# spatialcompare: Comparing spatial patterns

*Sriram Ganapathi Subramanian and Colin Robertson*

*2017-08-09*

## Area Overlap Statistic

The area overlap statistics include all the statistics described by (Maruca and Jacquez 2002) (https://link.springer.com/article/10.1007/s101090100075). These statistics are probabilistic pattern association tests that are appropriate when edge effects are present, polygon size is heterogeneous, and the number of polygons varies from one classification to another.

### Relative Area Overlap

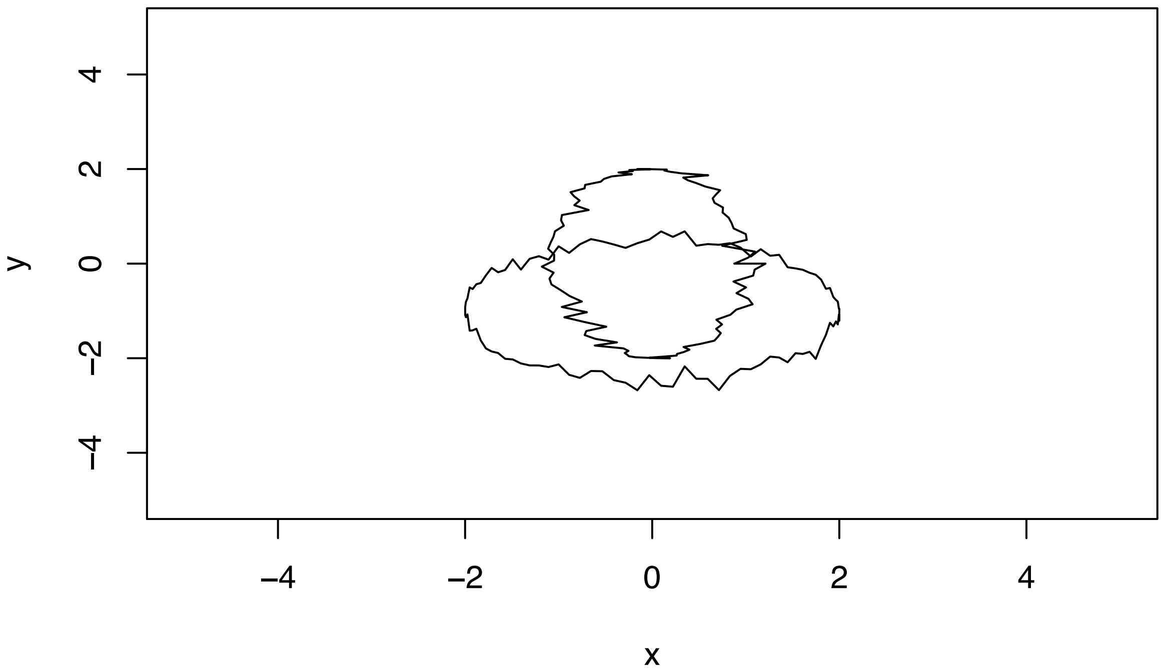
The relative area overlap is area of intersection as a fraction of the area of union of two polygons. For identical polygons the relative area overlap will be 1.

```r
library(gpclib)
library(sp)
```

```
## Warning: package 'sp' was built under R version 3.3.2
```

```r
library(spatialcompare)

set.seed(1234)
theta <- seq(0, 2 * pi, length = (100))
poly1 <- cbind(c(0 + 1 * cos(theta) + rnorm(100, sd = 0.1)), c(0 + 2 * sin(theta)))
poly2 <- cbind(c(0 + 2 * cos(theta)), c(-1 + 1.5 * sin(theta) + rnorm(100, sd = 0.1)))
x = 300
y = 300
plot(x, y, type = "n", xlim = c(-5, 5), ylim = c(-5, 5))
polygon(poly1)
polygon(poly2)
```




```
p1 <- as(poly1, "gpc.poly")
p2 <- as(poly2, "gpc.poly")
relative_area_overlap(p1, p2)
```

