coord1,coord2=coord2,
cut.points=NULL,max.dist=NULL,nbins=NULL)
## Estimating parameters
## Without specifying initial values for the parameters
param.variog <-
Variog.fit(emp.variog=variogram,variog.model="exponential",max.dist.fit=NULL,
init.val=NULL,fix.nugget=FALSE)
## Plotting the empirical variogram with the estimated parametric variogram superimposed
plot(variogram$bin.midpoints,variogram$empir.variog,xlab="Distance",ylab="Semi-variance")
lines(variogram$bin.midpoints,linesmodel(distance=variogram$bin.midpoints,
variog.model="exponential",param=c(param.variog$nugget,
param.variog$variance,param.variog$range)))


## Specifying the initial values for the parameters and keeping the nugget effect fixed
param.variog <-
```

```
Variog.fit(emp.variog=variogram,variog.model="exponential",max.dist.fit=NULL,
init.val=c(0,2,100),fix.nugget=TRUE)
## Plotting the empirical variogram with superimposed the estimated parametric variogram
plot(variogram$bin.midpoints,variogram$empir.variog,xlab="Distance",ylab="Semi-variance")
lines(variogram$bin.midpoints,linesmodel(distance=variogram$bin.midpoints,
variog.model="exponential",param=c(param.variog$nugget,
param.variog$variance,param.variog$range)))
```

Index