Spatial Statistics 57 (2023) 100768

Contents lists available at ScienceDirect

Spatial Statistics

journal homepage: www.elsevier.com/locate/spasta

Effects of geographically stratified random sampling initial solutions on solving a continuous surface p-median location problem using the ALTERN heuristic

Changho Lee [a], Daniel A. Griffith [a], Yongwan Chun [a,*], Hyun Kim [b]

[a] *Geospatial Information Sciences, School of Economic, Political, and Policy Sciences, The University of Texas at Dallas, Richardson, TX, USA*

[b] *Department of Geography and Sustainability, The University of Tennessee, Knoxville, TN, USA*

A R T I C L E   I N F O

Article history:
Received 31 July 2023
Accepted 1 August 2023
Available online 7 August 2023

Keywords:
Geographically stratified random sampling
p-median location model
Heuristics
Spatial autocorrelation

A B S T R A C T

In the fields of location theory and spatial optimization, heuristic algorithms have been developed to overcome the *NP*-hard nature of solutions to their problems, which results in an exponential increase in computation time. These algorithms aim to generate good initial solutions, narrow the solution space, and guide the search process to optimality. Geographically stratified random sampling (GSRS) can be regarded as a method to generate such high-quality initial solutions. This study investigates the application of GSRS to solving the p-median location problem on a continuous surface solution space punctuated with weighted demand points, and its impact on the performance of the popular ALTERN heuristic algorithm. Results demonstrate the effectiveness of GSRS in finding optimal p-median solutions, but only for smaller p values: the ALTERN heuristic with initial solutions generated by local spatial means from GSRS for these smaller p always produces optimal final solutions. In contrast, implementing a random search by executing a large number of random initial solutions often produces non-optimal results. Findings reported in this paper also highlight that sample size and degree of positive spatial autocorrelation (PSA) in the geographic distribution of weights influence how close final solutions are to

* Corresponding author.
E-mail addresses: changho.lee@utdallas.edu (C. Lee), dagriffith@utdallas.edu (D.A. Griffith), ywchun@utdallas.edu (Y. Chun), hkim56@utk.edu (H. Kim).

https://doi.org/10.1016/j.spasta.2023.100768

optimality for larger p. Increasing the sample size leads solutions to be concentrated near their optimal counterparts, as does increasing PSA levels.

1. Introduction

The *NP*-hard nature of location problems poses significant challenges in finding optimal solutions within a reasonable timeframe. These problems require exponential time in the number of medians and the number of demand points for exact solutions, making it impractical for many, if not most, real-world applications. To overcome this challenge, heuristic algorithms have been developed as efficient and effective alternatives. These algorithms focus on finding optimal or near-optimal approximate solutions by employing strategies such as generating good initial solutions, narrowing the solution space, and guiding the search process (Beheshti and Shamsuddin, 2013).

The selection of an appropriate initial solution is crucial to efficiently exploring a solution space. A well-informed initial solution can significantly improve the effectiveness of an algorithm by providing a starting point close to its desired optimal solution. Perttunen (1994) emphasizes the importance of an initial solution that is in proximity to the optimal solution. In the context of edge exchange heuristics for routing problems, that study investigates the impact of the initial solution's quality on performance. Computational experiments conducted with Euclidean traveling salesman problems reveal that judiciously constructed initial solutions outperform randomly generated solutions. This outcome highlights the significance of selecting a suitable initial solution, as it can influence the overall performance of an algorithm in terms of optimality and time efficiency.

One important variant of location problems is the p-median, which involves determining the optimal placement of a single or multiple facilities to minimize aggregate (weighted) travel distance from demand points to the selected p facilities. Due to its *NP*-hard nature, finding exact optimal solutions for large-sized p-median problems becomes computationally infeasible (Kariv and Hakimi, 1979). While random initial solutions are dominantly utilized in heuristics, they generally do not reflect important spatial patterns in their demand weights, such as spatial autocorrelation (SA). Instead, heuristic approaches commonly compare multiple solutions using different random initial solutions, and then report the best solution from a collection of, say, 10,000. However, leveraging prior knowledge, such as an initial solution, can be highly beneficial in efficiently searching for near-optimal solutions. By incorporating an initial solution that captures essential characteristics of the problem, the algorithm can more effectively navigate the solution space and improve the likelihood of finding optimal p-median solutions (Mu and Tong, 2018).

Using geographically (i.e., tessellation) stratified random sampling (GSRS) can furnish valuable prior knowledge for the initial solution when identifying a p-median solution. GSRS is a sampling technique that divides a study area into n distinct geographic strata, and then randomly selects a sample of size one from within each stratum. One effect of GSRS is to avoid concentrations of samples in a small area of a map, thus enhancing the spatial coverage of the sample, making it more representative. This sampling approach effectively captures the spatial variability of the area, reducing bias. Another effect is the grouping of similar values, for positive SA (PSA), such that repeated local samples mimic a single value with measurement error. Hence, GSRS exploits redundant information represented by PSA to effectively reduce the dimensionality of a p-median problem (after Griffith et al., 2022). In doing so, it enables the derivation of high-quality initial solutions that accurately represent the spatial distribution of demand locations in a p-median location problem.

This paper aims to investigate the effects of GSRS on the initial heuristic solution for identifying p-medians on continuous surfaces using the ALTERN heuristic (Cooper, 1964). Four research questions are formulated based upon simulation experiments. First, the paper explores the effectiveness of GSRS in achieving optimal p-median solutions for all demand locations. Second, it examines the impact of increasing the size of GSRS on the frequency of optimal solution achievement. Third, it summarizes an investigation about how different levels of PSA influence *p*-median solutions. Finally, it assesses how initial solutions derived from GSRS guide heuristic solutions toward optimality. Overall, the objective of this paper is to examine if integrating GSRS, varying sampling sizes and PSA levels, and executing the ALTERN heuristic can establish an effective approach for achieving optimal *p*-median solutions.

2. Literature review

The integration of GSRS into solving the *p*-median problem involves combining principles from spatial statistics and location theory. While GSRS and *p*-medians are popular in their respective research fields, namely spatial statistics and spatial optimization, there is a lack of research discussing their synthesis and potential synergies. One perspective connecting them discusses that GSRS offers the advantage of using a smaller sample size to represent a population. PSA latent in attribute variables geotagged to locations (e.g., weights), as well as direct interactions among these locations, can bolster this advantage (Griffith, 1992). By employing localized randomized sampling locations systematically spaced over a two-dimensional surface containing PSA, GSRS ensures a comprehensive geographic coverage and more precise descriptive statistics estimates, effectively representing the original geographic landscape with greater accuracy (Griffith, 2005). Therefore, opting for GSRS samples instead of using the entire set of demand locations has the potential to effectively reduce a problem size without greatly compromising its capability to yield optimal solutions for an original location problem.

