An adjusted spatial inconsistency index

Jeremy Gelb

02/04/2021

Contents

Introduction

This vignette is a short description of an adjusted version of the spatial inconsistency index proposed in the package **geocmeans**. For a published reference, please consider citing Gelb and Apparicio (2021).
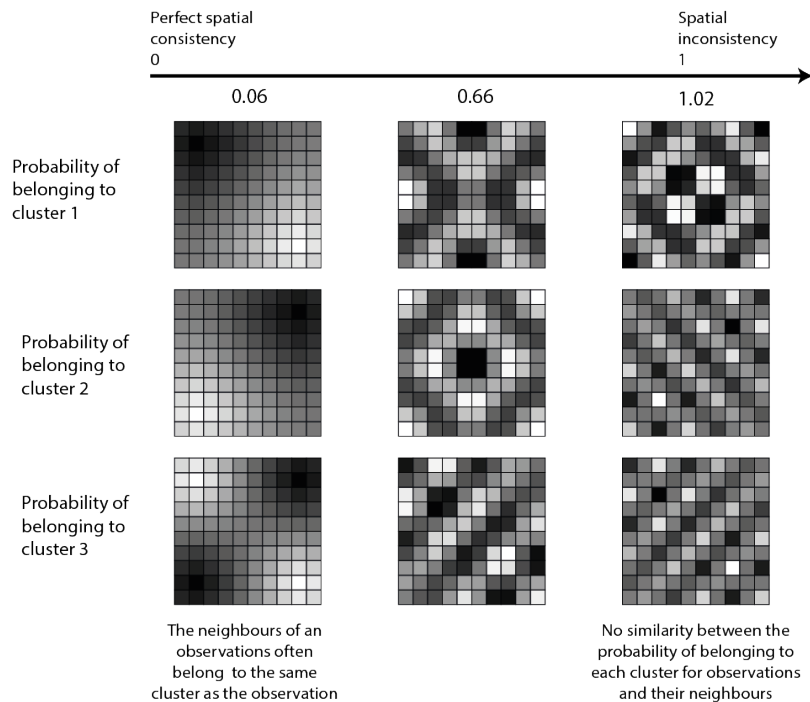
Most of the time, data with a spatial component show an autocorrelated structure: close observations tend to be more similar than distant ones. Classical unsupervised classification algorithms do not account for space and tend to produce scattered results. In other words, some close observations may end up in different clusters despite their differences not being too pronounced. Spatially aware clustering algorithms try to correct this by spatially smoothing the results.

Spatial inconsistency index

We proposed an index S_{cr} to measure the spatial inconsistency of a classification considering its membership matrix u and a spatial weight matrix w. It is simply the sum of the Euclidean distance of each observation k to its spatial neighbours l on the membership matrix, weighted by the spatial wight matrix.

$$S_{cr} = \sum_{k=1}^{k} \sum_{l=1}^{l} w_{kl} \left\| u_k - u_l \right\|^2$$

We also propose to standardize this sum by the total spatial inconsistency under the assumption of total spatial randomness (obtained by repeated permutation). A small value of this index indicates a strong spatial consistency as observations tend to belong to the same clusters as their neighbours. A value close to one indicates that the spatial inconsistency of the classification is close to what one can expect from a situation of total spatial randomness.

Perfect spatial consistency 0 — Spatial inconsistency 1

0.06　　0.66　　1.02

Probability of belonging to cluster 1

Probability of belonging to cluster 2

Probability of belonging to cluster 3

The neighbours of an observations often belong to the same cluster as the observation

No similarity between the probability of belonging to each cluster for observations and their neighbours

