



**HAL**  
open science

# Improvement of Spatial Autocorrelation, Kernel Estimation, and Modeling Methods by Spatial Standardization on Distance

Marc Souris, Florent Demoraes

► **To cite this version:**

Marc Souris, Florent Demoraes. Improvement of Spatial Autocorrelation, Kernel Estimation, and Modeling Methods by Spatial Standardization on Distance. *ISPRS International Journal of Geo-Information*, 2019, 8 (4), pp.199. 10.3390/ijgi8040199 . hal-02110138

**HAL Id: hal-02110138**

**<https://hal.science/hal-02110138>**

Submitted on 25 Apr 2019

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

Article

# Improvement of Spatial Autocorrelation, Kernel Estimation, and Modeling Methods by Spatial Standardization on Distance

Marc Souris <sup>1,2,\*</sup> and Florent Demoraes <sup>3</sup>

<sup>1</sup> UMR Unité des Virus Emergents (UVE: Aix-Marseille Univ—IRD 190—Inserm 1207—IHU Méditerranée Infection), 13005 Marseille, France

<sup>2</sup> RS&GIS FoS, School of Engineering and Technology, Asian Institute of Technology, P.O. Box 4, Klong Luang, Pathumthani 12120, Thailand

<sup>3</sup> Univ Rennes, CNRS, ESO—UMR 6590, F-35000 Rennes, France; florent.demoraes@univ-rennes2.fr

\* Correspondence: marc.souris@ird.fr

Received: 5 March 2019; Accepted: 22 April 2019; Published: 24 April 2019

**Abstract:** In a point set in dimension superior to 1, the statistical distribution of the number of pairs of points as a function of distance between the points of the pair is not uniform. This distribution is not considered in a large number of classic methods based on spatially weighted means used in spatial analysis, such as spatial autocorrelation indices, kernel interpolation methods, or spatial modeling methods (autoregressive, or geographically weighted). It has a direct impact on the calculations and the results of indices and estimations, and by not taking into account this distribution of the distances, spatial analysis calculations can be biased. In this article, we introduce a "spatial standardization", which corrects and adjusts the calculations with respect to the distribution of point pairs distances. As an example, we apply this correction to the calculation of spatial autocorrelation indices (Moran and Geary indices) and to trend surface calculation (by spatial kernel interpolation) on the results of the 2017 French presidential election.

**Keywords:** spatial analysis; spatial autocorrelation; spatial modeling; spatial kernel interpolation; standardization; SD-correction

---

## 1. Introduction

Many natural or anthropogenic phenomena show what is called "spatial dependence". When there is spatial dependence, different values of a spatially located variable  $X$  related to a phenomenon are not independent of each other, which means that two close values are more likely to resemble each other than two distant values. Thus, in many phenomena, variance increases as a function of distance. This spatial dependence is considered as the first law of geography—"everything is related to everything else, but near things are more related than distant things" [1]—and can be applied to many scientific fields, such as geology, botany, forestry, economics, epidemiology, meteorology, etc.

The concept of spatial autocorrelation represents the spatial dependence between numerical values of a spatially located variable. It allows the correlation between the values  $X_i$  of objects  $P_i$  to be formulated according to metric or topological relationships between objects:

"Given a set of  $n$  geographical units, the relationship observed for the  $n(n - 1)/2$  pairs of units between the differences in the values of a variable measured at these locations and a measure of geographical proximity is called spatial autocorrelation" [2–3].

To estimate and quantify this spatial autocorrelation, numerical indices similar to conventional correlation index in dimension 1 are used: a spatial autocorrelation index is a statistical measure of correlations between the values of spatially located objects using metric or topological relationships between these objects.

Testing the statistical significance of spatial autocorrelation is the most effective way to show the existence of spatial dependence. Autocorrelation indices allow to study the global or local spatial clustering or dispersion of values and to measure the role of adjacency or distance in spatial interactions. A spatial autocorrelation index can either be a global average measure or a localized measure involving only the vicinity of a place [4–7].

Many spatial autocorrelation indices have been developed over the last 70 years [2–3,8]. These indices are constructed from the geometric or topological relationships between pairwise objects  $(P_i, P_j)$  on the one hand; and the difference in the values  $X_i, X_j$  on the other hand. To calculate a numerical index, the geometric or topological relationship between pairwise objects must reflect the spatial dependence and be transformed into a numerical value. To do this, several options are possible, depending on how we want to take into account the spatial dependence [9–12]:

- By taking into account neighborhood relations. We can use the direct neighborhood between the two points—or centroids in case of polygons—of the pair (in the Voronoi sense) by assigning 1 if the two points are neighbors, 0 if they are not. When focusing on adjacency or adjacency relationships, the length of the common edge between objects can be used, either Voronoi tessellation in the case of a point pattern, or the length of the boundary in the case of adjacent polygons.
- By taking into account the distance between objects (represented by points or centroids  $P_i$ ). A distance function is used (Euclidean distance, Manhattan distance, distance along a valued network, etc.). The distance is often limited to a maximum distance ( $dmax$ ), called bandwidth, beyond which the value is 0, meaning that there is no spatial dependence beyond this distance.