Overton and Stehman (1993) further demonstrated that the GSRS design outperforms a strict systematic design in terms of variance estimation and overall precision. The strict systematic sample is obtained by randomly shifting a triangular grid, whereas the GSRS involves selecting one random point per stratum. The strata are formed by hexagons naturally created from the tessellation of a triangular grid. These researchers evaluate the performance of a proposed variance estimator using different surface model representations of a synthetic continuous response variable, such as linear, quadratic, and sinusoidal gradient surfaces. These surfaces represent spatial continuity and the presence of differing degrees of SA. The tessellation-stratified design avoids challenges associated with high-resolution patterns in strict systematic sampling, while still providing the advantage of equal representation across the spatial domain, similar to a systematic design. Its success lies in its ability to tap its design effect for possible improved precision; in the worst case scenario of a random mixture of attribute values (i.e., white noise rather than a geographic gradient), it renders the same precision as a classic random sample governed by the Central Limit Theorem.

Due to its advantages, GSRS has been widely utilized as an effective sampling methodology to capture the overall trend on a continuous surface. It has found applications in various fields such as ecology, remote sensing, and land cover monitoring. For instance, Stevens (1997) proposed GSRS as a suitable approach for sampling ecological resources like mineral reserves, vegetation cover, and chemical concentration in streams. In the field of remote sensing, Stehman (2009) demonstrated that GSRS improves upon random sampling, addressing biasing problems in accuracy assessment of land cover. Brink and Eva (2009) employed a 1% sampling rate within each stratum of GSRS to monitor land cover change in sub-Saharan regions over 25 years; this implementation overlooks the twofold trade-off benefits of gaining improved geographic coverage and reducing within-sample SA by employing smaller strata and drawing a single sample observation from each. Similarly, Hope et al. (2003) utilized GSRS-derived 4-by-4-km squares to incorporate spatial statistical analyses of biotic, abiotic, and human variables, covering the entire Central Arizona-Phoenix area.

Meanwhile, efforts to efficiently and effectively solve the *p*-median problem have been made to address the challenges posed by *NP*-hard problems and local optimality traps compromising heuristic approaches. Existing studies, such as Rosing et al. (1998), primarily focus on the *p*-median problems with discretized weights surfaces. These studies explore various solution techniques, including heuristic algorithms and integer linear programming. However, a lack of research exists that investigates the integration of the *p*-median problem with continuous surfaces, leaving room for further investigation and exploration in this area.

Rosing et al. (1998) conduct a comparative analysis between the Teitz & Bart heuristic (T&B; Teitz and Bart, 1968) and Tabu search (Rolland et al., 1997). The results show that Tabu search outperforms T&B in terms of both speed and solution quality (optimality or near-optimality). Additionally, these researchers introduce the concept of heuristic concentration, which yields better solutions compared to Tabu search, although they do not compare solution times. Heuristic concentration incorporates a component called the concentration set, derived from multiple runs of T&B, which has a high probability of including the facilities that constitute the optimal solution to the proposed problem (Rosing and ReVelle, 1997). Furthermore, Rosing and Hodgson (2002) utilize combinations within the concentration set to identify an optimal p-median solution, resulting in a significant reduction in the proposed problem size. The GSRS tactic follows this type of logic, using random sampling in a spatial statistics context to formulate an equivalent to this aforementioned concentration set.

Few attempts exist that integrate spatial statistics into solving location problems, aiming to leverage its advantages. Griffith et al. (2022) highlight the benefits of integrating SA principles into location problems. These benefits include addressing spatial outliers and hot/cold spots to reduce problem complexity, imputing missing data to enhance accuracy, and increasing the likelihood of achieving a global optimum through the use of GSRS, which to date has been limited to very small p cases. This paper builds upon these integration efforts by specifically focusing on the integration of GSRS, in terms of its management of SA, into location problems on continuous surfaces. The chief goal is to overcome particularly solution time and computer memory limitations imposed by *NP*-hard problems while shifting heuristic algorithm solutions to guaranteed optimality.

3. Methods

The principal idea of the proposed method is to generate good initial solutions for a heuristic algorithm with a smaller representative sample set using GSRS. Fig. 1 provides a 6-by-6 grid illustrative example of GSRS, with the simulation experiment utilizing various sample-size-increasing tessellation configurations (i.e., 6-by-6 to 10-by-10 grids), from which a single sample is to be drawn from each grid cell stratum, the specimen sample here denoted by red. The anticipation is that such a sample set represents a spatial pattern of its full parent dataset better than a non-spatial random sample set with the same size, because it tends to better reflect the spatial characteristics of demands with improved coverage of the study area and a reduction in sample SA (Griffith, 2005). For example, weights at two closely located demand points tend to have similar values if the weights surface has a high level of PSA. In contrast, a GSRS sample does not necessarily contain more information about spatial characteristics if the weights surface is completely spatially random. The effectiveness of GSRS as an initial solution for p-median heuristics is investigated with simulation experiments in this paper. These experiments consider the following factors: the level of SA in a weights surface of demands, and the number of strata, n, in GSRS.

3.1. Experimental design

The simulation experiment is designed to generate output allowing an investigation of two issues. First, it supports examining if p-median solutions using small size GSRS samples concentrate at their optimal solutions determined with their entire demand set. Because a Central Limit Theorem is in effect (e.g., classical random sampling occurs within each stratum), the expectation is that such solutions disperse around their optimal solutions, especially if the sample sets reflect the entire demand set well. Second, it empowers checking if p-median ALTERN solutions using GRSR samples can have a positive impact on the performance of a p-median heuristic for continuous space. For the first, multiple p-median solutions are calculated using GSRS, and then their spatial mean is obtained. For the second, ALTERN solutions are produced using these spatial means as an initial solution, and then the results are compared to ALTERN solutions with non-spatially random initial solutions. The final benchmark is a comparison between these and their exact solution counterparts.

The first step of the experiment begins by initializing 700 random demand locations within a unit square. Then weights values are assigned to these 700 demand points. Four different formulae

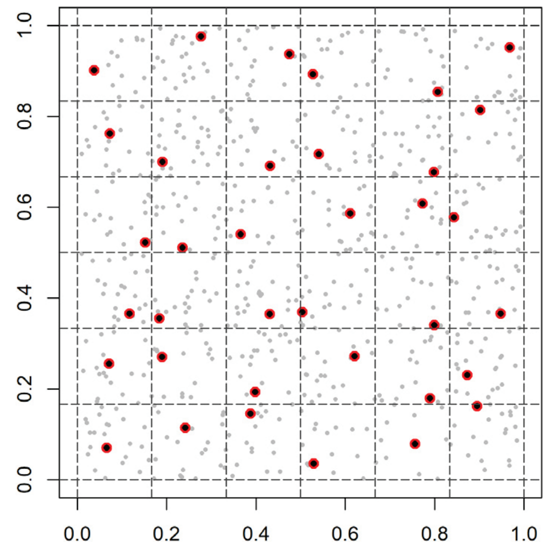

Fig. 1. A 6-by-6 GSRS illustration for a unit square geographic landscape.

