Package: mwa (via r-universe)

November 2, 2026

Title Causal Inference in Spatiotemporal Event Data

Version 0.5.1

Date 2026-01-22

Description Implementation of Matched Wake Analysis (mwa) for studying causal relationships in spatiotemporal event data, introduced by Schutte and Donnay (2014) <doi:10.1016/j.polgeo.2014.03.001>.

Maintainer Karsten Donnay <kdonnay@gmx.net>

License LGPL-3

Depends cem, rJava, MASS

SystemRequirements Java

NeedsCompilation no

Author Sebastian Schutte [aut], Karsten Donnay [aut, cre]

Config/pak/sysreqs default-jdk

Repository https://cran.r-universe.dev

Date/Publication 2026-01-22 10:49:24 UTC

RemoteUrl https://github.com/cran/mwa

RemoteRef HEAD

RemoteSha 8c56c56fcbb5dd2bed32889da02dd9061a1b947b

Contents

mwa-package . . . 2
matchedwake . . . 3
mwa_data . . . 7
plot.matchedwake . . . 8
print.matchedwake . . . 9
slideWakeMatch . . . 10
slidingWake . . . 11
summary.matchedwake . . . 12

Index **14**

mwa-package *Matched Wake Analysis (mwa): Analyzing causal relationships in spatiotemporal event data*

Description

The package is designed to analyze causal relationships in spatially and temporally referenced data. Specific types of events might affect subsequent levels of other events. To estimate the corresponding effect, treatment, control, and dependent events are selected from the empirical sample. Treatment effects are established through automated matching and a diff-in-diffs regression design. The analysis is repeated for various spatial and temporal offsets from the treatment events.

Details

The full functionality of **mwa** is given through `matchedwake`, which relies on a small set of auxiliary methods. Note that `print()`, `summary()` and `plot()` commands are overloaded to return outputs specific to class `matchedwake`. For performance reasons, the iterative counting is done in Java using the **rJava** interface.

IMPORTANT: The size of the Java heap space has to be set before **first** calling the package via `library(mwa)` since JVM size cannot change once it has been initialized. This also implies that R has to be restarted if another library was already using a JVM in order for the heap space option to have any effect. To set the heap space to 1 GB, for example, use `options(java.parameters = "-Xmx1g")` (512 MB is the default size).

Author(s)

Sebastian Schutte and Karsten Donnay

References

Schutte, S., Donnay, K. (2014). "Matched wake analysis: Finding causal relationships in spatiotemporal event data." *Political Geography* 41:1-10.

See Also

`matchedwake`, `slidingWake`, `slideWakeMatch`, `print.matchedwake`, `summary.matchedwake`, `plot.matchedwake`

Examples

```
# Loading sample data
data(mwa_data)

# Specify required parameters:
# - 2 to 10 days in steps of 2
t_window <- c(2,10,2)
# - 2 to 10 kilometers in steps of 2
spat_window <- c(2,10,2)
# - column and entries that indicate treatment events
```

```
treatment <- c("type","treatment")
# - column and entries that indicate control events
control  <- c("type","control")
# - column and entries that indicate dependent events
dependent <- c("type","dependent")
# - columns to match on
matchColumns <- c("match1","match2")

# Specify optional parameters:
# - use weighted regression (default estimation method is "lm")
weighted <- TRUE
# - temporal units
t_unit <- "days"
# - match on counts of previous treatment and control events
TCM <- TRUE


# Execute method:
results <- matchedwake(mwa_data, t_window, spat_window, treatment, control, dependent,
                       matchColumns, weighted = weighted, t_unit = t_unit, TCM = TCM)

# Plot results:
plot(results)

# Return detailed summary of results:
summary(results, detailed = TRUE)
```

`matchedwake` *Estimate Treatment Effect for Sliding Spatiotemporal Windows*

Description

This function performs the Matched Wake Analysis (mwa), which consists of two steps: counts for previous and posterior events are established for different spatial and temporal offsets from treatment and control events. After that, the treatment effect is estimated in a difference-in-differences regression design. For performance reasons, the iterative counting is done in Java using the **rJava** interface.

