

# Package: geosimilarity (via r-universe)

November 1, 2024

**Title** Geographically Optimal Similarity

**Version** 3.8

**Description** Understanding spatial association is essential for spatial statistical inference, including factor exploration and spatial prediction. Geographically optimal similarity (GOS) model is an effective method for spatial prediction, as described in Yongze Song (2022) <[doi:10.1007/s11004-022-10036-8](https://doi.org/10.1007/s11004-022-10036-8)>. GOS was developed based on the geographical similarity principle, as described in Axing Zhu (2018) <[doi:10.1080/19475683.2018.1534890](https://doi.org/10.1080/19475683.2018.1534890)>. GOS has advantages in more accurate spatial prediction using fewer samples and critically reduced prediction uncertainty.

**License** GPL-3

**Encoding** UTF-8

**Roxygen** list(markdown = TRUE)

**RoxygenNote** 7.3.2

**URL** <https://github.com/ausgis/geosimilarity>,  
<https://ausgis.github.io/geosimilarity/>

**BugReports** <https://github.com/ausgis/geosimilarity/issues>

**Depends** R (>= 4.1.0)

**Imports** stats, parallel, tibble, dplyr (>= 1.1.0), purrr, ggplot2,  
magrittr, ggrepel

**Suggests** knitr, cowplot, viridis, car, DescTools,  
PerformanceAnalytics, testthat (>= 3.0.0), rmarkdown

**LazyData** true

**VignetteBuilder** knitr

**Config/testthat/edition** 3

**Repository** <https://ausgis.r-universe.dev>

**RemoteUrl** <https://github.com/ausgis/geosimilarity>

**RemoteRef** HEAD

**RemoteSha** fd1ca0ce573c1d163087a29f8d6b6771a4a4daea

## Contents

gos . . . . .	2
gos_bestkappa . . . . .	3
grid . . . . .	5
removeoutlier . . . . .	5
zn . . . . .	6

<b>Index</b>	<b>7</b>
--------------	----------

---

gos	<i>geographically optimal similarity</i>
-----	--

---

### Description

Computationally optimized function for geographically optimal similarity (GOS) model

### Usage

```
gos(formula, data = NULL, newdata = NULL, kappa = 0.25, cores = 1)
```

### Arguments

formula	A formula of GOS model.
data	A data.frame or tibble of observation data.
newdata	A data.frame or tibble of prediction variables data.
kappa	(optional) A numeric value of the percentage of observation locations with high similarity to a prediction location. $kappa = 1 - tau$ , where tau is the probability parameter in quantile operator. The default kappa is 0.25, meaning that 25% of observations with high similarity to a prediction location are used for modelling.
cores	(optional) Positive integer. If cores > 1, a parallel package cluster with that many cores is created and used. You can also supply a cluster object. Default is 1.

### Value

A tibble made up of predictions and uncertainties.

pred	GOS model prediction results
uncertainty90	uncertainty under 0.9 quantile
uncertainty95	uncertainty under 0.95 quantile
uncertainty99	uncertainty under 0.99 quantile
uncertainty99.5	uncertainty under 0.995 quantile
uncertainty99.9	uncertainty under 0.999 quantile
uncertainty100	uncertainty under 1 quantile

## References

Song, Y. (2022). Geographically Optimal Similarity. *Mathematical Geosciences*. doi: 10.1007/s11004-022-10036-8.