(random, weak, moderate, and strong) are utilized in order to generate different levels of PSA; that is, surfaces with four different SA degrees are constructed in order to examine if the level of SA has an impact on p-median heuristic solution quality. Details about the weights surfaces are discussed in the next section. Subsequently, a GSRS sample set is drawn using square strata with a 6-by-6 tessellation (i.e., 36 strata establishing that $n = 36$). This action involves selecting one demand location from each square stratum. Then, a p-median solution for the small GSRS sample is computed. This exercise is repeated 100 times, which produces 100 p-median solutions, one for each of the 100 GSRS samples. This procedure also involves two more factors: conducting the same process with a range of other sizes of strata coarseness, specifically, 7-by-7, 8-by-8, 9-by-9, and 10-by-10 (i.e., $n = 49$, 64, 81, and 100 across the experimental steps); and, the p-median is solved for a variety of p values ranging from 2 to 15.

To assess the performance of the GSRS, each set of 100 replicate p-median solutions is compared with its exact optimal solution based upon the entire population of discrete 700 weights; construction of continuous weights surfaces merely ensures that each entry in a randomly selected set of 700 demand points has a weight attached to it affiliated with a given SA level. If GSRS subsets perform well, then their solutions concentrate close to their associated optimal solutions. Spatial means and standard distances respectively index the central tendencies and dispersions of these clusterings. To ensure robustness (i.e., engage the Law of Large Numbers), this entire process was repeated 1,000 times, resulting in 1,000 sets of spatial means. Standard distances measure similarities between these spatial means and their corresponding optimal solutions. This analysis allows an evaluation of GSRS effectiveness with regard to capturing the geographic distribution of all demand locations, as well as its ability to produce solutions that closely position themselves near their individual optimal solutions. Fig. 2a illustrates the first step of the experiment.

Design of the second step of the experiment is to assess the ALTERN algorithm performance based upon initial solutions using the spatial means obtained from the first step. Fig. 2b presents an overview of this process. The next execution is of the ALTERN heuristic for the p-median problem with $n = 700$ utilizing each of the 1,000 sets of spatial means as initial solutions. These are the sources of the output summarized and evaluated here. To establish a benchmark, a control set is constructed with non-spatial random initial solutions. The comparison between these results without using GSRS (a control set) and those results obtained using GSRS help uncover the impact of GSRS integration on the quality of enhanced ALTERN heuristic solutions.

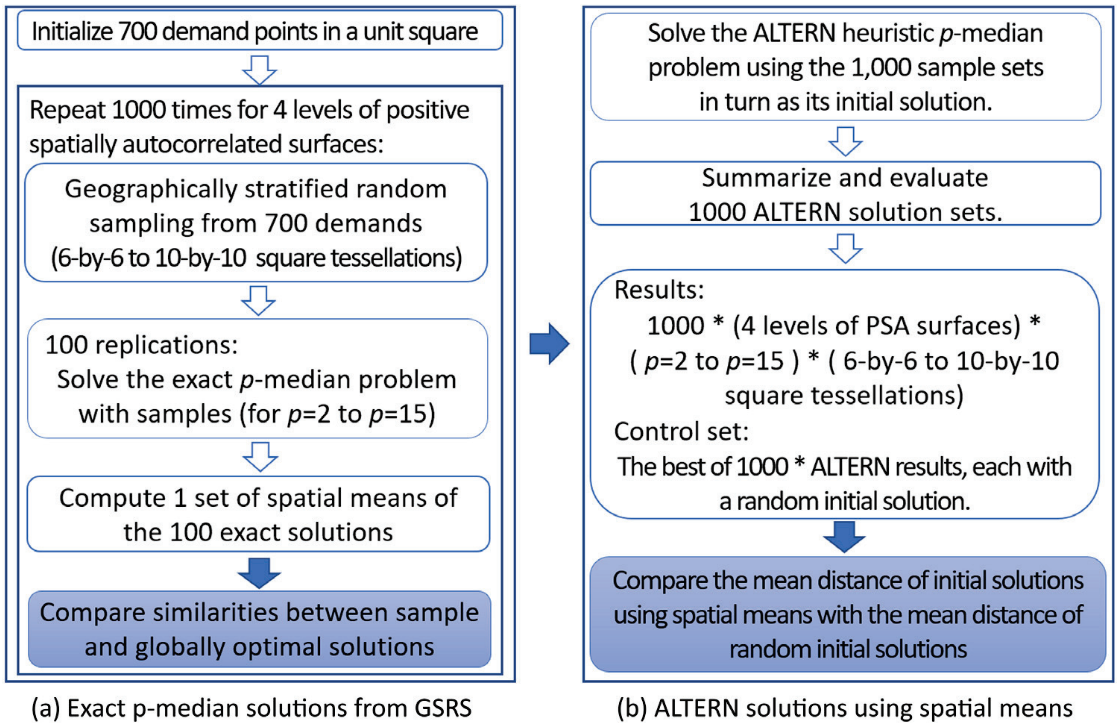

Fig. 2. A flow chart for the experimental design.

3.2. Location problem settings

The experimental environment was established based on the two designs devised by Rosing and Hodgson (2002) and Overton and Stehman (1993), which involve a set of different weights surfaces. In this experiment, four different types of weights surfaces are utilized, each embedding a different PSA level, from random, through weak and moderate, to strong. Fig. 3 displays the spatial distribution of demands and weights surfaces using the following four formulae:

$$\text{Random}: \ f_{\text{random}}(u, v) = \text{Poisson}(\mu = 4) + 1 \tag{1}$$

$$\text{Weak}: \ f_{\text{weak}}(u, v) = 0.5 \times [\sin(10u\pi) + \sin(10v\pi)] + 5 + 0.5 \times \text{Normal}(\sigma = 0, \mu = 1) \tag{2}$$

$$\text{Moderate}: \ f_{\text{moderate}}(u, v) = 10 \times \left[(u - 0.5)^2 + (v - 0.5)^2\right] + 3.5 + \text{Normal}(\sigma = 0, \mu = 1) \tag{3}$$

$$\text{Strong}: \ f_{\text{strong}}(u, v) = 5 \times (u + v) + 0.5 \times \text{Normal}(\sigma = 0, \mu = 1) \tag{4}$$

where u and v represent (u, v) coordinates of the 700 demand locations scattered across a 2-D square surface ranging from 0 to 1 for both u and v. The random weights surface is generated as Poisson pseudo-random numbers having a mean of 4, with one (+1) added to each in order to avoid zero weights; hence, the mean of the weights is five. The weak surface is generated with a cyclic pattern using sine and cosine functions. The moderate surface has a quadratic form with low weights around the center, and large weights around the edge, of the unit square. The strong surface is generated with a linear gradient coupled with a diagonal orientation aligning along a line spanning the lower left-hand corner to the upper right-hand corner. Note that all of the weights surfaces have a mean of approximately 5. In addition, the weights surfaces in Fig. 3b–3e are scientific visualizations via spatial interpolation (i.e., spherical variogram model in ordinary kriging interpolation) based upon the demand locations and their attached weights.