IMPORTANT: The size of the Java heap space has to be set before **first** calling the package via `library(mwa)` since JVM size cannot change once it has been initialized. This also implies that R has to be restarted if another library was already using a JVM in order for the heap space option to have any effect. To set the heap space to 1 GB, for example, use `options(java.parameters = "-Xmx1g")` (512 MB is the default size).

Usage

```
matchedwake(data, t_window, spat_window, treatment, control,
            dependent, matchColumns, t_unit = "days", estimation = "lm",
```

```
    formula = "dependent_post ~ dependent_pre + treatment",
           weighted = FALSE, estimationControls = c(), TCM = FALSE,
           deleteSUTVA = FALSE, alpha1 = 0.05, alpha2 = 0.1,
           match.default = TRUE, match.details = FALSE, ...)
```

Arguments

data	data.frame containing the observations. See Details.
t_window	specification of temporal windows in t_units. See Details.
spat_window	specification of spatial windows in kilometers. See Details.
treatment	vector of Strings identifying which type of events serve as treatments. See Details.
control	vector of Strings identifying which type of events serve as controls. See Details.
dependent	vector of Strings identifying which type of events are affected by treatment. See Details.
matchColumns	vector of Strings indicating the columns to match on. See Details.
t_unit	String specifying the temporal units to be used, either "days", "hours", "mins" or "secs". Default = "days". See Details.
estimation	String specifying method used for estimation, "lm", "att" or "nb". Default = "lm". See Details.
formula	String specifying the model used for estimation. Default = "dependent_post ~ dependent_pre + treatment". See Details.
weighted	Boolean specifying whether regression is weighted (only affects estimations using "lm" or "att"). Default = FALSE.
estimationControls	vector of Strings indicating additional control dimensions to be included in the estimation. See Details.
TCM	Boolean to select whether the method should match on counts of previous treatment and control instances. Default = FALSE.
deleteSUTVA	Boolean to select whether overlapping treatment and control episodes are deleted. Default = FALSE.
alpha1	first significance level used for the analysis and plots. Default = 0.05.
alpha2	second significance level used for the analysis and plots. Default = 0.1.
match.default	Boolean to select whether observations are matched using cem. Default = TRUE.
match.details	Boolean to select whether detailed matching statistics are returned from cem. Default = FALSE.
...	optional parameters that can be passed to the methods used for matching and estimation. See Details.

Details

The method expects `data` to be a `data.frame`. Dates must be given in column `timestamp` and formatted as a date string with format `"YYYY-MM-DD hh:mm:ss"`. Alternatively, a POSIX Date can be specified using the same format. `data` must also contain two entries called `lat` and `lon` for the geo location of each entry.

`t_window` specifies the minimal and maximal temporal window sizes and corresponding steps used in the iteration. Required syntax is `c(min_window, max_window, step_size)` with `step_size` in units of `t_unit`. The spatial window `spat_window` is specified in the same way with kilometers as units.

`treatment`, `control` and `dependent` define which category of events is considered to be treatment, control and dependent cases respectively. The required syntax is `c(column_name, value)` where `column_name` must be entered as String and `value` can be Numeric, Boolean, or a String.

`matchColumns` selects the columns in `data` used for matching. Matching variables are expected to be coded together with every treatment and control type event and are assumed to reflect a set of suitable matching variables (what is suitable will, of course, vary from case to case).

The optional argument `t_unit` specifies the temporal resolution for which the analysis is to be conducted, one of either `"days"`, `"hours"`, `"mins"` or `"secs"`. If the time stamps provided in `data` are more precise than the resolution they are truncated accordingly.

mwa estimates treatment effects using a diff-in-diffs regression design. By default this is specified as `"dependent_post ~ dependent_pre + treatment"` (where "pre" and "post" refer to pre and post intervention). Alternatively, `"dependent_post - dependent_pre ~ treatment"` is accepted. Only those two input specifications are allowed, any other input will result in an error.

