geographical analysis

Geographical Analysis (2024) **56**, 573–599

Measuring and Testing Multivariate Spatial Autocorrelation in a Weighted Setting: A Kernel Approach

François Bavaud

Department of Geography and Sustainability, Department of Language and Information Sciences, University of Lausanne, Lausanne, Switzerland

We propose and illustrate a general framework in which spatial autocorrelation is measured by the Frobenius product of two kernels, a feature kernel and a spatial kernel. The resulting autocorrelation index δ generalizes Moran's index in the weighted, multivariate setting, where regions, differing in importance, are characterized by multivariate features. Spatial kernels can traditionally be obtained from a matrix of spatial weights, or directly from geographical distances. In the former case, the Markov transition matrix defined by row-normalized spatial weights must be made compatible with the regional weights, as well as reversible. Equivalently, space is specified by a symmetric exchange matrix containing the joint probabilities to select a pair of regions. Four original weight-compatible constructions, based upon the binary adjacency matrix, are presented and analyzed. Weighted multidimensional scaling on kernels yields a low-dimensional visualization of both the feature and the spatial configurations. The expected values of the first four moments of δ under the null hypothesis of absence of spatial autocorrelation can be exactly computed under a new approach, invariant orthogonal integration, thus permitting to test the significance of δ beyond the normal approximation, which only involves its expectation and expected variance. Various illustrations are provided, investigating the spatial autocorrelation of political and social features among French departments.

Introduction

Spatial autocorrelation denotes the presence of a relation between the attributes or features characterizing a set of regions and the spatial disposal of the regions. Measuring and testing spatial autocorrelation is at the very heart of a large body of geographical analysis, and innumerable formal and empirical studies have addressed this issue.

I am much indebted to an anonymous reviewer whose remarks and suggestions have helped to clarify the manuscript

Correspondence : François Bavaud, Department of Geography and Sustainability, Department of Language and Information Science, University of Lausanne, Lausanne, Switzerland.
e-mail: fbavaud@unil.ch

Submitted: October 18, 2023. Revised version accepted: January 19, 2024.

doi: 10.1111/gean.12390

The mainstream framework on spatial autocorrelation, shared among the geographical community, has arguably been achieved and stabilized in the seminal work of Cliff and Ord (1973, 1981): Moran's I, the most popular measure of spatial autocorrelation, is expressed as a normalized, covariance-like similarity between the value of a numerical feature and the average value of its neighbors, where the neighborhood relation is specified by a continuous (as opposed to binary) nonnegative matrix of spatial weights. Testing the statistical significance of Moran's I is traditionally performed (a) either by computing its expected first moments under the null hypothesis H_0 of absence of autocorrelation, under the additional assumption that the feature under consideration is normally distributed (normal test); or (b), by permuting the values of the feature among the regions, and comparing the distribution of the permuted Moran's I to its observed value (permutation test).

Yet, as a general rule, regions differ in importance, and this circumstance requires the use of a *weighted formalism*, where the regional weights are typically proportional to the regional areas in Physical Geography, or to the number of inhabitants in Human Geography, or more precisely to the number of specified actors such as voters, active people, commuters etc. depending on the scenario under analysis. Dealing with weighted objects disqualifies the direct use of normal and permutation testing procedures: statistical distributions are bound to depend on the regional weights: for instance, the highest (and lowest) proportions of votes in favor of a given candidate in a presidential election are always attained in the smallest populated municipalities, and exchanging those extreme scores with those of the more populated municipalities in a permutation test seems questionable.

Also, a great deal of contexts require to simultaneously take into consideration numerous regional features, thus requiring a *multivariate approach* (see, e.g., Wartenberg 1985; Thioulouse, Chessel, and Champely 1995; Dray and Jombart 2011; Anselin 2019; Lin 2020; Eckardt and Mateu 2021, and references therein), as well as the definition of a sensible measure of inter-regional dissimilarity, aptly summarizing the contrasts between regions.

The present study adopts such a *weighted, multivariate* framework, and systematically investigates the properties of a multivariate, weighted index of spatial autocorrelation δ, which constitutes a direct generalization of Moran's I (section Generalizing Morans in the multivariate, weighted setting). δ expresses as a convex mixture (direct or spectral) of Moran's I (section Decompositions of δ), and turns out to linearly decompose into local indicators (Anselin 1995).

On one hand, components of the row-standardized matrix of spatial weights $\mathbf{W} = (w_{ij})$ represent a measure of influence of region j on region i, and constitute a central ingredient in autoregressive models and measures of spatial autocorrelation. On the other hand, row-standardized spatial weights formally constitute the transition matrix of a *Markov chain* (Bavaud 1998), and, given the major impact of regional weights on spatial analyses, as well as the extreme versatility in which spatial weights can be constructed, it seems highly commendable, for formal as well as conceptual reasons, to design the spatial weights in a *weight-compatible* way, that is such that the stationary distribution of the corresponding Markov chain precisely coincides with the regional weights, supposed constant (section Regional weights, adjusted, and reversible spatial weights). One will also require spatial weights $\mathbf{W}$ to be *reversible*, which constitutes a natural condition of symmetry (section Obtaining adjusted spatial weights from the adjacency matrix). The construction of such adjusted, reversible spatial weights permits in turn to define a *spatial kernel* by equation (21).

Kernels, well-known in Machine Learning, appear in Classical Data Analysis in the (weighted) *multidimensional scaling* (MDS) procedure, generalizing both Principal Component

Analysis and Correspondence Analysis (section Illustration: political and social autocorrelation among French departments). Weighted MDS is a weighted extension of the Torgerson–Gower classical MDS procedure, and consists in spectrally decomposing the matrix of weighted scalar products or *kernel*. The procedure yields a low-dimensional factorial visualization (factor scores) of the feature configuration, as well as the inertia expressed by each factor (scree plot). Applying the same procedure to spatial kernels yields a low-dimensional visualization of the spatial configuration (Fig. 1). It requires the spatial kernel to be positive semi-definite (p.s.d), which can always be insured, if necessary, by redefining the spatial weights as $\mathbf{W}^2$, the matrix square of the initial spatial weights $\mathbf{W}$, or, equivalently, by considering the matrix square of spatial kernel itself (section Spatial kernels: visualizing the geographic space from spatial weights).

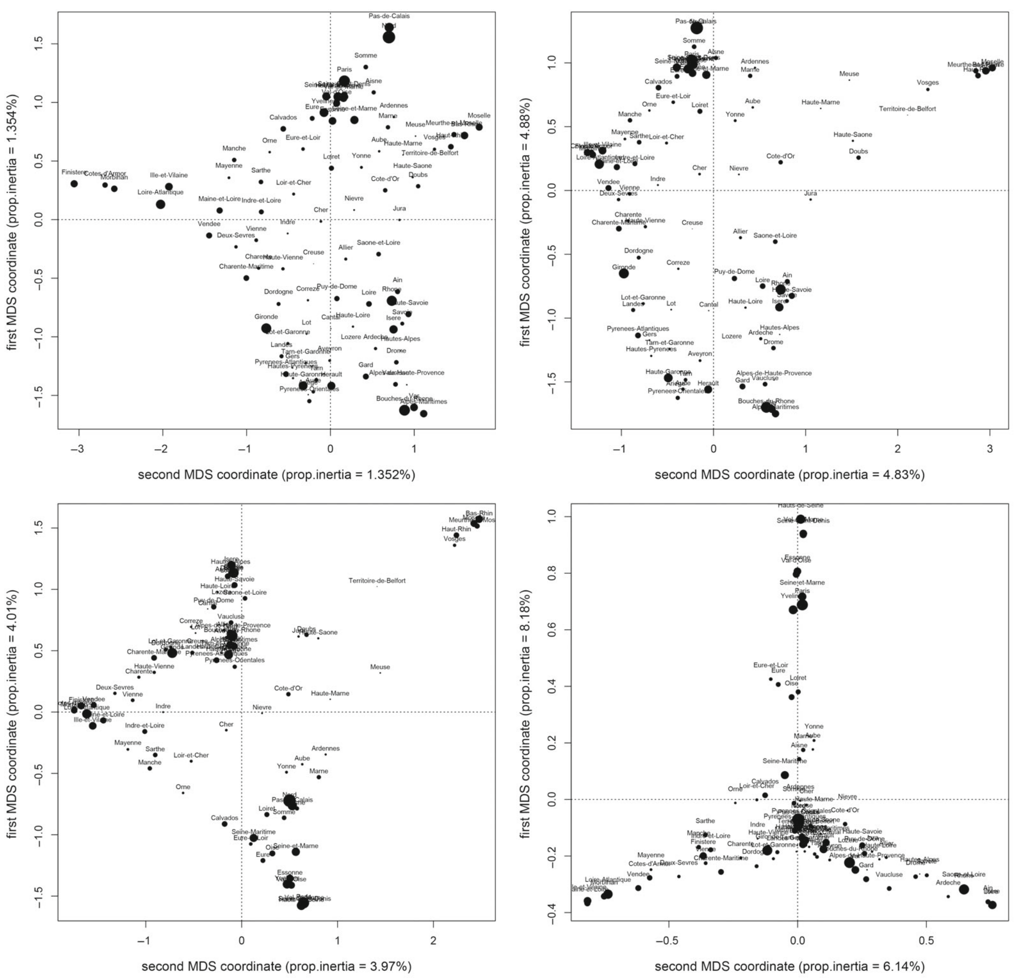

Figure 1. First spatial coordinates $\hat{x}_{i\alpha}$ (24) extracted from spatial weights $\mathbf{W}_a$ (8) (top, left), $\mathbf{W}_b^2$ (9) (top, right), $\mathbf{W}_c^2$ (10) with $g = 0.001$ (bottom, left) and $\mathbf{W}_c^2$ with $g = 0.1$ (bottom, right). Factors $\alpha = 1, 2$ have been swapped to ease the comparison with the geography of French departments, which is fairly well restituted in the top maps, but much less in the bottom maps.

Central to the article is the fact (Theorem 3) that the autocorrelation index δ is proportional to the Frobenius product of the features kernel and the spatial kernel: in the present approach, features and space kernels, respectively, determined from $\mathbf{D}_X$ and $\mathbf{W}$ (1), play a symmetric role.

Alternatively, spatial kernels can also be directly constructed from *geographic distances* $\mathbb{D}$ as in (32), that is without needing to introduce spatial weights $\mathbf{W}$: the proposed formalism seeks to extend and unify various approaches to spatial autocorrelation under a single umbrella.

Last but not least, the Frobenius product expression makes δ proportional to the (weighted version of the) so-called RV coefficient, proposed by Escoufier (1973) and Robert and Escoufier (1976) as a general measure of similarity between two multivariate configurations. Hence, any significance testing procedure for the RV coefficient can be directly translated into a significance testing for δ, and section Testing δ by invariant orthogonal integration relies upon a recent weighted, non-parametric approach (Bavaud 2023), namely *invariant orthogonal integration*, permitting to compute the first expected *moments*[1] of δ under H_0, yielding p-values beyond the normal approximation through the Cornish–Fisher expansion (45).

Section Illustration: political and social autocorrelation among French departments illustrates the formalism by defining three political and socioeconomic feature kernels on the 94 continental French departments, together with four spatial kernels, all based upon the adjacency matrix between departments. To each kernel corresponds a visualization of the (feature or spatial) configuration of departments, as well as a scree plot, permitting to determine the *spectral moments*, directly entering into the definition of the first expected moments for δ. Depending on the pair of kernels involved, its expected asymmetry and excess kurtosis can be large or small, as well as positive or negative (Table 3), with direct consequences on the p-values associated to the test of significance for δ (Table 4).

This article contains quite a few mathematical formulas, as possibly expected in an essay in quantitative geography. Yet, to resume, its underlying rationale is fairly straightforward: in a general weighted, multivariate setting, spatial autocorrelation δ simply expresses as a similarity index between a feature kernel $\mathbf{K}_X$ and a spatial kernel $\mathbf{K}_W$, proportional to their Frobenius product. The spectral decomposition of the kernels, provided they are p.s.d., permits to visualize the feature and the spatial configurations. This framework further allows a nonparametric significance testing procedure for spatial autocorrelation, based upon the spectral moments of the two kernels at stake. Most of the above is summarized in the following identities (26), (16) and (21) derived and detailed in the sequel:

$$\delta = \frac{\mathrm{Tr}(\mathbf{K}_W\mathbf{K}_X)}{\mathrm{Tr}(\mathbf{K}_X)} \qquad \mathbf{K}_X = -\frac{1}{2}\sqrt{\mathbf{\Pi}}\mathbf{H}\mathbf{D}_X\mathbf{H}^\top\sqrt{\mathbf{\Pi}} \qquad \mathbf{K}_W = \mathbf{\Pi}^{\frac{1}{2}}\mathbf{W}\mathbf{\Pi}^{-\frac{1}{2}} - \sqrt{\mathbf{f}}\sqrt{\mathbf{f}}^\top. \quad (1)$$

A multivariate, weighted measure of spatial autocorrelation: the δ index

Regional weights, adjusted, and reversible spatial weights

Consider a set of n regions differing in importance (by population, area, wealth, etc.), together with *regional weights* $\mathbf{f} \in \mathbb{R}^n$, where $f_i > 0$, obeying $f_\bullet = \sum_{i=1}^n f_i = 1$, reflects the relative importance of region i.

Row-standardized *spatial weights* $\mathbf{W} = (w_{ij}) \in \mathbb{R}^{n\times n}$ are nonnegative ($w_{ij} \geq 0$) and normalized to $w_{i\bullet} = \sum_{j=1}^n w_{ij} = 1$: they constitute the components of a Markov chain transition matrix $\mathbf{W}$, specifying the conditional probability to visit region j immediately after visiting region i (Bavaud 1998). In standard geographical and econometrical applications, $\mathbf{W}$ is *regular*, that is irreducible (all regions communicate) and aperiodic (return times are not restricted to multiples of $t = 2, 3, \dots$). As a consequence (see, e.g., Levin and Peres 2017), $\mathbf{W}$ possesses a *unique stationary distribution* $\boldsymbol{\pi}$ (with $\pi_i > 0$ and $\pi_\bullet = 1$) obeying $\mathbf{W}^\top \boldsymbol{\pi} = \boldsymbol{\pi}$.

