TESSELLATION
AS AN ALTERNATIVE AGGREGATION METHOD

Kateryna Zabarina (ORCID: 0000-0003-0241-6977)
Faculty of Economic Sciences
University of Warsaw, Poland
e-mail: kzabarina@wne.uw.edu.pl

Abstract: The sensitivity of statistical results to the choice of a particular zoning system is known as the Modifiable Areal Unit Problem (MAUP). Level of aggregation is a significant factor determining results. Moving from point data to areal data, one should take into account aggregation and existence of spatial autocorrelation in the data. Though usually Voronoi tessellation is used in different study fields, this paper suggests it can be an alternative aggregation method to connect point and areal data in economics. Paper shows the possibility to calculate spatial average from point data and vice versa.

Keywords: Voronoi tessellation, point data, regional data, random sample, spatial analysis, aggregation, MAUP, spatial average

JEL classification: C43, C63, C65, R12

INTRODUCTION

The sensitivity of statistical results to the choice of a particular zoning system is known as the Modifiable Areal Unit Problem (MAUP) [Briant et al. 2010]. Because of the MAUP, level of aggregation is a significant factor determining results [Pietrzak and Ziemkiewicz 2016], as it cuts spatial information and introduces errors [Jeffery et al. 2014]. Though spatial data are available either in point form (with geographical coordinates) or aggregated by administrative region (e.g. zip code) [Jeffery et al. 2014], aggregated data are more often used by researchers due to computational constraints [Steśniak and Jacobs-Crisioni, 2017]. To mitigate the MAUP, some advocate the use of methods that do not depend on spatial aggregation [Kwan 1998, Kwan and Weber 2008, Tobler 1989], but these

DOI: 10.22630/MIBE.2018.19.1.8
methods typically require substantial additional data, which are often unavailable [Steptniak and Jacobs-Crisioni 2017].


Firms’ locations are usually represented by a single point, and one can observe that they are correlated, i.e. dependent on each other. Tobler’s first law of geography says: “everything is related to everything else, but nearby things are more related than distant things” [Tobler 1979]. This fact implies spatial autocorrelation for the observations in a geographic space. It means that there is a relation between values monitored at the neighboring locations [Sharifzadeh and Shahabi 2004].

Movement from point data to areal data should take into account also aggregation method. Point data cannot be treated as simple regional data aggregation - aggregation hides the variance of point data and ’cuts-off’ the spatial information\(^1\). Most empirical work in economic geography relies on scattered geocoded data that are aggregated into discrete spatial units, such as cities or regions. However, the aggregation of spatial dots into boxes of different size and shape is not benign regarding statistical inference [Briant et al. 2010].

Point measures are usually obtained from continuous surface, that’s why one need to estimate the whole surface in order to get them. Aggregation from points to regions is needed, when value of some regional index has to be calculated. When spatial units do not have the same shape, averaging is less sensitive to changes in size than summation [Briant et al. 2010]. It can nonetheless be argued that administrative boundaries do not capture the essence of economic phenomena that often spill over boundaries [Briant et al. 2010].

Term ‘spatial average’ firstly appeared in the work of Weaver\(^2\). Later, the spatial averaging theorem was presented independently in 1967 by Anderson and

---

\(^1\) Regional aggregates do not reflect spatial location, spatial distributions etc.

Jackson [Anderson and Jackson, 1967], Slattery [Slattery 1976] and Whitaker [Whitaker, 1976] [all authors were mentioned in Hower, Whitaker 1985]. Problem of spatial average was widely discussed (among others) by Goodchild [Goodchild 1979]⁴ and Hower and Whitaker [Hower and Whitaker 1985]⁴. It became a popular tool in geography (especially hydrology⁵) and sensor network[Sharifzadeh and Shahabi 2004] studies. In both cases, area of Voronoi tessellation cell was used as weighting coefficient and it was proved that this approach performs better than arithmetical mean and captures stochasticity of data.

To obtain more precise results, it is better to used weighting coefficients showing relative area of smaller administrative unit to bigger one (which includes that smaller unit). But taking into account fact, that even in the smallest administrative unit there could be more than one firm, calculation of some aggregated index could be biased. As Voronoi tessellation covers all the plane without no overlaps and no gaps, and there is only one point in each Voronoi cell, it can be considered as a division of space, which captures relation between values observed in the neighbourhood⁶. Despite the fact that usually Voronoi tessellation is used in cellular biology, image compression or resources distribution studies, it can be an alternative method which connects point data and areal data in economics.

