Package 'BMEmapping'

May 13, 2025

Type Package

Title Spatial Interpolation using Bayesian Maximum Entropy (BME)

Version 0.3.0

Maintainer Kinspride Duah <kinspride2020@gmail.com>

Description Provides an accessible and robust implementation of core BME
methodologies for spatial prediction. It enables the systematic integration
of heterogeneous data sources including both hard data (precise
measurements) and soft interval data (bounded or uncertain observations)
while incorporating prior knowledge and supporting variogram-based spatial
modeling. The BME methodology is described in Christakos (1990)
<doi:10.1007/BF00890661> and Serre and Christakos (1999)
<doi:10.1007/s004770050029>.

License MIT + file LICENSE

Encoding UTF-8

LazyData true

URL https://github.com/KinsprideDuah/BMEmapping

BugReports https://github.com/KinsprideDuah/BMEmapping/issues

RoxygenNote 7.3.2

Suggests knitr, rmarkdown, testthat (>= 3.0.0)

Config/testthat/edition 3

Imports mvtnorm

Depends R (>= 3.5)

VignetteBuilder knitr

NeedsCompilation no

Author Kinspride Duah [aut, cre, cph] (ORCID:
<https://orcid.org/0009-0004-3990-7985>),
Yan Sun [aut]

Repository CRAN

Date/Publication 2025-05-13 08:40:13 UTC

Contents

Index   **9**

---

bme_cv   *Leave-one-out cross validation (LOOCV) at hard data locations.*

Description

bme_cv performs LOOCV to evaluate the prediction performance of the Bayesian Maximum Entropy (BME) spatial interpolation method using both hard and soft (interval) data.

For each hard data location, the function removes the observed value and predicts it using all remaining hard and soft data points. This is repeated for every hard data location. The predictions are either posterior means or posterior modes, depending on the type argument.

The function returns prediction results at each location, including the residuals (differences between observed and predicted values), and computes three performance metrics:

- **ME (Mean Error)** – measures prediction bias.
- **MAE (Mean Absolute Error)** – measures average magnitude of prediction error.
- **RMSE (Root Mean Squared Error)** – emphasizes larger errors and reflects prediction accuracy.

This function is useful for validating the BME interpolation method and tuning variogram parameters.

Usage

```
bme_cv(ch, cs, zh, a, b,
       model, nugget, sill, range, nsmax = 5,
       nhmax = 5, n = 50, zk_range = range(zh, a, b, -2, 2),
       type)
```

Arguments

ch	A matrix of spatial coordinates for hard data locations (each row is a location).
cs	A matrix of spatial coordinates for soft (interval) data locations.
zh	A numeric vector of observed values at the hard data locations.
a	A numeric vector of lower bounds for the soft interval data.
b	A numeric vector of upper bounds for the soft interval data.

model	A string specifying the variogram or covariance model to use (e.g., "exp", "sph", etc.).
nugget	A non-negative numeric value for the nugget effect in the variogram model.
sill	A numeric value representing the sill (total variance) in the variogram model.
range	A positive numeric value for the range (or effective range) parameter of the variogram model.
nsmax	An integer specifying the maximum number of nearby soft data points to include for estimation (default is 5).
nhmax	An integer specifying the maximum number of nearby hard data points to include for estimation (default is 5).
n	An integer indicating the number of points at which to evaluate the posterior density over zk_range (default is 50).
zk_range	A numeric vector specifying the range over which to evaluate the unobserved value at the estimation location (zk). Although zk is unknown, it is assumed to lie within a range similar to the observed data (zh, a, and b). It is advisable to explore the posterior distribution at a few locations using prob_zk() before finalizing this range. The default is c(min(zh, a, -2), max(zh, b, 2)).
type	A string indicating the type of BME prediction to compute: either "mean" for the posterior mean or "mode" for the posterior mode.

Value

A list with two elements:

results A data frame containing the coordinates, observed values, BME predictions (posterior mean or mode), posterior variance (if type = "mean"), residuals, and fold indices.

metrics A one-row data frame reporting the mean error (ME), mean absolute error (MAE), and root mean squared error (RMSE) from the cross-validation.

