

# Package ‘gateR’

November 30, 2020

**Type** Package

**Title** Flow/Mass Cytometry Gating via Spatial Kernel Density Estimation

**Version** 0.1.4

**Date** 2020-11-30

**Maintainer** Ian D. Buller <ian.buller@nih.gov>

**Description** Estimates statistically significant marker combination values within which one immunologically distinctive group (i.e., disease case) is more associated than another group (i.e., healthy control), successively, using various combinations (i.e., “gates”) of markers to examine features of cells that may be different between groups. For a two-group comparison, the ‘gateR’ package uses the spatial relative risk function that is estimated using the ‘sparr’ package. Details about the ‘sparr’ package methods can be found in the tutorial: Davies et al. (2018) <doi:10.1002/sim.7577>. Details about kernel density estimation can be found in J. F. Bithell (1990) <doi:10.1002/sim.4780090616>. More information about relative risk functions using kernel density estimation can be found in J. F. Bithell (1991) <doi:10.1002/sim.4780101112>.

**License** Apache License (>= 2.0)

**Encoding** UTF-8

**LazyData** true

**RoxygenNote** 7.1.1

**Depends** R (>= 3.5.0)

**Imports** fields, graphics, grDevices, maptools, pgirmess, raster, sp, sparr, spatstat, stats, tibble, utils

**Suggests** R.rsp, spelling, testthat

**VignetteBuilder** R.rsp

**Language** en-US

**URL** <https://github.com/Waller-SUSAN/gateR>

**BugReports** <https://github.com/Waller-SUSAN/gateR/issues>

**NeedsCompilation** no

**Author** Ian D. Buller [aut, cre, cph] (<<https://orcid.org/0000-0001-9477-8582>>),  
 Elena Hsieh [ctb] (<<https://orcid.org/0000-0003-3969-6597>>),  
 Debashis Ghosh [ctb] (<<https://orcid.org/0000-0001-6618-1316>>),  
 Lance A. Waller [ctb] (<<https://orcid.org/0000-0001-5002-8886>>),  
 NCI [cph]

**Repository** CRAN

**Date/Publication** 2020-11-30 18:20:03 UTC

## R topics documented:

gateR-package . . . . .	2
gating . . . . .	3
lotrrs . . . . .	7
randCyto . . . . .	9
rrs . . . . .	10
<b>Index</b>	<b>13</b>

---

gateR-package	<i>The gateR Package: Flow/Mass Cytometry Gating via Spatial Kernel Density Estimation</i>
---------------	--

---

## Description

Estimates statistically significant fluorescent marker combination values within which one immunologically distinctive group (i.e., disease case) is more associated than another group (i.e., healthy control), successively, using various combinations (i.e., "gates") of fluorescent markers to examine features of cells that may be different between groups.

## Details

For a two-group comparison, the 'gateR' package uses the spatial relative risk function that is estimated using the sparr package. Details about the sparr package methods can be found in the tutorial: Davies et al. (2018) <https://doi.org/10.1002/sim.7577>. Details about kernel density estimation can be found in J. F. Bithell (1990) <https://doi.org/10.1002/sim.4780090616>. More information about relative risk functions using kernel density estimation can be found in J. F. Bithell (1991) <https://doi.org/10.1002/sim.4780101112>.

This package provides a function to perform a gating strategy for flow cytometry data. The 'gateR' package also provides basic visualization for each gate.

Key content of the 'gateR' package include:

### Gating Strategy

**gating** Extracts cells within statistically significant combinations of fluorescent markers, successively, for a set of markers. Statistically significant combinations are identified using two-tailed

p-values of a relative risk surface assuming asymptotic normality. This function is currently available for two-level comparisons of a single condition (e.g., case/control) or two conditions (e.g., case/control at time 1 and time 2). Provides functionality for basic visualization and multiple testing correction.

**rrs** Estimates a relative risk surface and computes the asymptotic p-value surface for a single gate with a single condition. Includes features for basic visualization. This function is used internally within the **gating** function to extract the points within the significant areas. This function can also be used as a standalone function.

