

Article

Articulating Spatial Statistics and Spatial Optimization Relationships: Expanding the Relevance of Statistics

Daniel A. Griffith 

School of Economic, Political and Policy Sciences, University of Texas at Dallas, Richardson, TX 75080, USA; dagriffith@utdallas.edu

Abstract: Both historically and in terms of practiced academic organization, the anticipation should be that a flourishing synergistic interface exists between statistics and operations research in general, and between spatial statistics/econometrics and spatial optimization in particular. Unfortunately, for the most part, this expectation is false. The purpose of this paper is to address this existential missing link by focusing on the beneficial contributions of spatial statistics to spatial optimization, via spatial autocorrelation (i.e., dis/similar attribute values tend to cluster together on a map), in order to encourage considerably more future collaboration and interaction between contributors to their two parent bodies of knowledge. The key basic statistical concept in this pursuit is the median in its bivariate form, with special reference to the global and to sets of regional spatial medians. One-dimensional examples illustrate situations that the narrative then extends to two-dimensional illustrations, which, in turn, connects these treatments to the spatial statistics centrography theme. Because of computational time constraints (reported results include some for timing experiments), the summarized analysis restricts attention to problems involving one global and two or three regional spatial medians. The fundamental and foundational spatial, statistical, conceptual tool employed here is spatial autocorrelation: geographically informed sampling designs—which acknowledge a non-random mixture of geographic demand weight values that manifests itself as local, homogeneous, spatial clusters of these values—can help spatial optimization techniques determine the spatial optima, at least for location-allocation problems. A valuable discovery by this study is that existing but ignored spatial autocorrelation latent in georeferenced demand point weights undermines spatial optimization algorithms. All in all, this paper should help initiate a dissipation of the existing isolation between statistics and operations research, hopefully inspiring substantially more collaborative work by their professionals in the future.



Citation: Griffith, D.A. Articulating Spatial Statistics and Spatial Optimization Relationships: Expanding the Relevance of Statistics. *Stats* **2021**, *4*, 850–867. <https://doi.org/10.3390/stats4040050>

Academic Editor: Wei Zhu

Received: 18 September 2021

Accepted: 14 October 2021

Published: 19 October 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the author. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Keywords: centographic; location-allocation; spatial autocorrelation; spatial median; spatial optimization

1. Introduction

The history of statistics dates back to the fourth century B.C. [1] and earlier, with permutations and combinations being concepts discussed in the writings of those times. Even though modern statistics did not begin to develop until the mid-1600s, its evolution into mathematical statistics occurred several centuries later, in the early 1800s [2]. Meanwhile, pre-modern optimization developed with the apparently independent discovery of calculus by Newton and Gauss in the late 17th century, followed by Fermat and Lagrange (circa 1800) establishing function minimization/maximization, and, decades later, Riemann formulating the steepest descent method [3]. Kantorovich (linear programming), Dantzig (the simplex method), and von Neumann (duality theory) molded optimization's evolution into its modern-day, mathematical, computer-supported form. Although both disciplines have roots in permutations and combinations, their evolutionary trajectories essentially unfolded in parallel. The absence of an interface between the two also characterizes their respective sub-disciplines of spatial statistics and spatial optimization.

In contrast to this interface absence, inspecting today's academic degree and other programs of study at institutions of higher learning throughout the world reveals that one natural interdisciplinary educational coupling is statistics and operations research, raising questions about the continued absence of a more comprehensive subject matter interface between them. This academic home marriage occurs in schools across the various North American teaching/training institution tiers, ranging from Bowling Green State and Virginia Commonwealth University, through the University of North Carolina/Chapel Hill, to Harvard and MIT. This pairing is not exclusive to the United States and occurs in schools housed in other countries, such as Tel Aviv and Wuhan University, and in internationally renowned European universities such as Cambridge, Oxford, Edinburgh, and Vienna. All in all, the Quacquarelli Symonds (QS) 2021 World University Rankings by subject includes in its incomplete listing 221 enterprises with this pairing (See <https://www.topuniversities.com/university-rankings/university-subject-rankings/2021/statistics-operational-research> (accessed on 10 October 2021)). Surprisingly, although optimization sometimes utilizes probability to, for example, incorporate uncertainty into objective function formulations, state-of-the-art commercial software such as the IBM CPLEX Optimizer (this advanced computer package provides flexible, high-performance mathematical programming solving procedures for linear, mixed integer, quadratic, and quadratically constrained problems) lacks serious interfaces with theoretical and applied statistics. For the most part, the statistics and operations research sub-disciplines developed alongside and independent of each other, merely inadvertently sharing domains such as core quantitative foci. This lack of a synergistic interface, except perhaps in the realm of game theory, is especially acute in geography and the spatial sciences, with the development of spatial statistics/econometrics (i.e., addressing spatial autocorrelation in georeferenced data, namely the presence of, most often, similar attribute values clustering in geographic space; see Cliff and Ord [4], Paelinck and Klaassen [5], Cressie [6], and Griffith [7]) and spatial optimization (e.g., location covering, network routing [8,9]) having occurred side by side and nearly in isolation.

This paper seeks to facilitate the altering of this missing interface situation by helping to initiate transformative work as well as contribute to the sparse substantive literature at the interface of spatial statistics/econometrics and location science spatial optimization, which would hopefully encourage considerably more future collaboration and interaction between these two bodies of knowledge. Therefore, the primary objective of this paper is to begin to dissolve this isolation. Its methods are mathematical analysis and simulation experiments. The beneficial contributions of spatial statistics, via spatial autocorrelation, to spatial optimization outlined here aim to exemplify how their parent disciplines can cross-fertilize to profit both.

2. Background

The particular operations research problem addressed in this paper generalizes from one whose simplest version Fermat [10] and Weber [11] first studied: to determine the optimal locations $\{(U_j, V_j), j = 1, 2, \dots, p\}$ for p central facilities with unlimited capacity in order to serve a geographically distributed set of n demand points (finding a p -tuple of points on a continuous surface that jointly minimize the sum of weighted Euclidean distances to them from a designated set of discrete demand points), which may be stated in an objective function form, as follows:

$$\text{MIN} : \sum_{i=1}^n \sum_{j=1}^p \lambda_{ij} w_i \sqrt{(u_i - U_j)^2 + (v_i - V_j)^2} \text{ s.t. : } \sum_{j=1}^p \lambda_{ij} = 1, i = 1, 2, \dots, n, \quad (1)$$

where λ_{ij} denotes a set of p dichotomous 0–1 indicator variables for each location i , with $\lambda_{ij} = 1$ if its allocation is to central facility $j = 1, 2, \dots, p$, and 0 otherwise, (u_i, v_i) are the analyzed set of n demand point Cartesian coordinates, and $w_i > 0$ is the weight quantifying demand at point i ; $\lambda_{ij} \equiv 1$ if $p = 1$ (the problem examined by Fermat and Weber). The point $(U_j, V_j), j = 1, 2, \dots, p$, is also the bivariate global ($p = 1$) or regional ($p > 1$)

spatial median e.g., [12–14] for the demand points that constitute an allocation to central facility j (i.e., $\lambda_{ij} = 1$); this equivalence establishes an intrinsic link between spatial statistics and spatial optimization, specifically within the context of centrographic techniques [15]. Kuhn-Kuenne [16] devised a calculus-based algorithm for solving differential equations (given a set of n demand points $\{(u_1, v_1), (u_2, v_2), \dots, (u_n, v_n)\} \in \mathbb{R}^n \times \mathbb{R}^n$, then $(U, V) \in \mathbb{R}^n \times \mathbb{R}^n$, the spatial median is the expression (1) solution for this dataset if and only if $\sum_{i=1}^n \frac{w_i(u_i - U)}{\sqrt{(u_i - U)^2 + (v_i - V)^2}} = 0$ and $\sum_{i=1}^n \frac{w_i(v_i - V)}{\sqrt{(u_i - U)^2 + (v_i - V)^2}} = 0$; solving this pair of nonlinear equations requires recursion) in order to compute (U_1, V_1) for $p = 1$, which quantitative spatial scientists later adapted to extend to each of the $p > 1$ regional subsets of points in the more general p -median problem—initiating this computation with each region's spatial mean (i.e., the centrographic bivariate weighted arithmetic average $(\frac{\sum_{i=1}^n w_i u_i}{\sum_{i=1}^n w_i}, \frac{\sum_{i=1}^n w_i v_i}{\sum_{i=1}^n w_i})$) to avoid difficulties with algorithm nonlinearity convergence when computing for its spatial median; algorithm iterations involving reallocations and spatial median recalculations continue recursively until the p median centroid trajectories satisfy the invoked convergence criteria.

A paucity of articles deal with interfaces between spatial statistics/econometrics and spatial optimization, expressly in terms of this tremendously popular bivariate spatial median problem. Exceptions include: (1) the use of spatial statistics imputation (e.g., kriging, see [17,18]) to estimate missing weights, w_i , in expression (1)— $w_i = 0$ removes its affiliated demand point from the solving of its spatial optimization problem, motivating the imputation of missing values; (2) the employment of local indices of spatial autocorrelation [19,20] to identify potential spatial median solutions based upon hotspot/spatial outlier maps [21,22] shows considerable promise for guiding heuristic algorithms to better achieve globally optimal solutions; and (3) the use of spatial autocorrelation informed spatial sampling designs to help solve large spatial optimization problems, with specific reference to the location-allocation problem defined by expression (1), the overarching theme of this paper. Thus, to date, the interface between spatial statistics/econometrics and spatial optimization, similar to the existing treatment of synergies between its parent research fields of statistics and operations, generally tends to be overlooked by scholarly researchers and practitioners.