This function can be polynomial, for example,  $\max\left(0, 1 - \frac{d(P_i, P_j)^k}{dmax^k}\right)$  with  $k = \frac{1}{2}, 1, 2, \dots$ ; Gaussian, for example,  $\exp(-d(P_i, P_j)^2/dmax^2)$ ; sigmoid, etc. The maximum distance ( $dmax$ ) can be set for all pairs or it can be dependent on a density related parameter. For example,  $dmax$  can depend on the distance to the  $n$ -closest adjacent point to one of the points of the pair. It can be estimated by the range of the semi-variogram corresponding to the situation to be analyzed.

These values corresponding to the pairs of objects  $(P_i, P_j)$  are called spatial weights  $w_{ij}$ . They are often assembled in a matrix  $W$ , positive with a null diagonal, and symmetric if  $\forall i, j \ w_{ij} = w_{ji}$  (when spatial relationships are symmetrical). Spatial weights are fundamental in spatial autocorrelation calculations because they express, numerically, spatial dependence. When neighborhood relationships are considered, we talk about a matrix of contiguity. When distances are used, we talk about a matrix of distances.

Most of the autocorrelation indices are weighted means of all possible object pairs and are derived from the index described by Mantel [13]. They; therefore, implicitly assume that the phenomenon is stationary, which means that it corresponds to a global process and does not depend on the location.

The most commonly used is the Moran index [2–4]. It is defined as the mean of the products of the normalized values of points weighted by the spatial weight for the pair. The Moran index corresponds to a classical correlation index—Pearson index—extended to neighboring objects, and provided with the spatial weight  $W$ . It; thus, uses the following multiplicative model:

$$I_{Moran} = \frac{1}{S} \sum_{i,j} w_{ij} \left( \frac{X_i - m}{\sigma} \right) \left( \frac{X_j - m}{\sigma} \right) \quad (1)$$

where  $m$  is the mean of the  $X_i$  values of all objects,  $\sigma$  the standard deviation of the  $X_i$ ,  $w_{ij}$  the spatial weight of the point pair  $(P_i, P_j)$ , and  $S$  the sum of the spatial weights ( $S = \sum_{i,j} w_{ij}$ ).

The expected value of Moran's index under the null hypothesis of no spatial autocorrelation is:

$$E(I_{Moran}) = \frac{-1}{N-1} \quad (2)$$

where  $N$  is the sample size. Variance under the null hypothesis of no spatial autocorrelation depends on  $W$  and is given in [9].

In the literature, this index is often presented with the equivalent but less meaningful formula:

$$I_{Moran} = \frac{N}{S} \sum_i \sum_j w_{ij} (X_i - m)(X_j - m) / \sum_i (X_i - m)^2 \quad (3)$$

Another widely used spatial autocorrelation index, the Geary index [2,3,5], is built upon an additive rather than a multiplicative model. It is defined as the average of the squares of the differences in the normalized values of the pairs of points:

$$I_{Geary} = \frac{1}{S} \sum_{i,j} w_{ij} \left( \frac{X_i - m}{\sigma} - \frac{X_j - m}{\sigma} \right)^2 \quad (4)$$

Other global spatial autocorrelation indices do not use normalized variables, and are, therefore, less common: Black Black Seal, Black White Join, Knox [2].

Finally, some indices are used to estimate the local autocorrelation at a point  $P_i$  and are known as Local Indicator of Spatial Association (LISA) [6]. An example of these is the local Moran index:

$$I_{Moran}(P_i) = \frac{1}{S_i} \frac{(X_i - m)}{\sigma} \sum_{j,j \neq i} w_{ij} \left( \frac{X_j - m}{\sigma} \right), \quad \text{with } S_i = \sum_{j,j \neq i} w_{ij} \quad (5)$$

Another example of LISA index is the Getis–Ord index [14], which is constructed as the local Moran index, but in such a way as to make it a Z-score (i.e., number of standard deviations by which the value is above the mean, and frequently used to compare the value to a standard normal deviate) [2–3].

Other spatial analysis methods are based on an mean or sum of weighted values with weights being a function of distance. These include spatial kernel estimation and statistical modeling processes that take into account spatial autocorrelation or neighboring values, such as simultaneous autoregressive regression models (SAR), conditional autoregressive regression models (CAR), and geographically weighted regression models (GWR) [2–3,8,12, 15]):

- Spatial kernel estimation (Kernel estimation and Kernel Density estimation) extends to dimension 2 of the principles of classic one-dimensional kernel estimation. When the variable is numerical, the spatial interpolation by kernel calculates, at each point of a grid, the average of the values weighted by a function (referred to as kernel) of the distance to the grid point for all objects located at a distance lower than a given bandwidth distance  $dmax$  [16]. For example, commonly used kernel functions are linear function (e.g.,  $(dmax - d)/dmax$ ), quadratic function (e.g.,  $\left(\frac{dmax-d}{dmax}\right)^2$ ), or a Gaussian function (e.g.,  $\frac{1}{\sqrt{2\pi}} e^{-1/2\left(\frac{dmax-d}{dmax}\right)^2}$ ). When the variable is qualitative, the estimation of densities per kernel (kernel density estimation) consists of calculating, for each point of a grid, the weighted number of the objects located at a distance lower than a given distance  $dmax$ , each object being weighted by the kernel.
- Autoregressive spatial models (Autoregressive Regression, Simultaneous Autoregressive Regression, Conditional Autoregressive Regression, Generalized Additive Model, Structured Additive Regression) also use a spatial weight matrix constructed as for spatial autocorrelation indices [2–3][16–18]. For example, for autoregressive regressions, we have:

$$z_j = \sum_k x_{jk} \beta_k + \rho \sum_{i,i \neq j} w_{ij} z_i + \varepsilon_j \quad (Z = X\beta + \rho WZ + \varepsilon)$$

where  $z_j$  is the dependent variable at point  $P_j$ ,  $x_{jk}$  the independent variables at point  $P_j$ ,  $w_{ij}$  a spatial weight depending on the distance between points  $P_i$  and  $P_j$ , and  $\rho$  a parameter in the model reflecting the force and nature (attraction or repulsion) of spatial dependence. Spatial weights are

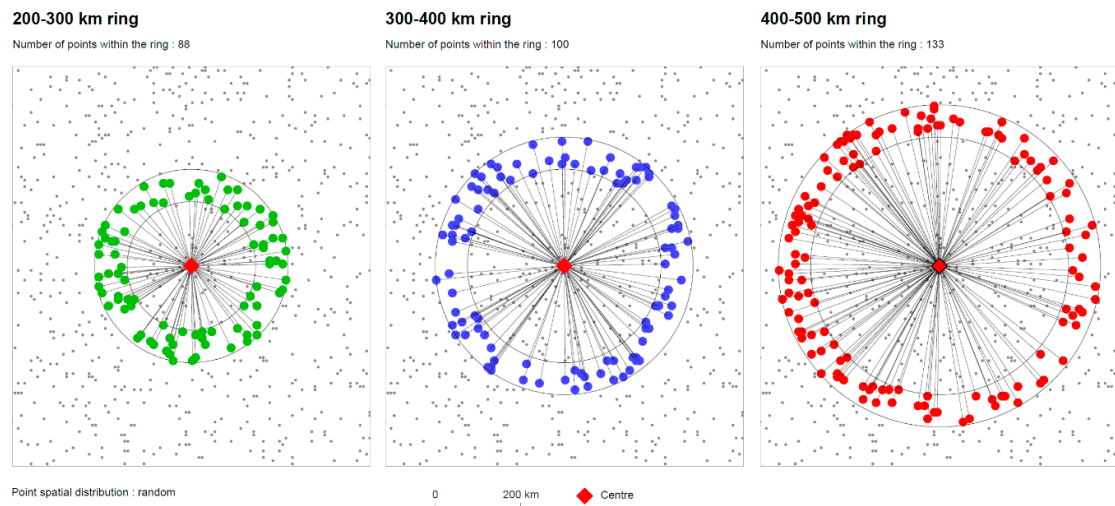
sometimes normalized to give relative and not absolute weights to the neighbors of a point  $P_j$ ; for all individuals  $P_j$ , the sum  $\sum_i w_{ij}$  of the weights of all its neighbors must then be equal to 1, and  $\sum_i w_{ij} z_i$  is just a weighted mean. In this case the matrix  $W$  is said to be standardized on the rows. If all weights are equal, this means adding the mean of the neighboring values to the model. The weights can also have an absolute influence. In this case, the more neighbors close to  $z_j$ , the higher the value of  $\sum_i w_{ij} z_i$  is. A weighted sum of the neighbors' values is added to the model and not a weighted average.

- The geographically weighted regression (GWR) models also use a spatial weight matrix. Here, the model's coefficients  $\beta$  are allowed to vary according to the location, in order to adapt the model locally to local spatial variations; these models aim at estimating regression parameters locally.
- Standardization on the rows of the distance matrix  $W$  (each weight being divided by the sum of the weights of its row) can also be used in the calculation of the Moran or Geary indices, which is equivalent to taking as an overall index the arithmetic mean of the local indices.

## 2. The Need of a Spatial Standardization

In any set of points, the distances between all pairs of points are not evenly distributed. In general, the number of pairs of points increases with distance, until it reaches a maximum and then decreases. This statistical distribution of distances between pair of points (onwards referred to as “inter-distances”) depends on the spatial distribution of the points.

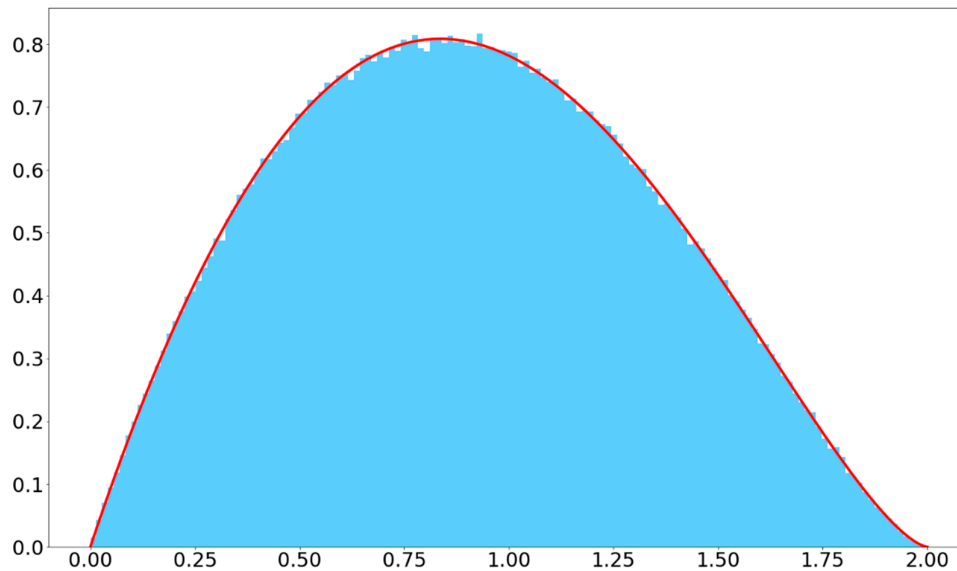
For example, when the points are independently and uniformly distributed in a disc  $D$  of radius  $R$  in dimension 2, the number of points in a radius ring  $[R, R + d]$  increases linearly with this radius following the surface of the ring, unlike in dimension 1 where it remains constant (Figure 1):