**lotrrs** Estimates a ratio of relative risk surfaces and computes the asymptotic p-value surface for a single gate with two conditions. Includes features for basic visualization. This function is used internally within the **gating** function to extract the points within the significant areas. This function can also be used as a standalone function.

### Flow Cytometry Data

**randCyto** A sample dataset containing information about flow cytometry data with two binary categorical variables. The data are a random subset of the 'extdata' data in the 'flowWorkspaceData' package found on Bioconductor <http://bioconductor.org/packages/release/data/experiment/html/flowWorkspaceData.html> and formatted for gateR input.

### Dependencies

The 'gateR' package relies heavily upon **sparr**, **sp**, and **raster**. For a two-level comparison, the spatial relative risk function uses the **risk** function. The **gating** function extracts cells within significant clusters using the **rasterToPolygons**, **unionSpatialPolygons**, and **over** functions. The calculation of a Bonferroni correction for multiple testing accounting for the spatial correlation of the estimated surface uses the **correlog** function. Basic visualizations rely on the **image.plot** function.

### Author(s)

Ian D. Buller

*Occupational and Environmental Epidemiology Branch, Division of Cancer Epidemiology and Genetics, National Cancer Institute, National Institutes of Health, Rockville, Maryland, USA.*

Maintainer: I.D.B. <ian.buller@nih.gov>

---

gating

*Gating strategy for mass cytometry data using spatial relative risk functions*

---

### Description

Extracts cells within statistically significant combinations of fluorescent markers, successively, for a set of markers. Statistically significant combinations are identified using two-tailed p-values of a relative risk surface assuming asymptotic normality. This function is currently available for two-level comparisons of a single condition (e.g., case/control) or two conditions (e.g., case/control at time 1 and time 2). Provides functionality for basic visualization and multiple testing correction.

**Usage**

```
gating(
  dat,
  vars,
  n_condition = c(1, 2),
  numerator = TRUE,
  alpha = 0.05,
  p_correct = "none",
  nbc = NULL,
  doplot = FALSE,
  rcols = c("#FF0000", "#cccccc", "#0000FF"),
  lower_lrr = NULL,
  upper_lrr = NULL,
  c1n = NULL,
  c2n = NULL,
  win = NULL,
  verbose = FALSE,
  ...
)
```

**Arguments**

<code>dat</code>	Input data frame flow cytometry data with the following features (columns): 1) ID, 2) Condition A ID, 3) Condition B ID (optional), and a set of markers.
<code>vars</code>	A vector of characters with the name of features (columns) within <code>dat</code> to use as markers for each gate. See details below.
<code>n_condition</code>	A numeric value of either 1 or 2 designating if the gating is performed with one condition or two conditions.
<code>numerator</code>	Logical. If TRUE (the default), cells will be extracted within all statistically significant numerator (i.e., case) clusters. If FALSE, cells will be extracted within all statistically significant denominator (i.e., control) clusters.
<code>alpha</code>	Numeric. The two-tailed alpha level for significance threshold (default is 0.05).
<code>p_correct</code>	Character string specifying whether to apply a correction for multiple comparisons including a Bonferroni correction <code>p_correct = "uncorrelated"</code> or a correlated Bonferroni correction <code>p_correct = "correlated"</code> . If <code>p_correct = "none"</code> (the default), then no correction is applied.
<code>nbc</code>	Optional. An integer for the number of bins when <code>p_correct = "correlated"</code> . Similar to <code>nbc</code> argument in <code>correlog</code> . The default is the average number of gridded knots in one-dimension (i.e., x-axis).
<code>doplot</code>	Logical. If TRUE, the output includes basic data visualizations.
<code>rcols</code>	Character string of length three (3) specifying the colors for: 1) group A (numerator), 2) neither, and 3) group B (denominator) designations. The defaults are <code>c("#FF0000", "#cccccc", "#0000FF")</code> or <code>c("red", "grey80", "blue")</code> .
<code>lower_lrr</code>	Optional, numeric. Lower cut-off value for the log relative risk value in the color key (typically a negative value). The default is no limit and the color key will include the minimum value of the log relative risk surface.