Consequently, the primary objective of this paper is to demonstrate selected noteworthy intellectual contributions that spatial statistics can make to spatial optimization through its spatial autocorrelation knowledge base and specialized sampling designs. In doing so, statistics is promoted as the truly interdisciplinary discipline that it is, which would aid in fostering statistics as a major supporter of and player in the science of big data.

3. Methods and Results

The condensed analysis here considers the first three possible cases, namely $p = 1, 2$, and 3, much like in the tradition of mathematical induction.

3.1. The $p = 1$ Spatial Median Problem

The bivariate spatial median has many names, including the following synonyms [14]: Euclidean median; the generalized Fermat point, reflecting its concept formation and beginnings; median center; minimum aggregate travel point; and the Weber problem, again reflecting its invention (a rediscovery) and origin. In addition, the spatial median is an extension of the univariate median [12]. One-dimensional (1-D) analyses in this and subsequent sections are for the unit interval $[0, 1]$, whereas two-dimensional (2-D) analyses are for the unit square. These former explorations are purely hypothetical, chiefly furnishing useful insights and proofs of concept, whereas these latter ones are for pragmatic endeavors.

3.1.1. The $p = 1$ Univariate Spatial Median Problem

For simplicity as well as illustrative purposes, the discussion here begins with a 1-D geographic landscape. GIScientists characterize this precise setting as the socially optimal Hotelling [23] problem, which recently received a relevant updating by Meager, Teo, and Xie [24] to exemplify the impacts of a wide range of non-uniform distributions of demand weights, w_i . For the basically straightforward $p = 1$ situation, any 1-D symmetric distribution of demand results in its spatial median being at the central position $\frac{1}{2}$ in the unit interval (Figure 1a–c). The respective skewed demand distribution of spatial medians (Figure 1d,e), while deviating from $\frac{1}{2}$, are closer to it than their individual modes are. Meanwhile, a multi-modal demand distribution (e.g., Figure 1f, an additive mixture of truncated exponential (weighted 0.5) and bell-shaped (weighted 0.2) probability density functions (pdfs) combined with a skewed beta (weighted 0.3) pdf) has a spatial median that is conditional upon the utilized types of mixed geographic demand distributions as well as their employed relative weights (which are non-negative and sum to one). These outcomes are equivalent to those of comparable univariate pdfs, with their spatial median proofs obtainable by completing the customary exercise of integrating an appropriate pdf to x , and then solving for $x = 0.5$ (see Table 1).

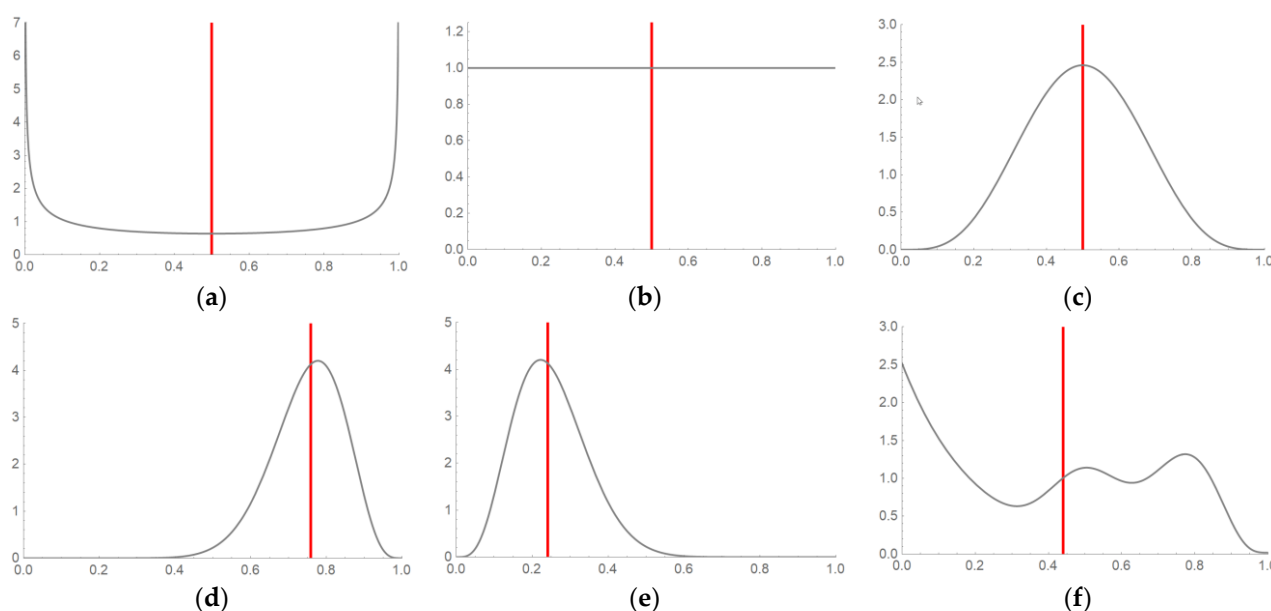


Figure 1. A representative range of possible distributions of demand across a unit interval. Gray curves denote the relative magnitude of demand at a location; red vertical lines denote the locations of $p = 1$ spatial medians. Top left (a): sinusoidal demand distribution. Top middle (b): uniform demand distribution. Top right (c): bell-shaped demand distribution. Bottom left (d): negative-skewed demand distribution. Bottom middle (e): positive-skewed demand distribution. Bottom right (f): multimodal demand distribution.

A simple random sample (with replacement for the infinite population) corroborates these Table 1 tabulations. Moreover, the central limit theorem for univariate medians of most random variables produces this quality of simulated results; the mixture outcome is a poorer replication because it contains considerably more variability (although its n is more than twice as large). Because these are order statistics, the sample size is $2n + 1$; because the statistic of interest is the median, the sample size is deliberately an odd integer. As is well known, the mean of the simulated data approximately equals the sample median (theoretically, these are identical); meanwhile, a consensus of expert opinion proposes that $n = 100$ is a sufficiently large sample size for the central limit theorem to be effective, regardless of the underlying random variable (the Cauchy is an exception to this rule, at least for the arithmetic mean). The examined mixture distribution in Table 1 goes beyond

the spectrum of the geographic distribution of demand categories investigated by Meager, Teo, and Xie [24] (p. 434).

Table 1. Single spatial median solutions for various distributions of demand across the unit interval.

Distribution of Demand	Relevant Integration Answers	Spatial Medial	Random Sample (n = 201; 10,000 Replications)
sinusoidal	$0.63662 \times \text{SIN}^{-1}(\sqrt{x}) = 0.5$	0.50000	0.49959 (0.05443)
uniform	$x = 0.5$	0.50000	0.50003 (0.03525)
bell-shaped	$70x^9 - 315x^8 + 540x^7 - 420x^6 + 126x^5 = 0.5$; only one real root	0.50000	0.50020 (0.01431)
left/negative skewed	$3060x^{19} - 12,920x^{18} + 20,520x^{17} - 14,535x^{16} + 3876x^{15} = 0.5$; only one real root	0.75846	0.75831 (0.00848)
right/positive skewed	$3060x^{19} - 45,220x^{18} + 311,220x^{17} - 1,322,685x^{16} + 3,879,876x^{15} - 8,314,020x^{14} + 13,430,340x^{13} - 16,628,040x^{12} + 15,872,220x^{11} - 11,639,628x^{10} + 6,466,460x^9 - 2,645,370x^8 + 755,820x^7 - 135,660x^6 + 11,628x^5 = 0.5$; only one real root	0.24154	0.24168 (0.00855)
tri-density function additive mixture	Solved with Mathematica 12.1 assuming $x \in \mathbb{R}^+$	0.44045	0.46118 (0.03576)

Note: Standard errors are in parentheses; for the simple, non-mixture, univariate random variables, they roughly equal their theoretical counterparts given by $1/[\text{pdf}(\text{median})\sqrt{800}]$.

3.1.2. The $p = 1$ Bivariate Spatial Median Problem

One important univariate $p = 1$ spatial median feature that directly transfers to the single global bivariate spatial median arises from symmetry in the distribution of demand across a 2-D geographic landscape, paralleling a 1-D geographic landscape symmetry consequence. The 2-D symmetric sinusoidal (i.e., resembling a bowl shape), uniform, and bell-shaped (resembling a mountain in the center of a unit square) geographic distribution of demand renders a spatial median of $(\frac{1}{2}, \frac{1}{2})$; Chen and Welsh [25] (p. 211) furnish a proof of this proposition—similar to the preceding 1-D geographic landscape, construction of this proof relies upon integration for its continuous case, and the sigma calculus for its discrete case.

