Assess congruence between two vectorial mosaics using vectorial Kappa (κ_v) in R

Vincent Bonhomme
French Institute of Pondicherry

December 13, 2012 - KappaV `0.1`

1 Preliminaries

This vignette intends to be a very short user manual to calculate vectorial Kappa (or κ_v) using the KappaV package in R.

It assumes a basic preliminary knowledge in R language. Hopefully, many ressources are available on the CRAN's homepage[1]. KappaV must of course be installed and loaded.If that's not the case, the two lines below will do the job:

```
install.packages("KappaV") # you only have to install package once
```

```
require(KappaV) # but loading them is mandatory for every new R session
```

Many approaches have been developped to compare two rasters, including the Cohen's Kappa index. A simple approach would be to rasterize the vectorial mosaics but it is neither elegant nor exact.

We[2], as co-authors of the companion paper we submitted in December 2012, extended the Cohen's Kappa index to directly use the vectorial mosaics as input data leading to the filling of a confusion matrix, and then to the κ_v value. The basic idead behind is to intersect polygons of the two mosaics, calculate areas of the resulting polygons and use them to fill the confusion matrix.

2 Calculate κ_v

I will use below two shapefiles that can be downloaded from my website there: `http://www.vincentbonhomme.fr/KappaV/shp.zip`.

This archive contains two shapefiles created with Voronoi tesselation using the DYPAL platform. Be sure to change the paths according to the location of your own shapefiles. We will use the path to the `.shp` files, and the companion files (`.dbf`, `.prj`,etc.). On my systems these paths are `"D:/R-Packages/KappaV/shp/voro1.shp"` and `"D:/R-Packages/KappaV/shp/voro2.shp"`, respectively.

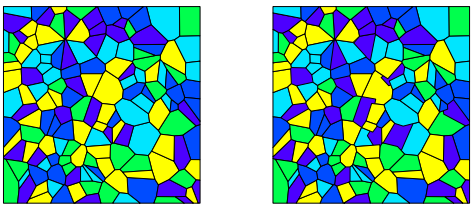

Figure 1: The two "Voronoi" shapes used

You will also need to specify, for each shapefile, which colums in your `.dbf` file contain the polygon IDs and the nominal variable of interest. If these informations are contained in columns which names are `ID` and `OD`, then you do not have to worry about that, these are the default values. For instance, in the example below, you can omit the last line. Then, the κ_v calculation is straightforward:

[1] http://cran.r-project.org/

[2] Myself, Mathieu Castets, Jules Morel and Cédric Gaucherel

```
kv <- KappaV("D:/R-Packages/KappaV/shp/voro1.shp",
             "D:/R-Packages/KappaV/shp/voro2.shp",
             "ID", "ID", "OS", "OS")
kv

$confusion.matrix
       0      1      2      3      4
0 179875      0      0      0      0
1  16283 175642      0      0      0
2  26912      0 209182      0      0
3  11615      0      0 178002      0
4  25541      0      0      0 257196

$kappa.v
   Kappa  Kappa.sd
0 0.9067 0.0003153
```

You can retrieve the confusion matrix (`kv$confusion.matrix`) only, or the value of κ_v (`kv$kappa.v)[1]`) or the associated standard deviation (`kv$confusion.matrix)[2]`).

That's all folks! If you have any trouble with this package, have a look to `?KappaV`, then feel free to contact me: `bonhomme.vincent@gmail.com`.