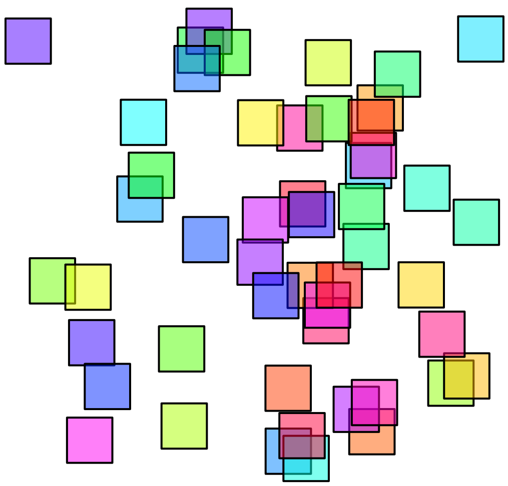
```
## [1] 0.3481681
```

#### Maximum Relative Area Overlap

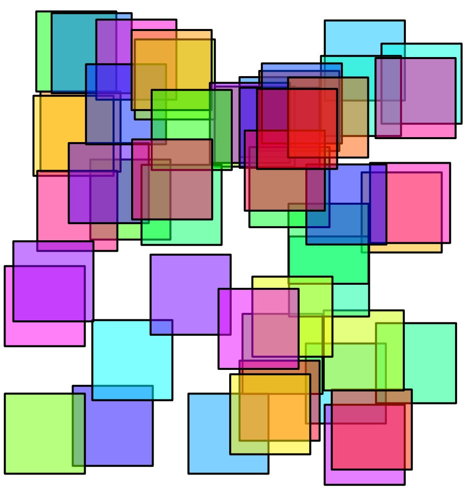
The maximum relative area overlap gives the the relative area overlap for the polygon which has maximum overlap with the candidate polygon. This can be useful when for example, multiple polygons overlap at the edges but a central polygon overlaps a much larger portion. This is common in image classification and segmentation problems and change analysis.

```
square <- t(replicate(50, {
    o <- runif(2)
    c(o, o + c(0, 0.1), o + 0.1, o + c(0.1, 0), o)
}))
square2 <- t(replicate(50, {
    o <- runif(2)
    c(o, o + c(0, 0.2), o + 0.2, o + c(0.2, 0), o)
}))
ID <- paste0("sq", seq_len(nrow(square)))
ID2 <- paste0("sq", seq_len(nrow(square2)))
# Create SP
polys <- SpatialPolygons(mapply(function(poly, id) {
    xy <- matrix(poly, ncol = 2, byrow = TRUE)
    Polygons(list(Polygon(xy)), ID = id)
}, split(square, row(square)), ID))
polys2 <- SpatialPolygons(mapply(function(poly, id) {
    xy <- matrix(poly, ncol = 2, byrow = TRUE)
    Polygons(list(Polygon(xy)), ID = id)
```

```r
}, split(square2, row(square2)), ID2))
# Create SPDF
polys.df <- SpatialPolygonsDataFrame(polys, data.frame(id = ID, row.names = ID))
polys.df2 <- SpatialPolygonsDataFrame(polys2, data.frame(id = ID, row.names = ID))
plot(polys.df, col = rainbow(50, alpha = 0.5))
```



```r
ids <- sapply(polys.df@polygons, function(p) p@ID)
plot(polys.df2, col = rainbow(50, alpha = 0.5))
```



```r
ids2 <- sapply(polys.df2@polygons, function(p) p@ID)

p <- polys2@polygons[[1]]@Polygons[[1]]@coords  #Sample Polygon for maximum relative area Overlap

maximum_rel_area_ovelap(polys, p)
```

```
## [1] 0.2267178
```

### Average Maximum Area Overlap

This function averages the maximum area overlap over the entire dataset. This is called the area overlap statistic in [maruca_area-based_2002].

```r
average_max_rel_area_overlap(polys,polys2)
```

```
## [1] 0.1935569
```

#### Simultaneous Area Overlap

Since overlap can be evaluated two ways (I:J and J:I) and is asymmetric, the simultaneous version gives a general overlap statistic for the two sets of polygons.

```r
simultaneous_area_overlap(polys,polys2)
```

```
## [1] 0.1940974
```

#### Weighted Average Relative Area Overlap

The average maximum relative overlap can also be computed as a weighted average where weights are given by the area of the focus polygon.

```r
weighted_avg_maximum_relative_area_overlap(polys,polys2)
```

```
## [1] 0.25
```