Three different estimation approaches can be chosen using `estimation`: a linear model (`"lm"`, **stats**), all models available through (`"att"`, **cem**) or a count dependent model (`"glm.nb"`, **MASS**). For regressions using `"lm"` or `"att"` `weighted` sets whether or not the regression is weighted by the number of treatment vs. control cases. Additional control variables can be specified via `estimationControls`. For example, if `estimationControls = c("covariate1")`, the package automatically modifies the default estimation formula to `"dependent_post ~ dependent_pre + covariate1 + treatment"` (analogously for the other specification). In this case the output then also not only returns the estimate and p value for `treatment` but further returns the coefficients and p values for all additional control variables.

The package supports full inheritance for optional arguments of the following methods: `cem` and `att` (**cem**), `lm` (**stats**), `glm.nb` (**MASS**). To guarantee unique inputs for each method, options have to entered into `matchedwake()` using a prefix that consists of the method name separated by ".". For example, in order for `cem` to return an exactly balanced dataset simply add `cem.k2k = TRUE` as optional argument.

Value

Returns an object of class `matchedwake`, which is a list of objects with the following slots:

`estimates` `data.frame` with estimates and p values for all spatial and temporal windows considered. For `estimation = "lm"` it also returns a pseudo R^2 value. If additional control dimensions were included in the estimation, it further returns the corresponding coefficients and p values.

matching	data.frame with detailed matching statistics for all spatial and temporal windows considered. Returns the number of control and treatment episodes, L1 metric, percent common support. All values are given both pre and post matching.
SUTVA	data.frame with detailed statistics on the degree of overlaps of the spatiotemporal cylinders. Returns the fraction of cases in which two or more treatment (or control) episodes overlap ("SO": same overlap) and the fraction of overlapping treatment and control episodes ("MO": mixed overlap). All values are given pre and post matching and for the full time window.
wakes	data.frame providing the information for the spatiotemporal cylinders (or wakes) for all spatial and temporal windows considered. Returns the eventID (i.e. the index of the event in the time-ordered dataset), treatment (1: treatment episode, 0: control episode), counts of dependent events, overlaps ("SO" and "MO") pre and post intervention, and the matching variables.
matched	data.frame with details on which observations were matched for all spatial and temporal windows considered. Returns the eventID (i.e. the index of the event in the time-ordered dataset), treatment (1: treatment episode, 0: control episode) and matched (1: matched, 0: not matched).
bins	data.frame with details for the bins used in cem matching for all spatial and temporal windows considered. Returns the bin information from cem (variable "breaks") for each matchColumns.
parameters	list of all arguments passed to the method.
call	the call.

Author(s)

Sebastian Schutte and Karsten Donnay.

References

Schutte, S., Donnay, K. (2014). "Matched wake analysis: Finding causal relationships in spatiotemporal event data." *Political Geography* 41:1-10.

See Also

mwa-package, slidingWake, slideWakeMatch, print.matchedwake, summary.matchedwake, plot.matchedwake

Examples

```
# Loading sample data
data(mwa_data)

# Specify required parameters:
# - 2 to 10 days in steps of 2
t_window <- c(2,10,2)
# - 2 to 10 kilometers in steps of 2
spat_window <- c(2,10,2)
# - column and entries that indicate treatment events
```

```
treatment <- c("type","treatment")
# - column and entries that indicate control events
control  <- c("type","control")
# - column and entries that indicate dependent events
dependent <- c("type","dependent")
# - columns to match on
matchColumns <- c("match1","match2")

# Specify optional parameters:
# - use weighted regression (default estimation method is "lm")
weighted <- TRUE
# - temporal units
t_unit <- "days"
# - match on counts of previous treatment and control events
TCM <- TRUE


# Execute method:
results <- matchedwake(mwa_data, t_window, spat_window, treatment, control, dependent,
                       matchColumns, weighted = weighted, t_unit = t_unit, TCM = TCM)

# Plot results:
plot(results)

# Return detailed summary of results:
summary(results, detailed = TRUE)
```

`mwa_data` *Data to Illustrate the Functionality of mwa*

Description

This artificial data set illustrates how **mwa** can be used to identify causal effects. Treatment, control, and dependent events are referenced in time and space. Increased levels of dependent events following treatments can be visually and numerically analyzed using the package.

Usage

```
data(mwa_data)
```

Format

A `data.frame` containing observations.