The product of two independent beta random variable probabilities furnishes the following joint density pdf:

$$p(u, v) = \frac{\Gamma(\alpha_u + \beta_u)}{\Gamma(\alpha_u)\Gamma(\beta_u)} \frac{\Gamma(\alpha_v + \beta_v)}{\Gamma(\alpha_v)\Gamma(\beta_v)} u^{\alpha_u-1} (1-u)^{\beta_u-1} v^{\alpha_v-1} (1-v)^{\beta_v-1}, \quad (2)$$

where Γ denotes the gamma function, $0 < u, v < 1$, and $\alpha_u, \alpha_v, \beta_u, \beta_v > 0$. This joint pair of univariate beta distribution functions differs from its potentially correlated bivariate counterpart with marginal univariate beta distribution functions presented by Olkin and Trikalinos [26], for example, which pertains to two conditionally dependent beta random variables. This paper employs Equation (2) because the simulation experiment sampling exercises require independent observations, and the manipulation of the four pdf parameters (i.e., $\alpha_u, \alpha_v, \beta_u, \beta_v$) is capable of generating an extensive variety of different geographic demand point distributions across a full unit square. Equation (2) implies that the benchmark

population spatial median for random weights has the pair of (U, V) Cartesian coordinates respectively approximated by $(3\alpha_u - 1)/(3\alpha_u + 3\beta_u - 2)$ and $(3\alpha_v - 1)/(3\alpha_v + 3\beta_v - 2)$.

An important technical complication for the 2-D $p = 1$ case is solving it when n is large-to-massively large; the required allocation of all points is to a single unknown spatial median, which constitutes the calculation goal. A latent positive spatial autocorrelation map pattern tends to move this optimal solution away from the central $(\frac{1}{2}, \frac{1}{2})$ position in the unit square. A sampling design can ignore this autocorrelation and focus on independence in the sample selection mechanism (hence, the Equation (2) specification); nevertheless, a spatial scientist must recognize and acknowledge the spatial autocorrelation's tendency to influence a spatial median location. Traditional mathematical statistical theory dictates the result for purely random map patterns of weights (i.e., they are identically distributed) through the calculus of expectations, namely the foregoing stated bivariate spatial median for a uniform geographic distribution of demand (i.e., each location's expected value is the same constant).

The simulation experimental design devised by Overton and Stehman [27] guided the research undertakings outlined here: spatial sampling of a linear gradient (inducing strong positive spatial autocorrelation), a quadratic gradient (inducing moderate positive spatial autocorrelation), and a periodic map pattern (inducing weak positive spatial autocorrelation) of the geographic distribution of attribute values (e.g., demand weights). Supplementing their point sampling from a uniform distribution is the second case of a skewed point distribution (utilizing a pair of beta random variables a la Equation (2)). The number of points is set to 50,000, and the minimum resample size is set to 100 (in keeping with the classical central limit theorem). Table 2 tabulations reveal two salient findings: (1) 1000 replications seem sufficient to realize the law of large numbers (also see Figure 2) for a spatial median analysis; and (2) the well-known variance suppression impact of spatial autocorrelation is detectable in the reported sampling standard errors (e.g., they are smaller than those for purely random samples) for the population of unconcentrated points.

Table 2. $P = 1$ simulation experiment output summaries; a single parent sample of $n = 50,000$, and 1000 resamples (without replacement) from it, with a size $n = 100$.

Demand Points Distribution	Spatial Median Coordinate	Map Pattern of Weights			
		Random	Linear Gradient	Quadratic Gradient	Periodic (i.e., SINE Function)
uniform—Beta(1, 1)	theoretical U, V	0.5, 0.5			
	complete data U	0.50352	0.59413	0.65100	0.55542
	complete data V	0.50119	0.59295	0.65240	0.61038
	sampled data U	0.50517 (0.06348)	0.59524 (0.05589)	0.65026 (0.05813)	0.55680 (0.05808)
	sampled data V	0.49924 (0.05947)	0.5929 (0.05795)	0.65114 (0.05694)	0.60905 (0.05421)
	theoretical U, V	0.35, 0.35			
skewed—Beta(9, 5)	complete data U	0.35139	0.37292	0.38368	0.36325
	complete data V	0.35265	0.37169	0.38448	0.37711
	sampled data U	0.35228 (0.02154)	0.37254 (0.02218)	0.38397 (0.02151)	0.36470 (0.02146)
	sampled data V	0.35359 (0.02395)	0.37155 (0.02086)	0.38406 (0.02247)	0.37783 (0.02105)

Note: Standard errors (the input for standard distances) are in parentheses. Map pattern generators (all demand weight specifications include adding 1 to eliminate the prospect of $w_i = 0$; all map-wide averages are between 5 and 11): (1) random—Poisson($\mu = 4$) + 1 ($\mu = 5$, $\sigma = 2$); linear gradient— $9(u + v) + 1 + 0.01 \times \text{Normal}(0,1)$ ($\mu = 10$, $\sigma \approx 11/3$); quadratic gradient— $5(u + v)^2 + 1 + 0.01 \times \text{Normal}(0,1)$ ($\mu \approx 6.83$, $\sigma \approx 4.20$); and periodic— $5[\text{SIN}(u\pi) + 2\text{SIN}(v\pi)] + 1 + 0.01 \times \text{Normal}(0,1)$ ($\mu \approx 10.55$, $\sigma \approx 3.44$).

The simulation experiment outcomes yield several important implications. Foremost of these is, as in a proof of concept demonstration, that solving the $p = 1$ median problem for randomly selected and substantially smaller subsets of demand points, and then averaging

these computations, offers a reliable alternative to solving the original excessively large spatial optimization problem; sampling theory productively applies to spatial optimization contexts, and, when possible, should be exploited to reduce demanding computational burdens affiliated with combinatorial geometry. This finding furnishes a useful tool for $p > 1$ median problems. In other words, although analysts may never find the substituting of sampling for solving the $p = 1$ median problem to be a real time saver, the very close correspondence reported here between a sample average and its complete data parent calculated spatial median confirms their substitutability, suggesting a hypothesis stating that this same substitution is possible for $p > 1$ median problems. Furthermore, the centrographic tools of standard distance [28] and standard deviational ellipse [29] may well allow insightful additional analyses of a set of sample spatial medians, which is an appealing issue for future research. Second, skewed point distributions in the plane as well as latent spatial autocorrelation in the geographic distribution of weights may counterbalance to some degree, and are potentially exploitable georeferenced data features when engaging in spatial optimization.

The resampling procedure utilized is as follows: initially, a sample size of 50,000 was drawn randomly from a unit square; each replication selected a random starting name in the interval [1] from a single random permutation of the originally sampled 50,000 point names, each choosing a 100-point sequence from the global permutation, beginning with that starting name.

3.2. The $p = 2$ Spatial Median Problem

Once $p > 1$, the spatial optimization problem then loosely resembles multivariate statistics cluster analysis problems, with regional subsets of demand points in spatial statistics mimicking group/cluster subsets of observations in conventional statistics. Similarly, interest shifts from global to p regional medians. One infamous complication here is that the objective function (1) minimum is not necessarily unique (e.g., for uniform distributions of weights and demand points, identical objective function optimal solutions can exhibit a north–south or an east–west orientation); this feature is more rampant for uniform distribution contexts (points, weights, or both), with the presence of non-zero spatial autocorrelation often partially alleviating its severity. Solving this $p = 2$ median problem is essentially reduced to the methodically iterative allocation of each of the n demand points to one of two non-overlapping coterminous geographic regions, after which the Kuhn–Kuenne [16] algorithm is separately applied to each subset of demand points to compute regional spatial medians. This iterative estimation process sequentially computes a pair of regional spatial medians and then allocates each of the n demand points to its closest spatial median, alternating between these two steps until no change occurs (i.e., convergence). This conceptualization is the basis of the ALTERN heuristic algorithm [30].

3.2.1. The $p = 2$ Univariate Spatial Median Problem

Once more, for illustrative purposes, the discussion here begins with a 1-D geographic landscape that involves scrutinizing an extension of the preceding Hotelling problem. Only for a uniform distribution of demand (Figure 3b) are the regional spatial medians evenly spaced at the first and third quartiles of the unit interval; unlike in the $p = 1$ case, as the symmetric geographic distribution of demand goes from sinusoidal to uniform to bell-shaped in this $p = 2$ case, the absolute location of the pair of regional spatial medians changes, repositioning from near the interval extremes to near the global spatial median location. Besides this tendency for a variable spacing of regional spatial medians that is a function of the form of the geographic distribution of demand (which reflects upon latent spatial autocorrelation), another important and noteworthy outcome highlights the persistent spacing gap between them. In all cases, the allocated demand's split percentage is 50–50. This scenario is the first hint of a link between spatial median problems and quantile and order statistics.