It is easy to calculate this index with the package **geocmeans** using the function *spConsistency*

```
library(spdep)
library(geocmeans)
library(ggplot2)

#preparing the data
data("LyonIris")

#selecting the columns for the analysis
AnalysisFields <-c("Lden","NO2","PM25","VegHautPrt","Pct0_14",
                   "Pct_65","Pct_Img","TxChom1564","Pct_brevet","NivVieMed")

#rescaling the columns
Data <- LyonIris@data[AnalysisFields]
for (Col in names(Data)){
  Data[[Col]] <- scale(Data[[Col]])
}

#creating the queen spatial weight matrix
Neighbours <- poly2nb(LyonIris,queen = TRUE)
WMat <- nb2listw(Neighbours,style="W",zero.policy = TRUE)

#applying SFCM algorithm
SFCM <- SFCMeans(Data, WMat, k = 4, m = 1.5, alpha = 0.7,
                tol = 0.0001, standardize = FALSE,
                verbose = FALSE, seed = 456)

#calculating the spatial inconsistency index
consistIndex <- spConsistency(SFCM$Belongings, WMat, nrep = 500)

ggplot() +
```
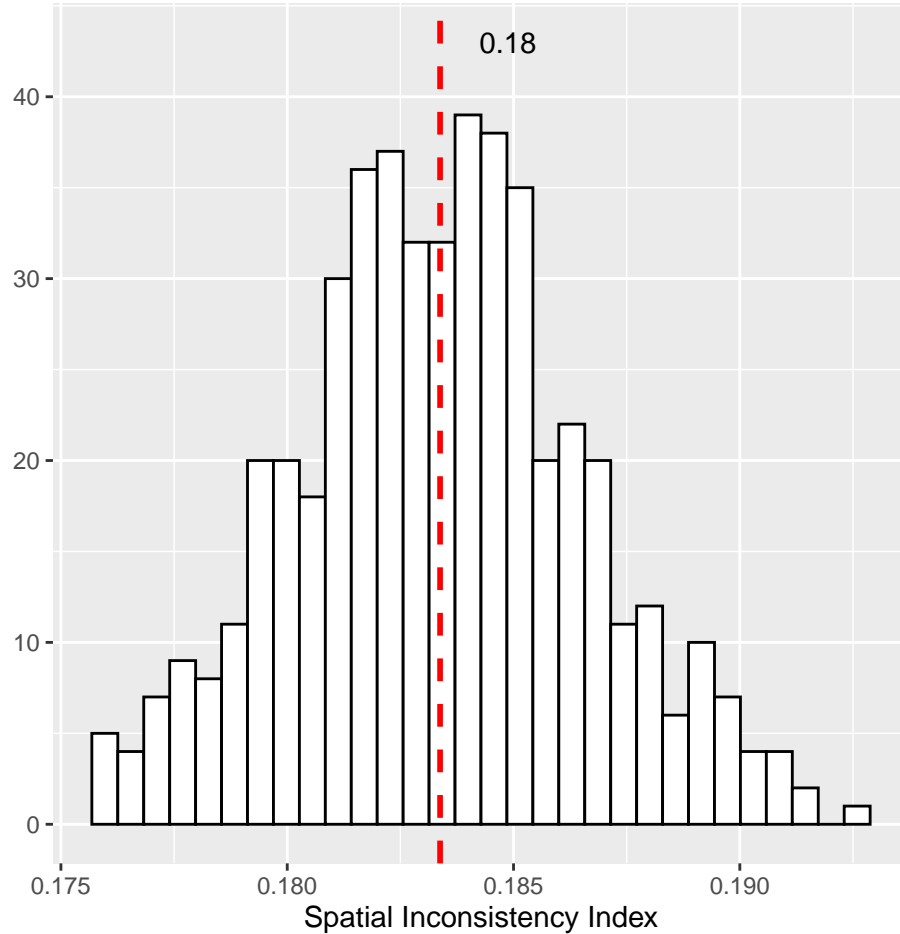
```
geom_histogram(aes(x = consistIndex$samples),
               bins = 30, fill = "white", color = "black") +
geom_vline(aes(xintercept = consistIndex$Mean),
           color = "red", linetype="dashed", size = 1) +
geom_text(aes(x = consistIndex$Mean+0.0015, y = 43,
              label = round(consistIndex$Mean,2))) +
labs(x = "Spatial Inconsistency Index", y = "")
```


So on average, for this classification, the spatial inconsistency index is about 0.18.

Adjusted spatial inconsistency index

The spatial inconsistency index presented above has two drawbacks :

- first, as the number of clusters increases, the index will increase as well because more borders will be needed to delimit them in space.
- second, two close observations should not have to belong to the same group if they are very dissimilar.