Current research suggests, that in order to obtain proper results, area of Voronoi cell is divided by the sum of all Voronoi cells in the certain region⁷. Such weighting coefficient can be used when calculating spatial average from point data or vice versa – having spatial average, it is possible to set separate values for points, using weighting coefficient. Using firm location data from Slaskie voivodeship, two abovementioned cases are described. Method will be used for calculating economic indexes determining firm location. All calculations are made using R software⁸ using randomly generated values⁹.

---

⁵ Precipitation spatial average. On-line access: https://earth.boisestate.edu/drycreek/education/hypsometric/
⁶ As it comes from Voronoi tessellation properties, each cell contains exactly one point and every other location in a given cell is closer to its corresponding point than to any other point (Sharifzadeh and Shahabi, 2004).
⁷ Dirichlet tessellation is computed exactly by the Lee-Schachter algorithm. When switching from bigger unit to smaller (for example, from voivodeship to powiat), area of tessellation cell is calculated with respect to the bigger unit and not to smaller one, so it makes sense to use a proportion.
⁸ Except from standard packages for spatial data and task view “Spatial”, packages deldir, spatstat, qleVisualize were used to calculation and visualization.
Tessellations were known for people since ancient times. The term comes from Latin ‘tessella’ – small square, (or Greek ‘τέσσερα’ - four) and means small cubical piece of clay, stone or glass used to make mosaics. But nowadays some new forms (such as Penrose tiling) are widely used [Boots et al. 1999].

One can define tessellation of d-dimensional Euclidean space, \( \mathbb{R}^d \), either as subdivision into d-dimensional, non-overlapping regions or a set of d-dimensional regions which cover space without overlaps and gaps. Formal definition of a tessellation is the following.

Let \( S \) be a closed subset of \( \mathbb{R}^d \), \( \mathcal{S} = \{ s_1, ..., s_n \} \), where \( s_i \) is the closed subset of \( S \) and \( s'_i \) is the interior of \( s_i \). If the elements of \( \mathcal{S} \) satisfy

\[
\begin{align*}
&s'_i \cap s'_j = \emptyset, \text{for } i \neq j, \\
&\bigcup_{i=1}^n s_i = S
\end{align*}
\]

then the set \( \mathcal{S} \) is called a tessellation of \( S \). First property means that the interiors of the elements of \( \mathcal{S} \) are disjoint, and second property means that collectively elements of \( \mathcal{S} \) fill the space [Boots et al. 1999]. In more simple way of explanation, spatial tessellation is a set of regions that are collectively exhaustive and mutually exclusive except for the boundaries.

Planar tessellations are composed of three elements of d (d \( \leq \) 2) dimensions: cells (2-d), edges (1-d), and vertices (0-d). In GIS these elements are usually referred to as polygons, lines (or arcs), and points respectively. All tessellations should be divided in three groups: uniform (or Archimedean) (consist of regular polygons\(^{11}\), which are not isohedral\(^{12}\)), regular (the three uniform tessellations which are also isohedral (i.e. those consisting of regular triangles, squares, or hexagons)) and irregular (normal\(^{13}\) tessellations consisting of only convex cells) [Boots et al. 1999]. The very best example of irregular tessellation is Voronoi tessellation\(^{14}\).

Generic definition of the Voronoi diagram is the following. Let \( S \) denote a set of \( n \) points (called sites) in the plane. Since the regions are coming from

---

\(^{9}\) There was no possibility to obtain any real numeric data except number of employees. However, distribution of employees was discrete with value 5 in around 80\% of observations.

\(^{10}\) When d=2, tessellation is called a planar tessellation.

\(^{11}\) A regular polygon is one with equal side length and equal internal angles.

\(^{12}\) If \( r_i \) denotes the number of edges meeting at the \( i^{th} \) corner of a cell in a monohedral tessellation, an isohedral tessellation is one in which the ordered sequence of values of \( r_i \) is the same for every cell.

\(^{13}\) A d-dimensional tessellation in which every s-dimensional element lies in the boundaries of (d-s+1) cells (0 \( \leq s \leq d-1 \)).