Spatial weights $\mathbf{W} = (w_{ij})$ are meant to model spatial proximity, spatial interaction, mutual influence, etc. between pairs i, j of regions, and their precise specification from inter-regional adjacencies, travel or cost times, migratory or economic flows, social interactions etc. can accommodate considerable flexibility. It is hence most natural (and we advise it much vigorously, when possible) to require $\boldsymbol{\pi}$ to coincide with the regional weights $\mathbf{f}$, that is to impose $\mathbf{W}^\top \mathbf{f} = \mathbf{f}$: such spatial weights $\mathbf{W}$ are said to be *weight-compatible* or *adjusted*. Dealing with weight-compatible spatial weights considerably simplifies the subsequent formalism and its interpretation. In particular, the distinction between Moran and Geary autocorrelation indices becomes immaterial (section Generalizing Moran's I in the multivariate, weighted setting). Geographically speaking, the condition of weight compatibility says that the initial distribution of regional importance $\mathbf{f}$ coincides with the distribution $\mathbf{W}^\top \mathbf{f}$ obtained after one transition, that is that the regional importance is invariant, with no regions expanding or shrinking. By contrast, and for instance, spatial weights directly constructed on migratory flows (see, e.g., Bavaud 1998, 2002), possess in general *two distinct regional weights*, namely initial $\mathbf{f}_{\text{initial}}$ and final $\mathbf{f}_{\text{final}}$, thus making the issue of weight compatibility more involved.

Let us finally require that $f_i w_{ij} = f_j w_{ji}$, which simply says that $\mathbf{W}$ is a *reversible* Markov chain. Equivalently, the associated *backwards* or *time reversed* Markov chain $\mathbf{W}^\star = \boldsymbol{\Pi}^{-1} \mathbf{W}^\top \boldsymbol{\Pi}$, giving the conditional probability $w^\star_{ji}$ that region i was visited immediately *before* visiting j, equals the forwards chain $\mathbf{W}$. The reversibility condition insures that all eigenvalues of $\mathbf{W}$ are real, and is trivially satisfied if one defines $\mathbf{W}$ from symmetric adjacencies or symmetric geographic distances, as performed in this article. Yet, spatial weights directly constructed from flows are generally non reversible, unless the flows are *quasi-symmetric*, that is follow a *Gravity modelling* pattern (Bavaud 2002).

Summarizing up, and for a fixed vector of regional weights $\mathbf{f} \in \mathbb{R}^n$, one considers *row-standardized, weight-compatible, reversible spatial weights* $\mathbf{W}$ of the form

$$\mathbf{W} \geq 0, \quad \mathbf{W}\mathbf{1}_n = \mathbf{1}_n, \quad \mathbf{W}^\top \mathbf{f} = \mathbf{f}, \quad \boldsymbol{\Pi}\mathbf{W} = \mathbf{W}^\top \boldsymbol{\Pi} \quad \text{where} \quad \boldsymbol{\Pi} = \text{diag}(\mathbf{f}) \in \mathbb{R}^{n \times n}, \tag{2}$$

where $\mathbf{1}_n \in \mathbb{R}^n$ denotes the unit vector.

The joint probability of visiting i, and then j (or the other way round) constitutes the components $e_{ij} = f_i w_{ij}$ of the the so-called *exchange matrix* (Berger and Snell 1957) $\mathbf{E} = (e_{ij})$, which obeys by construction

$$\mathbf{E} = \boldsymbol{\Pi}\mathbf{W} \geq 0, \quad \mathbf{E}\mathbf{1}_n = \mathbf{f}, \quad \mathbf{E}^\top = \mathbf{E}. \tag{3}$$

$\mathbf{W}$, similar to the symmetric matrix $\boldsymbol{\Pi}^{-\frac{1}{2}} \mathbf{E} \boldsymbol{\Pi}^{-\frac{1}{2}}$, possesses real eigenvalues λ_α in the interval $[-1, 1]$ (Perron-Frobenius theorem. See, e.g., Levin and Peres 2017).

Two limit spatial weights are worth considering:

$$\mathbf{W}_0 = \left(w^0_{ij}\right) = \mathbf{I}_n \quad \text{i.e. } w^0_{ij} = \delta_{ij} \qquad\qquad \mathbf{W}_\infty = \left(w^\infty_{ij}\right) = \mathbf{1}_n \mathbf{f}^\top \quad \text{i.e. } w^\infty_{ij} = f_j. \tag{4}$$

$\mathbf{W}_0$ is the *pure stayers* or *frozen* Markov chain with no interregional transitions, while $\mathbf{W}_\infty$ is the *perfectly mobile* Markov chain whose transitions do not dependent of the initial region. Both are totally uninformative regarding the spatial configuration. The components $f_i f_j$ of the perfectly mobile exchange matrix $\mathbf{E}_\infty = \mathbf{\Pi W}_\infty = \mathbf{f}\mathbf{f}^\top$ yield the independent pair selection probabilities resulting from sampling with replacement.

In the uniformly weighted case $\mathbf{f} = \mathbf{1}_n/n$, row-standardized spatial weights $\mathbf{W}$ are adjusted and reversible iff $\mathbf{W} = \mathbf{W}^\top$, that is iff $\mathbf{W}$ is symmetric.

Obtaining adjusted spatial weights from the adjacency matrix

Obtaining in full generality weight-compatible, reversible spatial weights $\mathbf{W}$ is an issue in itself, requiring some additional computing effort (see e.g., Bavaud 2014). This article restricts itself to symmetric constructions from the adjacency matrix (Definition 1) or from a matrix of geographical distances (Definition 2).

Consider the binary adjacency (or connectivity) matrix between neighboring regions $\mathbf{A} = (a_{ij}) \in \mathbb{R}^{n\times n}$, defined as

$$a_{ij} = \begin{cases} 1 & \text{if } i \text{ and } j \text{ are distinct and possess a common border,} \\ 0 & \text{otherwise; in particular, } a_{ii} = 0. \end{cases}$$

The natural random walk transition matrix

$$\mathbf{P} = (p_{ij}) \quad \text{with} \quad p_{ij} = \frac{a_{ij}}{a_{i\bullet}} \tag{5}$$

defines a nonnegative, row-normalized matrix of reversible spatial weights (and adopted as such by Thioulouse, Chessel, and Champely (1995)), corresponding to the simple random walk across adjacent regions. Its stationary distribution $\boldsymbol{\pi}$ is the normalized degree $\pi_i = a_{i\bullet}/a_{\bullet\bullet}$.

Yet, $\boldsymbol{\pi}$ typically differs from the given regional weights $\mathbf{f}$: the first distribution is a mesure of *regional centrality*, but the second one is a measure of *regional importance*. Said otherwise, $\mathbf{P}$ is adjusted to $\boldsymbol{\pi}$, but not to $\mathbf{f}$, and further modifications are required to obtain adjusted spatial weights $\mathbf{W}$.

Allowing, in addition to direct transitions between neighbors, null transitions within the same region, that is carefully blending the pure movers dynamics $\mathbf{P}$ with stayer components yields proposal (a) below. The latter, which emphasizes the role of the Laplacian operator, also obtains as the first-order expansion of the *continuous-time, diffusive* exchange matrix presented in Bavaud (2014); see also (7)

$$\mathbf{E}(t) := \mathbf{\Pi}^{1/2} \exp(-t\mathcal{L}\mathbf{A})\mathbf{\Pi}^{1/2} \qquad \text{where} \quad t \geq 0, \tag{6}$$

and whose associated kernel constitutes a weighted extension of the so-called *heat kernel* in Machine Learning (Kondor and Lafferty 2002). The diffusive dynamics (6) results from a *jump process*, where a unit initially in region i remains there for some random time t_i exponentially distributed with mean $E(t_i) = f_i/a_{i\bullet}$ (sojurn time), until it jumps to a different region j with probability p_{ij}.

Proposal (b) is presumably original in its presentation, but results from a direct application of the Metropolis–Hastings algorithm Hastings (1970), which precisely solves the issue of how to modify a Markov chain $\mathbf{P}$ so that the new Markov chain $\mathbf{W}$ possesses a given stationary distribution $\mathbf{f}$.

Finally, proposal (c) results from a simple iterative fitting procedure, known to converge with a unique solution if the initial matrix is not "too sparse" (see, e.g., Schneider and Zenios 1990; Ruschendorf 1995; Knight 2008, and references therein), thus generally requiring the initial addition of a small uniform ground $g > 0$ to $\mathbf{A}$. This modification, creating spatial interaction between initially noninteracting regional pairs, could be justified in the same way as unit fictitious flows are sometimes added to the observed flows in a Bayesian setting (see, e.g., LeSage and Thomas-Agnan 2015).

Definition 1 Three constructions yielding spatial weights from the adjacency matrix. Let $\mathbf{A} = (a_{ij}) \in \mathbb{R}^{n\times n}$ denote the symmetric binary adjacency matrix between regions:

(a) Linearized diffusive weights: define the *Laplacian* $\mathbf{LA}$ and the *adjusted Laplacian* $\mathcal{L}\mathbf{A}$ as

$$(\mathbf{LA})_{ij} = a_{i\bullet}\delta_{ij} - a_{ij}, \quad \mathcal{L}\mathbf{A} = \mathbf{\Pi}^{-\frac{1}{2}}\mathbf{LA}\mathbf{\Pi}^{-\frac{1}{2}}, \quad i.e.\ (\mathcal{L}\mathbf{A})_{ij} = \frac{a_{i\bullet}\delta_{ij} - a_{ij}}{\sqrt{f_i f_j}}. \tag{7}$$

where the Kronecker symbol δ_{ij} denotes the components of the identity matrix $\mathbf{I}_n \in \mathbb{R}^{(n\times n)}$. Define

$$\mathbf{W}_a = \mathbf{I}_n - t\mathbf{\Pi}^{-1}\mathbf{LA}, \qquad i.e.\ w^a_{ij} = \delta_{ij} + t\frac{a_{ij} - a_{i\bullet}\,\delta_{ij}}{f_i}. \tag{8}$$

where $0 < t \le t_1 := \min_{i=1}^n (f_i/a_{i\bullet})$.

(b) Metropolis–Hastings weights: let $\mathbf{P} = (p_{ij})$ be the random walk transition matrix (5), and define $\mathbf{\Gamma} = (\gamma_{ij})$ as $\gamma_{ij} := \min(f_i p_{ij}, f_j p_{ji})$, together with

$$\mathbf{E}_b = \mathbf{\Pi} - \mathbf{L\Gamma} \quad \text{i.e.} \quad e^b_{ij} = \delta_{ij}(f_i - \gamma_{i\bullet}) + \gamma_{ij} \quad \text{and} \quad \mathbf{W}_b = \mathbf{\Pi}^{-1}\mathbf{E}^b = \mathbf{I}_n - \mathbf{\Pi}^{-1}\mathbf{L\Gamma}. \tag{9}$$

(c) Iteratively fitted weights: Let $g > 0$ be a small quantity ($g \ll 1$) and define the modified adjacency matrix $\mathbf{A}_g = (a^g_{ij})$ as $a^g_{ij} = a_{ij} + g$, that is $\mathbf{A}_g = \mathbf{A} + g\,\mathbf{J}_n$ where $\mathbf{J}_n = \mathbf{1}_n\mathbf{1}_n^\top$ is the unit matrix. By iterative fitting, determine $\mathbf{c} \in \mathbb{R}^n$ such that the margins of the symmetric matrix $\mathbf{E}_c = \mathbf{C}\mathbf{A}_g\mathbf{C}$, where $\mathbf{C} = \text{diag}(\mathbf{c})$, yield the regional weights, that is in order to achieve $\mathbf{E}_c\mathbf{1}_n = \mathbf{f}$. Define

$$\mathbf{W}_c = \mathbf{\Pi}^{-1}\mathbf{E}_c, \qquad \text{that is,}\ w^c_{ij} = \frac{e^c_{ij}}{f_i}. \tag{10}$$

Theorem 1 $\mathbf{W}_a$, $\mathbf{W}_b$ *and* $\mathbf{W}_c$ *defined above constitute row-normalized, adjusted and reversible spatial weights.*

Each the three above matrices favor transitions between adjacent regions, as they must.

Proof of Theorem 1. Reversibility of $\mathbf{W}_a$, $\mathbf{W}_b$ and $\mathbf{W}_c$ is immediate. (a) w^a_{ij} in (8) is nonnegative for $i \ne j$. For $i = j$, nonnegativity holds iff $1 - t\,a_{i\bullet}/f_i \ge 0$, in view of $a_{ii} = 0$. Also, $\mathbf{LA1}_n = \mathbf{0}_n$, and thus $\mathbf{W}_a^\top\mathbf{f} = \mathbf{f} - t\mathbf{LA\Pi}^{-1}\mathbf{f} = \mathbf{f}$. (b) $\mathbf{L\Gamma1}_n = \mathbf{0}$, thus implying $\mathbf{W}_b^\top\mathbf{f} = \mathbf{f}$ by the same reasoning. e^b_{ij} in (9) is nonnegative for $i \ne j$. For $i = j$, nonnegativity holds since $\gamma_{i\bullet} = \sum_j \gamma_{ij} \le \sum_j f_i\, p_{ij} = f_i$. (c) Weight-compatibility and non-negativity follow directly form the iterative fitting procedure. □

Generalizing Moran's *I* in the multivariate, weighted setting

The multivariate profiles of the n regions under consideration may differ by countless features. A crucial responsibility of the analyst consists in aptly summarizing those regional contrasts by specifying a relevant $n \times n$ dissimilarity matrix $\mathbf{D} = (d_{ij})$, obeying

$$d_{ij} \ge 0, \quad d_{ij} = d_{ji}, \quad d_{ii} = 0. \tag{11}$$

We shall for the moment assume in addition $\mathbf{D}$ to be *squared Euclidean*, that is, of the form $d_{ij} = \|\mathbf{x}_i - \mathbf{x}_j\|^2$ for some $\mathbf{x}_i, \mathbf{x}_j \in \mathbb{R}^p$ (see section Feature kernels: visualizing regional dissimilarities).

