An Index for Quantifying Geometric Point Disorder in Geospatial Applications

S. Jones[1a*] and H.G. Momm[1b]

[1]Department of Geosciences, Middle Tennessee State University, Murfreesboro, Tennessee

*corresponding author: rj3h@mtmail.mtsu.edu

Abstract:

Many techniques have been developed to quantify different conceptualizations of self-interaction and patterns within spatial data. We propose a new metric and related algorithm that describes the geometric spatial disorder of geographic point sets, the "Index of Disorder" (IoD). The IoD algorithm was applied to synthetic and natural datasets and was shown to be able to differentiate between areas of high spatial disorder (randomly placed points) and low spatial disorder (e.g., curvilinear grids, wallpaper groups, and other repeating patterns). Because the IoD is a quantitative metric, it can be used on its own as an aid for identifying areas of unusually high or low spatial disorder or as enrichment for machine learning classification algorithms.

Keywords:

Point Patterns; Classification; Homogeneity, Algorithms; Forestry, Planning; Data Processing

One Sentence Summary:

The spatial disorder of any arbitrary point in a set of points can be quantified by comparing the relative positions of that point's neighbors to the relative positions of its neighbors' neighbors.

[a] Responsible for design, implementation and testing of algorithm, collection and processing of data, and drafting of manuscript.

[b] Responsible for supervision of algorithm design, reproduction of results, manuscript drafting and revision.

1 Introduction

The acquisition of large amounts of spatial data has generated the need for automated methods to convert this data into usable information. A significant number of established methods for pattern detection are available, such as spatial autocorrelation, outlier detection, clustering, hotspot detection, and regression (Shekhar, et al., 2011). However, there is a relative scarcity of methods for the analysis and identification of patterns purely comprised of the relationships of the coordinate positioning of points of geospatial data. Such “geometric patterns” are evident in the locational structure of many types of geospatial data, such as the neatly ordered placement of trees in an orchard or house centroids in a planned development. Geometric order contrasts with the relative disorder of naturally occurring features, which display little to no geometric structure as result of the relatively random positioning of each element. This structure, or lack of structure, is often immediately apparent to human investigators, yet is surprisingly difficult to define mathematically (Antuono, et al., 2014), and so there is a relative dearth of methods that can be used to analyze geometric patterns in geospatial contexts.

For example, land cover classification is one of the most widely used and accepted applications of geospatial technologies. Land cover classification studies have been conducted using a wide array of data sources, mostly raster grids derived from imagery (Lou, et al., 2016; Phiri & Morgenroth, 2017). The majority of landcover studies treated individual pixels as isolated data points, ignoring the information contained in the spatial relationships between pixels. Different set of methods have been developed to discriminate between features based on image texture and the frequency of tonal changes based on pre-determined texture operators (Momm, et al., 2009). Haralick textures were among the first proposed metrics that quantified the spatial relationships between adjacent pixels (Haralick, et al., 1973), and have since been used to increase the accuracy of land cover classification studies (Momm, et al., 2009). Subsequent developments, such as so-called contextual classification methods, further sought to quantify the information contained within spatial pixel-neighbor relationship (Swain & Stephen B. Vardman, 1981). Quantification of lacunarity (self-similarity) has also been used as a way to quantify scale-dependent heterogeneity in raster imagery as a way to aid differentiation of ordered and disordered land use (Dong, 2000). Additional work has shown that textural periodicity is suggestive of certain landcover types (Trias-Sanz, 2006). Over time, a number of additional quantitative metrics of landcover textures and patterns have been developed as well (Remmel & Csillag, 2003). Each of these methods extracts some meaning from the spatial relationships of data, but they are not suited for the detection of geometric patterns: each method is merely a statistical operator on a windowed subset of data, which inherently preserves some information about spatial relationships, but the information about the positioning of data points within the subset is ignored.

Furthermore, because imagery and elevation models are most easily represented as raster datasets, the bulk of methods developed to quantify spatial textures and patterns in geospatial contexts are applicable to raster grid data only. Yet not all geographic information can or should be represented as raster data, such as the coordinate locations of discrete features of interest. Extensive work has been done on the development of algorithms for detecting regularly repeating structures in point clouds, but these methods have been primarily applied in non-geographic contexts (Pauly, et al., 2008). Additionally, these algorithms do not make an attempt to quantify the level of regularity of individual points in the point cloud even though the regularity of the point cloud is detected quantitatively. Antuono and others suggested an algorithm to quantify disorder within simulated fluid-particle systems, but this algorithm makes the assumption that disorder is defined as the deviation from a grid-like structure, an assumption that is likely valid for fluid simulations but fails in the context of geographic systems where order and disorder might be more loosely defined (Antuono, et al., 2014). Because of this, geometric patterns of geospatial data is an aspect that is often ignored, and there is a deficiency of methods that can quantify the disorder of geospatial point data without relying on *a priori* assumptions about what patterns constitute "order".