**Figure 1.** Number of points in rings of increasing radius and same width, for points independently and uniformly distributed in a 2D space.

The distribution of the distances between all the point pairs in a disk  $D$  (inter-distances) is expressed by the following density function [18, 19] (Figure 2):

$$\forall d \in ]0; 2R], \quad f(d) = \frac{4d}{\pi R^2} \left( \arccos\left(\frac{d}{2R}\right) - \frac{d}{2R} \sqrt{1 - \left(\frac{d}{2R}\right)^2} \right) \quad (6)$$



**Figure 2.** Distribution of inter-distances inside the unit circle ( $R=1$ ) for independently and uniformly distributed set of points. In red, the curve for the theoretical probability density function; in light blue, a simulated distribution from values generated by a homogeneous Poisson model with density  $\rho = 1500$  inside the unit circle.

Yet, the simple observation that the distribution of inter-distances is not evenly distributed is not considered in common methods of spatial analysis, which use sum or mean of weighted values and distances to calculate spatial weights for characterizing and analyzing spatial dependence. Since the distribution of inter-distances is not uniform, calculations based on sums or means over all pairs of points favor the values of the pairs of points associated with the most frequent inter-distances, whereas spatial dependence should only be characterized as a function of distance. The influence of the distribution of the inter-distances considered in the calculation must therefore be eliminated in order to better capture and only measure spatial dependence. The spatial weight  $W$  used in the calculation does not address this problem, as it is constructed to model the spatial dependence and not the fact that some inter-distances are systematically more represented than others in the calculation of the index or estimate when this calculation is based on a sum.

Most of the time, the bandwidth  $d_{max}$  used in the calculation of the spatial weight is lower than the distance for which the number of inter-distances reaches a maximum. In this case, the number of inter-distances is increasing from 0 to  $d_{max}$ , and it is very likely that the influence of spatial weight in the calculation (which in general favors short distances) is cancelled out by not taking into account the distribution of inter-distances.

Therefore, in this article we propose an improvement for spatial analysis methods that use spatial weights based on distance and a sum or mean in the calculation, in order to take into account in the calculation the distribution of inter-distances. This improvement can be considered as a "spatial standardization", similar to the classic one-dimensional standardization (such as age standardization). It is different from the previously described standardization on the rows of the spatial weight matrix, which does not solve the distribution of inter-distances problem, since each row corresponds to a local index and faces the same problem (every local index calculation favors the values of the distant points).

### 3. Methods

We propose a correction (a spatial standardization on distance, onwards referred to as SD-correction) to adjust the calculation (of the indices or estimates) on statistical distribution of inter-distances in order to remove the influence of this distribution on the calculation. To do so, we propose to add a second weight for each pair of points  $(P_i, P_j)$  in the sum calculation. This second weight  $w'_{ij}$

corresponds for each inter-distance  $d(P_i, P_j)$  to the inverse of the relative influence of the inter-distance in the set of all inter-distances. It is given by the inverse  $1/p_{ij}$  of the probability  $p_{ij}$  of the inter-distance in the set of all inter-distances. It is calculated for each inter-distance from the total number of inter-distances and their distribution  $f$ .

This distribution function  $f$  can be given by the density function of inter-distances, when it is known, such as in the case above mentioned, where the spatial distribution of the point set is defined by a known spatial distribution. When the density function of the inter-distances is unknown, we suggest to approach this density function  $f$  calculating the relative number of inter-distances  $\frac{N(k)}{N}$  in each interval  $[d, d + h[$  with a given lag  $h$  and  $d = kh$  ( $k \in \mathbb{N}$ ),  $d$  varying between 0 and  $d_{max}$  (maximum limit of the inter-distances to be considered), where  $N$  is the total number of inter-distances between 0 and  $d_{max}$ , and by interpolation between the points  $(kh + h/2, N(k))$  by an affine piecewise function or by kernel interpolation, with a Gaussian function  $\frac{1}{h\sqrt{2\pi}} e^{-\frac{1}{2}(\frac{d}{h})^2}$  of standard deviation  $h$  as kernel. Lag  $h$  makes it possible to set the influence of standardization on the distance in the overall calculation. The weight  $w_{ij}$  of a pair of points  $(P_i, P_j)$  is then modified dividing it by the density function approximation  $f(d(P_i, P_j))$ .