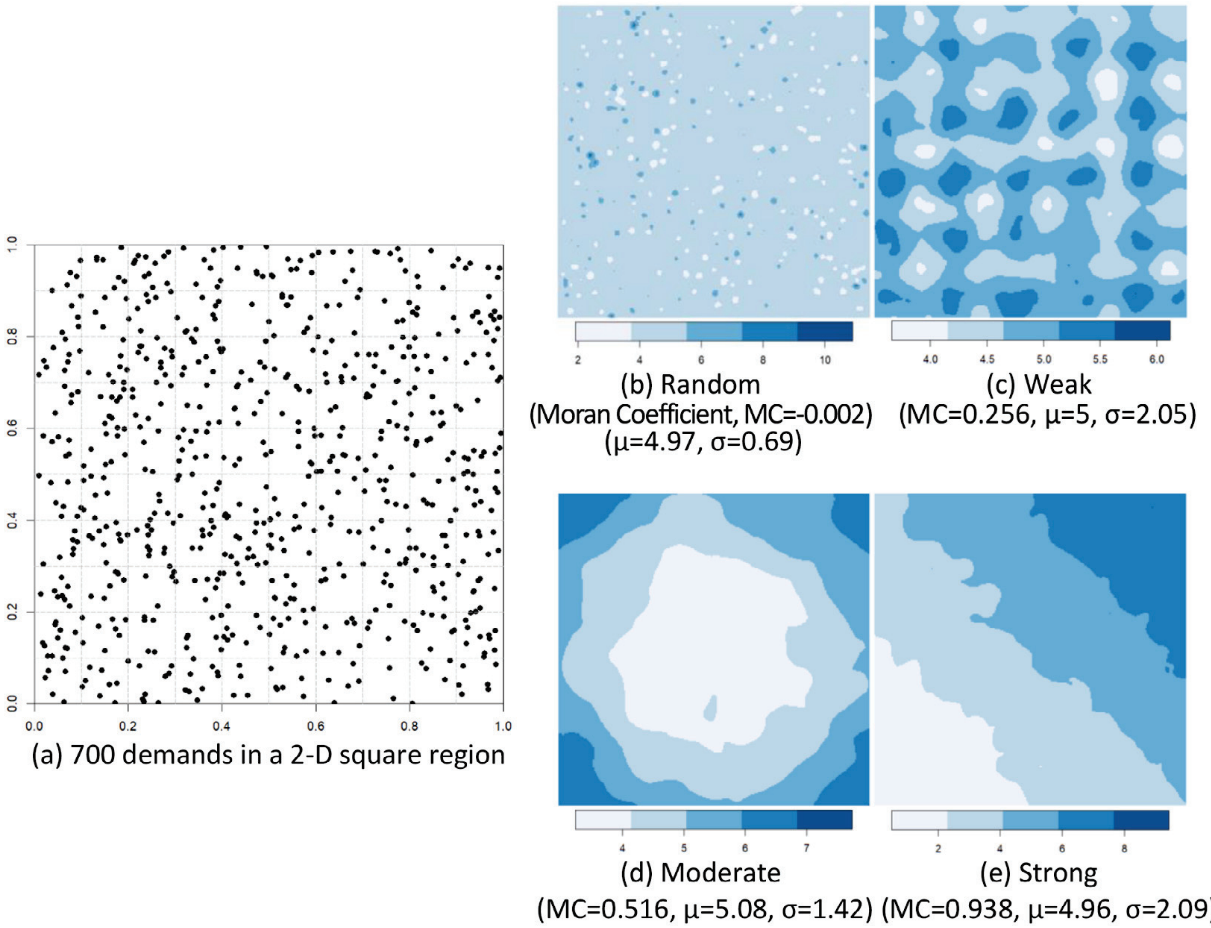

Fig. 3. Demand locations and four specimen weights surfaces.

3.3. The ALTERN heuristic algorithm: An overview

The ALTERN heuristic algorithm whose name derives from beginning with an initial allocation (i.e., grouping) of demand points, and then iteratively and alternately optimally locating p medians followed by re-*allocating* demand points to them, until no further changes occur—is capable of solving p-median problems in continuous space. Its nickname acronyms include ALT and ALA; it is based on Cooper's (1964) proposed ALTERNating (the source of the acronym) location–allocation algorithm. This algorithm follows two principles for solving the p-median problem. The first involves dividing the demand point set into p groups, effectively allocating the n demand points to an original set of p unknown spatial medians but p known subregions housing them. For each of these p subsets of demand points, the single facility Weber problem (Weber, 1922) can be solved, to determine the unknown median points. The 1-median problem may be formulated as the following Eq. (5), with a single median location denoted by (U, V):

$$MIN: Z = \sum_{i=1}^{n} w_i \sqrt{(u_i - U)^2 + (v_i - V)^2}, \tag{5}$$

where n sets of (u_i, v_i) and (U, V) are Cartesian coordinates.

Cooper's second principle states that when the locations of p spatial medians are known, the n demand points can be re-assigned to their closest medians to reveal an optimal allocation. The ALTERN heuristic algorithm alternates between these two principles until the computed objective function value with the sum of weighted distances no longer decreases This algorithm incorporates the Weber function that numerically solves a pair of partial differential equations (∂Z/∂U = 0 and ∂Z/∂V = 0) derived from Eq. (5) as the following Eq. (6), as described by Kuhn and Kuenne (1962):

$$U^{(\tau+1)} = \frac{\sum_{i=1}^{n} \frac{w_i u_i}{d_i^{(\tau)}}}{\sum_{i=1}^{n} \frac{w_i}{d_i^{(\tau)}}}, \textit{ and } V^{(\tau+1)} = \frac{\sum_{i=1}^{n} \frac{w_i v_i}{d_i^{(\tau)}}}{\sum_{i=1}^{n} \frac{w_i}{d_i^{(\tau)}}}, \qquad (6)$$
$$\textit{where } d_i^{(\tau)} = \sqrt{\left(u_i - U^{(\tau)}\right)^2 + \left(v_i - V^{(\tau)}\right)^2}$$

The distinguishing feature of the ALTERN algorithm is its ability to identify p-medians on a continuous surface, setting it apart from other heuristic algorithms based on integer linear programming techniques that require a discrete solution space. Its drawback is that they may not be optimal.