upper_lrr	Optional, numeric. Upper cut-off value for the log relative risk value in the color key (typically a positive value). The default is no limit and the color key will include the maximum value of the log relative risk surface.
c1n	Optional, character. The name of the level for the numerator of condition A. The default is null and the first level is treated as the numerator.
c2n	Optional, character. The name of the level for the numerator of condition B. The default is null and the first level is treated as the numerator.
win	Optional. Object of class <code>owin</code> for a custom two-dimensional window within which to estimate the surfaces. The default is <code>NULL</code> and calculates a convex hull around the data.
verbose	Logical. If <code>TRUE</code> will print function progress during execution. If <code>FALSE</code> (the default), will not print.
...	Arguments passed to <code>risk</code> to select bandwidth, edge correction, and resolution.

## Details

This function performs a sequential gating strategy for mass cytometry data comparing two levels with one or two conditions. Gates are typically two-dimensional space comprised of two fluorescent markers. The two-level comparison allows for the estimation of a spatial relative risk function and the computation of p-value based on an assumption of asymptotic normality. Cells within statistically significant areas are extracted and used in the next gate. This function relies heavily upon the `risk` function. Basic visualization is available if `doplot = TRUE`.

The `vars` argument must be a vector with an even-numbered length where the odd-numbered elements are the markers used on the x-axis of a gate and the even-numbered elements are the markers used on the y-axis of a gate. For example, if `vars = c("V1", "V2", "V3", and "V4")` then the first gate is "V1" on the x-axis and "V2" on the y-axis and then the second gate is "V3" on the x-axis and "V4" on the y-axis. Markers can be repeated in successive gates.

The `n_condition` argument specifies if the gating strategy is performed for one condition or two conditions. If `n_condition = 1`, then the function performs a one condition gating strategy using the internal `rrs` function, which computes the statistically significant areas (clusters) of a relative risk surface at each gate and selects the cells within the clusters specified by the numerator argument. If `n_condition = 2`, then the function performs a two conditions gating strategy using the internal `lotrrs` function, which computes the statistically significant areas (clusters) of a ratio of relative risk surfaces at each gate and selects the cells within the clusters specified by the numerator argument. The condition variable(s) within `dat` must be of class 'factor' with two levels. The first level is considered the numerator (i.e., "case") value and the second level is considered the denominator (i.e., "control") value. The levels can also be specified using the `c1n` and `c2n` parameters. See the documentation for the internal `rrs` and `lotrrs` functions for more details.

The p-value surface of the ratio of relative risk surfaces is estimated assuming asymptotic normality of the ratio value at each gridded knot. The bandwidth is fixed across all layers.

Provides functionality for a correction for multiple testing. If `p_correct = "uncorrelated"`, then a conventional Bonferroni correction is calculated by dividing the alpha level by the number of gridded knots across the estimated surface. The default in the `risk` function is a resolution of 128 x 128 or  $n = 16,384$  knots and a custom resolution can be specified using the `resolution` argument within the `risk` function. If `p_correct = "correlated"` (NOTE: May take a considerable amount of computation resources and time), then a Bonferroni correction that takes into account the spatial

correlation of the surface is calculated within the internal `pval_correct` function. The alpha level is divided by the minimum number of knots that are not spatially correlated. The minimum number of knots that are not spatially correlated is computed by counting the knots that are a distance apart that exceeds the minimum distance of non-significant spatial correlation based on a correlogram using the `correlog` function. If `p_correct = "none"` (the default), then the function does not account for multiple testing and uses the uncorrected alpha level. See the internal `pval_correct` function documentation for more details.

## Value

An object of class `list`. This is a named list with the following components:

- `obs` An object of class `'tibble'` of the same features as `dat` that includes the information for the cells extracted with significant clusters in the final gate.
- `n` An object of class `'list'` of the sample size of cells at each gate. The length is equal to the number of successful gates plus the final result.
- `gate` An object of class `'list'` of `'rrs'` objects from each gate. The length is equal to the number of successful gates.