For example, the corrected Moran index will be:

$$\hat{I}_{Moran} = \frac{1}{S'} \sum_{i,j} w_{ij} w'_{ij} \left( \frac{X_i - m}{\sigma} \right) \left( \frac{X_j - m}{\sigma} \right) \quad (7)$$

where  $w_{ij}$  is the spatial weight for spatial dependence,  $w'_{ij} = 1/p_{ij}$  and  $p_{ij}$  the probability of inter-distance  $d(P_i, P_j)$ , given by the probability density function approximation  $f(d(P_i, P_j))$  calculated as indicated above.  $S'$  is the sum of the weights  $w_{ij} w'_{ij}$  ( $S' = \sum_{i,j} \frac{w_{ij}}{p_{ij}}$ ).

Expected value of SD-corrected Moran's  $\hat{I}$  under null hypothesis of no spatial autocorrelation is equal to the expected value of original Moran's  $I$ , since the expected value of Moran's  $I$  does not depend on spatial weights.

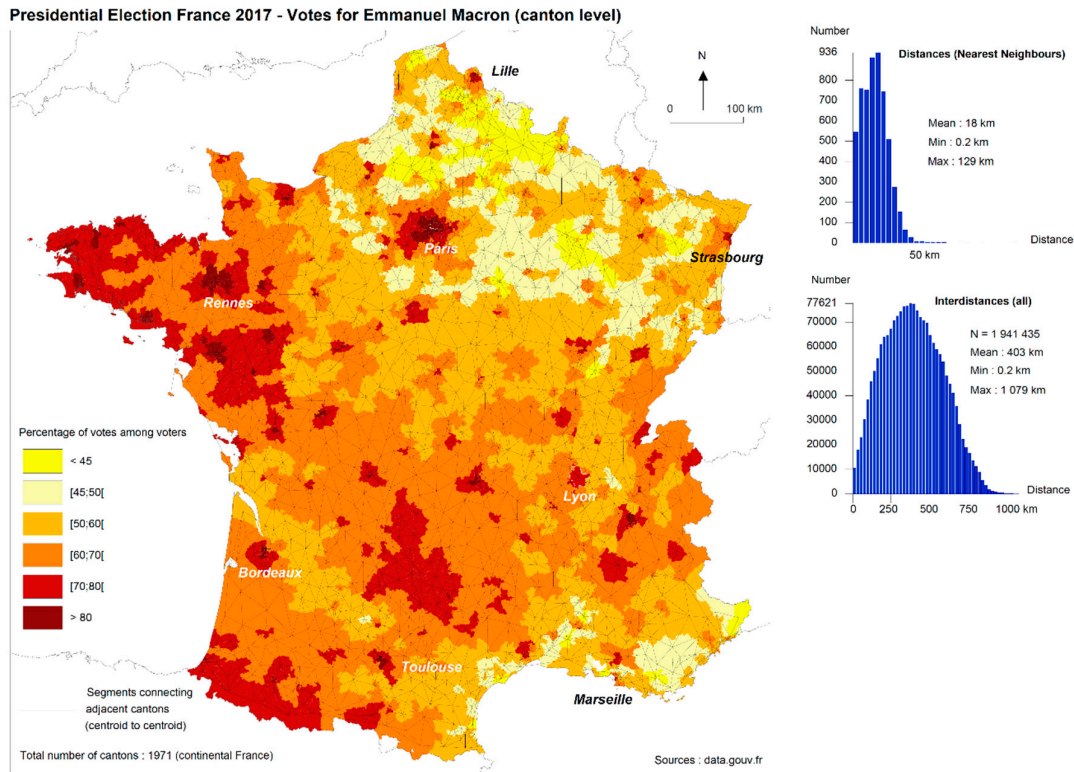
With the same notations, SD-corrected Geary index will be:

$$\hat{I}_{Geary} = \frac{1}{S'} \sum_{i,j} w_{ij} w'_{ij} \left( \frac{X_i - m}{\sigma} - \frac{X_j - m}{\sigma} \right)^2 \quad (8)$$

The SD correction applied to the calculation of the autocorrelation indices and spatial kernel estimations was implemented in SavGIS, a GIS freeware ([www.savgis.org](http://www.savgis.org)). The following example was computed in this GIS software.