For some time, Bavaud (2013, 2014) has proposed a weighted, multivariate generalization of Moran's I as the index

$$\delta = \frac{\Delta - \Delta_{\mathrm{loc}}}{\Delta} \in [-1, 1], \tag{12}$$

where the configuration *inertia* Δ and *local inertia* Δ_{loc} are

$$\Delta = \frac{1}{2}\sum_{i,j=1}^{n} f_i f_j d_{ij} \qquad \Delta_{\mathrm{loc}} = \frac{1}{2}\sum_{i,j=1}^{n} e_{ij} d_{ij}. \tag{13}$$

Here e_{ij} are the components of the exchange matrix $\mathbf{E}$ (3), giving the joint probability to select the pair i, j of regions. In the univariate case, $d_{ij} = (x_i - x_j)^2$ for the variable of interest $\mathbf{x} \in \mathbb{R}^n$, and

$$\Delta = \frac{1}{2}\sum_{i,j=1}^{n} f_i f_j (x_i - x_j)^2 = \sum_i f_i (x_i - \bar{x})^2 = \mathrm{var}(\mathbf{x}) \qquad \text{where} \quad \bar{x} = \sum_i f_i x_i.$$

The *lagged variable* $\mathbf{Wx}$ measures the average value among regional neighbors. Weight-compatibility $\mathbf{W}^{\top}\mathbf{f} = \mathbf{f}$ entails the identity $\overline{\mathbf{Wx}} = \bar{\mathbf{x}}$, a circumstance making the distinction between Moran and Geary indices irrelevant: the multivariate generalization of Geary index is simply $c = \Delta_{\mathrm{loc}}/\Delta = 1 - \delta \in [0, 2]$. Also, simple algebra demonstrates that

$$\Delta - \Delta_{\mathrm{loc}} = \mathrm{cov}(\mathbf{x}, \mathbf{Wx}) = \sum_{ij} f_i (x_i - \bar{x}) w_{ij} (x_j - \bar{x}) = \sum_{ij} e_{ij} (x_i - \bar{x})(x_j - \bar{x}).$$

In the unweighted univariate case, $f_i = 1/n$, and, for possibly unstandardized spatial weights, one gets the familiar Moran's index expression (Cliff and Ord 1973)

$$\delta = I = \frac{n}{\sum_{ij} w_{ij}} \; \frac{\sum_{ij} w_{ij} (x_i - \bar{x})(x_j - \bar{x})}{\sum_i (x_i - \bar{x})^2}, \tag{14}$$

where the first ratio in the r.h.s. is equal to one for row-standardized spatial weights.

In the weighted univariate case, identity $\delta = \mathrm{cov}(\mathbf{x}, \mathbf{Wx})/\mathrm{var}(\mathbf{x})$, involving weighted (co-)variances, shows δ to be identical to the slope a of the weighted least squares (WLS) regression $\widehat{\mathbf{Wx}} = a\mathbf{x} + b$ in the Anselin–Moran scatterplot, mapping the values of $\mathbf{x}$ versus those of $\mathbf{Wx}$, with an intercept given by $b = (I - 1)\bar{x}$ (see Figs. 2 and 3).

Feature kernels: visualizing regional dissimilarities

Since the regional dissimilarities (11), reflecting disparities among regional features, are assumed to be squared Euclidean (see the illustrations (34) and 35)), there exists a for each region a vector of regional coordinates $\tilde{\mathbf{x}}_i = (\tilde{x}_{i\alpha})$ such that $d_{ij} = \|\tilde{\mathbf{x}}_i - \tilde{\mathbf{x}}_j\|^2 = \sum_{\alpha \geq 1} (\tilde{x}_{i\alpha} - \tilde{x}_{j\alpha})^2$. Such regional coordinates will serve in turn to *visualize* the feature dissimilarities in dimensions $\alpha = 1, 2, \ldots$. The regional coordinates are far from unique, since their rotation and reflection yields the same dissimilarities (orthogonal invariance). In the quest for a unique solution, one further requires that, in addition, the first components $\alpha = 1, 2, \ldots$ of the regional coordinates express a maximal proportion of the weighted inertia Δ (13).

The solution to the above problem exactly reproduces the well-known Torgerson-Gower classical MDS procedure (see, e.g., Borg and Groenen 2005), slightly modified in the weighted setting (Bavaud 2023). Weighted MDS, aiming at obtaining regional coordinates associated to the feature dissimilarity $\mathbf{D}$, results from the spectral decomposition of feature *kernels*, constructed as follows:

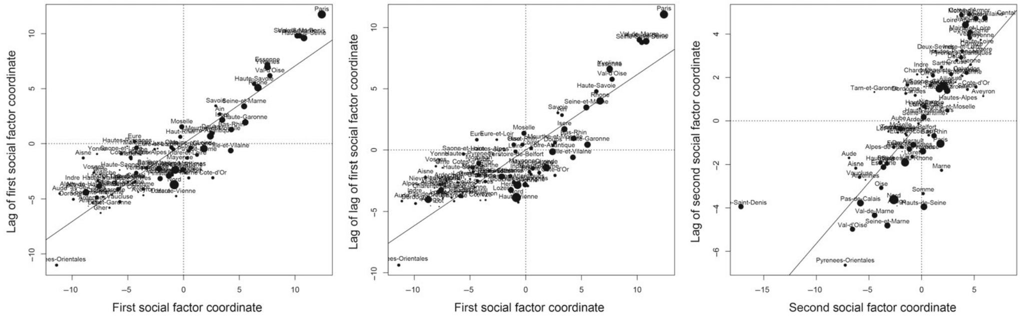

Figure 2. Moran–Anselin scatterplots for the two first social factorial coordinates (19) with spatial weights $\mathbf{W}_b$ (Metropolis–Hastings). Left: $\hat{\mathbf{y}}_1$. Middle: $\hat{\mathbf{y}}_1$ again, but with iterated weights $\mathbf{W}_b^2$. Right: $\hat{\mathbf{y}}_2$. The slopes of the regression lines give the corresponding Moran's I, with values (in order) 0.706, 0.616, and 0.564. The intercepts are zero, since factorial coordinates are centered.

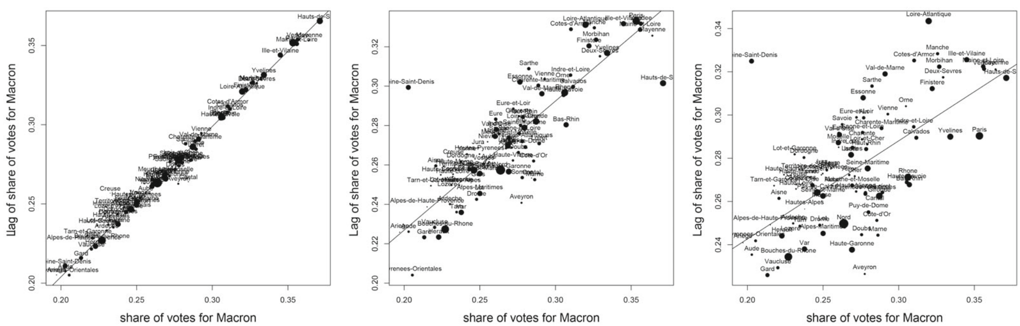

Figure 3. Moran–Anselin scatterplots for the share of votes for Macron (first round of the French presidential election, 2022), with spatial weights specified as $\mathbf{W}_a$ (diffusive, left), $\mathbf{W}_b$ (Metropolis–Hastings, middle) and $\mathbf{W}_c$ (iterative fitting, right). The larger $\mathrm{Tr}(\mathbf{W})$, the smaller the dispersion around the regression line, whose slope δ figures in the last line of Table 1.

1. First, define the weighted *centering matrix* $\mathbf{H} = \mathbf{I}_n - \mathbf{1}_n\mathbf{f}^\top$, where $\mathbf{I}_n \in \mathbb{R}^{n\times n}$ is the identity matrix and $\mathbf{1}_n \in \mathbb{R}^n$ the unit vector. The centering matrix is a projection ($\mathbf{H}^2 = \mathbf{H}$), but $\mathbf{H}^\top \neq \mathbf{H}$, unless $\mathbf{f}$ is uniform.
2. Second, compute the symmetric matrix $\mathbf{B}$ of *scalar products* by double centering:

$$\mathbf{B} = -\frac{1}{2}\mathbf{H}\,\mathbf{D}\,\mathbf{H}^\top. \tag{15}$$

3. Third, compute the diagonal matrix $\mathbf{\Pi} = \mathrm{diag}(\mathbf{f})$ and define the *kernel* $\mathbf{K} \in \mathbb{R}^{n\times n}$ as the symmetric matrix of *weighted scalar products*:

$$\mathbf{K} = \sqrt{\mathbf{\Pi}}\,\mathbf{B}\sqrt{\mathbf{\Pi}} = -\frac{1}{2}\sqrt{\mathbf{\Pi}}\,\mathbf{H}\,\mathbf{D}\mathbf{H}^\top\sqrt{\mathbf{\Pi}}, \quad \text{that is} \quad K_{ij} = \sqrt{f_i f_j}B_{ij}. \tag{16}$$

By construction, the kernel $\mathbf{K}$ is *symmetric*, and *centered* in the sense

$$\mathbf{K}\sqrt{\mathbf{f}} = \mathbf{0}_n \qquad (\text{in view of } \mathbf{B}\mathbf{1}_n = \mathbf{0}_n), \tag{17}$$

Table 1. Values of the autocorrelation index δ (12), together with its standardized value z (41)

	$\mathbf{W}_a$		$\mathbf{W}_b$		$\mathbf{W}_c$		$\mathbf{W}_{\mathbb{D}}$	
	δ	z	δ	z	δ	z	δ	z
X	0.96	10.51	0.75	15.59	0.55	14.21	0.094	17.57
Y	0.94	7.78	0.67	10.88	0.42	8.94	0.078	11.73
x	0.95	5.64	0.66	7.14	0.45	6.64	0.131	14.50

Note: Rows refer to the three sets of features considered in section Illustration: political and social autocorrelation among French departments. Columns refer to the three spatial weights considered in Definition 1 with $s = s_2$ for $\mathbf{W}_a$, $g = 0.001$ for $\mathbf{W}_c$, as well as $\mathbf{W}_{\mathbb{D}}$ in Definition 2 with $c = c_1$.

as well as p.s.d, the latter property characterizing the squared Euclidean nature of $\mathbf{D}$.

4. Fourth, perform the spectral decomposition of the kernel $\mathbf{K}$. (17) shows $\mathbf{K}$ to possess a *trivial eigenvalue*denoted $\lambda_0 = 0$ associated to the *trivial eigenvector* denoted $\mathbf{u}_0 = \sqrt{\mathbf{f}}$. The remaining $n-1$ nontrivial eigenvalues are ordered as $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_{n-1} \geq 0$, among which $r = \text{rank}(\mathbf{K})$ are strictly positive. The spectral decomposition reads

$$\mathbf{K} = \mathbf{U}\boldsymbol{\Lambda}\mathbf{U}^\top, \tag{18}$$

where $\boldsymbol{\Lambda} = \text{diag}(\lambda_1, \ldots, \lambda_{n-1})$ and $\mathbf{U} = (\mathbf{u}_1 | \ldots | \mathbf{u}_{n-1}) \in \mathbb{R}^{n\times(n-1)}$ contains the associated $n-1$ nontrivial eigenvectors.

5. Finally, the sought-after feature regional coordinates obtain as

$$\tilde{\mathbf{x}} = \boldsymbol{\Pi}^{-\frac{1}{2}}\mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}} \qquad \text{that is,} \quad \tilde{x}_{i\alpha} = \frac{1}{\sqrt{f_i}} u_{i\alpha}\sqrt{\lambda_\alpha} \qquad \text{for } \alpha = 1, \ldots, n-1. \tag{19}$$

By construction, $\Delta = \text{Tr}(\mathbf{K}) = \sum_{\alpha=1}^{n-1} \lambda_\alpha$, and λ_α / Δ gives the proportion of inertia explained by factor α.

Kernels play a central role in the forthcoming developments, permitting in particular to provide an alternative expression for the autocorrelation index δ (section Expressing δ by kernels). Note that, whether $\mathbf{D}$ is squared Euclidean or not, definition (16) always yields a well-defined kernel $\mathbf{K} = (K_{ij})$ obeying $\mathbf{K} = \mathbf{K}^\top$ and $\mathbf{K}\sqrt{\mathbf{f}} = \mathbf{0}_n$. From $\mathbf{K}$ (respectively $\mathbf{B}$), the regional weights $\mathbf{f}$ obtain as the square of the unique positive normalized eigenvector $\sqrt{\mathbf{f}}$ in the null space of $\mathbf{K}$ (respectively $\mathbf{f}$ in the null space of $\mathbf{B}$), and dissimilarities can be recovered as

$$d_{ij} = B_{ii} + B_{jj} - 2B_{ij} = \frac{K_{ii}}{f_i} + \frac{K_{jj}}{f_j} - 2\frac{K_{ij}}{\sqrt{f_i f_j}}. \tag{20}$$

In conclusion, any weighted configuration $(\mathbf{f}, \mathbf{D})$ yields a unique, symmetric, centered kernel $\mathbf{K}$, and conversely. Moreover, $\mathbf{K}$ is p.s.d. iff $\mathbf{D}$ is squared Euclidean, that is iff the configuration can be visualized by means of (19).

Spatial kernels: visualizing the geographic space from spatial weights

Spatial weights $\mathbf{W}$ quantify neighborhood relations between regions, and a possible extraction of *spatial kernels* $\mathbf{K}_W$, apt to provide, by applying the weighted MDS of the previous section, a visualization of the *spatial* configuration, is provided by the following:

Theorem 2 whose proof is immediate. *Let the row-standardized spatial weights matrix* $\mathbf{W} \in \mathbb{R}^{n\times n}$ *be adjusted and reversible, that is satisfy (2) and (3). Define the associated spatial kernel* $\mathbf{K}_W$ *as the "Pearson residuals" of the exchange matrix* $\mathbf{E}$, *that is,*

$$\mathbf{K}_W = \mathbf{\Pi}^{-\frac{1}{2}}(\mathbf{E} - \mathbf{f}\mathbf{f}^\top)\mathbf{\Pi}^{-\frac{1}{2}} = \mathbf{\Pi}^{\frac{1}{2}}\mathbf{W}\mathbf{\Pi}^{-\frac{1}{2}} - \sqrt{\mathbf{f}}\sqrt{\mathbf{f}}^\top, \quad \text{i.e. } K_{ij}^W = \frac{e_{ij} - f_i f_j}{\sqrt{f_i f_j}} = \frac{\sqrt{f_i}\, w_{ij}}{\sqrt{f_j}} - \sqrt{f_i f_j}. \tag{21}$$

Then $\mathbf{K}_W$ *is symmetric and centered* ($\mathbf{K}_W\sqrt{\mathbf{f}} = \mathbf{0}_n$). *Its nontrivial eigenvalues* μ_α ($\alpha = 1, \dots, n-1$) *coincide with the non-trivial eigenvalues of* $\mathbf{W}$. *Yet, to the trivial eigenvalue* $\mu_0 = 0$ *of* $\mathbf{K}_W$ *corresponds the Perron-Frobenius eigenvalue of 1 for* $\mathbf{W}$. *In consequence,* $\mathrm{Tr}(\mathbf{W}) = \mathrm{Tr}(\mathbf{K}) + 1$.