4. Results and discussion

This section presents the results of these aforementioned experiments that underscore the effectiveness of GSRS samples in achieving optimal solutions for the p-median location problem on a continuous surface. It begins with the results of examining whether or not p-median solutions with GSRS subsets concentrate around the global optimal solutions with all demand points. Then it presents how initial solutions generated using subset solutions with GSRS contribute to ensuring the optimality of the heuristic solutions for the entire set of n demand points.

4.1. Results of the first step in the experiment: Spatial patterns of spatial means using GSRS

The p-median solution results obtained with GSRS reveal two noteworthy outcomes. First, spatial means of the p-median locations using these samples tend to densely concentrate on or near their corresponding optimal locations. Fig. 4 illustrates this result, where the spatial means (blue dots) closely align with their respective optimal locations (red circles) for the example of $p = 4$ medians obtained when strong PSA prevails. Second, as the sample size (i.e., number of strata) increases, the sample p-median solutions tend to increase their concentration around their corresponding spatial means. In Fig. 4, the point clouds in four pastel colors are the p-median solutions obtained from GSRS samples that progressively get closer to their affiliated spatial means (blue dots) as the number of strata increases. When the number of strata is 81 and 100, the four separate point clouds become conspicuously distinct. In contrast, for smaller numbers of strata, these point clouds tend to constitute overlapping groups. As an aside, the pastel colors were arbitrarily chosen to represent four groups corresponding to the four globally optimal solutions (calculated with the entire set of demand points). These point clouds disperse around both the spatial means and the globally optimal solutions.

The two previous findings show that increasing the strata size for GSRS can lead to a convergence of solutions toward optimality. This pattern is also observed for the other p-median cases (i.e., $p = 2, \ldots, 15$). Fig. 5 showcases a decrease in the mean of the standard distances (SDs) for all cases as the number of strata in GSRS increases. The mean of SD represents the average dispersion of a solution; more specifically, the average value of SD calculated for each allocation to p groups.

Furthermore, an increase in the PSA level consistently leads to decreases in the mean SD for each p median, except for $p = 6$ paired with a moderate surface, and $p = 5$ paired with a strong surface (as shown in Fig. 5). Notably, the SD mean is highest for the random surface, whereas the strongest SA surface exhibits its smallest values. Values for the weak and moderate SA surfaces show similar or slightly different tendencies. These findings signify a significant contribution of GSRS, particularly when a surface displays a high PSA level, to computing excellent quality initial ALTERN p-median solutions, and subsequently their globally optimal ALTERN counterparts.

These results show that overall, the quality of the solutions obtained with GSRS is extremely encouraging. Individual solutions are dispersed but they form sampling distribution clusters around their optimal solutions. Because the affiliated spatial means are very close to their globally optimal solutions for the entire set of demand points, findings reported here suggest that spatial scientists should use the spatial means of a large number of sample p-median solutions acquired with GSRS as initial solutions when executing the ALTERN heuristic for large-to-massively-large n, and $p \leq 15$.

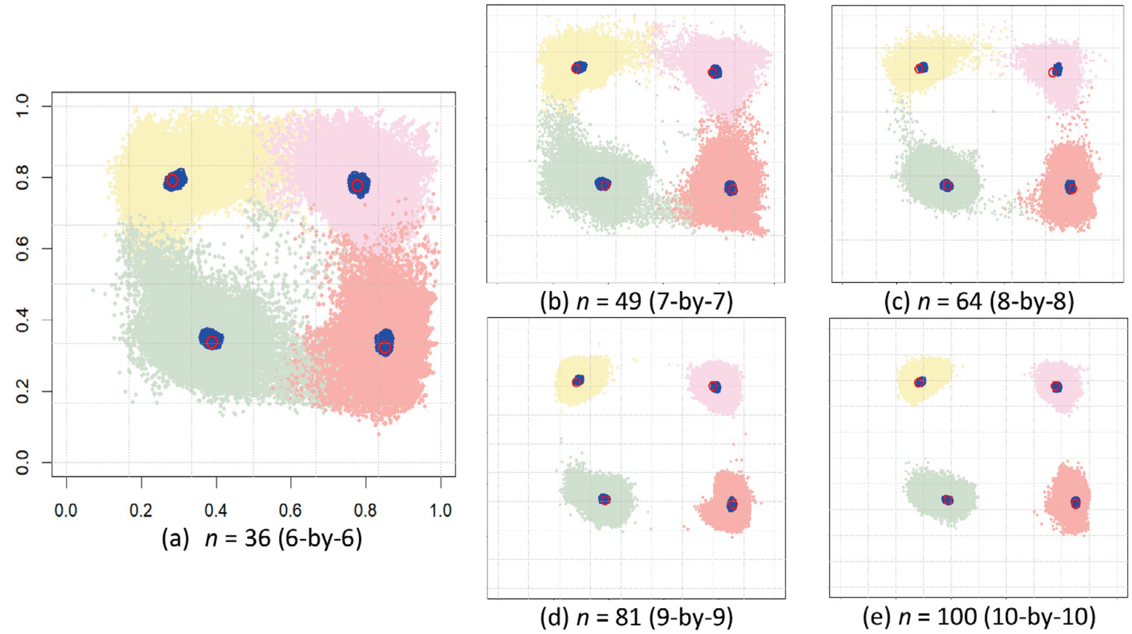

Fig. 4. Exact 4-median solutions obtained using GSRS and the spatial distributions of spatial means for the strong PSA surface. Blue dots denote the spatial means, whereas red circles denote the globally optimal solutions. The other pastel-colored dots denote the exact sample solutions.. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