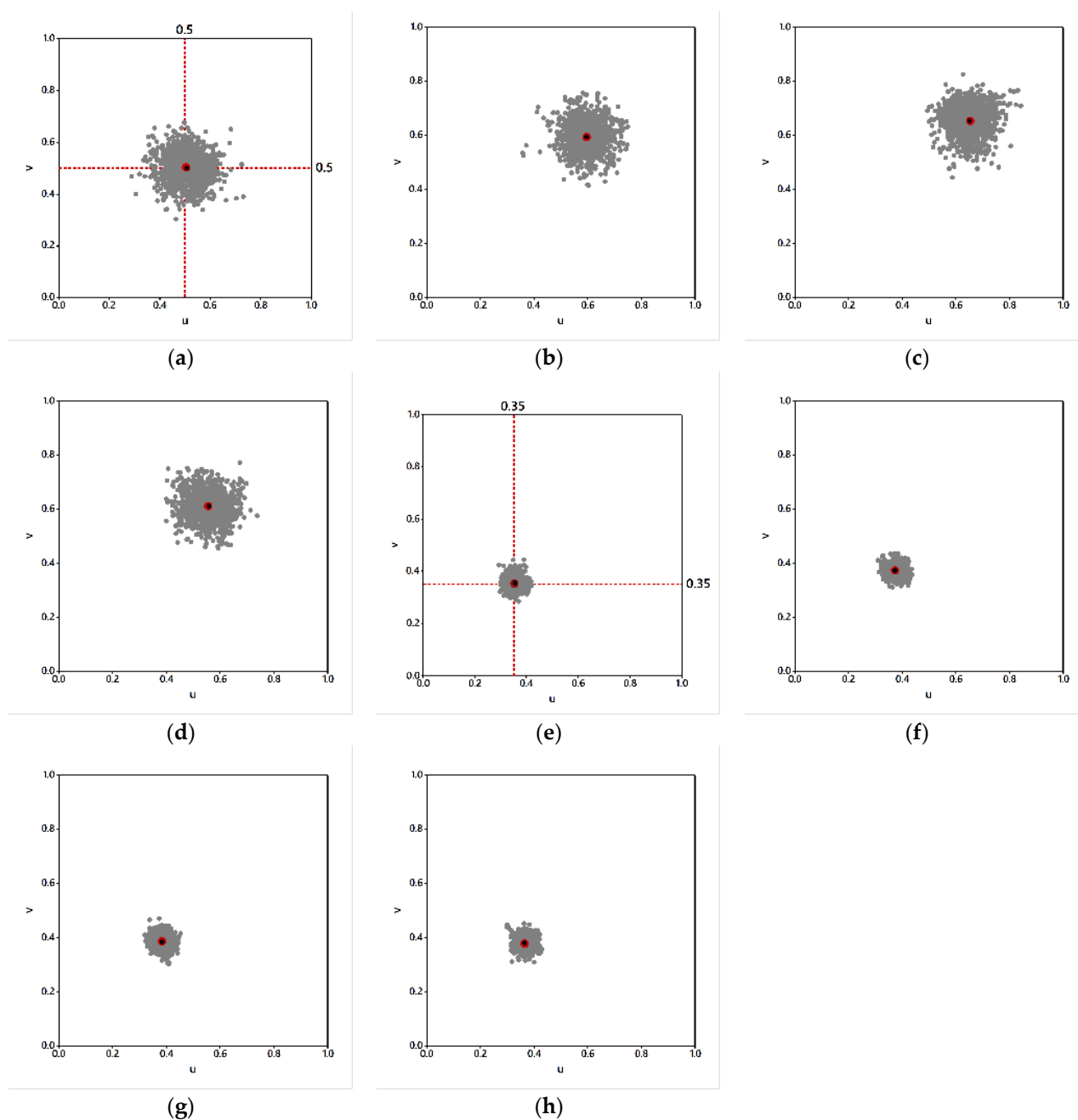


Figure 2. Sample spatial median geographic distributions (gray, red, and black filled circles respectively denote sample ($n = 1000$), complete data, and averaged sample spatial medians); red crosshair reference lines denote theoretical marginal and bivariate spatial medians—in all cases, complete data and average spatial medians essentially collocate. Top left (a): random independent weights. Top middle (b): linear gradient of weights. Top right (c): quadratic gradient of weights. Middle left (d): periodic geographic distribution of weights. Middle (e): random independent weights. Middle right (f): linear gradient of weights. Bottom left (g): quadratic gradient of weights. Bottom middle (h): periodic geographic distribution of weights. (a–d): uniform points distribution; (e–h): skewed points distribution.

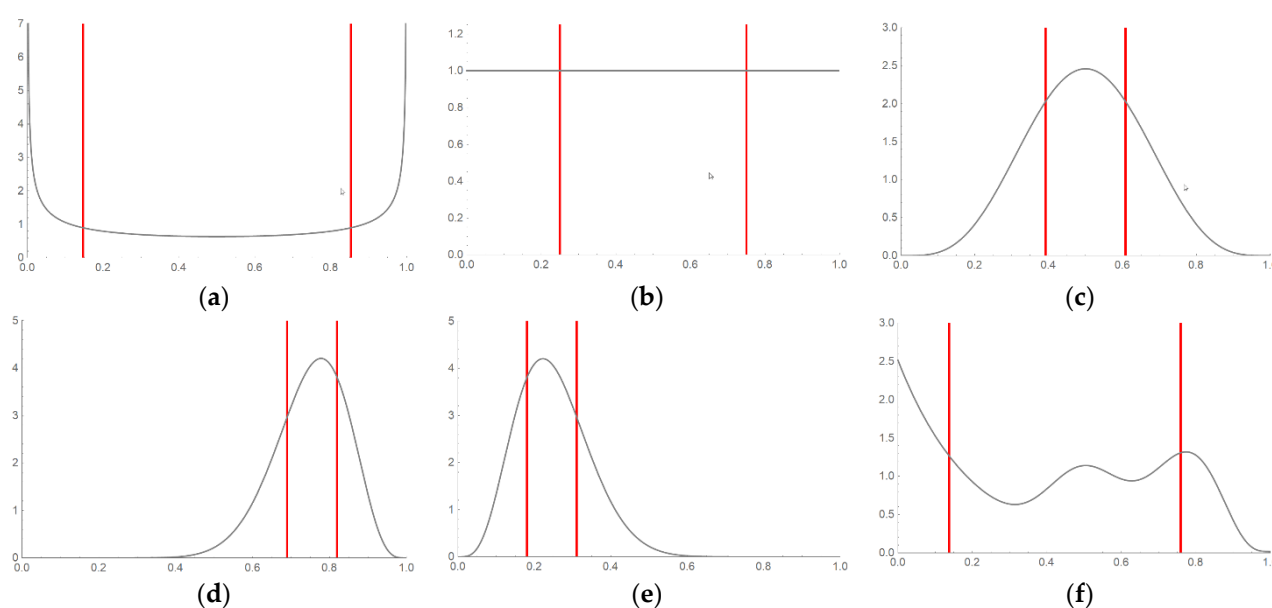


Figure 3. A representative range of possible distributions of demand across a unit interval. Gray curves denote the relative magnitude of demand at a location; red vertical lines denote the locations of $p = 2$ regional spatial medians. Top left (a): sinusoidal demand distribution. Top middle (b): uniform demand distribution. Top right (c): bell-shaped demand distribution. Bottom left (d): negative-skewed demand distribution. Bottom middle (e): positive-skewed demand distribution. Bottom right (f): multimodal demand distribution.

3.2.2. The $p = 2$ Bivariate Spatial Median Problem

One important univariate $p = 2$ spatial median feature that directly transfers to $p = 2$ bivariate spatial median regional pairs arises from the separation tendency in a distribution of demand across a 2-D geographic landscape, paralleling a 1-D geographic landscape spacing consequence (Figure 3); in other words, the pair of spatial medians fail to form a geographic cluster. This dispersion propensity, often positioning two regional spatial medians on opposite sides of a geographical landscape's global $p = 1$ spatial median, can exploit latent spatial autocorrelation, which is a georeferenced data property that frequently curtails the potential and/or magnitude of multiple optimal solutions. Another Overton and Stehman [27] spatial sampling protocol reconnaissance retrieves appropriate geographic tessellation stratified design guidelines effective within spatial autocorrelation situations that help ensure a variety of random sampling can sustain spatial median spacing. Because positive spatial autocorrelation signifies redundant attribute data information in nearby locations, geographic tessellation-based stratified sampling has values within its areal unit polygons that are relatively similar, producing repeated samples that are far more alike than their unconstrained random sample equivalents. Exploiting this similarity facet is not automatically a helpful strategy for $p = 1$ because a set of repeated samples performs better when solely concentrating on its single global spatial median; simulation experiments supplementing those of the preceding $p = 1$ expose a larger gap between average geographic tessellation stratified vis-à-vis unconstrained simple random sample spatial medians and their respective complete data counterparts, as visualized in Figure 2. Regardless, given the $p = 1$ solution success with $n = 100$, this section employs a 10-by-10 regular square tessellation superimposed upon the unit square, with the sampling design requiring a single drawing from each of the 0.1-by-0.1 grid squares; for a uniform distribution of points, the risk of zero-demand point quadrats is rather minimal when n is sizeable (e.g., 50,000). Overton and Stehman [27] argue for hexagonal rather than square geographic strata polygons; implementing their scheme provides better geographic coverage of a unit square, but is far more complicated, particularly with regard to edge effects, and if implemented, should merely enhance simulation experiment findings summarized in this

section. Furthermore, whereas $n = 50,000$ is effectively not an extremely large problem for $p = 1$, it is for $p = 2$: the solution time for an extremely efficient algorithm (TWIN [31]) coupled with contemporary desktop computer resources is extrapolated to be at least 78 days (Figure 4a; exponentially increasing time), and as many as 226 (Figure 4b; n^3 -order increasing time), with a prediction error of 1–2 days.