Also, $\mathbf{K}_W$ *is p.s.d. iff* $\mathbf{W}$ *(or equivalently* $\mathbf{E}$*) is p.s.d., that is iff all the eigenvalues* μ_α *of* $\mathbf{W}$ *are non-negative.*

Note that off-diagonal spatial weight matrices are never p.s.d. since $\mathrm{Tr}(\mathbf{W}) = \sum_{\alpha\geq 1}\mu_\alpha = 0$. In definition (1), $\mathbf{W}_a$ in (8) can readily seen to be positive semi-definite (p.s.d.) iff $0 < t \leq t_2$ where $t_2 := 1/\ell$ and ℓ denotes the largest eigenvalue of the adjusted Laplacian $\mathcal{L}\mathbf{A}$; also $t_2 \leq t_1 = \min_{i=1}^n(f_i/a_{i\bullet})$ as a consequence of the Schur-Horn theorem (see, e.g., Bhatia 2001) applied on the p.s.d. matrix $\mathcal{L}\mathbf{A}$. By contrast, $\mathbf{W}_b$ in (9) is generally not p.s.d., and $\mathbf{W}_c$ in (10) cannot be p.s.d. if g is small enough.

This being so, any adjusted and reversible spatial weights matrix $\mathbf{W}$ can be made p.s.d. by simply replacing $\mathbf{W}$ with its matrix square $\mathbf{W}^2$, which is adjusted and reversible as well, with eigenvalues $\mu_\alpha^2 \geq 0$. This amounts in replacing $\mathbf{K}_W$ by its square $\mathbf{K}_W^2$ in view of the identity

$$\mathbf{K}_W^q = \mathbf{\Pi}^{\frac{1}{2}}\mathbf{W}^q\mathbf{\Pi}^{-\frac{1}{2}} - \sqrt{\mathbf{f}}\sqrt{\mathbf{f}} \quad q = 1, 2 \dots . \tag{22}$$

If the spatial weights $\mathbf{W}$ are p.s.d., the spectral decomposition

$$\mathbf{K}_W = \mathbf{V}\mathbf{M}\mathbf{V}^\top \qquad \text{with} \quad \mathbf{V} = (v_{i\alpha}) \in \mathbb{R}^{n\times(n-1)}, \mathbf{V}^\top\mathbf{V} = \mathbf{I}_n \quad \text{and} \quad \mathbf{M} = \mathrm{diag}(\mu_1, \dots, \mu_{n-1}), \tag{23}$$

yields, as in (19), *spatial factorial coordinates* of the form

$$\hat{\mathbf{X}} = \mathbf{\Pi}^{-\frac{1}{2}}\mathbf{V}\mathbf{M}^{\frac{1}{2}} \qquad \text{i.e.} \quad \hat{x}_{i\alpha} = \frac{1}{\sqrt{f_i}} v_{i\alpha}\sqrt{\mu_\alpha} \qquad \text{for } \alpha = 1, \dots, n-1. \tag{24}$$

The present use of spectral decomposition for visualizing spatial networks, illustrated in Figs. 1 and 4 below, follows a long geographical tradition (see, e.g., Boots 1982; Griffith 1996; Dray, Legendre, and Peres-Neto 2006; Demšar et al. 2013).

The spatial kernels (21) associated to the two limit Markov chains (4), as well as the associated dissimilarities 20 are

$$\mathbf{K}_0 = \mathbf{I}_n - \sqrt{\mathbf{f}}\sqrt{\mathbf{f}} \quad , \quad d_{ij}^0 = \begin{cases} \frac{1}{f_i} + \frac{1}{f_j}, & \text{for } i \neq j, \\ 0, & \text{otherwise.} \end{cases} \quad , \qquad \mathbf{K}_\infty \equiv \mathbf{D}_\infty \equiv \mathbf{0}_{n\times n} \tag{25}$$

The nontrivial eigenvalues of $\mathbf{K}_0$ are $\mu_1 = \cdots = \mu_{n-1} = 1$ (constant scree graph), and $\mathbf{D}_0$ is the weighted extension of the so-called discrete distances, characterizing a maximally incompressible configuration (see also section Testing δ by invariant orthogonal integration).

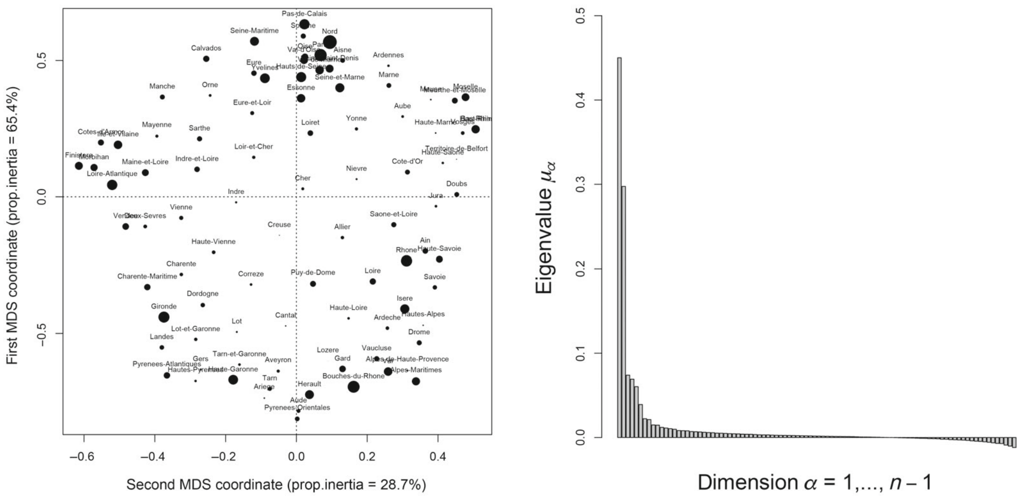

Figure 4. Left: MDS on $K_{\mathbb{D}}^2$, where $K_{\mathbb{D}}$ is the kernel associated to the shortest-path distances $\mathbb{D}$ on the adjacency network. Right: eigenvalues $\{\mu_\alpha\}$ of $K_{\mathbb{D}}$, some of which are negative, thus ruining the squared Euclidean nature of $\mathbb{D}$. Spectral moments are $\overline{\boldsymbol{\mu}} = 0.023$, $v(\lambda) = 0.011$, $a(\boldsymbol{\mu}) = 6.45$ and $\gamma(\boldsymbol{\mu}) = 43.20$.

Expressing δ by kernels

The following result together with its implications constitutes the main motivation for the present article:

Theorem 3 *Let* $\mathbf{K}_X$ *in (16) be the standard kernel associated to the squared Euclidean feature dissimilarity, denoted as* $\mathbf{D}_X$*, and let* $\mathbf{K}_W$ *in (21) be the spatial kernel associated to the adjusted and reversible spatial weights* $\mathbf{W}$*. Then the δ index of spatial autocorrelation (12) expresses as*

$$\delta = \frac{\mathrm{Tr}(\mathbf{K}_W\ \mathbf{K}_X)}{\mathrm{Tr}(\mathbf{K}_X)}. \tag{26}$$

Proof. By (16), and using $\mathrm{Tr}(\mathbf{QR}) = \mathrm{Tr}(\mathbf{RQ})$ and $\mathbf{H}^\top \boldsymbol{\Pi} \mathbf{H} = \boldsymbol{\Pi} - \mathbf{f}\mathbf{f}^\top$,

$$\mathrm{Tr}(\mathbf{K}_X) = \mathrm{Tr}(\boldsymbol{\Pi}\ \mathbf{B}_X) = \frac{1}{2}\mathrm{Tr}([\mathbf{f}\mathbf{f}^\top - \boldsymbol{\Pi}]\mathbf{D}_X) = \frac{1}{2}\mathrm{Tr}(\mathbf{D}_X\ \mathbf{f}\mathbf{f}^\top) = \frac{1}{2}\sum_{ij} f_i f_j D_{ij}^X = \Delta.$$

On the other hand, using $\mathbf{H}^\top \mathbf{f} = \mathbf{0}_n$ and $\mathbf{H}^\top \mathbf{E} \mathbf{H} = \mathbf{E} - \mathbf{f}\mathbf{f}^\top$,

$$\mathrm{Tr}(\mathbf{K}_W\ \mathbf{K}_X) = \mathrm{Tr}(\mathbf{E}\ \mathbf{B}_X) = \frac{1}{2}\mathrm{Tr}([\mathbf{f}\mathbf{f}^\top - \mathbf{E}]\mathbf{D}_X) = \Delta - \Delta_{\mathrm{loc}}.$$

□

Expression (26) demonstrates that the regional features on one hand and the regional spatial configuration on the other hand, whose interplay constitutes the study of spatial autocorrelation, can be expressed by the *same formalism*, namely by means of kernels. Also, (26) bears an obvious relationship to the weighted extension of the so-called *RV coefficient* (Bavaud 2023), defined as the cosine similarity between the vectorized matrices $\mathbf{K}_W$ and $\mathbf{K}_X$, namely

$$\mathtt{RV}_{WX} = \mathtt{RV} = \frac{\mathrm{Tr}(\mathbf{K}_W\ \mathbf{K}_X)}{\sqrt{\mathrm{Tr}(\mathbf{K}_W^2)\ \mathrm{Tr}(\mathbf{K}_X^2)}} = \delta\ \frac{\mathrm{Tr}(\mathbf{K}_X)}{\sqrt{\mathrm{Tr}(\mathbf{K}_W^2)\ \mathrm{Tr}(\mathbf{K}_X^2)}} \quad \text{i.e.} \quad \delta = \mathtt{RV}\ \frac{\sqrt{\mathrm{Tr}(\mathbf{K}_W^2)\ \mathrm{Tr}(\mathbf{K}_X^2)}}{\mathrm{Tr}(\mathbf{K}_X)}. \tag{27}$$

The original "random vector" RV coefficient has been proposed by Escoufier (1973) and Robert and Escoufier (1976) as a symmetric, broadly applicable measure of similarity between two multivariate configurations, boiling down to the square of the correlation for two univariate configurations. Its significance testing, which has attracted numerous studies, can be directly transposed into a significance testing procedure for δ (section Testing δ by invariant orthogonal integration).

By the Cauchy–Schwarz inequality, $\mathtt{RV} \in [-1, 1]$, with $\mathtt{RV} \in [0, 1]$ if both $\mathbf{K}_W$ and $\mathbf{K}_X$ are p.s.d. By contrast, and provided $\mathbf{K}_X$ is p.s.d., $\delta \in [-1, 1]$ as a consequence of the Perron–Frobenius theorem (see, e.g., Levin and Peres 2017), implying that all eigenvalues μ_α of the reversible $\mathbf{W}$ are real, and contained in $[-1, 1]$. In addition $\delta \in [0, 1]$ if both $\mathbf{K}_W$ and $\mathbf{K}_X$ are p.s.d. Note in passing that while δ on one hand, and the RV and the Pearson correlation on the other hand share the same range of values $[-1, 1]$, they do it for completely different reasons. In particular, taking too literally the analogy between δ and a correlation coefficient arguably does more harm than good: in the simplest case of two univariate kernels, one finds $\mathrm{Tr}(\mathbf{K}_X\ \mathbf{K}_Y) = \mathrm{cov}^2(x, y)$.

Alike the RV, the δ coefficient of spatial is a measure of similarity between a feature configuration as depicted in Fig. 5 and a spatial configuration as depicted in Fig. 1. Yet, taken alone, the value of δ is very little informative. In particular, as illustrated in Table 1, δ mechanically increases with $\mathrm{Tr}(\mathbf{W})$, and the knowledge of its expected value $\mathbb{E}(\delta)$ under the null hypothesis, as well as of its higher-order expected moments, provided in section Testing δ by invariant orthogonal integration, is crucial to make δ helpful and interpretable.

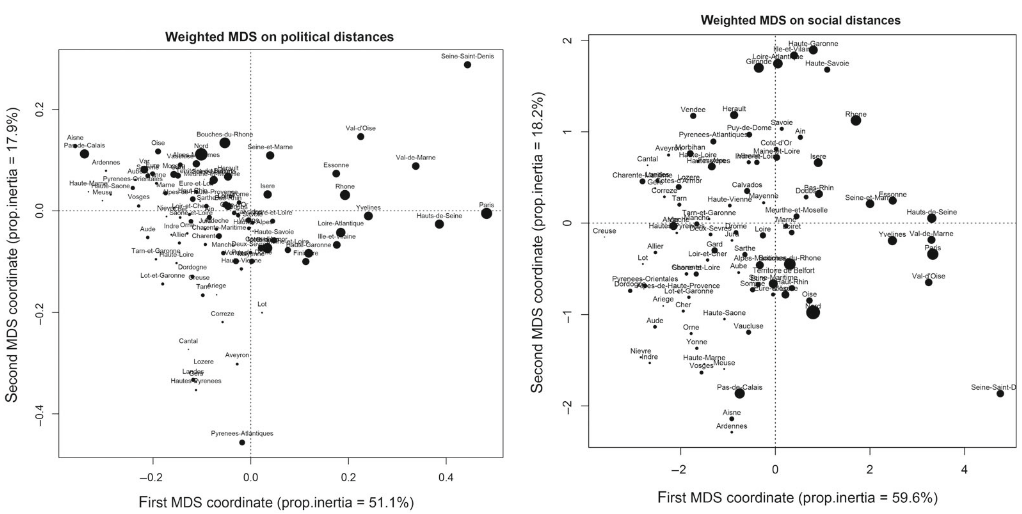

Figure 5. Left: weighted MDS of the political configuration $(\mathbf{f}, \mathbf{D}_X)$, whose first factor $\tilde{\mathbf{x}}_1$ roughly expresses a political right-left gradient. Right: weighted MDS of the social configuration $(\mathbf{f}, \mathbf{D}_Y)$, whose first factor $\tilde{\mathbf{x}}_1$ opposes aging departments (left) to departments with a large foreign population and large natural balance (right).

Nonpositive semi-definite kernels

Obtaining factor coordinates permitting to *visualize* the features configuration by (19), or the spatial configuration by (24) requires $\mathbf{K}_X$, respectively $\mathbf{K}_W$, to be p.s.d. By contrast, the possible nonpositive semi-definite nature of $\mathbf{K}_X$ and/or $\mathbf{K}_W$ poses no difficulty whatsoever in *defining* and *testing* the spatial autocorrelation index δ (26): in (16), $\mathbf{D}$ can very well represent a dissimilarity matrix which is not squared Euclidean, and in (21), $\mathbf{W}$ can very well contain negative eigenvalues, as it is the case for traditional off-diagonal spatial weights obeying $\mathrm{Tr}(\mathbf{W}) = 0$.