The objects of class `'rrs'` is similar to the output of the `risk` function with two additional components:

- `rr` An object of class `'im'` with the relative risk surface.
- `f` An object of class `'im'` with the spatial density of the numerator.
- `g` An object of class `'im'` with the spatial density of the denominator.
- `P` An object of class `'im'` with the asymptotic p-value surface.
- `lrr` An object of class `'im'` with the log relative risk surface.
- `alpha` A numeric value for the alpha level used within the gate.

## Examples

```
if (interactive()) {
  ## Single condition, no multiple testing correction
  test_gate <- gateR::gating(dat = randCyto,
    vars = c("arcsinh_CD4", "arcsinh_CD38",
             "arcsinh_CD8", "arcsinh_CD3"),
    n_condition = 1,
    p_correct = "none")
}
```

lotrrs

*A single gate for two conditions***Description**

Estimates a ratio of relative risk surfaces and computes the asymptotic p-value surface for a single gate with two conditions. Includes features for basic visualization. This function is used internally within the [gating](#) function to extract the points within the significant areas. This function can also be used as a standalone function.

**Usage**

```
lotrrs(
  dat,
  alpha = 0.05,
  p_correct = "none",
  nbc = NULL,
  doplot = FALSE,
  rcols = c("#FF0000", "#cccccc", "#0000FF"),
  lower_lrr = NULL,
  upper_lrr = NULL,
  c1n = NULL,
  c2n = NULL,
  win = NULL,
  verbose = FALSE,
  ...
)
```

**Arguments**

<code>dat</code>	Input data frame flow cytometry data with five (5) features (columns): 1) ID, 2) Condition A ID, 3) Condition B ID, 4) Marker A as x-coordinate, 5) Marker B as y-coordinate.
<code>alpha</code>	Numeric. The two-tailed alpha level for significance threshold (default is 0.05).
<code>p_correct</code>	Character string specifying whether to apply a correction for multiple comparisons including a Bonferroni correction <code>p_correct = "uncorrelated"</code> or a correlated Bonferroni correction <code>p_correct = "correlated"</code> . If <code>p_correct = "none"</code> then no correction is applied.
<code>nbc</code>	Optional. An integer for the number of bins when <code>p_correct = "correlated"</code> . Similar to <code>nbc</code> argument in <a href="#">correlog</a> . The default is the average number of gridded knots in one-dimension (i.e., x-axis).
<code>doplot</code>	Logical. If TRUE, the output includes basic data visualizations.
<code>rcols</code>	Character string of length three (3) specifying the colors for: 1) group A (numerator), 2) neither, and 3) group B (denominator) designations. The defaults are <code>c("#FF0000", "#cccccc", "#0000FF")</code> or <code>c("red", "grey80", "blue")</code> .

lower_lrr	Optional, numeric. Lower cut-off value for the log relative risk value in the color key (typically a negative value). The default is no limit and the color key will include the minimum value of the log relative risk surface.
upper_lrr	Optional, numeric. Upper cut-off value for the log relative risk value in the color key (typically a positive value). The default is no limit and the color key will include the maximum value of the log relative risk surface.
c1n	Optional, character. The name of the level for the numerator of condition A. The default is null and the first level is treated as the numerator.
c2n	Optional, character. The name of the level for the numerator of condition B. The default is null and the first level is treated as the numerator.
win	Optional. Object of class <code>owin</code> for a custom two-dimensional window within which to estimate the surfaces. The default is <code>NULL</code> and calculates a convex hull around the data.
verbose	Logical. If <code>TRUE</code> will print function progress during execution. If <code>FALSE</code> (the default), will not print.
...	Arguments passed to <code>risk</code> to select bandwidth, edge correction, and resolution.

## Details

This function estimates a ratio of relative risk surfaces and computes the asymptotic p-value surface for a single gate with two conditions using three successive `risk` functions. A relative risk surface is estimated for Condition A at each level of Condition B and then a ratio of the two relative risk surfaces is computed.

$$RR_{ConditionB1} = \frac{ConditionA2ofB1}{ConditionA1ofB1}$$

$$RR_{ConditionB2} = \frac{ConditionA2ofB2}{ConditionA1ofB2}$$

$$\ln(rRR) = \ln\left(\frac{RR_{ConditionB2}}{RR_{ConditionB1}}\right)$$

The p-value surface of the ratio of relative risk surfaces is estimated assuming asymptotic normality of the ratio value at each gridded knot. The bandwidth is fixed across all layers. Basic visualization is available if `doplot = TRUE`.