Table 3 tabulations reveal four salient findings: (1) As before, roughly 1000 replications seem sufficient to realize the law of large numbers (also see Figure 5) for a spatial median analysis; (2) The tessellation stratification reduces spatial autocorrelation in individual samples (they are spread out by design), which is detectable in the lack of standard error suppression (e.g., compare the appropriate matching entries in Tables 2 and 3). Unconstrained simple random sampling (utilizing the $p = 1$ resampling protocol), ignoring and failing to acknowledge latent spatial autocorrelation, produces standard errors that are at least twice as large as those for geographically tessellated stratified random sampling with purely random weights; (3) The regional spatial median gap appears to be roughly on the order of one-third of the largest distance in the geographic landscape (i.e., $(\sqrt{2})/3 \approx 0.5$); and (4) The pull by the solution's non-uniqueness seems somewhat conspicuous.

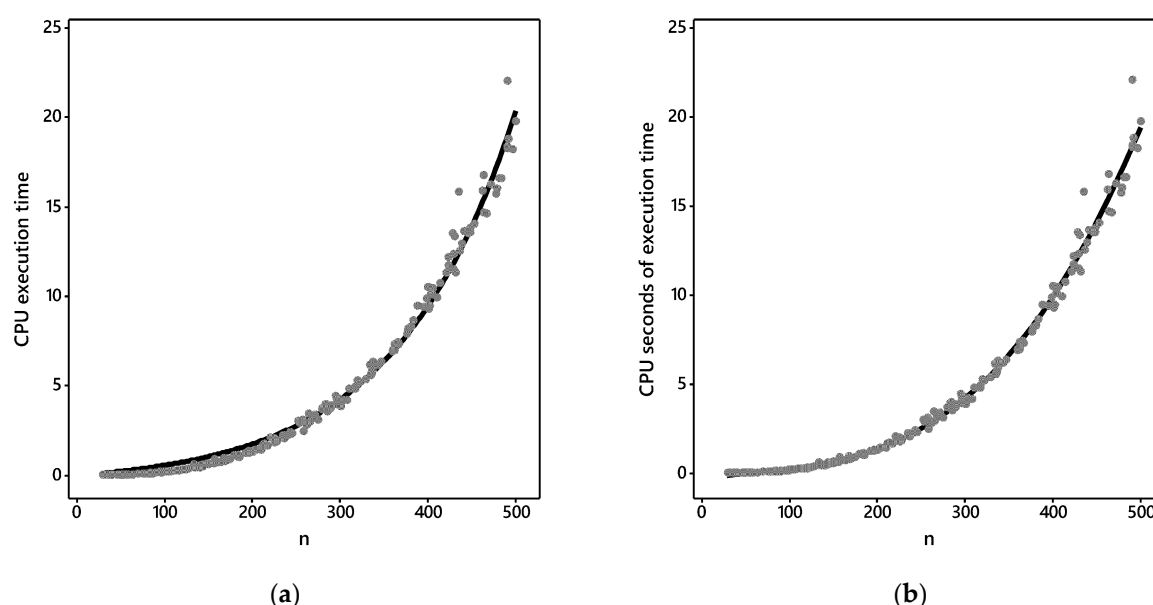


Figure 4. Central processing unit (CPU) timing (in seconds) experiments ($n = 208$ after four infeasible/excessive anomaly removals) for $p = 2$ solutions (random samples from a uniform distribution of points and $w_i \equiv 1$) with randomly selected $n \in (25, 500)$; predicted–observed correlations are at least 0.995. Left (a): an exponential trend line. Right (b): an n^3 trend line.

A visual comparison of Figures 2 and 5 divulges a tendency for the 2-D sampling distributions of points for $p = 1$ to be more circular (i.e., symmetric), and for $p = 2$ to be more elliptical (i.e., skewed), in its geographic dispersion. Figure 6 magnifies this skewness feature for Figure 5b. This distortion appears to arise from the non-uniqueness property of $p = 2$ solutions, which tends to create a more circular pattern of potential optimal solutions in a square/circle shaped geographic landscape. Removing samples struggling to constitute alternative clusters reduced each simulation replication size respectively from 1500 to 1115; 1150; 934; and 1500. An examination of Figure 5d, which deliberately had no centrifugal stragglers removed, underscores a penchant for $p = 2$ regional spatial median formation bifurcations (e.g., north–south versus east–west, or northeast–southwest versus northwest–southeast). Nonetheless, when selecting the majority from the concentration samples, the average sample medians—the statistics of interest here—are very accurate (Figure 5). As an aside, an alternative approach could be to identify both solution pairs, when the true pair of regional spatial medians is unknown, and choose between them

by allocating each demand point to its closest median (the Voronoi/Dirichlet/Thiessen principle) in order to calculate the pair of regional objective functions (expression (1)), and then select the pair of spatial medians with the smallest total objective function value.

Table 3. Uniform distribution of demand points $p = 2$ simulation experiment output summaries; a single parent sample of $n = 500$, and 1000 resamples (without replacement) from it of size $n = 100$.

Demand Points Distribution	Spatial Median Coordinate		Map Pattern of Weights			
			Random	Linear Gradient	Quadratic Gradient	Periodic (i.e., SINE Function)
region #1	theoretical U	triangle	0.29289			
		rectangle	0.25000			
	theoretical V	triangle	0.29289			
		rectangle	0.50000			
	complete data U		0.74921	0.50940	0.58642	0.29617
	complete data V		0.48371	0.78234	0.81068	0.68327
	sampled data U		0.74979 (0.02094)	0.51026 (0.02161)	0.58840 (0.04718)	0.29731 (0.01607)
	sampled data V		0.48451 (0.05633)	0.78300 (0.01095)	0.81223 (0.01116)	0.69387 (0.02193)
region #2	theoretical U	triangle	0.70711			
		rectangle	0.75000			
	theoretical V	triangle	0.70711			
		rectangle	0.50000			
	complete data U		0.24599	0.70223	0.75412	0.77668
	complete data V		0.50031	0.32890	0.37663	0.55974
	sampled data U		0.24982 (0.02075)	0.70190 (0.02322)	0.74355 (0.03638)	0.77080 (0.01317)
	sampled data V		0.50255 (0.06146)	0.32999 (0.01625)	0.37797 (0.03768)	0.54873 (0.02427)

Note: Standard errors (the input for standard distances) are in parentheses. See Table 2 for the specifications of the map pattern generator, and the definitions of demand point distribution. Resampling procedure: initially, a 10-by-10 square grid tessellation superimposed on a unit square created geographic strata, and a sample of size five was drawn randomly from each strata; each replication randomly selected one sample from each strata.

3.3. The $p = 3$ Spatial Median Problem

Prevailing computer technology also enables investigations of modest-size $p = 3$ problems for which the calculation of exact solutions is possible for a rather small-size n (e.g., as large as 100, but requiring considerable computer execution time, far more than for the preceding $p = 1$ or 2 cases; see Figure 7). The combinatorial nature of the general p -median problem is such that for $p = 2$, TWAIN evaluates $n(n - 1)/2$ possible combinations (e.g., 4950 for $n = 100$); $p = 3$ for n between 9 and 10 achieves a similar order of magnitude—for $n = 500$ and $p = 2$ (124,750 evaluations), the $p = 3$ value of n is between 12 and 13; these computations refer to $n(n - 1)(n - 2)/6$. In other words, extending the method indorsed in this paper to $p > 3$ presents a challenge of devising more ingenious sampling strategies that dramatically reduce severe solution time requirements. Continuing the quest for a proof of concept by coalescing the solving of spatial optimization problems with a proposal from a consensus of expert opinions on the central limit theorem, which states that a minimum sample size should be at least 30, and the stratified random sample size experiments for this $p = 3$ section that used $n = 36$ for a 6-by-6 strata grid—a square grid—preserves sampling balance; thus, the $p = 3$ timing experiments (Figure 7) endorse

this sample size as having an affiliated efficient CPU execution time. Furthermore, the experiments utilized $n = 72$ (because its solution CPU execution time estimate is 2–3 min) and 1000 resamplings (because their cumulative solution CPU execution time estimate is 1–2 h). The preceding average resample solutions of $p = 1$ and $p = 2$ suggest that this number is sufficient, at least for exploratory work.