## Examples

```
data("zn")
# log-transformation
hist(zn$Zn)
zn$Zn <- log(zn$Zn)
hist(zn$Zn)
# remove outliers
k <- removeoutlier(zn$Zn, coef = 2.5)
dt <- zn[-k,]
# split data for validation: 70% training; 30% testing
split <- sample(1:nrow(dt), round(nrow(dt)*0.7))
train <- dt[split,]
test <- dt[-split,]
system.time({
g1 <- gos(Zn ~ Slope + Water + NDVI + SOC + pH + Road + Mine,
         data = train, newdata = test, kappa = 0.25, cores = 1)
})
test$pred <- g1$pred
plot(test$Zn, test$pred)
cor(test$Zn, test$pred)
```

---

gos\_bestkappa

*function for the best kappa parameter*

---

## Description

Computationally optimized function for determining the best kappa parameter for the optimal similarity

## Usage

```
gos_bestkappa(
  formula,
  data = NULL,
  kappa = seq(0.05, 1, 0.05),
  nrepeat = 10,
  nsplit = 0.5,
  cores = 1
)
```

**Arguments**

formula	A formula of GOS model.
data	A <code>data.frame</code> or <code>tibble</code> of observation data.
kappa	(optional) A numeric value of the percentage of observation locations with high similarity to a prediction location. $kappa = 1 - tau$ , where $tau$ is the probability parameter in quantile operator. $kappa$ is 0.25 means that 25% of observations with high similarity to a prediction location are used for modelling.
nrepeat	(optional) A numeric value of the number of cross-validation training times. The default value is 10.
nsplit	(optional) The sample training set segmentation ratio, which in $(0, 1)$ . Default is 0.5.
cores	(optional) Positive integer. If $cores > 1$ , a <code>parallel</code> package cluster with that many cores is created and used. You can also supply a cluster object. Default is 1.

**Value**

A list.

`bestkappa` the result of best kappa

`cvrms` all RMSE calculations during cross-validation

`cvmean` the average RMSE corresponding to different kappa in the cross-validation process

`plot` the plot of rmse changes corresponding to different kappa

**References**

Song, Y. (2022). Geographically Optimal Similarity. *Mathematical Geosciences*. doi: 10.1007/s11004-022-10036-8.

**Examples**

```
data("zn")
# log-transformation
hist(zn$Zn)
zn$Zn <- log(zn$Zn)
hist(zn$Zn)
# remove outliers
k <- removeoutlier(zn$Zn, coef = 2.5)
dt <- zn[-k,]
# determine the best kappa
system.time({
b1 <- gos_bestkappa(Zn ~ Slope + Water + NDVI + SOC + pH + Road + Mine,
  data = dt,
  kappa = c(0.01, 0.1, 1),
  nrepeat = 1,
  cores = 1)
})
b1$bestkappa
```

```
b1$plot
```

---

grid	<i>spatial grid data of explanatory variables.</i>
------	--

---

**Description**

spatial grid data of explanatory variables.

**Usage**

```
grid
```

**Format**

grid: A tibble of grided trace element explanatory variables with 13132 rows and 12 variables, where the first column is ID.

**Author(s)**

Yongze Song <yongze.song@outlook.com>

---

removeoutlier	<i>removing outliers.</i>
---------------	---------------------------

---

**Description**

Function for removing outliers.

**Usage**

```
removeoutlier(x, coef = 2.5)
```

**Arguments**

x	A vector of a variable
coef	A number of the times of standard deviation. Default is 2.5.

**Value**

Location of outliers in the vector

**Examples**

```
data("zn")
# log-transformation
hist(zn$Zn)
zn$Zn <- log(zn$Zn)
hist(zn$Zn)
# remove outliers
k <- removeoutlier(zn$Zn, coef = 2.5)
k
```

---

zn

*spatial datasets of trace element Zn.*

---

**Description**

spatial datasets of trace element Zn.

**Usage**

zn

**Format**

zn: A tibble of trace element Zn with 894 rows and 12 variables

**Author(s)**

Yongze Song <yongze.song@outlook.com>

# Index

- \* **dataset**
  - grid, 5
  - zn, 6
- \* **elements**
  - grid, 5
  - zn, 6
- \* **trace**
  - grid, 5
  - zn, 6
  
- gos, 2
- gos\_bestkappa, 3
- grid, 5
  
- removeoutlier, 5
  
- zn, 6