To adjust the index, we propose to replace the spatial weight matrix used in the calculus by a semantico-spatial matrix. In this matrix, the spatial weight between two neighbours is replaced by the inverse of the Euclidean distance between their attributes.

$$W_{kl} = \frac{1}{||k - l||^2}$$

Thus, if two neighbours do not belong to the same cluster, but are very dissimilar, this new semantico-spatial weight matrix will limit the weight of their dissimilarity in the spatial inconsistency index. Note, the new spatial matrix must be standardized globally (and not by row) to consider the differences between observations for the full dataset and not only at the local level. This second index is noted S_{mcr}.

It is easy to calculate this adjusted version with the package **geocmeans** by using the function *adjustSpatialWeights*.

```r
WMat2 <- adjustSpatialWeights(Data, WMat$neighbours, style = "C")
consistIndex2 <- spConsistency(SFCM$Belongings, WMat2, nrep = 500)

ggplot() +
  geom_histogram(aes(x = consistIndex2$samples),
                 bins = 30, fill = "white", color = "black") +
  geom_vline(aes(xintercept = consistIndex2$Mean),
             color = "red", linetype="dashed", size = 1) +
  geom_text(aes(x = consistIndex2$Mean+0.0015, y = 43,
                label = round(consistIndex2$Mean,2))) +
  labs(x = "Adjusted Spatial Inconsistency Index", y = "")
```
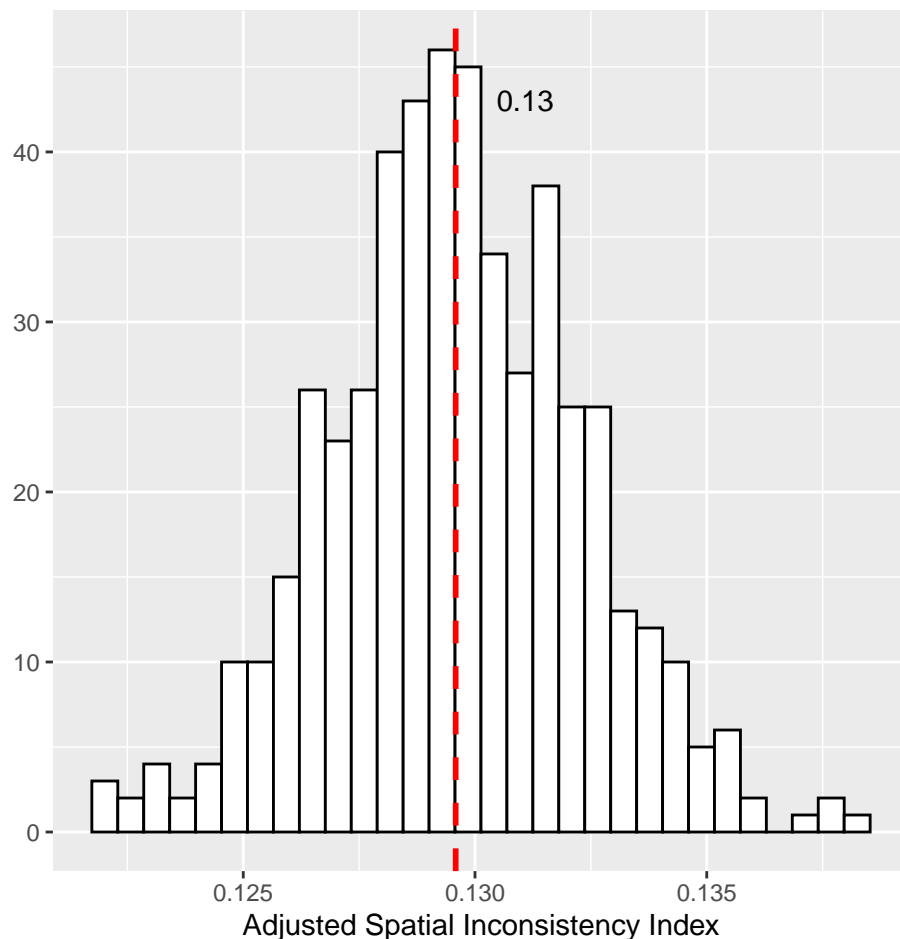
Gelb, Jérémy, and Philippe Apparicio. 2021. "Apport de La Classification Floue c-Means Spatiale En géographie: Essai de Taxinomie Socio-résidentielle Et Environnementale à Lyon." *Cybergeo: European Journal of Geography.*