Provides functionality for a correction for multiple testing. If `p_correct = "uncorrelated"`, then a conventional Bonferroni correction is calculated by dividing the alpha level by the number of gridded knots across the estimated surface. The default in the `risk` function is a resolution of 128 x 128 or  $n = 16,384$  knots and a custom resolution can be specified using the `resolution` argument within the `risk` function. If `p_correct = "correlated"` (NOTE: May take a considerable amount of computation resources and time), then a Bonferroni correction that takes into account the spatial correlation of the surface is calculated within the internal `pval_correct` function. The alpha level is divided by the minimum number of knots that are not spatially correlated. The minimum number of knots that are not spatially correlated is computed by counting the knots that are a distance apart that exceeds the minimum distance of non-significant spatial correlation based on a correlogram using the `correlog` function. If `p_correct = "none"`, then the function does not account

for multiple testing and uses the uncorrected alpha level. See the internal `pval_correct` function documentation for more details.

The two condition variables (Condition A and Condition B) within `dat` must be of class 'factor' with two levels. The first level in each variable is considered the numerator (i.e., "case") value and the second level is considered the denominator (i.e., "control") value. The levels can also be specified using the `c1n` and `c2n` parameters.

## Value

An object of class 'list' where each element is a object of class 'rrs' created by the `risk` function with two additional components:

- `rr` An object of class 'im' with the relative risk surface.
- `f` An object of class 'im' with the spatial density of the numerator.
- `g` An object of class 'im' with the spatial density of the denominator.
- `P` An object of class 'im' with the asymptotic p-value surface.
- `lrr` An object of class 'im' with the log relative risk surface.
- `alpha` A numeric value for the alpha level used within the gate.

## Examples

```
test_lotrrs <- lotrrs(dat = randCyto)
```

---

randCyto

*Subset of the 'extdata' data in the 'flowWorkspaceData' package*

---

## Description

A sample dataset containing information about flow cytometry data with two binary conditions and four markers. The data are a random subset of the 'extdata' data in the 'flowWorkspaceData' package found on Bioconductor <http://bioconductor.org/packages/release/data/experiment/html/flowWorkspaceData.html> and formatted for 'gateR' input. The selected markers are arcsinh transformed.

## Usage

```
randCyto
```

## Format

A data frame with 11763 rows and 7 variables:

- id** cell ID number
- g1** binary condition #1
- g2** binary condition #2

**arcsinh\_CD4** arcsinh-transformed CD4  
**arcsinh\_CD38** arcsinh-transformed CD38  
**arcsinh\_CD8** arcsinh-transformed CD8  
**arcsinh\_CD3** arcsinh-transformed CD3

### Source

<https://github.com/Waller-SUSAN/gateR/blob/master/README.md>

### Examples

```
head(randCyto)
```

---

rrs

*A single gate for a single condition*

---

### Description

Estimates a relative risk surface and computes the asymptotic p-value surface for a single gate with a single condition. Includes features for basic visualization. This function is used internally within the `gating` function to extract the points within the significant areas. This function can also be used as a standalone function.

### Usage

```
rrs(
  dat,
  alpha = 0.05,
  p_correct = "none",
  nbc = NULL,
  doplot = FALSE,
  rcols = c("#FF0000", "#cccccc", "#0000FF"),
  lower_lrr = NULL,
  upper_lrr = NULL,
  c1n = NULL,
  win = NULL,
  verbose = FALSE,
  ...
)
```

### Arguments

<code>dat</code>	Input data frame flow cytometry data with four (4) features (columns): 1) ID, 2) Condition A ID, 3) Marker A as x-coordinate, 4) Marker B as y-coordinate.
<code>alpha</code>	Numeric. The two-tailed alpha level for significance threshold (default is 0.05).