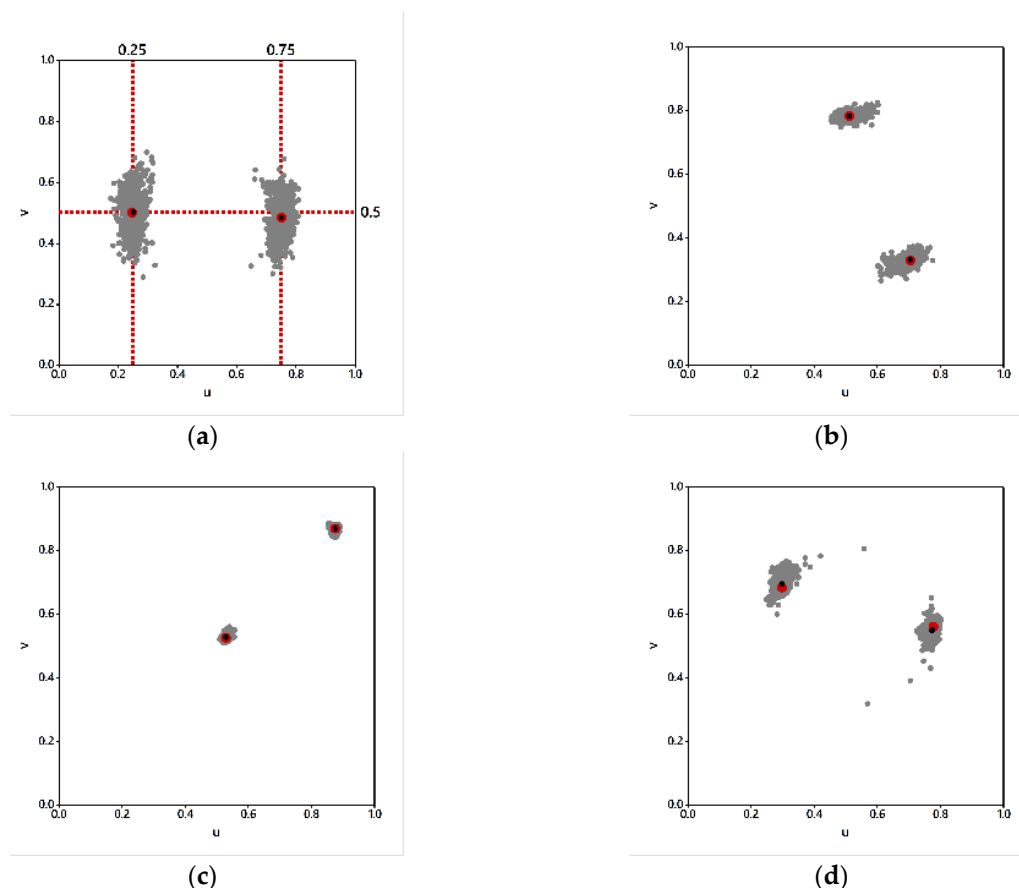


Figure 5. Sample spatial median geographic distributions (gray, red, and black filled circles respectively denote sample ($n \approx 1000$), complete data, and average spatial medians) for a uniform distribution of demand points; red crosshair reference lines denote theoretical bivariate spatial medians—in all cases, complete data and average spatial medians essentially collocate, whereas the slight deviation of the empirical spatial medians from their theoretical counterparts is due to sampling error in the single parent sample displayed. Top left (a): random independent weights. Top right (b): linear gradient of weights. Bottom left (c): quadratic gradient of weights. Bottom right (d): periodic geographic distribution of weights.

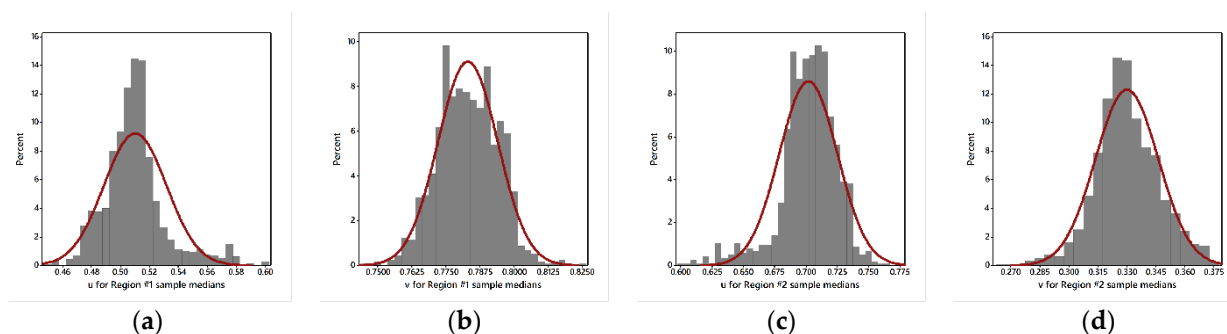


Figure 6. Skewness in the sampling distribution of $p = 2$ linear gradient ($n = 1150$) estimated spatial median coordinates (U_j , V_j); separate regional sample U_j and V_j coordinate histograms (gray bars) with superimposed normal curves (denoted by red lines). Left (a) Region #1 U_1 . Left middle (b): Region #1 V_1 . Right middle (c): Region #2 U_2 . Right (d): Region #2 V_2 .

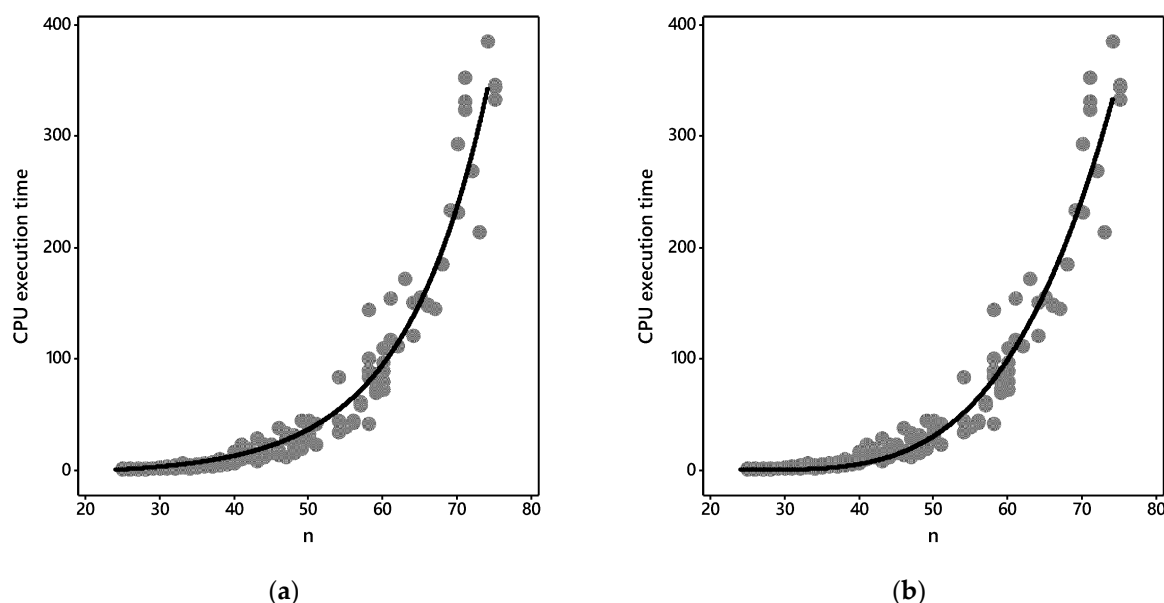


Figure 7. Central processing unit (CPU) timing (in seconds) experiments (197 after removing three infeasible/excessive anomalies) for $p = 3$ solutions (random samples from a uniform distribution of points and $w_i \equiv 1$) with randomly selected n ranging from 25 to 75; predicted–observed correlations are at least 0.95. Left (a): an exponential trend line function. Right (b): an $n^{3.7}$ trend line function.

3.3.1. The $p = 3$ Univariate Spatial Median Problem

Yet again, and primarily for illustrative purposes, the discussion here begins with a 1-D geographic landscape that involves scrutinizing an extension of the preceding Hotelling problem. For this $p = 3$ case, only for a uniform distribution of demand (Figure 8b) are the three regional spatial medians evenly spaced at the first, third, and fifth sextiles of the unit interval; similar to the $p = 2$ case, as the symmetric geographic distribution of demand goes from sinusoidal to uniform to bell-shaped, the absolute location of the extreme pair of spatial medians changes, repositioning from nearer the interval extremes to nearer the global spatial median location, which is also a regional spatial median for $p = 3$. In addition to this tendency for a variable spacing of regional medians that is a function of the form of the geographic distribution of demand (which reflects upon latent spatial autocorrelation), another important and noteworthy outcome highlights the persistent spacing gap between them, mirroring what also occurs with $p = 2$ solutions. A new manifestation is that the final total demand assigned to each spatial median is not necessarily the same (this allocation outcome characterizes only Figure 8b). This aspect complicates the notion of regional medians by introducing the need to estimate allocated demand as well as locations (see Table 4)—the second set of the preceding location-allocation problem. These unequal total demand share allocations (which do not include drastic deviations from uniformity) preserve the partitioning property of the Voronoi/Dirichlet/Thiessen polygon geographic landscape, stating that the boundary between two geographically juxtaposed regions is half the distance separating their spatial medians [32,33]; geometrically speaking, this is a perpendicular bisector location for 2-D settings. Table 5 tabulates the demand allocations for the three scenarios (Figures 1, 3 and 8) presented in this paper. This scenario also reinforces the preceding revelation that a link exists between spatial median problems and quantile and order statistics.