If $\mathbf{K}_X$ is not p.s.d., the plain discarding of its negative eigenvalues, set to zero, is sometimes advocated in MDS, and justified by the the fact that the resulting *truncated matrix* constitutes the closest p.s.d. approximation of $\mathbf{K}_X$ in the Frobenius norm (Higham 1988; Cheng and Higham 1998). Another popular modification advocated in non-weighted MDS consists in adding a constant squared Euclidean dissimilarity among all pairs of distinct objects (see, e.g., Borg and Groenen 2005). In the weighted setting, this boils down to consider the modified kernel $\mathbf{K}_X + c\,\mathbf{K}_0$ where $\mathbf{K}_0$ is defined in (25) and c is not less than the absolute value of the smallest eigenvalue of $\mathbf{K}_X$; this modification seems to be optimal in the 2-norm (Higham 1988; Cheng and Higham 1998). Finally, replacing the kernel by its matrix square $\mathbf{K}_X^2$ evidently yields a p.s.d. kernel, associated to a *flattened* configuration in which directions of large (resp. small) dispersions are amplified (resp. downscaled).

If $\mathbf{K}_W$ or equivalently $\mathbf{W}$ is not p.s.d., plain spectral truncation will in general ruin the Markovian nature of the modified spatial weights. However, the modification $s\,\mathbf{K}_W + (1-s)\mathbf{K}_0$, or equivalently $s\,\mathbf{W} + (1-s)\mathbf{I}_n$, we shall refer to as *stayers mixing*, possess many virtues: it insures positive semi-definiteness, for $s \in (0, 1/(1-\mu_{n-1})]$ (note that $1/(1-\mu_{n-1}) \geq 0.5$), while respecting the Markovian character of the modified transition matrix, which obtains as a mixture of the original matrix (selected with probability s) and the pure stayers matrix $\mathbf{W}_0 = \mathbf{I}_n$ (selected with probability $1-s$). Furthermore, stayers mixing possesses the pleasant property of keeping the values of the standardized autocorrelation index z defined below in (41) unchanged, as well as the expected skewness and excess kurtosis of δ, as stated in Theorem 5 below. In other words, the stayers mixing modification enables an Euclidean visualization of the spatial configuration, without affecting the tests (normal or Cornish–Fisher) of spatial autocorrelation. Finally, the simple *iteration*, already encountered and regularly used in this article when needed, replaces the spatial weights matrix $\mathbf{W}$ by $\mathbf{W}^2$. This transformation enables Euclidean spatial visualization, although flattened, respecting the reversible, weight-compatible Markovian nature of the weights, and endowed with an intuitively simple geographical interpretation ("the neighbors of the neighbors").

A toy example

Consider a territory made of $n = 4$ regions with weights $\mathbf{f}^\top = (0.4, 0.1, 0.3, 0.2)$ and characterized by a single attribute $\mathbf{x}$. In example A (Fig. 6, left), Markov transitions occur between distinct,

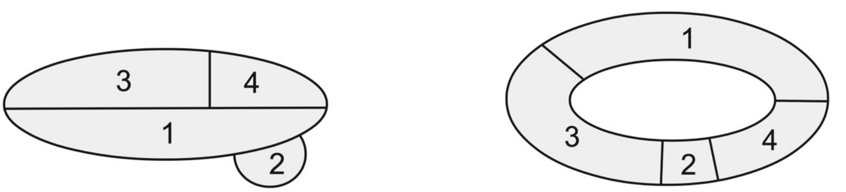

Figure 6. Left: spatial layout for examples A and B. Right: spatial layout for example C.

adjacent regions, with an intensity roughly proportional to the size of the common boundary. In example B, either no transition occurs with probability 0.5, or a transition following the previous scheme occurs with probability 0.5. In example C (Fig. 6, right), Markov transitions occur between distinct, adjacent regions, with an intensity roughly proportional to the area of the destination region. Numerically,

$$\mathbf{x}_A = \begin{pmatrix} 4 \\ 9 \\ 1 \\ 1 \end{pmatrix} \quad \mathbf{W}_A = \begin{pmatrix} 0 & 1/4 & 1/2 & 1/4 \\ 1 & 0 & 0 & 0 \\ 2/3 & 0 & 0 & 1/3 \\ 1/2 & 0 & 1/2 & 0 \end{pmatrix} \quad \mathbf{E}_A = \begin{pmatrix} 0 & 0.1 & 0.2 & 0.1 \\ 0.1 & 0 & 0 & 0 \\ 0.2 & 0 & 0 & 0.1 \\ 0.1 & 0 & 0.1 & 0 \end{pmatrix}$$

$$\mathbf{x}_B = \begin{pmatrix} 4 \\ 9 \\ 1 \\ 1 \end{pmatrix} \quad \mathbf{W}_B = \begin{pmatrix} 1/2 & 1/8 & 1/8 & 1/4 \\ 1/2 & 1/2 & 0 & 0 \\ 1/3 & 0 & 1/2 & 1/6 \\ 1/4 & 0 & 1/4 & 1/2 \end{pmatrix} \quad \mathbf{E}_B = \begin{pmatrix} 0.2 & 0.05 & 0.1 & 0.05 \\ 0.05 & 0.05 & 0 & 0 \\ 0.1 & 0 & 0.15 & 0.05 \\ 0.05 & 0 & 0.05 & 0.1 \end{pmatrix}$$

$$\mathbf{x}_C = \begin{pmatrix} 3 \\ 3 \\ 1 \\ 1 \end{pmatrix} \quad \mathbf{W}_C = \begin{pmatrix} 0 & 0 & 0.6 & 0.4 \\ 0 & 0 & 0.6 & 0.4 \\ 0.8 & 0.2 & 0 & 0 \\ 0.8 & 0.2 & 0 & 0 \end{pmatrix} \quad \mathbf{E}_C = \begin{pmatrix} 0 & 0 & 0.24 & 0.16 \\ 0 & 0 & 0.06 & 0.04 \\ 0.24 & 0.06 & 0 & 0 \\ 0.16 & 0.04 & 0 & 0 \end{pmatrix}.$$

Spatial weights $\mathbf{W}$ are readily checked to be row-normalized. The corresponding exchange matrices $\mathbf{E} = \mathbf{\Pi W}$, whose components specify the joint probability to visit a pair of regions, are symmetric with margins $\mathbf{f}$, thus proving that $\mathbf{W}$ is weight-compatible and reversible.

By construction, the univariate feature kernels (16) contain a single nonzero, positive eigenvalue. Also, the spatial kernel (21) turns out to be p.s.d. for configuration B, but not for configurations A and C which involve negative eigenvalues. Computing the autocorrelation indices (26) yields $\delta_A = 0.13$ and $\delta_B = 0.57$. Also, the minimal autocorrelation value is attained with $\delta_C = -1$, which can be proved to necessarily involve *bipartite* spatial networks, partitioning the regions into two complementary sets of same importance, namely $\{1, 2\}$ and $\{3, 4\}$ in Fig. 6 right, so that nonzero transitions necessarily occur between pairs of regions belonging to complementary sets.

Note $\mathbf{W}_B$ to result from the stayers mixture of $\mathbf{W}_A$ with $t = 0.5$ (section Nonpositive semi-definite kernels). As a consequence (Theorem 5), their standardized autocorrelation index (41) coincide: $z_A = z_B = 1.98$, denoting positive autocorrelation, with a two-tailed p-value of 0.048 in the normal approximation test (44)). By contrast, $z_C = -0.37$ (negative autocorrelation, with $p = 0.025$).

Decompositions of δ

δ can be expressed as convex mixtures of Moran's I. First, simple algebra demonstrates that, for features dissimilarities $\mathbf{D}_Y$ of the form (35),

$$\delta = \frac{\sum_{k=1}^{p} \mathrm{var}(y_k) I(y_k)}{\sum_{k=1}^{p} \mathrm{var}(y_k)} = \frac{1}{\Delta_Y} \sum_{k=1}^{p} \mathrm{var}(y_k) I(y_k), \tag{28}$$

where $I(y_k)$ denotes Moran's associated to the kth feature, and $\mathrm{var}(y_k)$ its variance. Also,

$$\delta = \frac{\sum_{\alpha=1}^{n-1} \lambda_\alpha I(\hat{y}_\alpha)}{\sum_{\alpha=1}^{n-1} \lambda_\alpha} = \frac{1}{\Delta_Y} \sum_{\alpha=1}^{n-1} \lambda_\alpha I(\hat{y}_\alpha), \tag{29}$$

where $\hat{y}_{i\alpha} = u_{i\alpha}\sqrt{\lambda_\alpha}/\sqrt{f_i}$ are the feature factorial coordinates in (19). In particular, the Moran indices $I(\hat{y}_\alpha)$ coincide with the slopes of the corresponding Moran–Anselin regression lines (see Fig. 2). Also, *local indicators of spatial autocorrelation* (Anselin 1995) δ_i can be constructed as follow:

$$\delta = \sum_{i=1}^{n} f_i\ \delta_i \qquad \text{where} \quad \delta_i = \frac{1}{f_i \Delta_Y} \sum_{j=1}^{n} K_{ij}^W K_{ij}^Y = \frac{1}{\Delta_Y} (\mathbf{W}\ \mathbf{B}_Y)_{ii}, \tag{30}$$

where $\mathbf{B}_Y$ is the matrix of scalar products (15) associated to $\mathbf{D}_Y$, and the last identity follows form (21). Finally, δ can be further doubly decomposed into local and factorial components as

$$\delta = \frac{1}{\Delta_Y} \sum_{i=1}^{n} f_i \sum_{\alpha=1}^{n-1} \lambda_\alpha \delta_{i\alpha} \qquad \text{where} \quad \delta_{i\alpha} = \sum_{j=1}^{n} \frac{u_{i\alpha} w_{ij} u_{j\alpha}}{\sqrt{f_i f_j}}. \tag{31}$$

A direct approach to spatial autocorrelation in presence of geographic distances

Consider finally the frequent encountered case where space is specified by a matrix of symmetric *geographic distances* $\mathbb{D} = (\mathbb{d}_{ij}) \in \mathbb{R}^{n\times n}$, such as road distances, travel costs or travel times, obeying (11). Those distances are in general metric, that is, obey the triangular inequality, but not squared Euclidean. In any case, the construction (16) of the feature kernel can be reused to define a spatial kernel together with a spatial autocorrelation index δ (26) as

$$\delta = \frac{\mathrm{Tr}(\mathbf{K}_{\mathbb{D}}\ \mathbf{K}_X)}{\mathrm{Tr}(\mathbf{K}_X)} \qquad \text{where} \quad \mathbf{K}_{\mathbb{D}} = -\frac{1}{2}\sqrt{\mathbf{\Pi}}\ \mathbf{H}\ \mathbb{D}\mathbf{H}^\top \sqrt{\mathbf{\Pi}}, \tag{32}$$

without requiring any matrix of spatial weights. In particular,

$$\mathrm{Tr}(\mathbf{K}_{\mathbb{D}}) = \Delta = \frac{1}{2} \sum_{i,j=1}^{n} f_i f_j\ \mathbb{d}_{ij} = \sum_{\alpha=1}^{n-1} \mu_\alpha$$

mesures the overall spatial dispersion, where $\{\mu_\alpha\}$ denote the non-trivial eigenvalues of $\mathbf{K}_{\mathbb{D}}$. In general, some eigenvalues are negative (see Fig. 4 right), that is $\mathbb{D}$ is not squared Euclidean. Again, if one wishes to visualize the geographical configuration by MDS, one can replace $\mathbf{K}_{\mathbb{D}}$ by $\mathbf{K}_{\mathbb{D}}^2$, whose effect is to transform μ_α into $\mu_\alpha^2 \geq 0$ while keeping the eigenvectors of $\mathbf{K}_{\mathbb{D}}$ unchanged.

δ in (32) is a perfectly valid index of spatial autocorrelation, whose expected moments and significance testing are provided by Theorem 4 below (where $\mathbf{K}_W$ is simply replaced by the spatial kernel $\mathbf{K}_{\mathbb{D}}$) and (45). However, its value clearly depends upon the units used to express geographical distances: for instance, δ becomes a thousand times larger when $\mathbb{D}$ is measured in meters rather than in kilometers, and $\delta \leq 1$ does not hold in general. To control for this arbitrariness, let us replace $\mathbb{D}$ and $\mathbf{K}_{\mathbb{D}}$ by $c\ \mathbb{D}$ and $c\ \mathbf{K}_{\mathbb{D}}$, respectively, where the positive parameter c modulates the distance scale.

It is natural to investigate whether a matrix of spatial weights can be obtained from the $c\ \mathbf{K}_{\mathbb{D}}$ by inverting relation (21). This motivates the following definition:

Definition 2 Obtaining spatial weights from geographic distances Let $\mathbf{W}_{\mathbb{D}} = (w_{ij}^{\mathbb{D}})$ where

$$w_{ij}^{\mathbb{D}} = f_j + c\,\frac{\sqrt{f_j}}{\sqrt{f_i}}\,K_{ij}^{\mathbb{D}} = (1 + c\,B_{ij}^{\mathbb{D}})f_j \quad \text{where} \quad c \in (0, c_1] \quad \text{and} \quad c_1 = \frac{-1}{\min\limits_{ij} B_{ij}^{\mathbb{D}}} > 0, \tag{33}$$

where $\mathbf{B}_{\mathbb{D}}$ is the matrix of scalar products (15) associated to $\mathbb{D}$.

By construction, $\mathbf{W}_{\mathbb{D}}$ is a nonnegative, row-normalized, adjusted and reversible matrix of spatial weights.

Illustration: political and social autocorrelation among French departments

Consider the $n = 94$ continental French departments (that is excluding Corsica and overseas territories). The inter-departmental binary adjacency matrix $\mathbf{A} = (a_{ij})$ happens to yield a maximum degree of $\max_i a_{i\bullet} = 10$ (Seine-et-Marne), a minimum degree of $\min_i a_{i\bullet} = 2$ (Alpes-Maritimes, Finistère, Haute-Savoie, Moselle, Pas-de-Calais, Pyrénées-Orientales), and an average degree of $a_{\bullet\bullet}/n = 5.06$.

Define the regional weights $\mathbf{f}$ as the corresponding proportion of voters in the 2022 presidential primary election (INSEE 2022). They range from $\min_i f_i = 0.0014$ (Lozère) to $\max_i f_i = 0.037$ (Nord).

The pair $(\mathbf{f}, \mathbf{A})$ permits to define the three weight-compatible, reversible spatial weights of Definition 1: linearized diffusive weights $\mathbf{W}_a$ (8) (with $t = t_2$, defined after Theorem 2), Metropolis–Hastings weights $\mathbf{W}_b$ (9), and iteratively fitted weights $\mathbf{W}_c$ (10) (with $g = 0.001$).