\(^{14}\) The same as Voronoi diagram, Voronoi decomposition, Voronoi partition, Dirichlet tessellation, Thiessen polygons.
intersecting \( n - 1 \) half planes, they are convex polygons. Thus, the boundary of a region consists of at most \( n - 1 \) edges (maximal open straight-line segments) and vertices (their endpoints). Each point on an edge is equidistant from exactly two sites, and each vertex is equidistant from at least three. As a consequence, the regions are edge to edge and vertex to vertex, that is to say, they form a polygonal partition of the plane. This partition is called the Voronoi diagram, \( V(S) \), of the finite point-set \( S \) (Figure 1) [Aurenhammer 1991].

Figure 1. Voronoi diagram for eight sites in the plane


Current study bases on Voronoi tessellation, because it allows to cover all plane instead of regular tessellation. Thiessen polygons have wide usage: in natural sciences (cellular biology [Bock et al. 2009, Saribudak et al. 2016], hydrodynamics [Springel 2010], physics [Kasim 2017]), medicine [Sanchez-Gutierrez et al. 2016], image compression [Du et al. 2006], resource distribution studies [Liu et al. 2009]. In geographical analysis, it may represent administrative units, census tracts, postal zones, or electoral and school districts [Sadahiro 2010]. In this study Voronoi tessellation is suggested to be an instrument measuring different economic indexes.

DATASET AND PRELIMINARY STAGE OF PROPOSED APPROACH

Usually spatial data are presented as areal (one value for administrative unit) or point data (in case of firms, for example). Using weighting coefficients, one can split value of areal index between each single point or vice versa – calculate spatial average using data from each single one.

To conduct the study data from Slaskie voivodeship\(^{15}\) were used. Data present 504917 firms, with information about their location (address, geographical

\(^{15}\) NUTS2 unit.
coordinates, administrative unit), economic sector, number of employees. However, for the purpose of study we generate normally distributed data\textsuperscript{16}.

In case of Poland, sometimes data are presented on powiat\textsuperscript{17} level. But when plotting tessellation and powiat structure together, one can observe difficulty in understanding which tessellation cell belongs to which powiat – administrative boundaries do not capture the essence of economic phenomena that often spill over boundaries [Briant et al. 2010]. Moreover, administrative units have a tendency to change their boundaries (in contrast to tessellation cells for the same sample). Figure 2 presents randomly selected 100 firms\textsuperscript{18} from Slaskie voivodeship (square dots), with Voronoi tessellation based on that point (thin lines) and powiats (thick lines).

Figure 2. Tessellation and powiat structure of Slaskie voivodeship, plotted together

\textbf{Tessellation, spsample = Random}

Source: own calculations

Proposed approach is the following. We first choose random sample of 1000 points. After it, we do a tessellation (Figure 3) in order to obtain Voronoi cell area\textsuperscript{19}, which will be used in further calculations.

\textsuperscript{16} Data were generated using normal distribution with mean=50 and standard deviation=15
\textsuperscript{17} Smaller unit of voivodeship; there are 380 powiats in Poland.
\textsuperscript{18} 100 firms from abovementioned dataset; hereafter ‘points’.
\textsuperscript{19} Also known as Dirichlet weight or Dirichlet area; in R ‘dirichletWeights()’ and ‘tile.areas()’ give the same results.
As it was stated in the introduction, this study presents two different methods of aggregation – calculating spatial average from point data and calculating point values having spatial average. Detailed description is in section below.

**PROPOSED APPROACH**

Proposed approach is described on the Slaskie voivodeship in Poland. Using this method, it is possible to make calculations for administrative units of different size and shape. Two different aspects of the proposed method are described below.

**Calculating spatial average from point data**

R software allows user to create a point pattern inside specified window and set values to each point\(^\text{20}\). Figure 4 shows initial state for this part of approach – point pattern with single value for each single point, where bigger circle size means bigger value in a certain point.