Source

Monte Carlo Simulations. See supplementary information of Schutte and Donnay (2014).

References

Schutte, S., Donnay, K. (2014). "Matched wake analysis: Finding causal relationships in spatiotemporal event data." *Political Geography* 41:1-10.

`plot.matchedwake` *Plot Function for Objects of Class* `matchedwake`

Description

Overloads the default `plot()` for objects of class `matchedwake`. Returns a contour plot: The lighter the color the larger the estimated treatment effect. The corresponding standard errors are indicated by shading out some of the estimates: No shading corresponds to $p < alpha1$ for the treatment effect in the diff-in-diffs analysis. Dotted lines indicate p-values between `alpha1` and `alpha2` and full lines indicate $p > alpha2$. The cells indicating effect size and significance level are arranged in a table where each field corresponds to one specific combination of spatial and temporal sizes.

Usage

```
## S3 method for class 'matchedwake'
plot(x, zlim = NA, plotNAs = TRUE, adjust = "none", ...)
```

Arguments

`x`	object of class `matchedwake`.
`zlim`	Manually sets the range of the color map of the contour plot, required format is c(MINIMUM,MAXIMUM). `Default = NA`, i.e. the range is automatically set from the MINIMUM and MAXIMUM values of the estimates.
`plotNAs`	Boolean indicating whether or not to visualize NA estimates as "no effect" (i.e. 0). `Default = TRUE`.
`adjust`	Select p-value adjustment method for multiple testing, `Default = "none"`; takes values "none", "holm", "hochberg", "hommel", "bonferroni", "BH", "BY" or "fdr", see '?p.adjust' (**stats**) for details (Credit: David Armstrong, github.com/davidaarmstrong)
`...`	further arguments passed to or from other methods.

Details

The argument `adjust` allows to adjust p-values for multiple comparisons implementing the functionality of the `p.adjust` function of package **stats**. Given the nature of individual statistical estimation per unique combination of spatial and temporal windows, results might be vulnerable to multiple inference problems that require correction. Users can use any of the different corrections implemented in `p.adjust`, depending on which they think is most appropriate for their application at hand. By default no correction is applied.

Author(s)

Sebastian Schutte and Karsten Donnay.

References

Schutte, S., Donnay, K. (2014). "Matched wake analysis: Finding causal relationships in spatiotemporal event data." *Political Geography* 41:1-10.

See Also

matchedwake

`print.matchedwake` *Print Function for Objects of Class* `matchedwake`

Description

Overloads the default `print()` for objects of class `matchedwake`.

Usage

```
## S3 method for class 'matchedwake'
print(x, ...)
```

Arguments

`x`	object of class `matchedwake`.
`...`	further arguments passed to or from other methods.

Value

Returns a `data.frame` with all significant results (significance level is `alpha1` as retrieved from `x$parameters`).

Author(s)

Sebastian Schutte and Karsten Donnay.

References

Schutte, S., Donnay, K. (2014). "Matched wake analysis: Finding causal relationships in spatiotemporal event data." *Political Geography* 41:1-10.

See Also

matchedwake

slideWakeMatch *Auxiliary Function to Match Data and Estimate Treatment Effects*

Description

Method takes the output of slidingWake, matches observations using cem and estimates treatment effects using linear models (lm or att) or a count dependent variable model (glm.nb).

Usage

```
slideWakeMatch(wakes, alpha1, matchColumns, estimation, formula, weighted,
               estimationControls, TCM, match.default, match.details, ...)
```

Arguments

wakes	data.frame. See "wakes" in the description of matchedwake for details.
alpha1	significance level used for the analysis and plots. Default = 0.05.
matchColumns	vector of Strings indicating the columns to match on.
estimation	String specifying method used for estimation.
formula	String specifying the model used for estimation.
weighted	Boolean specifying whether regression is weighted.
estimationControls	vector of Strings indicating additional control dimensions to be included in the estimation.
TCM	Boolean to select whether the method should match on counts of previous treatment and control instances.
match.default	Boolean to select whether observations are matched using cem.
match.details	Boolean to select whether detailed matching statistics are returned from cem. Default = FALSE.
...	optional parameters that can be passed to the methods used for matching and estimation.