In addition, a fourth geographical distance weights $\mathbf{W}_{\mathbb{D}}$ can be derived from (33) (with $c = c_1$) by choosing $\mathbb{D}$ as the shortest-path distance on the adjacency graph of French departments, that is, by defining $\mathbb{d}_{ij} = \mathbb{d}_{ji}$ as the minimum number of crossed inter-departmental borders needed to attain department j from department i. Its maximum turns out to be $\max_{ij} \mathbb{d}_{ij} = 11$, attained along the North-East South-West direction.

Among those four spatial weights, only $\mathbf{W}_a$ is p.s.d., and the first spatial coordinates (24), resulting from the eigen-decomposition of $\mathbf{K}_W$ in (21), are depicted in Fig. 1 (top left). The other spatial weights are first squared, then treated similarly, with resulting spatial coordinates depicted in Figs. 1 and 4.

Consider in addition three feature kernels, respectively, denoted $\mathbf{K}_X$, $\mathbf{K}_Y$, and $\mathbf{K}_x$, based upon three squared Euclidean dissimilarities between departments $\mathbf{D}$, namely:

(X) *political data* : the $n \times m$ contingency table $\mathbf{N} = (n_{ik})$ counts the number of votes n_{ik} obtained in department i for candidate k, among the $m = 12$ candidates of the 2022 presidential primary election (INSEE 2022). Define, here in what follows, the regional weights $\mathbf{f}$ as the proportion of voters in each department, and define the squared Euclidean political distance $\mathbf{D}_X$ as chi-square dissimilarity between departments (see, e.g., Lebart, Morineau, and Warwick 1984; Greenacre 2017), that is:

$$f_i = \frac{n_{i\bullet}}{n_{\bullet\bullet}}, \qquad d_{ij}^X = \sum_{k=1}^{m} \frac{n_{\bullet\bullet}}{n_{\bullet k}} \left(\frac{n_{ik}}{n_{i\bullet}} - \frac{n_{jk}}{n_{j\bullet}} \right)^2. \tag{34}$$

(Y) *social data* : the $n \times p$ matrix $\mathbf{Y} = (y_{ik})$ contains $p = 5$ variables expressed by percentages: “natural demographic balance,” “migratory demographic balance,” “population over 65,” “foreign population,” and “inactive young people” around 2015 (INSEE 2022). Those sociodemographic profiles directly serve in defining a rudimentary squared Euclidean social distance $\mathbf{D}_Y$ between departments as

$$d_{ij}^{Y} = \sum_{k=1}^{p} (y_{ik} - y_{jk})^2. \quad (35)$$

The feature dissimilarities $\mathbf{D}_X$ and $\mathbf{D}_Y$ are multidimensional: depicting all the nonzero corresponding MDS coordinates (19) requires more than one dimension. In addition, one considers the univariate case

(x) *share of votes for the candidate Macron*: denoted $\mathbf{x} \in \mathbb{R}^n$, with components $x_i = n_{i\,\text{Macron}}/n_{i\bullet}$, and corresponding dissimilarities $d_{ij}^{x} = (x_i - x_j)^2$.

Weighted MDS on the political dissimilarities $\mathbf{D}_X$ above turns out to be equivalent to Correspondence Analysis (see, e.g., Lebart, Morineau, and Warwick 1984; Greenacre 2017), while weighted MDS social dissimilarities $\mathbf{D}_Y$ amont to weighted Principal Component Analysis. Their optimal visualization in dimensions $\alpha = 1, 2$ is depicted in Fig. 5.

The corresponding values of the autocorrelation index δ (12) depend on both the set of regional features and on the choice of spatial weights (Table 1). The larger $\text{Tr}(\mathbf{W})$, the larger δ in general: here $\text{Tr}(\mathbf{W}_a) = 74.65$, $\text{Tr}(\mathbf{W}_b) = 24.29$, $\text{Tr}(\mathbf{W}_c) = 0.20$ and $\text{Tr}(\mathbf{W}_{\mathbb{D}}) = 2.16$ (the tendency is however inverted between $\mathbf{W}_c$ and $\mathbf{W}_{\mathbb{D}}$).

Testing δ by invariant orthogonal integration

As stated in the introduction, the two main strategies used to assess the significance of (univariate) spatial autocorrelation, namely permutation tests and normal tests, are difficult to justify in the weighted setting, where the distribution of features is bound to depend upon the region sizes (see however, e.g., Tiefelsdorf and Griffith 2007; Chun 2008; Griffith 2010; Bavaud 2013; Zhang and Lin 2016, and references therein for addressing some of the above issues).

Recently, Bavaud (2023) has proposed a new approach for testing the RV coefficient, the *invariant orthogonal integration*, based on the simple rationale that, under the null hypothesis H_0 of absence of relationship between the two multivariate configurations at stake, any relative orientation between the two nontrivial eigen-spaces is equally likely. Invariant orthogonal integration, rooted in random matrix theory, involves integrating over Haar mesures on the orthogonal groups, and the computation, under H_0, of the integer moments of the form $\mathbb{E}(\text{Tr}^q(\mathbf{K}_W\ \mathbf{K}_X))$ in the present context, become rapidly entangled for growing q. Yet, those moments have been successfully computed for $q = 1, 2, 3, 4$ (Bavaud 2023), permitting in turn to compute the first moments (mean, variance, skewness, and excess kurtosis) of the RV coefficient.

Although the computation of the above moments were initially proposed in the context of p.s.d. kernels, nothing in the formalism prevents from dealing with negative eigenvalues. Also, as the quantities in (27) relating RV and δ are simple constants unaffected by orthogonal integration, moments for δ directly obtain in the following theorem:

Theorem 4 First moments of the δ coefficient, directly adapted from corollary 1 in (Bavaud 2023) *Under invariant orthogonal integration, the null expectation, variance, skewness and excess kurtosis of δ (26) are, in order,*

$$\mathbb{E}(\delta) = \frac{\mathrm{Tr}(\mathbf{K}_W)}{n-1} = \frac{\mathrm{Tr}(\mathbf{W})-1}{n-1} = \overline{\boldsymbol{\mu}}. \tag{36}$$

$$\begin{aligned}\mathbb{V}\mathrm{ar}(\delta) &= \frac{2}{(n-2)(n-1)^2(n+1)}\left[(n-1)\mathrm{Tr}(\mathbf{K}_W^2)-\mathrm{Tr}^2(\mathbf{K}_W)\right] \\ &\left[\frac{(n-1)\mathrm{Tr}(\mathbf{K}_X^2)-\mathrm{Tr}^2(\mathbf{K}_X)}{\mathrm{Tr}^2(\mathbf{K}_X)}\right] =\end{aligned} \tag{37}$$

$$\begin{aligned}&= \underbrace{\frac{2}{n^2-1}\left[\mathrm{Tr}(\mathbf{W}^2)-1-\frac{(\mathrm{Tr}(\mathbf{W})-1)^2}{n-1}\right]}_{\mathbb{V}\mathrm{ar}(I)}\underbrace{\left[\frac{1}{n-2}\left(\frac{n-1}{\nu_X}-1\right)\right]}_{=:\kappa_X} \\ &= \frac{2}{(n-2)(n+1)}\,v(\boldsymbol{\mu})\,\frac{v(\boldsymbol{\lambda})}{\overline{\boldsymbol{\lambda}}^2}.\end{aligned} \tag{38}$$

$$\mathbb{A}(\delta) = \frac{\mathbb{E}(\delta_c^3)}{\mathbb{E}^{\frac{3}{2}}(\delta_c^2)} = \frac{\sqrt{8(n-2)(n+1)}}{(n-3)(n+3)}a(\boldsymbol{\mu})a(\boldsymbol{\lambda}). \tag{39}$$

$$\begin{aligned}\mathbb{\Gamma}(\delta) = \frac{\mathbb{E}(\delta_c^4)}{\mathbb{E}^2(\delta_c^2)}-3 &= \frac{3\,(n-2)(n+1)}{(n-4)(n-3)(n-1)n(n+3)(n+5)} \\ &\left\{4(n^2-n+2)\,\gamma(\boldsymbol{\mu})\,\gamma(\boldsymbol{\lambda})+(4n^2-8n\right. \\ &\left.+52)(\gamma(\boldsymbol{\mu})+\gamma(\boldsymbol{\lambda}))-\frac{4(5n^3-57n^2+27n+169)}{(n-2)(n+1)}\right\}.\end{aligned} \tag{40}$$

In the above expressions, $\boldsymbol{\mu}$ denotes the vector of the $n-1$ non-trivial eigenvalues of the spatial kernel $\mathbf{K}_W$, identical to those of $\mathbf{W}$, and $\boldsymbol{\lambda}$ denotes the vector of the $n-1$ nontrivial eigenvalues of the features kernel $\mathbf{K}_X$. The *spectral moments*, *spectral variance*, *spectral skewness* and the *spectral excess spectral kurtosis* are respectively denoted as

$$\overline{\boldsymbol{\mu}^q} = \frac{\sum_{\alpha=}^{n-1}\mu_\alpha^q}{n-1}, \qquad v(\boldsymbol{\mu}) = \frac{\sum_{\alpha=}^{n-1}(\mu_\alpha-\overline{\boldsymbol{\mu}})^2}{n-1} = \overline{\boldsymbol{\mu}_c^2}, \qquad a(\boldsymbol{\mu}) = \frac{\overline{\boldsymbol{\mu}_c^3}}{(\overline{\boldsymbol{\mu}_c^2})^{\frac{3}{2}}}, \qquad \gamma(\boldsymbol{\mu}) = \frac{\overline{\boldsymbol{\mu}_c^4}}{(\overline{\boldsymbol{\mu}_c^2})^2}-3,$$

where $\mu_\alpha^c = \mu_\alpha - \overline{\boldsymbol{\mu}}$. Spectral moments for $\boldsymbol{\lambda}$ are defined analogously. By construction,

$$\overline{\boldsymbol{\mu}^q} = \frac{\mathrm{Tr}\left(\mathbf{K}_W^q\right)}{n-1} = \mathrm{tr}\left(\mathbf{K}_W^q\right),$$

where $\mathrm{tr}(\mathbf{A}) = \mathrm{Tr}(\mathbf{A})/(n-1)$ denotes the normalized trace. Centered spectral moments can be transformed into normalized traces, and conversely. Specificaly, $\overline{\boldsymbol{\mu}_c^2} = \mathrm{tr}(\mathbf{K}_W^2) - \mathrm{tr}^2(\mathbf{K}_W)$, $\overline{\boldsymbol{\mu}_c^3} = \mathrm{tr}(\mathbf{K}_W^3) - 3\,\mathrm{tr}(\mathbf{K}_W^2)\,\mathrm{tr}(\mathbf{K}_W) + 2\,\mathrm{tr}^3(\mathbf{K}_W)$ and $\overline{\boldsymbol{\mu}_c^4} = \mathrm{tr}(\mathbf{K}_W^4) - 4\,\mathrm{tr}(\mathbf{K}_W^3)\,\mathrm{tr}(\mathbf{K}_W) + 6\,\mathrm{tr}(\mathbf{K}_W^2)\,\mathrm{tr}^2(\mathbf{K}_W) - 3\,\mathrm{tr}^4(\mathbf{K}_W)$, thus permitting to express, if wished, identities (39) and (40) in terms of traces of powers of $\mathbf{K}_W$ and $\mathbf{K}_X$.

Identities $\mathrm{Tr}(\mathbf{K}_W^q) = \mathrm{Tr}(\mathbf{W}^q) - 1$, resulting from (22), have been used in (36) and (38). The centered autocorrelation index δ_c appearing in (39) and (40) and the standardized autocorrelation index z appearing in Table 1 and (45) are respectively defined as

$$\delta_c = \delta - \mathbb{E}(\delta) = \frac{1}{\mathrm{Tr}}(\mathbf{K}_X)\left(\mathrm{Tr}(\mathbf{K}_W\ \mathbf{K}_X) - \frac{\mathrm{Tr}(\mathbf{K}_W)\ \mathrm{Tr}(\mathbf{K}_X)}{n-1}\right) \qquad z = \frac{\delta - \mathbb{E}(\delta)}{\sqrt{\mathbb{V}\mathrm{ar}(\delta)}}. \tag{41}$$

Inequality $\delta_c > 0$ (respectively $\delta_c < 0$) characterizes *positive* (respectively *negative*) spatial autocorrelation.

Next, the quantity v_X in (38) denotes the *effective dimensionality* of the multivariate features, defined as

$$v_X = \frac{\mathrm{Tr}^2\left(\mathbf{K}_X\right)}{\mathrm{Tr}\left(\mathbf{K}_X^2\right)} = \frac{\left(\sum_{\alpha\geq 1}\lambda_\alpha\right)^2}{\sum_{\alpha\geq 1}\lambda_\alpha^2} = (n-1)\frac{\overline{\lambda}^2}{\overline{\lambda^2}}. \tag{42}$$

Provided $\mathbf{K}_X$ is p.s.d. (that is $\mathbf{D}_X$ is squared Euclidean), the effective dimensionality ranges from its minimal value $v_X = 1$, attained for univariate features, to its maximum value $v_X = n-1$, attained iff the dissimilarities are proportional to the weighted discrete distances (25). As a consequence, the quantity $\kappa_X = [(n-1)/v_X - 1]/(n-2)$ in (38) (behaving as $\kappa_X \cong 1/v_X$ for n large) ranges from its maximum value $\kappa_X = 1$ (univariate features) to its minimum value $\kappa_X = 0$ (incompressible configuration).

Finally, the quantity $\mathbb{V}\mathrm{ar}(I)$ in (38) exactly matches the expected null variance of the weighted Moran's I under normal approximation (see eq. (14) in Bavaud 2013), and identity

$$\mathbb{V}\mathrm{ar}(\delta) = \mathbb{V}\mathrm{ar}(I)\ \kappa_X, \tag{43}$$

in (38) says that, in comparison to the univariate case, the variance of the multivariate δ index, depending only upon the spatial weights, is *downsized* by a factor $\kappa_X \in [0, 1]$ depending only upon the effective dimensionality v_X of the regional features. By contrast, $\mathbb{E}(\delta) = \mathbb{E}(I)$ (see eq. (12) in Bavaud 2013), that is the multivariate nature of δ does not affects its expected value under H_0.

Interestingly enough, one can demonstrate that $\mathbb{V}\mathrm{ar}(I) = 0$ iff the spatial weights are a convex mixture of the limit weights (4), that is iff $\mathbf{W} = s\,\mathbf{W}_0 + (1-s)\,\mathbf{W}_\infty$ for some $s \in [0, 1]$.