#### 4. Example

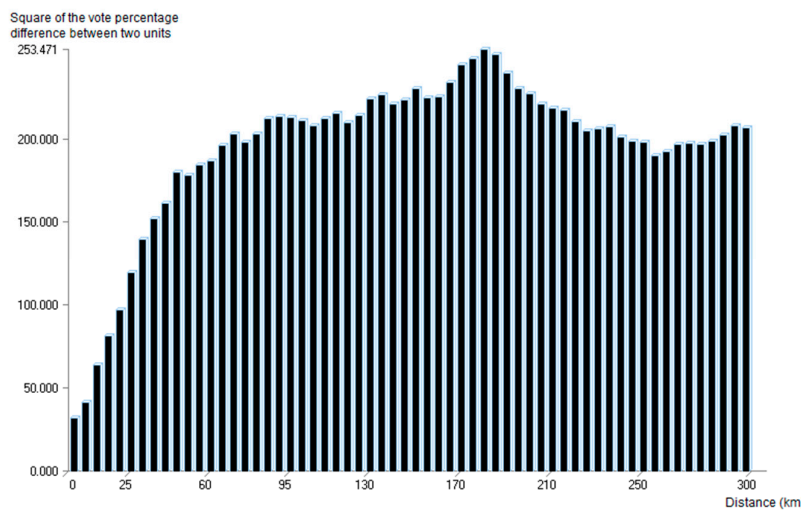
As an example, we applied this SD-correction to the spatial analysis of the results of the presidential elections that took place in France in April 2017. The variable analyzed is the percentage of votes won by Emmanuel Macron in the second round in continental France, aggregated by electoral canton (continental France is divided into 1971 electoral cantons). The data used is available on the open data website of the French government (<https://www.data.gouv.fr/fr/datasets/elections-legislatives-des-11-et-18-juin-2017-resultats-du-2nd-tour>). As it has been demonstrated in many countries, electoral behavior generally shows a certain continuity in space, although discrepancies may exist between neighboring spatial units. This variable is, therefore, well suited for spatial autocorrelation analysis and calculations. Figure 3 highlights a contrast between cities, largely in favor of Emmanuel Macron, and rural areas, especially the north-east of France and the Mediterranean coast in which the extreme right-wing candidate won the highest number of votes. On the right side of Figure 3, graphs respectively provide the distribution of the distances between nearest neighbors (the average distance between adjacent cantons is 18 km, centroid to centroid), and the distribution of inter-distances (distances between every pair of points).



**Figure 3.** Votes for Emmanuel Macron (%) at the second round of the presidential election in France, canton level (2017) (source: data.gouv.fr and Institut Géographique National-IGN).

#### 4.1. Spatial Autocorrelation Indices

From these percentages, we calculated the spatial autocorrelation Moran and Geary indices, without and with SD-correction. The semi-variogram of the percentage of votes in favor of Emmanuel Macron shows a range of spatial dependence lower than 250 km (Figure 4). We scanned the Moran and Geary indices with  $d_{max}$  varying from 25 to 250 km (Figure 5, Table 1).

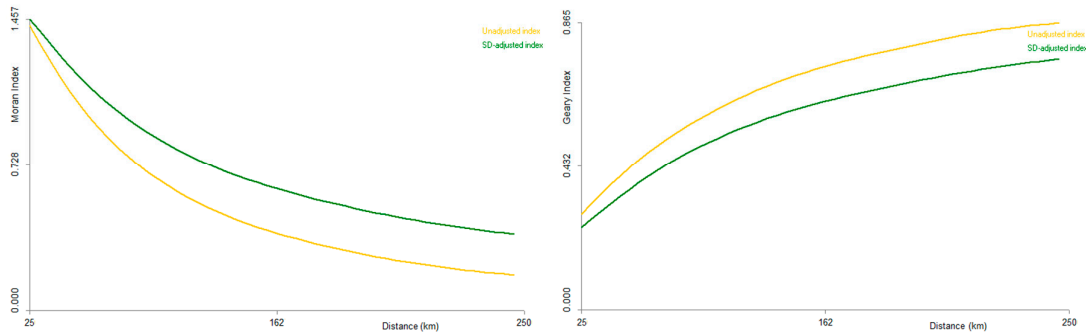


**Figure 4.** Semi-variogram of the percentage of votes cast in favor of Emmanuel Macron.

Figure 5 and Table 1 show that the value of the Moran or Geary index with SD-correction (in green) shows a stronger spatial autocorrelation than the value of the uncorrected index (in yellow).



Figure 5 also indicates that the values of both indices (Moran or Geary), both corrected and uncorrected, show a steadily decreasing autocorrelation when bandwidth  $dmax$  increases. This is logical, as the spatial dependence of values between spatial units decreases when distance increases, and considering more distant pairs may decrease the global weighted mean.



**Figure 5.** Moran (left) and Geary (right) autocorrelation indices with bandwidth  $dmax$  varying from 25 to 250 km. In yellow without SD-correction, in green with SD-correction.

**Table 1.** Uncorrected and SD-corrected Moran index values for growing bandwidths.

Bandwidth (km)	Number of Pairs	Uncorrected Moran Index	Z-Score	Standard Deviation	SD-Corrected Moran Index	Z-Score	Standard Deviation
25	12,804	1.43	138.78	0.0104	1.46	33.46	0.0433
50	34,012	1.01	165.45	0.0062	1.14	45.47	0.0251
75	62,930	0.73	156.81	0.0047	0.92	89.67	0.0101
100	101,544	0.55	145.10	0.0036	0.76	128.16	0.0059
150	206,304	0.35	134.88	0.0025	0.57	149.73	0.0038
200	337,288	0.24	124.62	0.0019	0.45	150.47	0.0029
250	484,394	0.18	113.65	0.0016	0.37	152.98	0.0025
300	644,016	0.14	110.83	0.0013	0.32	153.41	0.0020