Examples

```
data("utsnowload")
ch <- data.matrix(utsnowload[2:10, c("latitude", "longitude")])
cs <- data.matrix(utsnowload[68:232, c("latitude", "longitude")])
zh <- c(utsnowload[2:10, c("hard")])
a <- c(utsnowload[68:232, c("lower")])
b <- c(utsnowload[68:232, c("upper")])
bme_cv(ch, cs, zh, a, b, model = "exp", nugget = 0.0953, sill = 0.3639,
        range = 1.0787, type = "mean")
```

---

bme_predict   *Bayesian Maximum Entropy (BME) Spatial Interpolation*

Description

`bme_predict` performs BME spatial interpolation at user-specified estimation locations. It uses both hard data (precise measurements) and soft data (interval or uncertain measurements), along with a specified variogram model, to compute either the posterior mean or mode and associated variance for each location. This function enables spatial prediction in settings where uncertainty in data must be explicitly accounted for, improving estimation accuracy when soft data is available.

Usage

```
bme_predict(x, ch, cs, zh, a, b,
            model, nugget, sill, range, nsmax = 5,
            nhmax = 5, n = 50, zk_range = range(zh, a, b, -2, 2),
            type)
```

Arguments

x	A two-column matrix of spatial coordinates for the estimation locations.
ch	A two-column matrix of spatial coordinates for hard data locations.
cs	A two-column matrix of spatial coordinates for soft (interval) data locations.
zh	A numeric vector of observed values at the hard data locations.
a	A numeric vector of lower bounds for the soft interval data.
b	A numeric vector of upper bounds for the soft interval data.
model	A string specifying the variogram or covariance model to use (e.g., "exp", "sph", etc.).
nugget	A non-negative numeric value for the nugget effect in the variogram model.
sill	A numeric value representing the sill (total variance) in the variogram model.
range	A positive numeric value for the range (or effective range) parameter of the variogram model.
nsmax	An integer specifying the maximum number of nearby soft data points to include for estimation (default is 5).
nhmax	An integer specifying the maximum number of nearby hard data points to include for estimation (default is 5).
n	An integer indicating the number of points at which to evaluate the posterior density over zk_range (default is 50).
zk_range	A numeric vector specifying the range over which to evaluate the unobserved value at the estimation location (zk). Although zk is unknown, it is assumed to lie within a range similar to the observed data (zh, a, and b). It is advisable to explore the posterior distribution at a few locations using prob_zk() before finalizing this range. The default is c(min(zh, a, -2), max(zh, b, 2).
type	A string indicating the type of BME prediction to compute: either "mean" for the posterior mean or "mode" for the posterior mode.

Value

A data frame with either 3 or 4 columns, depending on the prediction type. The first two columns contain the geographic coordinates. If type = "mean", the third and fourth columns represent the posterior mean and its associated variance, respectively. If type = "mode", only a third column is returned for the posterior mode.

Examples

```
data("utsnowload")
x <- data.matrix(utsnowload[1, c("latitude", "longitude")])
ch <- data.matrix(utsnowload[2:67, c("latitude", "longitude")])
cs <- data.matrix(utsnowload[68:232, c("latitude", "longitude")])
zh <- c(utsnowload[2:67, c("hard")])
a <- c(utsnowload[68:232, c("lower")])
b <- c(utsnowload[68:232, c("upper")])
bme_predict(x, ch, cs, zh, a, b, model = "exp", nugget = 0.0953,
            sill = 0.3639, range = 1.0787, type = "mean")
```

---

casnowload *California Snow Load Data*

Description

A subset of data from the 7964 measurement locations included in the 2020 National Snow Load Study. This data is basically on reliability-targeted snow loads (RTSL) in the state of California.

Usage

```
casnowload
```

Format

A data frame with 346 rows and 8 columns.