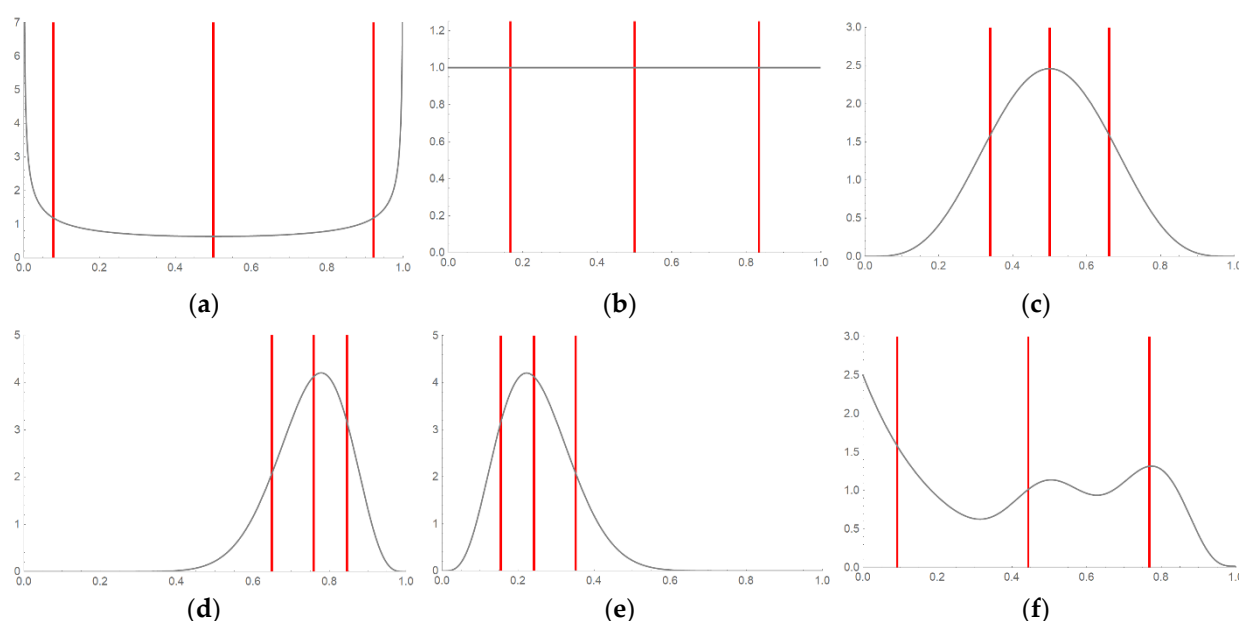


Figure 8. A representative range of possible distributions of demand across a unit interval. Gray curves denote the relative magnitude of demand at a location; red vertical lines denote the locations of $p = 3$ regional spatial medians. Top left (a): sinusoidal demand distribution. Top middle (b): uniform demand distribution. Top right (c): bell-shaped demand distribution. Bottom left (d): negative-skewed demand distribution. Bottom middle (e): positive-skewed demand distribution. Bottom right (f): multimodal demand distribution.

Table 4. The iterative solution for Figure 7d, initiated by assuming equal total regional demand.

Iteration	Regional Demand			Regional Spatial Median			Objective Function
	Region #1	Region #2	Region #3	Region #1	Region #2	Region #3	
0	0.500	0.000	0.500	0.137	0.440	0.759	0.250
1	0.387	0.276	0.337	0.097	0.440	0.766	0.117
2	0.381	0.292	0.327	0.095	0.440	0.767	0.115
3	0.374	0.298	0.328	0.093	0.440	0.767	0.114
4	0.371	0.301	0.328	0.092	0.440	0.767	0.114
5	0.371	0.305	0.324	0.092	0.440	0.767	0.114

Note: The objective function is the sum of the squared differences between the actual and ideal/uniform (i.e., $1/3$) allocations of demand to spatial medians.

Table 5. Percentage of allocations of demand to each spatial median for Figures 1, 3 and 8.

Geographic Distribution of Weights	$p = 1$	$p = 2$		$p = 3$		
		Region #1	Region #2	Region #1	Region #2	Region #3
sinusoidal	100	50	50	36.15	27.70	36.15
uniform	100	50	50	33.33	33.33	33.33
bell-shaped	100	50	50	30.92	38.16	30.92
negatively skewed	100	50	50	31.95	38.72	29.33
positively skewed	100	50	50	29.33	38.72	31.95
irregular multimodal	100	50	50	37.11	30.47	32.42

Note: Bold denotes an equal distribution of shares of the regional demand.

3.3.2. The $p = 3$ Bivariate Spatial Median Problem

Two important univariate $p = 3$ spatial median features that directly transfer to a $p = 3$ bivariate spatial median regional triplet are their dispersion tendency across a 2-D geographic landscape, often in the form of a triangular arrangement, and the proliferation of unequal total regional demand allocations to regional spatial medians. This dispersal is reminiscent of the packing of a plane when constructing urban geography central place structures (e.g., [34,35]), which, for a uniform distribution of demand points, results in a hexagonal lattice spacing of numerous regional spatial medians. The possible non-uniqueness of a solution complication also persists in this context, more problematic than for the $p = 2$ case, with spatial autocorrelation once again moderating it.

The preceding $p = 2$ sampling solution exploits spatial autocorrelation via a geographic tessellation stratified random design, and is then improved by applying the ALTERN heuristic to its result. The $p = 3$ procedure exemplified in this section builds upon this approach by applying ALTERN to each individual sample rather than to a sample average. This application is more effective because of the increased complexity companion accompanying increasing p . As mentioned previously, the simulation experiment performed for this section's analysis utilized a 6-by-6 sampling grid (to be close to the minimum sample size of 30), $n = 72$ (to balance the stratified sampling with two observations per areal unit), and 1000 replications (to allow the law of large numbers to take effect) because of timing requirements. In addition, for comparison purposes, the execution of the ALTERN heuristic from random allocation initiations furnishes benchmark optima. Table 6 tabulates selected results for this numerical example; Figure 9 portrays the simulated sample of sampling distributions for the objective function value pairs. Solution frequencies appearing in Table 6 substantiate the contention that exploiting spatial autocorrelation in the geographic distribution of weights is an effective strategy for solving spatial optimization L-A problems. Although the ALTERN heuristic is extremely fast and could easily be executed 10,000 times without truly taxing computing resources, repeatedly solving the exact $p > 2$ problem for a 6-by-6 stratification grid is not an onerous task, at least for $p = 3$. The tessellation stratified random sampling design ensures a wide geographic landscape coverage when computing a solution, a property that random sampling fails to guarantee. Furthermore, safeguarding the representativeness of each sample (via spatial autocorrelation and geographic stratification), together with the feasibility of each initial solution (e.g., regional groupings of demand points are coterminous), tends to render globally optimal solutions. Adding to these implications, Figure 9 unveils a strong tendency for spatial autocorrelation to concentrate local optima near their corresponding global optima, as measured by objective function values.

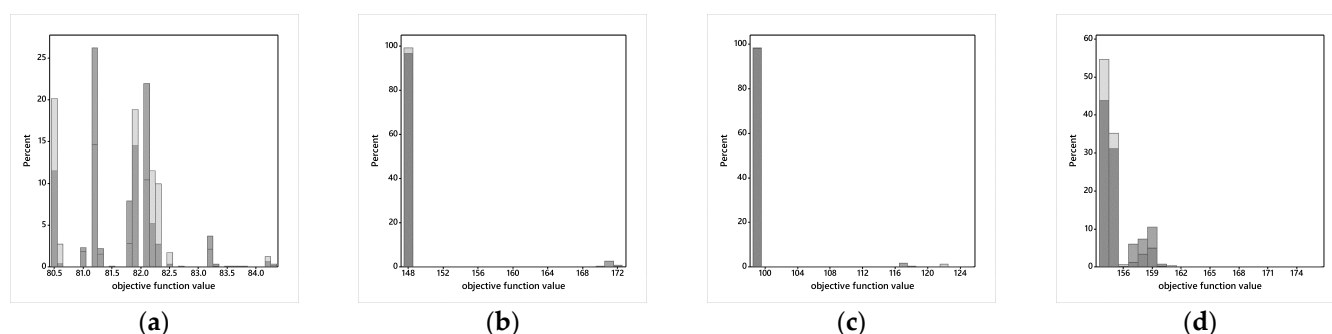


Figure 9. Histograms of the output of superimposed simulation objective function values for the two solution strategies summarized in Table 6; light gray denotes “sampling + ALTERN,” and dark gray denotes “random initiation of ALTERN.” Left (a): random weights. Left middle (b): linear gradient of weights. Right middle (c): quadratic gradient of weights. Right (d): periodic geographic distribution of w .

Table 6. A summary of $p = 3$ spatial median solution quality frequency of occurrence simulation experiment output; $n = 72$; 1000 replications.

Scheme	Solution Type	Sampling + ALTERN		Random Initiation of ALTERN	
		Frequency	Objective Function	Frequency	Objective Function
random	exact	1	80.5101	1	80.5101
	best sample	43	80.5101	92	80.5101
	2nd ranked sample	219	82.1280	123	81.1721
linear gradient	exact	1	147.9170	1	147.9170
	best sample	608	147.9170	0	147.9170
	2nd ranked sample	282	148.4167	873	148.4207
quadratic gradient	exact	1	98.6672	1	98.6672
	best sample	725	98.6672	0	98.6672
	2nd ranked sample	252	99.4786	669	98.6812
periodic geographic distribution	exact	1	154.0689	1	154.0689
	best sample	176	154.0689	47	154.0689
	2nd ranked sample	176	154.8659	141	154.8659

Note: No exact solution sets in these case studies contain multiple, non-unique global optima.

4. Discussion and Conclusions

To summarize, 1-D and 2-D statistical analyses can be illuminating for spatial statistical problems, transcending the conventional univariate and multivariate statistical specialties. These usages take interested scholars of the statistics discipline to the realms of centrography and spatial statistics, addressing ideas such as spatial mean, spatial median, standard distance, and a range of 2-D probability functions and random variables. This paper employs 2-D independent variables—and hence joint pdfs that are products of their pairs of marginal pdfs—to describe geographic distributions of discrete points across a continuous landscape, as well as to describe attribute values attached to these individual geocoded locations. The toy 1-D constructions assist in elucidating more real-world, tailored 2-D constructions, with the spatial medians treated here ranging from a single global to three regional locations.