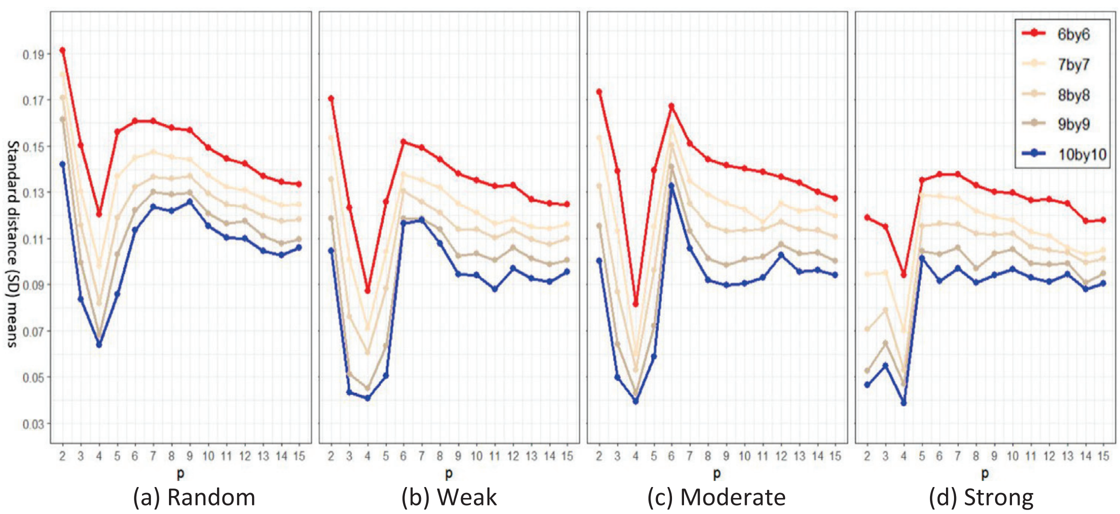

Fig. 5. Standard distance (SD) means for four weights surfaces.

4.2. Results of the second step of the experiment: Global optimality of heuristic solutions using GSRS

Results from the second experimental step clearly indicate that initial solutions generated using GSRS spatial means contribute noticeably to ensuring optimality of the final ALTERN heuristic solutions. Fig. 6 provides a clear demonstration of the effectiveness of GSRS spatial means as excellent initial solutions for the ALTERN algorithm, exemplified by the $p = 5$ case coupled with a 6-by-6 strata tessellation. Fig. 6b illustrates that the ALTERN heuristic achieves near-perfect optimality in 1,000 runs when using GSRS spatial means (Fig. 6a) as initial solutions, with only a few exceptions. In contrast, when random initial solutions are utilized (Fig. 6c), the resulting

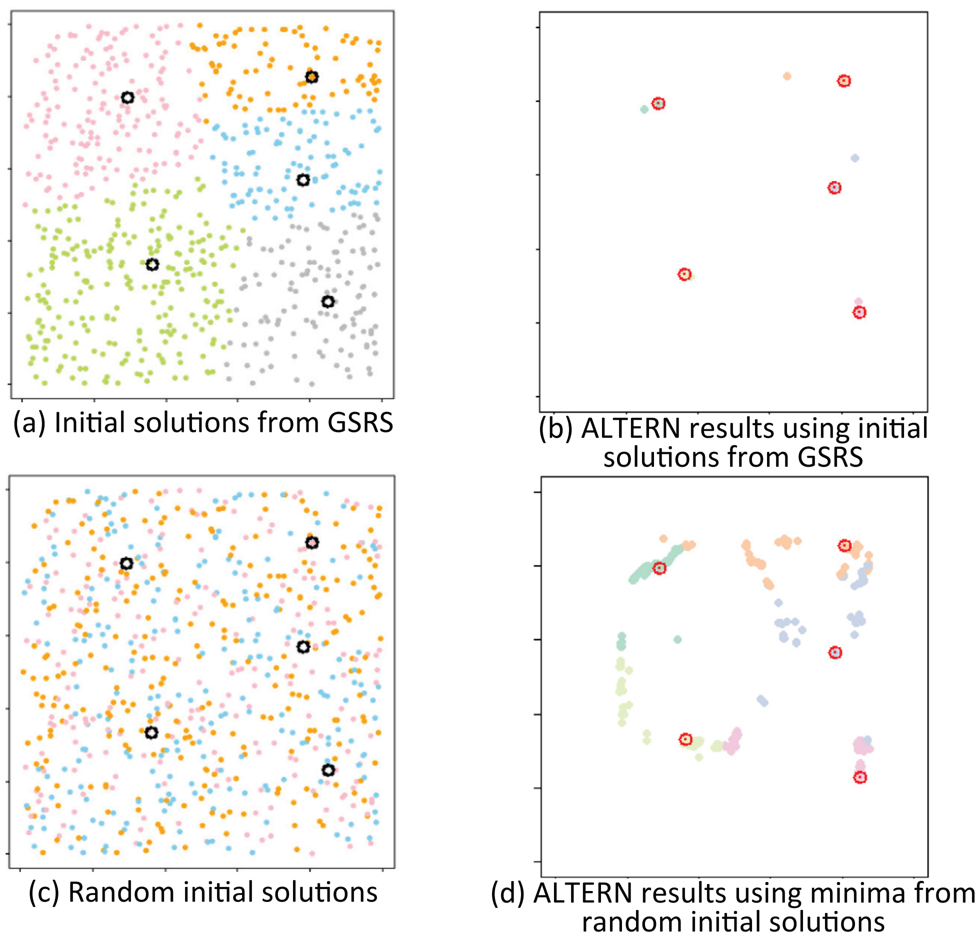

Fig. 6. Initial and final solutions from the ALTERN heuristic for a strong PSA surface. In (a) and (c), black circles denote optimal locations, and colored dots denote initial solutions, with the same color indicating the same allocation as an initial solution. In (b) and (d), red circles denote optimal locations, and colored dots denote the ALTERN solutions. GSRS is executed using a 6-by-6. square tessellation.. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

ALTERN solutions exhibit higher variance in their geographic distributions, with their empirical minima deviating significantly from their respective optimal solutions (Fig. 6d).

For all p values, ranging from 2 to 15, across four weights surfaces and various numbers of strata in GSRS, the ALTERN solutions consistently demonstrate near-perfect optimality. Optimality is assessed based on a threshold of 0.093, to isolate rounding errors based upon the variance of random distances between two points in a unit square (Bäsel, 2021), applied to the means of objective function distances,[1] the average distance between solutions with samples and their optimal counterparts. Table 1 provides a tabulation of the optimality frequencies observed for the ALTERN solutions. These results indicate that the ALTERN solutions achieve optimality frequencies exceeding the threshold, except for the case of $p = 6$ using an initial solution generated from a 10-by-10 square tessellation on a moderate PSA surface. However, even in this exceptional case, the frequency of achieving optimality is remarkably high, with 971 out of 1,000 runs resulting in

[1] The threshold of 0.093 was established with a simulation experiment calculating average distance between two randomly located points (10,000,000 replications) in: a unit square and circle which captures all polygons in the transition from one to the other; and, in a rectangle and ellipse analogous to expanding a landscape shape respectively from a square and a circle. The area was kept constant, which means the dimensions of the landscapes expanded or contracted from 1. In all cases, the simulated 2.5% critical value was around 0.093.

Table 1
The frequencies of global optimality with a threshold of 0.093 under 1000 ALTERN runs.