STATION Name of the snow measuring station

LATITUDE Latitude coordinate position

LONGITUDE Longitude coordinate position

ELEVATION Elevation of the measring station (measured in meters)

RTSL The hard data RTSL value

LOWER The lower endpoint RTSL

UPPER The upper endpoint RTSL

TYPE Type of snow measurement, WESD is direct and SNWD is indirect measurement. Direct measurements are hard data and have the lower, upper and center values are the same. Indirect measurements have LOWER < RTSL < UPPER.

Source

---

prob_zk	*Posterior Density Estimation at a Single Location*

Description

Computes the posterior and plots probability density function (PDF) at a single unobserved spatial location using the Bayesian Maximum Entropy (BME) framework. This function integrates both hard data (precise measurements) and soft data (interval or uncertain observations), together with a specified variogram model, to numerically estimate the posterior density across a range of possible values. Optionally displays a plot of the posterior density function for the specified location.

Usage

```
prob_zk(x, ch, cs, zh, a, b,
        model, nugget, sill, range, nsmax = 5,
        nhmax = 5, n = 50, zk_range = range(zh, a, b, -2, 2),
        plot = FALSE)
```

Arguments

x	A two-column matrix of spatial coordinates for a single estimation location.
ch	A two-column matrix of spatial coordinates for hard data locations.
cs	A two-column matrix of spatial coordinates for soft (interval) data locations.
zh	A numeric vector of observed values at the hard data locations.
a	A numeric vector of lower bounds for the soft interval data.
b	A numeric vector of upper bounds for the soft interval data.
model	A string specifying the variogram or covariance model to use (e.g., "exp", "sph", etc.).
nugget	A non-negative numeric value for the nugget effect in the variogram model.
sill	A numeric value representing the sill (total variance) in the variogram model.
range	A positive numeric value for the range (or effective range) parameter of the variogram model.
nsmax	An integer specifying the maximum number of nearby soft data points to include for estimation (default is 5).
nhmax	An integer specifying the maximum number of nearby hard data points to include for estimation (default is 5).
n	An integer indicating the number of points at which to evaluate the posterior density over zk_range.

zk_range	A numeric vector specifying the range over which to evaluate the unobserved value at the estimation location (zk). Although zk is unknown, it is assumed to lie within a range similar to the observed data (zh, a, and b). It is advisable to explore the posterior distribution at a few locations using prob_zk() before finalizing this range. The default is c(min(zh, a, -2), max(zh, b, 2).
plot	Logical; if TRUE, plots the posterior density curve.

Value

Two elements:

data frame A data frame with two columns: zk_i (assumed zk values) and prob_zk_i (corresponding posterior densities).

plot An optional plot of posterior density of the estimation location.

Examples

```
data("utsnowload")
x <- data.matrix(utsnowload[1, c("latitude", "longitude")])
ch <- data.matrix(utsnowload[2:67, c("latitude", "longitude")])
cs <- data.matrix(utsnowload[68:232, c("latitude", "longitude")])
zh <- utsnowload[2:67, "hard"]
a <- utsnowload[68:232, "lower"]
b <- utsnowload[68:232, "upper"]
prob_zk(x, ch, cs, zh, a, b, model = "exp", nugget = 0.0953, sill = 0.3639,
        range = 1.0787, plot = TRUE)
```

---

utahsnowload	*A detrended reliability-targeted design ground snow loads in Utah*

Description

This dataset contains detrended reliability-targeted design ground snow load measurements from 232 locations in state of Utah. Of these, 65 sites report precise measurements, treated as hard data, while the remaining 167 sites report imprecise measurements, represented as interval (soft) data. The dataset is structured such that the first 67 rows contain hard (point) measurements, and the remaining rows represent soft data using lower and upper interval bounds. For a detailed explanation of the dataset and its use, refer to the related version described in Duah et al. (2025) doi:10.1016/j.spasta.2025.100894

Usage

utsnowload

Format

A data frame with 232 rows and 5 variables:

latitude Latitude coordinate position

longitude Longitude coordinate position

hard The hard data value

lower The lower endpoint of the soft-interval

upper The upper endpoint of the soft-interval

Source

Index