#### Weighted Simultaneous Area Overlap

```r
weighted_simultanous_area_overlap(polys,polys2)
```

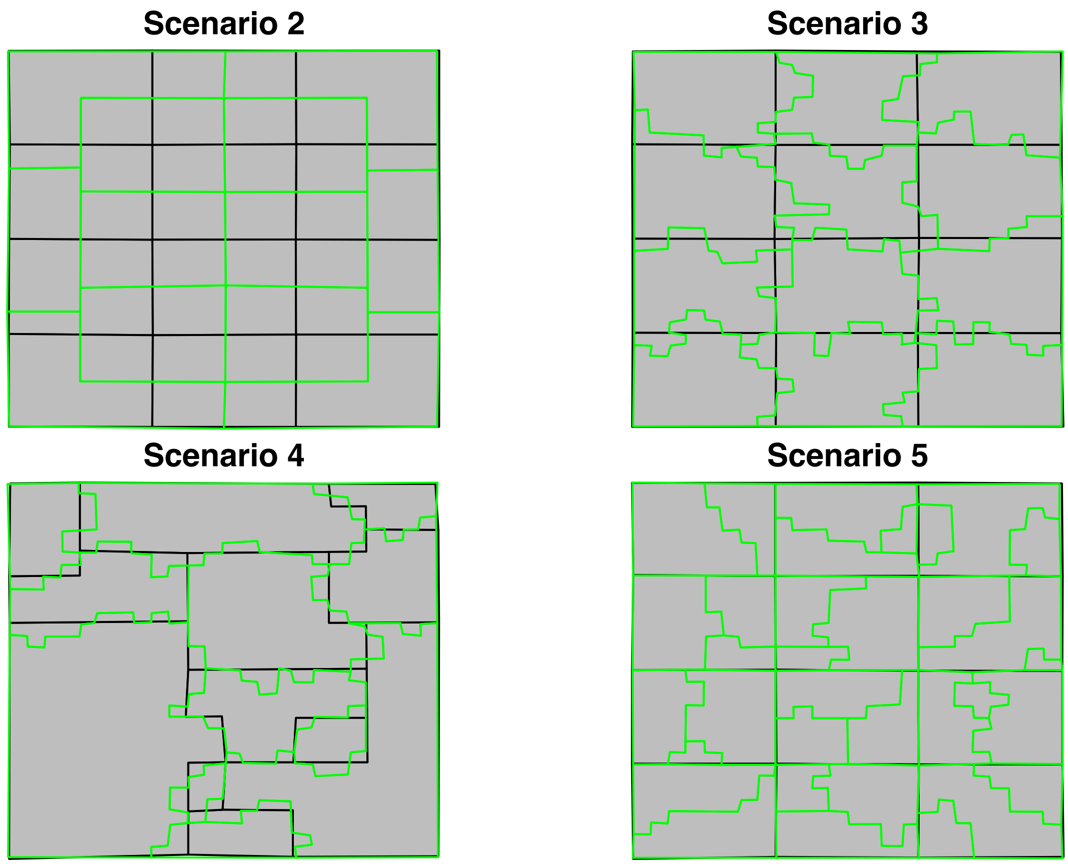
```
## [1] 0.8387082
```

#### Sample Analysis

Data from (Maruca and Jacquez 2002) were re-created to demonstrate metric values for overlap statistics.

Here is some fake data replicating the four offset scenarios (scenarios 2-5) in (Maruca and Jacquez 2002) and a table giving the values of the statistics computing using the spatialcompare package.

```r
par(mar = c(0.001, 0.1, 0.8, 0.1), mfrow = c(2, 2))
plot(subset(scenarios, Scenario == 2 & Poly == 1), col = "grey", main = "Scenario 2")
plot(subset(scenarios, Scenario == 2 & Poly == 2), add = TRUE, border = "green")
plot(subset(scenarios, Scenario == 3 & Poly == 1), col = "grey", main = "Scenario 3")
plot(subset(scenarios, Scenario == 3 & Poly == 2), add = TRUE, border = "green")
plot(subset(scenarios, Scenario == 4 & Poly == 1), col = "grey", main = "Scenario 4")
plot(subset(scenarios, Scenario == 4 & Poly == 2), add = TRUE, border = "green")
plot(subset(scenarios, Scenario == 5 & Poly == 1), col = "grey", main = "Scenario 5")
plot(subset(scenarios, Scenario == 5 & Poly == 2), add = TRUE, border = "green")
```




```
knitr::kable(df)
```