---

\(^{20}\) Known in R as ‘marks’. It is possible to create multiple marks, however, in this study for simplicity only one mark dimension was used.
The next step is very easy – using weighting coefficient (calculated as area of certain Voronoi cell over sum of all Voronoi cells’ areas), we can calculate spatial average as a sumproduct of weights and values (code in R is given below):

```r

d_weights <- dirichletWeights(X.rand, exact=TRUE) #X.rand – is a point pattern for which we calculate Voronoi cells areas
suma <- 0 # initial value of spatial average
for (i in 1:length(X.rand$marks)){
  suma[i] <- X.rand$marks[i]*(d_weights[i]/sum(d_weights))
} #sumproduct of value and weighting coefficient
sum(suma) # summing up all sumproducts
```

After one iteration on 1000 randomly chosen points, spatial average (52.35) is bigger than arithmetic mean (50.76). However, arithmetic mean doesn’t take into account spatial nature of data. Also, it produces more ‘flat’ value, especially in samples with huge outliers.

After simulation of 500 iterations (sampling 1000 points in each, and then calculating arithmetic mean and spatial average) we clearly see that in all cases spatial average exceeds arithmetical mean (see figure 5). Comparison of obtained values shows, that arithmetical mean is ‘flatter’, it doesn’t characterize the real situation and doesn’t take into account spatial nature of data.
Let’s show on the simple example why location matters. Consider a sample of 10 points, for which we calculate both arithmetic and spatial average. On the next step, we change location of one of the points to a different random (only location, so value of parameter in an old point is set to the new point), and it can be observed, that location matters, so spatial average changed, but arithmetic didn’t. Figure 6a shows the values of both indexes in the first step, 6b – values of them after changing location of 1 point\textsuperscript{21}.

**Figure 6.** Comparison of measures in case when location of one data point is different

\begin{align*}
\text{Spatial average} &= 54,989 \\
\text{Arithmetic average} &= 52,4336 \\
\text{(a)}
\end{align*}

\begin{align*}
\text{Spatial average} &= 46,13369 \\
\text{Arithmetic average} &= 52,4336 \\
\text{(b)}
\end{align*}

Source: own calculations

\textsuperscript{21} For the purpose of experiment, only location and then tessellation is shown. Plotting points’ value results in a mess on the graph.
As it is possible to calculate spatial average for big unit (as voivodeship), it is possible to do the same for smaller units (for example, powiats).

**Calculating point values having areal index**

This subsection describes how to calculate values for point data having spatial average. We still consider the same point pattern, however, without marks. Assume that there is some spatial average given. Having Voronoi cells areas, in the loop we set the new values for each point (code in R is given below):

```r
marks(X.rand) <- 0 # setting values to zero
some_index <- 95 # assumed value of spatial average
new_marks <- matrix(0,1000,1)
d_weights <- dirichletWeights(X.rand, exact=TRUE)
for (i in 1:length(X.rand$marks)){
  new_marks[i] <- (d_weights[i]/sum(d_weights))*some_index
}
marks(X.rand) <- new_marks
```

Arithmetical mean approach in this case resulted in value of 0.095 for each single point, which doesn’t reflect spatial features of sample – points’ location and neighbourhood structure, in contrary to spatial average approach. Figure 7 shows the result of above described procedure.

Figure 7. Point pattern for calculating point values from spatial average (second aspect of approach) – results
As it is possible to calculate point values from spatial average for one big unit, it is possible to do the same for other big unit as well, as for small ones.

CONCLUSIONS

This paper studies Voronoi tessellation from a different perspective – as an alternative aggregation method which can connect point data and areal data. It is shown that having Voronoi cell area, it is possible to calculate a weighting coefficient (as ratio of one Voronoi cell area to sum of all areas) and use it for computing spatial average from point data or vice versa.

Study results show, that proposed method performs better in terms of capturing spatial features of the data. Arithmetic mean flats the value of sample average, however spatial approach allows to set proper values for points, taking into account their location.

The results of the study will be helpful in economic studies for calculating different economic indexes. Furthermore, proposed approach could be useful in firm location studies – having some local economic index as a feature of region, it could be splitted between firms and used as an explanatory feature.

REFERENCES


Tessellation as an Alternative Aggregation Method


Openshaw S., Taylor P. J. (1979) A Million or so Correlation Coefficients: Three Experiments on the MAUP. In Wrigley N (ed), Statistical Applications in the Spatial Sciences, London Pion, 127-144.


Tessellation as an Alternative Aggregation Method


https://earth.boisestate.edu/drycreek/education/hypsometric/
https://www.slideshare.net/jluchette/gis-spatially-weighted-averages
http://www.mcs.anl.gov/research/projects/mct/mct_APIs/node15.html