Initial solution	PSA	Sample size	p = 2	3	4	5	6	7	8	9	10	11	12	13	14	15
ALTERN using GSRS	Strong	100	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000
		81	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000
		64	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000
		49	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000
		36	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000
	Moderate	100	1000	1000	1000	1000	**971**	1000	1000	1000	1000	1000	1000	1000	1000	1000
		81	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000
		64	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000
		49	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000
		36	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000
	Weak	100	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000
		81	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000
		64	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000
		49	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000
		36	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000
	Random	100	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000
		81	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000
		64	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000
		49	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000
		36	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000
ALTERN with random initial solution	Strong	NA	675	**1000**	**1000**	594	756	430	912	794	857	952	950	954	963	982
	Moderate	NA	573	950	978	736	201	494	688	844	932	965	983	980	898	861
	Weak	NA	608	984	999	821	515	393	564	812	909	967	959	**1000**	932	904
	Random	NA	592	949	984	692	622	564	473	547	807	901	988	947	967	851

NA: GSRS is not applicable. One random initial solution is used.

optimal solutions. These frequencies are obtained by conducting 1,000 executions of the ALTERN algorithm, each with 1,000 different sets of spatial means. In contrast, the ALTERN solutions using random initial solutions exhibit lower frequencies of global optimality, with exceptions occurring for $p = 3$ and 4 on a strong PSA surface, and $p = 13$ medians on a weak PSA surface.

Taking a microscopic perspective, the majority of cases demonstrate that ALTERN heuristic final solutions with initial exact solutions from GSRS outperform those with random initial solutions across various conditions, including the number of p, PSA levels, and the number of strata in GSRS. However, deviant cases exist, namely $p = 4$ on a random surface, and $p = 3$ and 4 on a strong PSA surface, where the means of objective function distances for ALTERN algorithm output with GSRS initial solutions are higher than those with random initial solutions; outcomes presumably attributable to rounding errors. Fig. 7 provides a portrayal of this finding.

Regarding PSA levels, the exceptional objective function distance means occur for different p medians across all surfaces. On a random surface, $p = 4$ and 13 exhibit relatively high peaks in their objective function distance means. On a weak PSA surface, $p = 7$ shows a high peak, while on a moderate PSA surface, $p = 7$ and 13 have high peaks. On a strong PSA surface, $p = 3$, 7, and 8 display relatively high peaks. These irregular peaks behavior indicates that change in the number of p medians to be calculated has a more pronounced influence on the irregular patterns of objective function distance means than change in the geographic distribution of weights' PSA levels.

When considering the number of GSRS strata, many cases utilizing a 6-by-6 square tessellation show the lowest objection function distance means. The red dots and lines in Fig. 7 denote results obtained using a 6-by-6 square tessellation for $p = 7$ on a random surface, $p = 8$ and 13 on a weak PSA surface, $p = 6$, 9, and 14 on a moderate PSA surface, and $p = 4$, 9, 11, and 14 on a

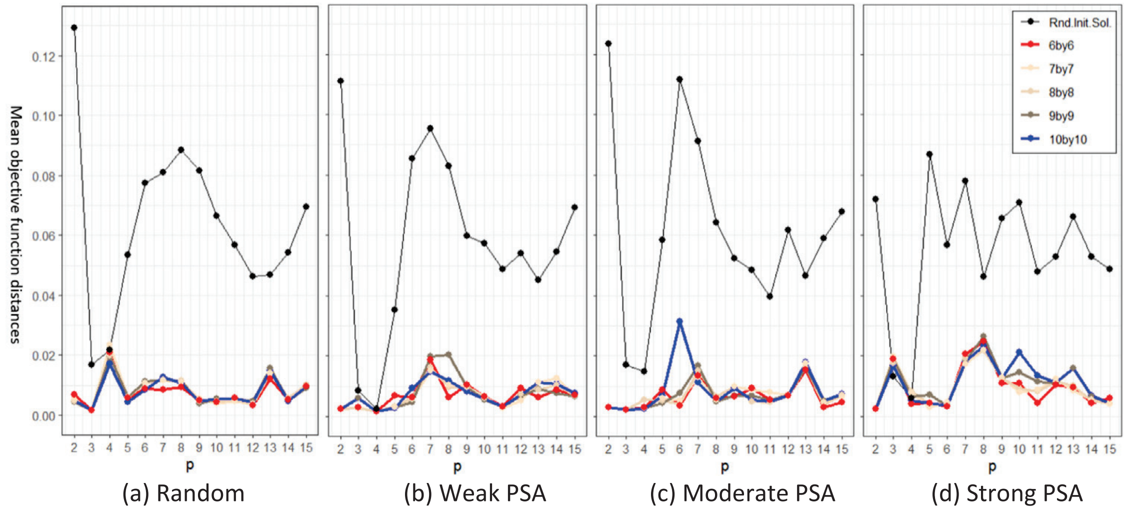

Fig. 7. Mean objective function distances for ALTERN heuristic solutions and their optimal counterparts for the four weights surfaces. The black dots and lines denote the results obtained from ALTERN solutions with random initial solutions, and the other dots and lines denote the impact of varying the number of tessellation strata.. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

strong PSA surface, all of which outperform the results obtained with other numbers of strata. This outcome suggests that a 6-by-6 square tessellation may be sufficient to obtain high-quality initial solutions when initializing the ALTERN heuristic algorithm. This conjecture warrants extensive future research.

Considering both the number of strata and PSA levels, the objective function distance means on a random surface remains similar regardless of the sizes used, whereas the means on PSA surfaces show variation depending on the number of strata. However, no regular patterns of change in these means are observed for all surfaces and p medians when increasing the GSRS strata number.

In summary, the experiment undertaken for this research yields three important findings. First, the use of GSRS proves to be highly beneficial in generating initial solutions of excellent quality, particularly when a surface exhibits PSA (the preponderance of GIScience cases). The spatial means obtained from GSRS closely align with the optimal locations of p-medians, indicating their effectiveness in capturing an underlying spatial pattern. Additionally, as the number of strata increases, the sample p-median solutions appear to converge on both their optimal solutions and their spatial means, further emphasizing the contribution of GSRS to securing more efficient optimal location–allocation solutions.

Second, results summarized in this paper highlight that the initial solutions generated by GSRS spatial means significantly improve the ability of the ALTERN heuristic to identify globally optimal solutions. Across various scenarios, including different numbers of p, weights surface PSA levels, and strata numbers, ALTERN algorithmic solutions consistently achieve optimality, as assessed by a very small threshold buffer applied to the objective function distance means to account for rounding error. Finally, results show that a 6-by-6 square tessellation for GSRS yields particularly high-quality initial solutions. This finding suggests that a 6-by-6 square tessellation may be sufficient for obtaining optimal solutions when initializing the ALTERN heuristic, at least for $p = 2, \ldots, 15$.