| Scenario | Ai_Value | Aj_Value | Aij_Value |
|---|---|---|---|
| 1. Perfect overlap-unweighted | 1.0000000 | 1.0000000 | 1.0000000 |
| 1. Perfect overlap-weighted | 1.0000000 | 1.0000000 | 1.0000000 |
| 2. Maximally offset-unweighted | 0.2972698 | 0.3172427 | 0.3072563 |
| 2. Maximally offset-weighted | 0.1441215 | 0.5378970 | 0.6579387 |
| 3. Good overlap - polygon areas similar -unweighted | 0.7400701 | 0.7400701 | 0.7400701 |
| 3. Good overlap - polygon areas similar -weighted | 0.7893637 | 0.7893637 | 0.8695256 |
| 4. Good overlap - polygon areas different -unweighted | 0.7593413 | 0.7593413 | 0.7593413 |
| 4. Good overlap - polygon areas different -weighted | 0.7686126 | 0.7686126 | 0.9115579 |
| 5. Overlapping partitions - different spatial scales -unweighted | 0.3473764 | 0.5363995 | 0.3966868 |
| 5. Overlapping partitions - different spatial scales -weighted | 0.6936571 | 0.6936571 | 0.7675578 |

## Quantity Disagreement

The Quantity Disgreement is a statistic introduced in [jr_death_2011] ("http://www.tandfonline.com/doi/abs/10.1080/01431161.2011.552923"). This statistic along with the allocation disagreement statistic are presented as an alternative to the kappa indices in the paper.

Setting up the data

```
A1 = matrix(c(0, 0, 0, 0, 0, 0,0,0,0),nrow=3, ncol=3, byrow = TRUE)
A2 = matrix(c(0, 0, 0, 0, 0, 0,0,0,0),nrow=3, ncol=3, byrow = TRUE)
```

```r
refer = matrix(c(0, 0, 0, 0, 0, 0,1,1,1),nrow=3, ncol=3, byrow = TRUE)
comparison = matrix(c(0, 0, 1, 0, 0, 0,1,0,0),nrow=3, ncol=3, byrow = TRUE)
A3 = matrix(c(0,1,1,0,0,0,1,0,0), nrow = 3, ncol = 3, byrow = TRUE)
A4 = matrix(c(0,0,0,1,0,0,1,1,1),nrow = 3, ncol = 3, byrow = TRUE)
```

Quantity Disagreement

```r
quantity_disagreement(comparison,refer)
```

```
## [1] 1
```

```r
quantity_disagreement(A1,refer)
```

```
## [1] 3
```

```r
quantity_disagreement(A2,refer)
```

```
## [1] 3
```

```r
quantity_disagreement(A3,refer)
```

```
## [1] 0
```

```r
quantity_disagreement(A4,refer)
```

```
## [1] 1
```

## Allocation Disagreement

The Allocation Disgreement is a statistic introduced in the paper "Death to Kappa: birth of quantity disagreement and allocation disagreement for accuracy assessment" ("http://www.tandfonline.com/doi/abs/10.1080/01431161.2011.552923"). This statistic along with the quantity disagreement statistic are presented as an alternative to the kappa indices in the paper.

```r
allocation_disagreement(comparison,refer)
```

```
## [1] 2
```

```r
allocation_disagreement(A1,refer)
```

```
## [1] 0
```

```r
allocation_disagreement(A2,refer)
```

```
## [1] 0
```

```r
allocation_disagreement(A3,refer)
```

```
## [1] 4
```

```r
allocation_disagreement(A4,refer)
```

```
## [1] 0
```

Maruca, S., and G. Jacquez. 2002. "Area-Based Tests for Association Between Spatial Patterns." *Journal of Geographical Systems* 4: 69–83. https://link.springer.com/article/10.1007/s101090100075.