<code>p_correct</code>	Character string specifying whether to apply a correction for multiple comparisons including a Bonferroni correction <code>p_correct = "uncorrelated"</code> or a correlated Bonferroni correction <code>p_correct = "correlated"</code> . If <code>p_correct = "none"</code> then no correction is applied.
<code>nbc</code>	Optional. An integer for the number of bins when <code>p_correct = "correlated"</code> . Similar to <code>nbclass</code> argument in <code>correlog</code> . The default is the average number of gridded knots in one-dimension (i.e., x-axis).
<code>doplot</code>	Logical. If TRUE, the output includes basic data visualizations.
<code>rcols</code>	Character string of length three (3) specifying the colors for: 1) group A (numerator), 2) neither, and 3) group B (denominator) designations. The defaults are <code>c("#FF0000", "#cccccc", "#0000FF")</code> or <code>c("red", "grey80", "blue")</code> .
<code>lower_1rr</code>	Optional, numeric. Lower cut-off value for the log relative risk value in the color key (typically a negative value). The default is no limit and the color key will include the minimum value of the log relative risk surface.
<code>upper_1rr</code>	Optional, numeric. Upper cut-off value for the log relative risk value in the color key (typically a positive value). The default is no limit and the color key will include the maximum value of the log relative risk surface.
<code>c1n</code>	Optional, character. The name of the level for the numerator of condition A. The default is null and the first level is treated as the numerator.
<code>win</code>	Optional. Object of class <code>owin</code> for a custom two-dimensional window within which to estimate the surfaces. The default is NULL and calculates a convex hull around the data.
<code>verbose</code>	Logical. If TRUE will print function progress during execution. If FALSE (the default), will not print.
<code>...</code>	Arguments passed to <code>risk</code> to select bandwidth, edge correction, and resolution.

## Details

This function estimates a relative risk surface and computes the asymptotic p-value surface for a single gate and single condition using the `risk` function. Bandwidth is fixed across both layers (numerator and denominator spatial densities). Basic visualization is available if `doplot = TRUE`.

Provides functionality for a correction for multiple testing. If `p_correct = "uncorrelated"`, then a conventional Bonferroni correction is calculated by dividing the alpha level by the number of gridded knots across the estimated surface. The default in the `risk` function is a resolution of 128 x 128 or  $n = 16,384$  knots and a custom resolution can be specified using the `resolution` argument within the `risk` function. If `p_correct = "correlated"` (NOTE: May take a considerable amount of computation resources and time), then a Bonferroni correction that takes into account the spatial correlation of the surface is calculated within the internal `pval_correct` function. The alpha level is divided by the minimum number of knots that are not spatially correlated. The minimum number of knots that are not spatially correlated is computed by counting the knots that are a distance apart that exceeds the minimum distance of non-significant spatial correlation based on a correlogram using the `correlog` function. If `p_correct = "none"`, then the function does not account for multiple testing and uses the uncorrected alpha level. See the internal `pval_correct` function documentation for more details.

The condition variable (Condition A) within `dat` must be of class 'factor' with two levels. The first level is considered the numerator (i.e., "case") value and the second level is considered the denominator (i.e., "control") value. The level can also be specified using the `c1n` parameter.

**Value**

An object of class 'list' where each element is a object of class 'rrs' created by the [risk](#) function with two additional components:

`rr` An object of class 'im' with the relative risk surface.

`f` An object of class 'im' with the spatial density of the numerator.

`g` An object of class 'im' with the spatial density of the denominator.

`P` An object of class 'im' with the asymptotic p-value surface.

`lrr` An object of class 'im' with the log relative risk surface.

`alpha` A numeric value for the alpha level used within the gate.

**Examples**

```
test_rrs <- rrs(dat = randCyto)
```

# Index

## \* datasets

randCyto, 9

## \* package

gateR-package, 2

correlog, 3, 4, 6–8, 11

gateR (gateR-package), 2

gateR-package, 2

gating, 2, 3, 3, 7, 10

image.plot, 3

lotrrs, 3, 7

over, 3

randCyto, 3, 9

raster, 3

rasterToPolygons, 3

risk, 3, 5, 6, 8, 9, 11, 12

rrs, 3, 10

sp, 3

sparr, 3

unionSpatialPolygons, 3