In the unweighted case, $f_i = 1/n$ and reversibility then implies $\mathbf{W} = \mathbf{W}^\top$. The auxiliary coefficients occurring in the derivation of Cliff and Ord (1981) pp. 35–36 under normal assumption then read, assuming row-standardized spatial weights, $S_0 = n$, $S_1 = 2\,\mathrm{Tr}(\mathbf{W}^2)$ and $S_2 = 4n$, yielding (in our notations) the well-known results (see p. 42 and eq. (2.35) in Cliff and Ord 1981)

$$\mathbb{E}_{\mathrm{CO}}(I) = \frac{-1}{n-1}\ , \qquad \mathbb{V}\mathrm{ar}_{\mathrm{CO}}(I) = \frac{2}{n^2-1}\left[\mathrm{Tr}(\mathbf{W}^2) - \frac{n}{n-1}\right],$$

which exactly coincides with (36) and (38), under the additional and popular assumption, yet much too restrictive in general, that spatial weights are off-diagonal, that is $\mathrm{Tr}(\mathbf{W}) = 0$.

Under *normal approximation*, the one-tailed significance of δ can be tested by rejecting the hypothesis of absence of spatial autocorrelation at level α if

$$z \geq u_{1-\alpha} \quad \text{where} \quad z = \frac{\delta - \mathbb{E}(\delta)}{\sqrt{\mathbb{V}\mathrm{ar}(\delta)}} \qquad \text{and } u_{1-\alpha} \text{ is the corresponding standard normal quantile.} \tag{44}$$

The second-order Cornish–Fisher cumulant expansion permits to approximatively redress the normal quantiles by taking into account the skewness and the "taildeness" of a non-normal distribution (see, e.g., Kendall and Stuart 1977; Amédée-Manesme, Barthélémy, and Maillard 2019). δ is statistically significant at level α if (one-tailed test)

$$z = \underbrace{\frac{\delta - \mathbb{E}(\delta)}{\sqrt{\mathbb{V}\text{ar}(\delta)}}}_{\text{z-score}} \geq \underbrace{u_{1-\alpha}}_{\text{standard normal quantile}}$$

$$+ \underbrace{\frac{\mathbb{A}(\delta)}{6}(u_{1-\alpha}^2 - 1) + \frac{\mathbb{\Gamma}(\delta)}{24}(u_{1-\alpha}^3 - 3u_{1-\alpha}) - \frac{\mathbb{A}^2(\delta)}{36}(2u_{1-\alpha}^3 - 5u_{1-\alpha})}_{\text{correction to the normal distribution}}. \quad (45)$$

Theorem 5 Invariance of z, $\mathbb{A}(\delta)$ and $\mathbb{\Gamma}(\delta)$ under stayers mixing. *Consider the modified spatial kernel under stayers mixing* $\widetilde{\mathbf{K}}_W = s\, \mathbf{K}_W + (1-s)\mathbf{K}_0$ *for some* $s \in (0, 1]$, *or equivalently* $\widetilde{\mathbf{W}} = s\, \mathbf{W} + (1-s)\mathbf{I}_n$, *called lazy Markov chain for s = 1/2. The resulting standardized autocorrelation index (41) as well as the third and fourth expected moments (39) and (40) are invariant under the transformation:*

$$\widetilde{z} = z \qquad \widetilde{\mathbb{A}}(\delta) = \mathbb{A}(\delta) \qquad \widetilde{\mathbb{\Gamma}}(\delta) = \mathbb{\Gamma}(\delta).$$

Proof. Under stayers mixing, the autocorrelation index (26) transforms as $\widetilde{\delta} = s\, \delta + (1-s)$, and the spatial eigenvalues as $\widetilde{\mu}_\alpha = s\, \mu_\alpha + (1-s)$. Also, $\mathbb{E}(\widetilde{\delta}) = s\, \mathbb{E}(\delta) + (1-s)$ and $v(\widetilde{\boldsymbol{\mu}}) = s^2 v(\boldsymbol{\mu})$, thus proving $\widetilde{z} = z$. Furthermore, the standardized eigenvalues $\mu_\alpha^s = (\mu_\alpha - \overline{\boldsymbol{\mu}})/\sqrt{v(\boldsymbol{\mu})}$ are left unchanged, and so are the spectral skewness $a(\boldsymbol{\mu})$ and excess spectral kurtosis $\gamma(\boldsymbol{\mu})$, and finally the expected third and fourth moments (39) and (40). □

Illustration, continued

The scree plots associated to the three feature kernels $\mathbf{K}_X$ (first round of the 2022 presidential election among the 94 French departments), $\mathbf{K}_Y$ (sociodemographic features) and $\mathbf{K}_x$ (votes for Macron) are depicted in Fig. 7. They all display small effective dimensionality ν, large skewness $a(\lambda)$ and huge excess kurtosis $\gamma(\lambda)$, as quantified in Table 2 left.

Fig. 8 depicts the scree plots associated to the three spatial kernels $\mathbf{K}_a$, $\mathbf{K}_b$ and $\mathbf{K}_c$ (Definition 1), displaying moderate skewness $a(\boldsymbol{\mu})$ and excess kurtosis $\gamma(\boldsymbol{\mu})$ of either sign

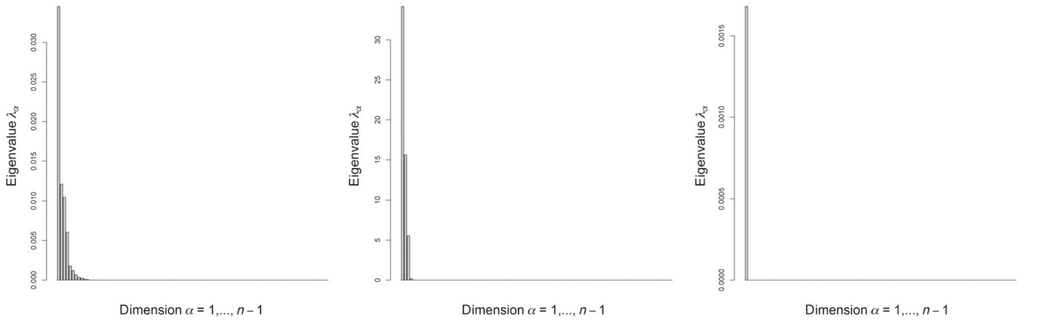

Figure 7. Scree plots of the nontrivial eigenvalues of the p.s.d. feature kernels $\mathbf{K}_X$ (left), $\mathbf{K}_Y$ (middle) and $\mathbf{K}_x$ (right). They all display small effective dimensionality ν, large skewness $a(\lambda)$ and huge excess kurtosis $\gamma(\lambda)$ (see Table 2 left).

Table 2. Spectral Moments and Other Quantities Entering in Theorem 4

	$\bar{\lambda}$	ν	κ	$a(\lambda)$	$\gamma(\lambda)$
X	0.00073	3.06	0.32	7.25	56.6
Y	0.59738	2.14	0.46	7.54	58.9
x	0.00002	1	1	9.49	88.0
	$\mathbb{E}(\delta) = \bar{\boldsymbol{\mu}}$	$\mathbb{V}\text{ar}(I)$	$a(\boldsymbol{\mu})$	$\gamma(\boldsymbol{\mu})$	
$\mathbf{W}_a$	0.7920	0.00078	-1.54	2.79	
$\mathbf{W}_b$	0.2504	0.00322	0.43	−1.10	
$\mathbf{W}_c$	−0.0086	0.00486	0.28	−0.79	
$\mathbf{W}_{\mathbb{D}}$	0.0124	0.000067	6.45	43.20	

Note: Left: features kernels. Right: spatial weights of Definitions 1 and 2, as in Table 1.

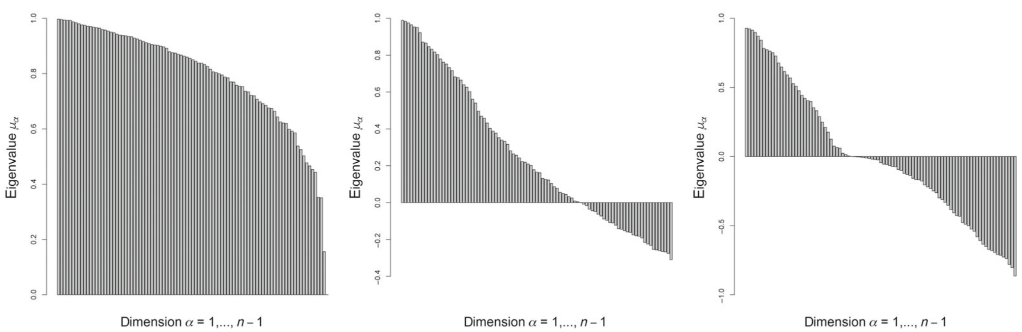

Figure 8. Scree plots of the nontrivial eigenvalues of the spatial weighs $\mathbf{W}_a$ (left), $\mathbf{W}_b$ (middle) and $\mathbf{W}_c$ (right), whose negative values characterize the non-p.s.d. nature of $\mathbf{W}_b$ and $\mathbf{W}_c$. They all display moderate skewness $a(\lambda)$ and excess kurtosis $\gamma(\lambda)$, whose signs are however contrasted (see Table 2 right): $a_a < 0$; $a_b, a_c > 0$ and $\gamma_a > 0$; $\gamma_b, \gamma_c < 0$.

(Table 2 right). By contrast, the skewness and excess kurtosis associated to the scree graph of $\mathbf{K}_{\mathbb{D}}$ (Fig. 4 right and Definition 2) are very large and positive.

Table 3 displays the expected skewness $\mathbb{A}(\delta)$ (39) and the expected excess kurtosis $\mathbb{\Gamma}(\delta)$ (40) for each combination between the feature kernels (rows) and the spatial kernels (columns). As expected from the magnitude of the spectral moments of Table 2, they take on large values (hence making the Cornish–Fisher correction important) only for the fourth spatial kernel under consideration.

Finally, Table 4 displays the p-values, computed by the Keisan online calculator, of the one-tailed significance test of the autocorrelation index δ, namely p_{normal} under normal approximation (44) and p_{CF} under the Cornish–Fisher correction (45). Both are extremely small by conventional testing standards. Also, p_{normal} and p_{CF} only differ by a few orders of magnitude, with the spectacular exception of the fourth spatial kernel (last block) where the difference is huge. All this directly reflects the magnitude of the values $\mathbb{A}(\delta)$ and $\mathbb{\Gamma}(\delta)$ in Table 3.

In general, $p_{\text{normal}} < p_{\text{CF}}$, except for the boxed values of the first block, as a consequence of the competition between $\mathbb{A}(\delta) < 0$, making a large positive z-score less probable, and $\mathbb{\Gamma}(\delta) > 0$, making it more probable.

Table 3. Expected Skewness $\mathbb{A}(\delta)$ (39) and Excess Kurtosis $\mathbb{\Gamma}(\delta)$ (40) of the Autocorrelation Index, under the Three Sets of Feature Kernels (rows) and the Four Spatial Weights (columns) Described in Table 1.

	$\mathbf{W}_a$		$\mathbf{W}_b$		$\mathbf{W}_c$		$\mathbf{W}_{\mathbb{D}}$	
	$\mathbb{A}(\delta)$	$\mathbb{\Gamma}(\delta)$	$\mathbb{A}(\delta)$	$\mathbb{\Gamma}(\delta)$	$\mathbb{A}(\delta)$	$\mathbb{\Gamma}(\delta)$	$\mathbb{A}(\delta)$	$\mathbb{\Gamma}(\delta)$
X	−0.331	0.289	0.093	−0.0095	0.060	0.0142	1.055	3.396
Y	−0.344	0.300	0.097	−0.0098	0.062	0.0148	1.097	3.529
x	−0.433	0.447	0.121	−0.0140	0.079	0.0227	1.381	5.244

Note: The null distribution of δ is slightly negatively skewed for $\mathbf{W}_a$, very slightly platykurtic for $\mathbf{W}_b$, as well as positively skewed and strongly leptokurtic for $\mathbf{W}_{\mathbb{D}}$.

Table 4. One-Tailed Significance Test of the Autocorrelation Index δ, under the Three Sets of Feature Kernels (rows) and the Four Spatial Weights Described in Table 1.

	$\mathbf{W}_a$		$\mathbf{W}_b$		$\mathbf{W}_c$		$\mathbf{W}_{\mathbb{D}}$	
	p_{normal}	p_{CF}	p_{normal}	p_{CF}	p_{normal}	p_{CF}	p_{normal}	p_{CF}
X	$4 \cdot 10^{-26}$	$2 \cdot 10^{-24}$	$4 \cdot 10^{-55}$	$2 \cdot 10^{-51}$	$4 \cdot 10^{-46}$	$7 \cdot 10^{-34}$	$2 \cdot 10^{-69}$	$4 \cdot 10^{-7}$
Y	$4 \cdot 10^{-15}$	$2 \cdot 10^{-17}$	$7 \cdot 10^{-28}$	$2 \cdot 10^{-24}$	$2 \cdot 10^{-19}$	$2 \cdot 10^{-18}$	$4 \cdot 10^{-32}$	$2 \cdot 10^{-5}$
x	$9 \cdot 10^{-9}$	$2 \cdot 10^{-11}$	$5 \cdot 10^{-13}$	$2 \cdot 10^{-11}$	$2 \cdot 10^{-11}$	$8 \cdot 10^{-10}$	$6 \cdot 10^{-48}$	$2 \cdot 10^{-5}$

Note: Here p_{normal} and p_{CF}, respectively, denote the p-values under normal approximation (44) and under the Cornish–Fisher correction (45).

Taking a step back, the main issues addressed in this article have been exemplified in depth in sections Illustration: political and social autocorrelation among French departments and Illustration, continued together with Tables and Figures, illustrating the computation of various quantities at stake, as well as depicting configurations and Moran scatterplots. This should hopefully encourage further analyses on new datasets and new contexts along the same lines, but let us admit that the tangible progress demonstrated by our new proposal is not that dramatic in the present example: in particular, the autocorrelation is here extremely significant under both the normal approximation and the Cornish–Fisher correction, and the values of the expected skewness and excess kurtosis are fairly moderate here – but this does not constitute a general rule, of course: it is easy to design "realistic" feature configurations and spatial weights possessing large spectral skewness and/or excess kurtosis.

Conclusion

This article addresses spatial autocorrelation, a time-honored theme of spatial analysis (see, e.g., the special issue, Volume 41, of Geographical Analysis, and the Editorial of Griffith 2009), and revisits a few recurrent topics, such as the eigen-decomposition and visualization of spatial weights, Moran–Anselin scatterplots and local indicators of spatial association, and the null distribution of Moran's *I*. It also, besides the weighted and multivariate generalizations, presumably innovates by insisting on designing weight-compatible spatial weights, on using a kernel-based formalism, and on proposing a new nonparametric procedure for testing the significance of spatial autocorrelation, exact up to the fourth order.