To quantify the geometric order/disorder of geospatial point datasets, a new metric and related algorithm is described. The proposed metric, referred to as "Index of Disorder" (IoD), quantitatively describes the relative geometric spatial order/disorder of geographic point sets. Calculation and assignment of an IoD score to each point allows the identification of areas of relatively high or low point pattern disorder. The main objectives of this study are three-fold: (1) describe the IoD assumptions, parameters, and calculation, (2) discuss the interpretation of IoD scores, and (3) assess the IoD's performance through its application to synthetic and natural geospatial datasets.

2 Methods

2.1 Algorithm Description

The IoD is designed to quantify similarities in point spatial pattern by comparing multiple "neighborhoods" of points; in which a neighborhood is defined by a set of all points within a user-specified distance of an individual point. In each pair of neighborhood analysis, an IoD sub-score is calculated for the point being investigated. The IoD for a point is obtained by calculating the IoD sub-score for that point and each of its neighbors and taking the mean of these scores.

If two neighborhoods are similar, then the relative positions of corresponding points will be close together; conversely, if two neighborhoods are dissimilar, then corresponding points will be far apart. A point set that is "ordered" will have a neighborhood that is, on average, similar to the neighborhoods of all of its neighbors, and thus the typical point-pair deviation will be small. Because this is a relative rather than absolute metric, an arbitrary point set could potentially have any IoD value since the IoD quantifies the disorder of a point relative to the other points in the set and in accordance with the algorithm's parameterization. Consequently, ordered points do not necessarily have a low IoD but rather have an IoD that is lower than the IoD of disordered points in the same dataset at the scale of interest.

The calculation of the IoD sub-score between two points p_1 and p_2, which are part of the set of points P, is as follows (Figure 1). A parent point p_1 is defined (red point in Figure 2A) and its neighbors N_1 selected based on user-provided threshold distance (Figure 2B). Similarly, a child point p_2 is defined and its neighbors N_2 selected (blue points in Figure 2C). Absolute X and Y coordinates values for all points in neighborhoods N_1 and N_2 are transformed to a relative coordinate system in which both parent and child points are located at the origin (Figure 2D). Each point in N_1 is assigned to a unique point in N_2 such that the mean cost function for point assignment is minimized (Figure 2E). The mean value of the cost function is the IoD sub-score (Figure 1). The process continues by recording the sub-score and repeating the operation by selecting a new child point (Figure 2F). Optionally, prior to point pair assignment, it is possible to apply a second coordinate transformation to N_2 to further minimize the cost of point assignment (Figure 1). The IoD sub-score is not the IoD itself, but rather an intermediate parameter used in the calculation of the IoD for a given point. A low IoD sub-score implies geometric similarity between the neighborhoods of two particular points, whereas a low IoD implies the relative presence of a geometric ordered pattern around a given point.

2.1.1 Point Pair Assignment

Calculation of the IoD sub-score for two sets of points involves point pair assignment, where every point in one set (neighborhood N_1) is assigned to a point in the other set (neighborhood N_2) with one-to-one correspondence. Additionally, a goodness-of-fit metric is calculated for every point-pair assignment that describes how good the assignment is. The objective is to find an optimal set of point-pair assignments such that that average cost of assignment is minimized. The cost function for point-pair assignment is arbitrary. A readily apparent cost function for point-pair assignment is the Euclidean distance between the two members of a point pair, two points that are close together are a more obvious match than those that are far apart. Other cost functions such as the square root of the distance or sigmoidal functions (Figure 3) that relate Euclidean distance to a score are viable as well.

The assignment of correspondence, or assignment problem, is formulated in such way that there exists a complete bipartite graph composed of two sets of parent and child vertices (Figure *2*E). Each candidate solution connects a parent to a child and has an associated cost $c(i,j)$. If the sets are not of equal size, then the graph is incomplete, but the graph can be made complete by adding some number of arbitrary vertices to the smaller set and connecting them to the larger set with an arbitrarily large assignment cost. When a solution is found that minimizes the assignment costs, the arbitrary vertices and edges are removed. Thus, the larger set will have some number of unpaired points.

Solving the assignment problem through brute force implies a time complexity of $O(n!)$, where n is the number of points being matched, though many algorithms have been proposed to improve computing efficiency. The point pair assignment algorithm used in this study is the Hungarian method as implemented in the scientific computing library *scipy* (Oliphant, 2006). The Hungarian method solves the assignment problem in polynomial time $O(n^3)$, (Kuhn, 1955). The algorithm works by reframing the problem as an *n x n* cost matrix *M* where the cost between P_i and P_j is $c(i,j)$. The solution is the set *S* of *n* entries that minimizes the sum of the set while respecting one-to-one correspondence. The Hungarian algorithm takes advantage of the fact that subtracting a constant from a row or column does not change the set of optimal entries, and so by repeatedly subtracting values from rows and columns the assignment problem can be reduced to a form where some of the matrix entries are 0 (Kuhn, 1955).