Details

See the description of matchedwake for details.

Value

Returns a list with the following slots:

estimates	data.frame with estimates and p values for all spatial and temporal windows considered.
matching	data.frame with detailed matching statistics for all spatial and temporal windows considered.

SUTVA	data.frame with detailed statistics on the degree of overlaps of the spatiotemporal cylinders.
wakes	data.frame providing the information for the spatiotemporal cylinders (or wakes) for all spatial and temporal windows considered.
matched	data.frame with details on which observations were matched for all spatial and temporal windows considered.
bins	data.frame with details for the bins used in cem matching for all spatial and temporal windows considered.

See the description of matchedwake for details.

Author(s)

Sebastian Schutte and Karsten Donnay.

References

Schutte, S., Donnay, K. (2014). "Matched wake analysis: Finding causal relationships in spatiotemporal event data." *Political Geography* 41:1-10.

See Also

matchedwake, slidingWake

---

slidingWake *Auxiliary Function to Iterate Through Sliding Spatiotemporal Windows*

Description

Method iterates through all spatial and temporal window sizes specified and counts dependent events with a given spatial window and for a given temporal window (symmetrically in forward and backward direction in time). For performance reasons, the iterative counting is done in Java using the **rJava** interface.

IMPORTANT: The size of the Java heap space has to be set before **first** calling the package via `library(mwa)` since JVM size cannot change once it has been initialized. This also implies that R has to be restarted if another library was already using a JVM in order for the heap space option to have any effect. To set the heap space to 1 GB, for example, use `options(java.parameters = "-Xmx1g")` (512 MB is the default size).

Usage

```
slidingWake(data, t_unit, t_window, spat_window, treatment, control,
            dependent, matchColumns, estimationControls)
```

Arguments

data	data.frame containing the observations.
t_unit	String specifying the temporal units to be used.
t_window	specification of temporal windows in t_units.
spat_window	specification of spatial windows in kilometers.
treatment	vector of Strings identifying which type of events serve as treatments.
control	vector of Strings identifying which type of events serve as controls.
dependent	vector of Strings identifying which type of events are affected by treatment.
matchColumns	vector of Strings indicating the columns to match on.
estimationControls	vector of Strings indicating additional control dimensions to be included in the estimation.

Details

See the description of matchedwake for details.

Value

Returns a data.frame. See "wakes" in the description of matchedwake for details.

Author(s)

Sebastian Schutte and Karsten Donnay.

References

Schutte, S., Donnay, K. (2014). "Matched wake analysis: Finding causal relationships in spatiotemporal event data." *Political Geography* 41:1-10.

See Also

matchedwake, slideWakeMatch

---

summary.matchedwake *Summary Function for Objects of Class* matchedwake

---

Description

Overloads the default summary() for objects of class matchedwake.

Usage

```
## S3 method for class 'matchedwake'
summary(object, detailed = FALSE, ...)
```

Arguments

object	object of class matchedwake.
detailed	Boolean indicating whether or not a detailed summary should be returned. Default = TRUE.
...	further arguments passed to or from other methods.

Value

Returns a data.frame with an overview of all significant results (significance level is alpha1 as retrieved from x$parameters). If detailed = TRUE this overview includes a number of matching statistics and statistics on overlaps of the spatiotemporal cylinders. If additional control dimensions were included in the estimation, it also provides an overview of the corresponding coefficients and p values for all significant results.

Author(s)

Sebastian Schutte and Karsten Donnay.

References

Schutte, S., Donnay, K. (2014). "Matched wake analysis: Finding causal relationships in spatiotemporal event data." *Political Geography* 41:1-10.

See Also

matchedwake

Index

att, *5*, *10*

cem, *5*, *10*

glm.nb, *5*, *10*

lm, *5*, *10*

matchedwake, *2*, 3, *8–13*
mwa-package, 2
mwa_data, 7

p.adjust, *8*
plot.matchedwake, *2*, *6*, 8
print.matchedwake, *2*, *6*, 9

slideWakeMatch, *2*, *6*, 10, *12*
slidingWake, *2*, *6*, *10*, 11, *11*
summary.matchedwake, *2*, *6*, 12