Many workers have in particular addressed the issue of the validity of the normal approximation for the null distribution of Moran's I, or equivalently of the RV coefficient (see, e.g., Heo and Ruben Gabriel 1998; Tiefelsdorf, Griffith, and Boots 1999; Tiefelsdorf 2002; Josse, Pagès, and Husson 2008; Zhang et al. 2009; Westerholt 2022, and references therein). Expressions (39) and (40) precisely provide a transparent framework for analyzing the possible causes for nonnormality: the skewness and excess kurtosis turn out to symmetrically decompose into two separated, distinct contributions, namely one contribution depending on the scree plot of the feature kernel, and the other on the scree plot of the spatial kernel or spatial weights.

P.s.d. kernels constitute weighted similarity measures between observational units, and have been systematically exploited in Machine Learning, to represent numeric and categorical data, but also time series, strings and texts, networks, hard and soft clusterings, etc. (see, e.g., Lodhi et al. 2002; Hofmann, Schölkopf, and Smola 2008; Guyon et al. 2010, and references therein). Their amazing versatility and expressivity has been first identified in Classical Data Analysis (see, e.g., De la Cruz and Holmes 2011; Purdom 2011, and references therein). In the present context, to each kernel $\mathbf{K}$ corresponds a unique feature configuration $(\mathbf{f}, \mathbf{D})$ and vice versa, but also a unique geographical configuration $(\mathbf{f}, \mathbb{D})$, as well as a family of spatial weights (33). As demonstrated here, those fortunate circumstances permit to define and test spatial autocorrelation by formally treating regional features and space on the same footing.

In the present approach, geographical space is primarily defined from a probability distribution $\mathbf{f}$, measuring the importance of the regions, as well as a pair probability distribution $\mathbf{E} = \mathbf{\Pi W}$ measuring the regional interaction, whose symmetry amounts in dealing with an unoriented weighted network, equivalently specified by the symmetric spatial kernel $\mathbf{K}_W$. Considering weight-compatible oriented networks $\mathbf{E}^\top \neq \mathbf{E}$ with $\mathbf{E}^\top \mathbf{1}_n = \mathbf{E}\mathbf{1}_n = \mathbf{f}$ has no direct influence on the autocorrelation, since, as noticed ever since Cliff and Ord (1973) in their own formalism, $\mathrm{Tr}(\mathbf{K}_X, \mathbf{K}_W) = \mathrm{Tr}(\mathbf{K}_X, \frac{1}{2}[\mathbf{K}_W + \mathbf{K}_W^\top])$, unless the features kernel $\mathbf{K}_X$ is asymmetric as well, in which case a quantity of a new kind would emerge. The introduction of higher-order (triadic, tetradic etc.) weight-compatible probability distributions could prove of interest in itself, in particular regarding the issue of flow autocorrelation.

Since spatial kernels can also be obtained directly from geographic distances (33), the present formalism contributes to extend and unify the conceptual and operational scope of geographical space, and spatial autocorrelation in particular. In presence of a unique, invariant regional weights $\mathbf{f}$, the weight-compatibility requirement $\mathbf{W}^\top \mathbf{f} = \mathbf{f}$, little or not addressed in the current literature, may seem quite restrictive, but we insist again on its invaluable formal and conceptual benefits. And, naturally, all the exposed material remains valid if one wishes to consider regions of equal importance $f_i = 1/n$.

A great deal of modeling flexibility remains within the limits of the proposed formalism, in particular regarding the treatment of geographic distances $\mathbb{D} = (\mathbb{d}_{ij})$: in definition (1), a_{ij} can be replaced by *any distance-deterrence function* $g(\mathbb{d}_{ij})$, where $g(\mathbb{d}) \geq 0$ is decreasing in $\mathbb{d}$, typically exponentially or algebraically. Conversely, the geographic distances $\mathbb{d}_{ij}$ in (32) can be replaced by *any increasing function* $h(\mathbb{d}_{ij})$ with $h(\mathbb{d}) \geq 0$ and $h(0) = 0$, such as the *radial basis transformation* $h(\mathbb{d}) = 1 - \exp(-\lambda \mathbb{d})$ with $\lambda > 0$.

Obtaining the *exact distribution* of δ under H_0, using the newly proposed non-parametric invariant orthogonal integration procedure, may reveal itself out of reach (Bavaud 2023). Yet, the first four expected moments of δ can be exactly computed, thus granting a fairly precise statistical significance testing procedure via the Cornish–Fisher expansion. Considering the first two expected moments only amounts in resorting to the usual normal approximation, whose present extension to weighted observations endowed with multivariate profiles seems original in itself.

[Correction added on 22 March 2024 after first online publication: On page 5, first paragraph, $\mathbf{W}_0$ has been added to the start of the first sentence.]

Acknowledgement

Open access funding provided by Université de Lausanne.

Endnote

1 Note that, throughout this article, the term "moments" will loosely refer to as the expected (un-)centered and/or as the (un-)standardized moments of δ such as its mean, variance, skewness, or excess kurtosis-the latter being, technically, its fourth standardized cumulant. By contrast, "spectral moments" will refer to the observed moments of the kernel eigenvalues.

References

Amédée-Manesme, C.-O., F. Barthélémy, and D. Maillard. (2019). "Computation of the Corrected Cornish-Fisher Expansion Using the Response Surface Methodology: Application to Var and Cvar." *Annals of Operations Research* 281(1), 423–53.

Anselin, L. (1995). "Local Indicators of Spatial Association - LISA." *Geographical Analysis* 27(2), 93–115.

Anselin, L. (2019). "A Local Indicator of Multivariate Spatial Association: Extending Geary'sC." *Geographical Analysis* 51(2), 133–50.

Bavaud, F. (1998). "Models for Spatial Weights: A Systematic Look." *Geographical Analysis* 30(2), 153–71.

Bavaud, F. (2002). "The Quasi-Symmetric Side of Gravity Modelling." *Environment and Planning A* 34(1), 61–79.

Bavaud, F. (2013). "Testing Spatial Autocorrelation in Weighted Networks: The Modes Permutation Test." *Journal of Geographical Systems* 15(3), 233–47.

Bavaud, F. (2014). "Spatial Weights: Constructing Weight-Compatible Exchange Matrices from Proximity Matrices." In *Geographic Information Science*, 81–96, edited by M. Duckham, E. Pebesma, K. Stewart, and A. U. Frank. Cham: Springer International Publishing.

Bavaud, F. (2023). "Exact First Moments of the RV Coefficient by Invariant Orthogonal Integration." *Journal of Multivariate Analysis* 198, 105227.

Berger, J., and J. L. Snell. (1957). "On the Concept of Equal Exchange." *Behavioral Science* 2(2), 111–8.

Bhatia, R. (2001). "Linear Algebra to Quantum Cohomology: The Story of Alfred Horn's Inequalities." *The American Mathematical Monthly* 108(4), 289–318.

Boots, B. N. (1982). "Comments on the Use of Eigenfunctions to Measure Structural Properties of Geographic Networks." *Environment and Planning A* 14(8), 1063–72.

Borg, I., and P. J. Groenen. (2005). *Modern Multidimensional Scaling: Theory and Applications*. New York: Springer Science & Business Media.

Cheng, S. H., and N. J. Higham. (1998). "A Modified Cholesky Algorithm Based on a Symmetric Indefinite Factorization." *SIAM Journal on Matrix Analysis and Applications* 19(4), 1097–110.

Chun, Y. (2008). "Modeling Network Autocorrelation within Migration Flows by Eigenvector Spatial Filtering." *Journal of Geographical Systems* 10, 317–44.

Cliff, A. D., and J. K. Ord. (1973). *Spatial Autocorrelation*. London: Pion.

Cliff, A. D., and J. K. Ord. (1981). *Spatial Processes, Models and Applications*. London: Pion.

De la Cruz, O., and S. Holmes. (2011). "The Duality Diagram in Data Analysis: Examples of Modern Applications." *The Annals of Applied Statistics* 5(4), 2266–77.

Demšar, U., P. Harris, C. Brunsdon, A. S. Fotheringham, and S. McLoone. (2013). "Principal Component Analysis on Spatial Data: An Overview." *Annals of the Association of American Geographers* 103(1), 106–28.

Dray, S., and T. Jombart. (2011). “Revisiting Guerry’s Data: Introducing Spatial Constraints in Multivariate Analysis.” *The Annals of Applied Statistics* 5(4), 2278–99.
Dray, S., P. Legendre, and P. R. Peres-Neto. (2006). “Spatial Modelling: A Comprehensive Framework for Principal Coordinate Analysis of Neighbour Matrices (Pcnm).” *Ecological modelling* 196(3-4), 483–93.
Eckardt, M., and J. Mateu. (2021). “Partial and Semi-Partial Statistics of Spatial Associations for Multivariate Areal Data.” *Geographical Analysis* 53(4), 818–35.
Escoufier, Y. (1973). “Le Traitement des Variables Vectorielles.” *Biometrics* 29(4), 751–60.
Greenacre, M. (2017). *Correspondence Analysis in Practice*. New York: Chapman and Hall.
Griffith, D. A. (1996). “Spatial Autocorrelation and Eigenfunctions of the Geographic Weights Matrix Accompanying Geo-Referenced Data.” *Canadian Geographer/Le Géographe canadien* 40(4), 351–67.
Griffith, D. A. (2009). “Celebrating 40 Years of Scientific Impacts by Cliff and Ord.” *Geographical Analysis* 41(4), 343–6.
Griffith, D. A. (2010). “The Moran Coefficient for Non-normal Data.” *Journal of Statistical Planning and Inference* 140(11), 2980–90.
Guyon, I., A. Saffari, G. Dror, and G. Cawley. (2010). “Model Selection: Beyond the Bayesian/Frequentist Divide.” *Journal of Machine Learning Research* 11(1), 61–87.
Hastings, W. K. (1970). “Monte Carlo Sampling Methods Using Markov Chains and their Applications.” *Biometrika* 57(1), 97–109.
Heo, M., and K. Ruben Gabriel. (1998). “A Permutation Test of Association between Configurations by Means of the RV Coefficient.” *Communications in Statistics - Simulation and Computation* 27(3), 843–56.
Higham, N. J. (1988). “Computing a Nearest Symmetric Positive Semidefinite Matrix.” *Linear Algebra and its Applications* 103, 103–18.
Hofmann, T., B. Schölkopf, and A. J. Smola. (2008). “Kernel Methods in Machine learning.” *The Annals of Statistics* 36(3), 1171–220.
INSEE. (2022). Institut National de la Statistique et des Études Économiques (National Institute of Statistics and Economic Studies). https://www.insee.fr/en/accueil. Accessed 2022.
Josse, J., J. Pagès, and F. Husson. (2008). “Testing the Significance of the RV Coefficient.” *Computational Statistics & Data Analysis* 53(1), 82–91.
Kendall, M., and A. Stuart. (1977). “The Advanced Theory of Statistics.” In *Distribution Theory* Vol 1. London: *Griffin*.
Knight, P. A. (2008). “The Sinkhorn–Knopp Algorithm: Convergence and Applications.” *SIAM Journal on Matrix Analysis and Applications* 30(1), 261–75.
Kondor, R. I., and J. Lafferty. (2002). “Diffusion Kernels on Graphs and Other Discrete Structures.” In *Proceedings of the 19th International Conference on Machine Learning* Vol 2002, 315–22. Sydney: Carnegie Mellon University.
Lebart, L., A. Morineau, and K. Warwick. (1984). *Multivariate Descriptive Statistical Analysis: Correspondence Analysis and Related Techniques for Large Matrices. John Willey*. Chichester: Wiley and Son.
LeSage, J. P., and C. Thomas-Agnan. (2015). “Interpreting Spatial Econometric Origin-Destination Flow Models.” *Journal of Regional Science* 55(2), 188–208.
Levin, D. A., and Y. Peres. (2017). *Markov Chains and Mixing Times*, Vol 107. Providence, RI: American Mathematical Society.
Lin, J. (2020). “A Local Model for Multivariate Analysis: Extending Wartenberg’s Multivariate Spatial Correlation.” *Geographical Analysis* 52(2), 190–210.
Lodhi, H., C. Saunders, J. Shawe-Taylor, N. Cristianini, and C. Watkins. (2002). “Text Classification Using String Kernels.” *Journal of Machine Learning Research* 2, 419–44.
Purdom, E. (2011). “Analysis of a Data Matrix and a Graph: Metagenomic Data and the Phylogenetic Tree.” *The Annals of Applied Statistics* 5(4), 2326–58.
Robert, P., and Y. Escoufier. (1976). “A Unifying Tool for Linear Multivariate Statistical Methods: The RV-Coefficient.” *Journal of the Royal Statistical Society: Series C (Applied Statistics)* 25(3), 257–65.
Ruschendorf, L. (1995). “Convergence of the Iterative Proportional Fitting Procedure.” *The Annals of Statistics*, 23(4), 1160–74.

Schneider, M. H., and S. A. Zenios. (1990). "A Comparative Study of Algorithms for Matrix Balancing." *Operations research* 38(3), 439–55.

Thioulouse, J., D. Chessel, and S. Champely. (1995). "Multivariate Analysis of Spatial Patterns: A Unified Approach to Local and Global Structures." *Environmental and Ecological Statistics* 2, 1–14.

Tiefelsdorf, M. (2002). "The Saddlepoint Approximation of Moran's I's and Local Moran's I_i's Reference Distributions and Their Numerical Evaluation." *Geographical Analysis* 34(3), 187–206.

Tiefelsdorf, M., and D. A. Griffith. (2007). "Semiparametric Filtering of Spatial Autocorrelation: The Eigenvector Approach." *Environment and Planning A* 39(5), 1193–221.

Tiefelsdorf, M., D. A. Griffith, and B. Boots. (1999). "A Variance-Stabilizing Coding Scheme for Spatial Link Matrices." *Environment and Planning A* 31(1), 165–80.

Wartenberg, D. (1985). "Multivariate Spatial Correlation: A Method for Exploratory Geographical Analysis." *Geographical Analysis* 17(4), 263–83.

Westerholt, R. (2022). "A Simulation Study to Explore Inference about Global Moran's I with Random Spatial Indexes." *Geographical Analysis* 55(4), 621–650.

Zhang, T., and G. Lin. (2016). "On Moran's I Coefficient under Heterogeneity." *Computational Statistics & Data Analysis* 95, 83–94.

Zhang, H., J. Tian, J. Li, and J. Zhao. (2009). "RV-Coefficient and its Significance Test in Mapping Brain Functional Connectivity." In *Em Medical Imaging 2009: Biomedical Applications in Molecular, Structural, and Functional Imaging* Vol 7262, 627–35. Bellingham, Washington: SPIE.