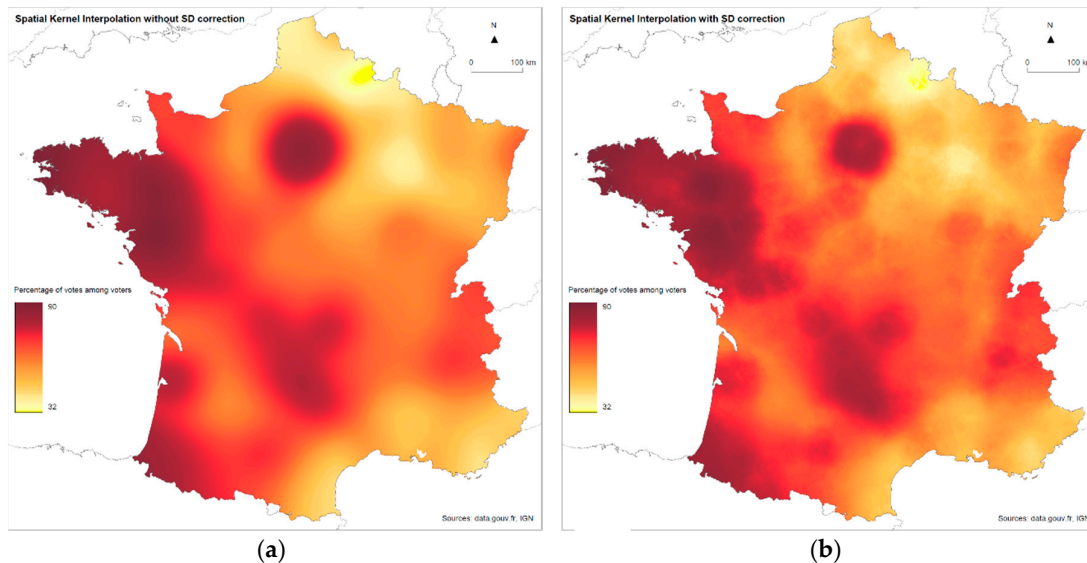
The statistical significance of the rejection of the null hypothesis ( $H_0$ ) of no spatial autocorrelation (provided here by the Z-score corresponding to the observed index value) is essential to conclude that autocorrelation is really effective. All these indices, both uncorrected and SD-corrected, show a very high probability of existence of spatial autocorrelation, as it can be seen in Table 1, with very high values of Z-Score for the indices for all tested bandwidth, corresponding to very low p-values for  $H_0$  hypothesis. We also note in this example that the significance of the SD-corrected index increases with the bandwidth, while the significance of the unadjusted index decreases (from a bandwidth at 75 km). We can also note that variance is higher for SD-corrected Moran's  $\hat{I}$  than original Moran's  $I$ .

If the values are randomly assigned to the geographical units to destroy spatial autocorrelation, both uncorrected and SD-corrected indices show similar values: SD-correction only acts in presence of spatial autocorrelation.

#### 4.2. Spatial Kernel Interpolation

We can see in Figure 6 that for the same kernel (in this case a Gaussian function with  $h = 200$  km), the SD-correction increases the accuracy of the interpolation. The SD-correction gives much more details and the output map is less smoothed. Without SD-correction, there are more pairs of

points associated with large distances within the limit of  $h$ , and these pairs of points have a higher influence in the calculation. This reduces the influence of the less frequent pairs of points and leads to much more averaged results based on pairs of points of larger distance. It produces a much smoother but less accurate trend surface.



**Figure 6.** Spatial kernel interpolation (Gaussian function,  $h = 200$  km) applied to the votes for Emmanuel Macron at the second round of presidential elections in France (2017): (a) Left map without SD-correction; (b) Right map with SD-correction.

## 5. Discussion and Conclusion

This article reviewed one of the foundations of spatial analysis. While the decrease in spatial weights with increasing distance has already been largely analyzed and discussed in the literature, the uneven distribution of pairs of points as a function of distance has been so far neglected. Yet, this uneven distribution of pairs of points has a direct impact on the calculation of a large number of methods in spatial analysis, such as spatial autocorrelation indices, kernel interpolation methods, or spatial modeling methods. Since the statistical distribution of the inter-distances is not uniform, all methods which rely on the calculation of a sum or a mean of pairs of points values favor the values of the pairs of points which are more numerous, while spatial dependence should only be assessed as a function of distance. When a calculation involves a sum or a mean on the weighted pairs, the influence of distance which derived from the uneven distribution of pairs of points must be corrected to not over- or under-represent certain inter-distances, as distance is precisely the main explanatory variable for spatial dependence. To address this issue, we introduced the concept of "spatial standardization" and a new weight  $w'_{ij} = 1/p_{ij}$ , where  $p_{ij}$  is the probability of inter-distance  $d(P_i, P_j)$ , given by the density function of inter-distances in the set of points, in the range of the bandwidth.

Logically, and as shown with the example above, the effect of the SD-correction becomes more and more effective when bandwidth increases and when the number of inter-distances involved in the calculation is growing. Indeed, in the calculation of uncorrected indices, the relative increase in the number of long inter-distances compared to short inter-distances reduce the influence of spatial dependence, since the spatial weight (which models the spatial dependence between two objects) decreases with distance. When the distance increases, so does the number of pairs of high-distance points, and so does their influence in the calculation. The effect of spatial dependence in the calculation is; therefore, reduced. The SD-correction aims at balancing this influence. The corrected index or estimation shows stronger autocorrelation values in comparison to the uncorrected index or estimation, by giving back weight to the inter-distances less frequent in the calculation, and;

therefore, in general, to the short distance pairs—precisely those that show, in the presence of spatial dependence, the strongest correlation between their values. SD-correction reinforces the purpose of the autocorrelation indices to capture and measure spatial autocorrelation when the calculation involves a weighted sum or mean of pairs of points values. Therefore, the use of spatial standardization is appropriate to all these situations.