In conclusion, a primary goal of this paper was to help fill the existing gap in the scholarly literature at the interface of spatial statistics and spatial optimization, with a secondary aim of encouraging considerably more future collaboration and interaction within this nascent domain. The research summarized here builds upon the fundamental and foundational spatial statistical concept of spatial autocorrelation: geographically informed sampling designs acknowledging a non-random mixture of geographic demand weight values that manifests itself as local homogeneous spatial clusters of these values can help spatial optimization techniques determine spatial optima, at least for L-A problems. One valuable discovery during this investigation is that existing but ignored spatial autocorrelation latent in georeferenced demand point weights undermines spatial optimization algorithms; unnoticed spatial autocorrelation seems to obscure their targets on the one hand, and on the other hand, they distract their trajectories from tracing a path to global optima by prodding them to stray to local optima along the way.

The existence of non-unique global optima for a spatial optimization problem is a serious complication apparently assuaged, at least to some degree, by spatial autocorrelation.

This obstacle may be a function of relatively small n and/or p , a uniform distribution of demand, the shape of a geographic landscape, the prevailing of zero spatial autocorrelation (an unlikely condition), or some combination of these and other factors. The reported cataloging in Table 6 raises existence questions about this possibility for practical empirical settings. Nevertheless, this paper documented success in solving very modest-size spatial optimization problems in order to obtain sizeable solutions for spatial optimization problems (which Figures 4 and 7 timing experiment scatterplots disclose as problems that quickly become intractable, and eventually infeasible) merely by exploiting spatial autocorrelation. This offers an innovative passageway—to fruitful future spatial statistical research endeavors in particular, and to those that are all-purpose statistical in general.

Future research also needs to expand p beyond 3, and establish how to generalize and upscale the findings encapsulated in this paper. In addition, other spatial optimization problems warrant comparable analytical attention, as do connections with other statistical science subfields such as quantile and order statistics.

Funding: This research was supported by the National Science Foundation, grant BCS-1951344. Any opinions, findings, and conclusions or recommendations expressed in this article are those of the author, and do not necessarily reflect the views of the National Science Foundation.

Institutional Review Board Statement: Not applicable.

Data Availability Statement: IMSL pseudo-random number generator routines produced all simulated data. Selected checks utilized SAS pseudo-random number generator routines.

Acknowledgments: The author is an Ashbel Smith Professor of geospatial information sciences.

Conflicts of Interest: The author declares no conflict of interest. The United States National Science Foundation had no role in: the design of the study; the generation, analyses, or interpretation of the data; the writing of the manuscript, or the decision to publish the article.

References

1. Raju, C. Probability in ancient India. In *Handbook of the Philosophy of Science*; Bandyopadhyay, P., Forster, M., Eds.; Volume 7: Philosophy of Statistics; Elsevier: Amsterdam, The Netherlands, 2011; pp. 1175–1195.
2. Stigler, S. *The History of Statistics: The Measurement of Uncertainty before 1900*; Belknap/Harvard: Cambridge, MA, USA, 1986.
3. Smith, D. *History of Mathematics*; Dover: New York, NY, USA, 1986; Volume 1–2.
4. Cliff, A.; Ord, J. *Spatial Autocorrelation*; Pion: London, UK, 1973.
5. Paelinck, J.; Klaassen, L. *Spatial Econometrics*; Saxon House: Farnborough, UK, 1979.
6. Cressie, N. *Statistics for Spatial Data*; Wiley: New York, NY, USA, 1991.
7. Griffith, D. A family of correlated observations: From independent to strongly interrelated ones. *Stats* **2020**, *3*, 14. [[CrossRef](#)]
8. Ghosh, A.; Rushton, G. *Spatial Analysis and Location-Allocation Models*; Van Nostrand Reinhold: New York, NY, USA, 1987.
9. Church, R.; Murray, A. *Location Covering Models: History, Applications and Advancements*; Springer: Cham, Switzerland, 2018.
10. Fermat, P. *Essai sur les Maximas et les Minimas*, in *Œuvres Fermat, Publiées par Lessoins de P. Tannery et C. Henry Sous les Auspices du Ministère de L'instruction Publique*; Gauthier-Villars et Fils: Paris, France, 1629; pp. 1891–1912.
11. Weber, A. *Über den Standort der Industrien*; J.C.B. Mohr: Tübingen, Germany, 1909; English translation: *The Theory of the Location of Industries*; Chicago University Press: Chicago, IL, USA, 1929.
12. Small, C. A survey of multidimensional medians. *Int. Stat. Rev.* **1990**, *58*, 263–277. [[CrossRef](#)]
13. Ninimaa, A. Bivariate generalizations of the median. In *Multivariate Statistics and Matrices in Statistics, Proceedings of the 5th Tartu Conference, Tartu-Pühajärve, Estonia, 23–28 May, 1994*; Tiit, E., Kollo, T., Niemi, H., Eds.; De Gruyter: Boston, MA, USA, 1995; pp. 163–180. [[CrossRef](#)]
14. Eftelioglu, E. Geometric median. In *Encyclopedia of GIS*, 2nd ed.; Shekhar, S., Xiong, H., Zhou, X., Eds.; Springer: Cham, Switzerland, 2017; pp. 701–704.
15. Kellerman, A. *Centrographic Measures in Geography*; GeoAbstracts, University of East Anglia: Norwich, UK, 1981.
16. Kuhn, H.; Kuenne, R. An efficient algorithm for the numerical solution of the generalized Weber problem in spatial economics. *J. Reg. Sci.* **1962**, *4*, 21–33.
17. Griffith, D. Using estimated missing spatial data in obtaining single facility location-allocation solutions. *l'Espace Géographique* **1997**, *26*, 173–182. [[CrossRef](#)]
18. Griffith, D. Using estimated missing spatial data with the 2-median model. *Ann. Oper. Res.* **2003**, *122*, 233–247. [[CrossRef](#)]
19. Anselin, L. Local indicators of spatial association—LISA. *Geogr. Anal.* **1995**, *27*, 93–115. [[CrossRef](#)]
20. Ord, J.; Getis, A. Local spatial autocorrelation statistics: Distributional issues and an application. *Geogr. Anal.* **1995**, *27*, 286–306. [[CrossRef](#)]

21. Griffith, D.; Chun, Y. Spatial autocorrelation in spatial interactions models: Geographic scale and resolution implications for network resilience and vulnerability. *Netw. Spat. Econ.* **2015**, *15*, 337–365. [[CrossRef](#)]
22. Griffith, D.; Paelinck, J. Chapter 2.6: Relationships between spatial autocorrelation and solutions to location-allocation problems. In *Morphisms for Quantitative Spatial Analysis*; Advanced Studies in Theoretical and Applied Econometrics Series; Springer: Berlin, Germany, 2018; pp. 18–22.
23. Hotelling, H. Stability in competition. *Econ. J.* **1929**, *39*, 41–57. [[CrossRef](#)]
24. Meager, K.; Teo, E.; Xie, T. Socially-optimal locations of duopoly firms with non-uniform consumer densities. *Theor. Econ. Lett.* **2014**, *4*, 431–445. [[CrossRef](#)]
25. Chen, L.-A.; Welsh, A. Distribution-function-based bivariate quantiles. *J. Multivar. Anal.* **2002**, *83*, 208–231. [[CrossRef](#)]
26. Olkin, I.; Trikalinos, T. Constructions for a bivariate beta distribution. *Stat. Probab. Lett.* **2015**, *96*, 54–60. [[CrossRef](#)]
27. Overton, S.; Stehman, S. Properties of designs for sampling continuous spatial resources from a triangular grid. *Commun. Stat.* **1993**, *22*, 251–264. [[CrossRef](#)]
28. Grekousis, G. Chapter 3: Analyzing geographic distributions and point patterns. In *Spatial Analysis Methods and Practice: Describe–Explore–Explain through GIS*; Cambridge University Press: Cambridge, UK, 2020; pp. 147–206. [[CrossRef](#)]
29. Wang, B.; Shi, W.; Miao, Z. Confidence analysis of standard deviational ellipse and its extension into higher dimensional Euclidean space. *PLoS ONE* **2015**, *10*, e0118537. [[CrossRef](#)] [[PubMed](#)]
30. Cooper, L. Heuristic methods for location-allocation problems. *SIAM Rev.* **1964**, *6*, 37–53. [[CrossRef](#)]
31. Ostresh, L. An efficient algorithm for solving the two center location-allocation problem. *J. Reg. Sci.* **1975**, *15*, 209–216. [[CrossRef](#)]
32. Hyson, C.; Hyson, W. The economic law of market areas. *Q. J. Econ.* **1950**, *64*, 319–324. [[CrossRef](#)]
33. Okabe, A.; Boots, B.; Sugihara, K.; Chiu, S. *Spatial Tessellations: Concepts and Applications of Voronoi Diagrams*, 2nd ed.; Wiley: Chichester, UK, 2000.
34. Dacey, M. The geometry of central place theory. *Geogr. Annaler. Ser. B Hum. Geogr.* **1965**, *47*, 111–124. [[CrossRef](#)]
35. Leepmeier, M. The Voronoi Cell in a saturated Circle Packing and an elementary proof of Thue’s theorem. *arXiv* **2019**, arXiv:1905.05837.