5. Conclusion

This paper highlights the significance of GSRS in finding optimal p-median solutions and improving the performance of the ALTERN heuristic algorithm for continuous space optimal solutions. The findings and contributions reflect upon improved computational efficiency by integrating spatial statistics with spatial optimization techniques. On the one hand, GSRS is shown to be effective in helping secure optimal p-median solutions for a large number of demand locations. By using spatial means generated through GSRS as initial solutions, the ALTERN heuristics algorithm consistently produces optimal solutions (except for rounding error). Increasing the number of GSRS strata results in higher quality initial solutions, as spatial means move closer to their corresponding optimal solutions. This outcome emphasizes the importance of larger sample sizes in generating high-quality initial solutions. Meanwhile, the presence of PSA contributes to the concentration and proximity of solutions vis-a-vis optimality, although the level of PSA may not consistently reveal improvement gains. This contention suggests that SA facilitates the identification of optimal *p*-median locations, in turn possibly influencing the ALTERN heuristic algorithm's performance. In other words, initial solutions established with GSRS significantly contribute to ensuring that ALTERN heuristic solutions are globally optimal, whereas random initial solutions, for example, often yield non-optimal results. These GSRS-derived initial solutions provide a well-informed starting point for a heuristic algorithm.

On the other hand, this paper provides valuable insights into considerations such as the number of GSRS strata, PSA levels latent in geographic distributions of weights, and the quality of initial solutions leading to optimal solutions. These findings contribute to a better understanding of *p*-median heuristics and offer practical guidance for efficiently obtaining high-quality solutions in real-world scenarios.

However, there are certain limitations to consider in the interpretation of the results. First, the *p*-median simulation tests were conducted using a specific set of 700 demand locations, which may introduce bias attributable to the underlying demand points location pattern, regardless of its selection randomness. Future tests should include a broader and more diverse set of random demand locations to ensure generalizability. Additionally, future research should explore discrete heuristic methods for solving larger *p*-median problems to provide insights into the performance of different algorithms in scenarios with a higher number (i.e., $p > 15$) of medians.

Acknowledgment

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

References

Bäsel, U., 2021. The moments of the distance between two random points in a regular polygon. arXiv preprint, http://arxiv.org/abs/2101.03815.

Beheshti, Z., Shamsuddin, S.M.H., 2013. A review of population-based meta-heuristic algorithm. Int. J. Adv. Soft Comput. Appl. 5 (1), 1–35.

Brink, A.B., Eva, H.D., 2009. Monitoring 25 years of land cover change dynamics in Africa: A sample based remote sensing approach. Appl. Geogr. 29 (4), 501–512. http://dx.doi.org/10.1016/J.APGEOG.2008.10.004.

Cooper, L., 1964. Heuristic methods for location-allocation problems. SIAM Review 6 (1), 37–53. http://dx.doi.org/10.1137/1006005.

Griffith, D.A., 1992. What is spatial autocorrelation? Reflections on the past 25 years of spatial statistics. Espac. Geogr. 21 (3), 265–280. http://dx.doi.org/10.3406/spgeo.1992.3091.

Griffith, D.A., 2005. Effective geographic sample size in the presence of spatial autocorrelation. Ann. Assoc. Am. Geogr. 95 (4), 740–760. http://dx.doi.org/10.1111/j.1467-8306.2005.00484.x.

Griffith, D.A., Chun, Y., Kim, H., 2022. Spatial autocorrelation informed approaches to solving location–allocation problems. Spat. Stat. 50, 100612. http://dx.doi.org/10.1016/J.SPASTA.2022.100612.

Hope, D., Gries, C., Zhu, W., Fagan, W.F., Redman, C.L., Grimm, N.B., Nelson, A.L., Martin, C., Kinzig, A., 2003. Socioeconomics drive urban plant diversity. Proc. Natl. Acad. Sci. 100 (15), 8788–8792. http://dx.doi.org/10.1073/pnas.1537557100.

Kariv, O., Hakimi, S.L., 1979. An algorithmic approach to network location problems. I: The p-centers. SIAM J. Appl. Math. 37 (3), 513–538.

Kuhn, H.W., Kuenne, K.E., 1962. An efficient algorithm for the numerical solution of the generalized Weber problem in spatial economics. J. Reg. Sci. 4 (2), 21–33. http://dx.doi.org/10.1111/j.1467-9787.1962.tb00902.x.

Mu, W., Tong, D., 2018. A spatial-knowledge-enhanced heuristic for solving the p-median problem. Trans. GIS 22 (2), 477–493. http://dx.doi.org/10.1111/tgis.12322.

Overton, W.S., Stehman, S.V., 1993. Properties of designs for sampling continuous spatial resources from a triangular grid. Comm. Statist. Theory Methods 22 (9), 251–264. http://dx.doi.org/10.1080/03610928308831175.

Perttunen, J., 1994. On the significance of the initial solution in travelling salesman heuristics. J. Oper. Res. Soc. 45 (10), 1131–1140. http://dx.doi.org/10.2307/2584476.

Rolland, E., Schilling, D.A., Current, J.R., 1997. An efficient tabu search procedure for the p-median problem. European J. Oper. Res. 96 (2), 329–342. http://dx.doi.org/10.1016/S0377-2217(96)00141-5.

Rosing, K.E., Hodgson, M.J., 2002. Heuristic concentration for the p-median: An example demonstrating how and why it works. Comput. Oper. Res. 29 (10), 1317–1330. http://dx.doi.org/10.1016/S0305-0548(01)00033-8.

Rosing, K.E., ReVelle, C.S., 1997. Heuristic concentration: Two stage solution construction. European J. Oper. Res. 97 (1), 75–86. http://dx.doi.org/10.1016/S0377-2217(96)00100-2.

Rosing, K.E., ReVelle, C.S., Rolland, E., Schilling, D.A., Current, J.R., 1998. Heuristic concentration and Tabu search: A head to head comparison. European J. Oper. Res. 104 (1), 93–99. http://dx.doi.org/10.1016/S0377-2217(97)00310-X.

Stehman, S.V., 2009. Sampling designs for accuracy assessment of land cover. Int. J. Remote Sens. 30 (20), 5243–5272. http://dx.doi.org/10.1080/01431160903131000.

Stevens, D.L., 1997. Variable density grid-based sampling designs for continuous spatial populations. Environmetrics 8, 167–195. http://dx.doi.org/10.1002/(SICI)1099-095X(199705)8:3<167::AID-ENV239>3.0.CO;2-D.

Teitz, M.B., Bart, P., 1968. Heuristic methods for estimating the generalized vertex median of a weighted graph. Oper. Res. 16 (5), 955–961. http://dx.doi.org/10.1287/opre.16.5.955.

Weber, A., 1922. Ueber den standort der industrien. J.C.B. Mohr (Paul Siebeck), Tübingen.