In the case of the Moran index, we also saw in our example that the SD-correction increased the variance of the index under the null hypothesis. The variance of the Moran index depends on spatial weights [10]. The SD-correction adjusts the weights by rebalancing the relative value of the weights according to the distribution of the inter-distances. It thus increases the variance of spatial weights, which is reflected in the variance of the index itself.

Corrected estimations or indices may be more sensitive than uncorrected estimations to the values of the shortest inter-distances, as the corrected weight of these inter-distances is the product of spatial weight (in general higher for short inter-distances in order to capture spatial autocorrelation) and correction weight (depends on points spatial distribution, but almost always higher for short and long inter-distances). This remark on variance also shows that the proposed SD-correction reinforces the ability of the corrected autocorrelation indices to capture and measure spatial autocorrelation, giving more weight to shortest inter-distances in the result, but resulting in an increase of the variance.

In conclusion, this article shows that it is important to implement the SD-correction in all spatial analysis methods, models, and estimates that involve spatial autocorrelation calculations based on sums or means of distance-weighted values.

**Author Contributions:** Conceptualization, Marc Souris; formal analysis, Marc Souris; investigation, Marc Souris; methodology, Marc Souris; software, Marc Souris; validation, Florent Demoraes; writing—original draft preparation, Marc Souris and Florent Demoraes; writing—review and editing, Marc Souris and Florent Demoraes.

**Funding:** This research received no external funding.

**Conflicts of Interest:** The authors declare no conflicts of interest.

## References

1. Tobler W. A computer movie simulating urban growth in the Detroit region. *Econ. Geography Suppl.* **1970**, *46*, 234–240.
2. Shabenberger, O.; Gotway, C. *Statistical Methods for Spatial Data Analysis*; Chapman & Hall: London, UK, 2005.
3. Souris, M. *Epidemiology and Geography. Principles, Methods and Tools of Spatial Analysis*; Wiley-ISTE: London, UK, 2019. *Epidemiologie et Géographie, Principes, Méthodes et Outils de L'analyse Spatiale*; ISTE: London, UK, 2019, pour la version française.
4. Moran, P. The interpretation of statistical maps. *J. R. Stat. Soc. Ser. B* **1948**, *10*, 243–251.
5. Geary, R.C. The contiguity ratio and statistical mapping. *Inc. Stat.* **1954**, *5*, 115–145.
6. Anselin, L. Local indicators of spatial association—LISA. *Geogr. Anal.* **1995**, *27*, 93–115.
7. Getis, A.; Ord, J.K. The analysis of spatial association by use of distance statistic. *Geogr. Anal.* **1992**, *24*, 189–206.
8. Fotheringham, S.; Rogerson, P.A. *The Sage Handbook of Spatial Analysis*; Sage: London, UK; Los Angeles, CA, USA, 2009.
9. Cliff, A.D.; Ord, J.K. *The Problem of Spatial Autocorrelation*; Scott, A.J., Ed.; Studies in Regional Science; Pion: London, UK, 1969; pp. 25–55.
10. Cliff, A.D.; Ord, J.K. *Spatial Processes: Models and Applications*; Pion Limited: London, UK, 1981.
11. Upton, G.J.G.; Fingleton, B. *Spatial Data Analysis by Example*; Wiley: New York, NY, USA, 1985.
12. Anselin, L.; Bera, A.K. Spatial dependence in spatial regression model, with an introduction to spatial econometrics. In *Handbook of applied Economic Statistics*; Ullah, A., Giles, D.E., Eds.; Marcel Dekker: New York, NY, USA, 1988; pp. 237–289.
13. Mantel, N. The detection of disease clustering and a generalized regression approach. *Cancer Res.* **1967**, *27*, 209–220.

14. Getis, A.; Ord, J.K. Local spatial statistics: An overview.: In *Spatial Analysis: Modeling in A GIS Environment*; Longley, P., Batty, M., Eds.; John Wiley & Sons: New York, NY, USA, 1996; pp. 261–277.
15. Droesbeke, J.J.; Lejeune, M.; Saporta, M. *Analyse Statistique des Données Spatiales*; Technip: Paris, France, 2006.
16. Bowman, A.W.; Azzalini, A. *Applied Smoothing Techniques for Data Analysis*; Oxford University Press: London, UK, 1997.
17. Dormann, C.; McPherson, J.; Araújo, M.; Bivand, R.; Bolliger, J.; Carl, G.; Davies, R.G.; Hirzel, A.; Jetz, W.; Kissling, W.D.; et al. Methods to account for spatial autocorrelation in the analysis of species distributional data: A review. *Ecography* **2007**, *30*, 609–628.
18. Alagar, V.S. The distribution of the distance between random points. *J. Appl. Probab.* **1976**, *13*, 558–566.
19. Lellouche, S.; Souris, M. Distribution of distances between elements in a compact set. Unpublished, manuscript in preparation.



© 2019 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).