2.1.2 Point Set Realignment

In some cases it may be desirable to realign neighborhoods for better correspondence (point sets registration) if it is suspected that spatial patterns in the data may become offset, rotated, scaled, or flipped (Figure 4). Existing point registration methods require a cost function to evaluate candidate solutions. In this study, the same cost function described for point pair assignment was used in the point set realignment.

Although multiple methods are available, and any of them could have been used, the Iterative Closest Point (ICP) method was selected (Besl & McKay, 1992). The algorithm iteratively registers points by assigning all points in N_1 to points in N_2 and then finds the least squares of the rigid transformation that minimizes the point-pair assignment cost. The process is repeated until the point-pair deviation falls below a threshold or the algorithm exceeds the pre-determined number of iterations.

2.1.3 Scoring Function

Once two neighborhoods have been assigned to one another, a function must be defined that relates the displacement between two assigned points with a score describing the quality of the match, which we call the “assignment score” for that point. It is convenient to define a scoring function that has a lower bound of 0 (to describe no disorder) and an upper bound of 1 (to describe maximal disorder), though such a range is arbitrary. Using the same function used in the point pair assignment ensures that the calculated point correspondence results in the minimum score possible.

2.1.4 Unpaired points

Often neighborhoods being compared do not have the same number of points. This will result in unpaired points. Because the distance between a point and its assigned partner is used to calculate the point’s assignment score, a decision must be made regarding how to calculate assignment scores for unpaired points.

The simplest option is to ignore unpaired points when calculating the mean assignment cost. An obvious drawback is when assigning a neighborhood with few points to another with many points what could lead to a small IoD despite the fact that no true point correspondence necessarily exists.

A second option is to penalize unmatched points by some arbitrary amount, which would prevent IoD depression when sparse neighborhoods are compared to dense neighborhoods. However, there are some circumstances where two neighborhoods may show the same pattern, but the pattern in one neighborhood is truncated because e.g., the neighborhood's parent point lies at the edge of a pattern (Figure 4). In this case, penalizing unpaired points would raise the IoD even though true pattern correspondence exists.

A final and generally superior option is to punish unpaired points if and only if they are within the convex hull of the subset of points that have an assignment. This allows for patterns in compared neighborhoods to be of unequal spatial extent while preventing noisy points from reducing the IoD. In the case of 2-dimensional points, this can be visualized as the area enclosed by a rubber band stretched over the assigned set (Figure 4). This was the option adopted.

2.2 Theoretical Evaluation

Synthetically generated patterns and their IoD scores were evaluated (Figure 5). The inner blue circle describes the neighborhood size and the red circle denotes the radius outside of which the pattern becomes increasingly perturbed with noise of increasing strength outside an arbitrary distance from the origin. Thus, the order of the pattern is entirely preserved within the arbitrary radius, but the system becomes increasingly disordered beyond that radius. Multiple patterns were investigated: square grid (Figure *5*A), rectangular grid modified with a sinusoidal function (Figure *5*B), pattern formed by overlaying concentric circles with equal linear point densities (Figure *5*C), pattern formed by overlapping a square grid with a copy of the same grid rotated 45 degrees (Figure *5*D), pattern formed by overlapping two offset rectangular grids (Figure *5*E), pattern formed by overlapping three offset rectangular grids (Figure *5*F).

The IoD for all synthetic datasets were evaluated based on a sigmoidal function for point assignment and assignment scoring. Each synthetic point set was evaluated twice: with and without iterative closest point realignment, but with no other changes to algorithm parameters (Table 1). When ICP realignment was used, the same sigmoidal function used for point assignment and assignment scoring was used to calculate correspondence within the ICP algorithm. The values adopted for the theoretical evaluation were a neighborhood radius of 15 units, deviation value at which sigmoidal scoring function assigns a score of 0.5 (K_m) of 3 units, degree of cooperativity of the sigmoidal scoring function of 0.5, assignment score given to unpaired points of 1, punishment of points outside of the convex hull set to false, and Euclidian distance selected as cost function.

2.3 Applied Evaluations

The IoD was applied to three geospatial datasets, two consisting of tree crown locations and one consisting of building centroids. Each dataset is known to contain at least one area of intrinsic order. The ability of the IoD to differentiate disordered from ordered areas was quantified using Cohen's kappa coefficient (κ) of agreement.

2.3.1 Tree Crowns

Three-dimensional point clouds derived from LiDAR (Light Detection and Ranging) were acquired from United States Geological Survey's (USGS) National Map repository for two study areas. The LiDAR-processing software suite LAStools was used to generate multiple rasterized elevation models (Isenburg, 2019). Tree crowns were automatically extracted using the Laplace of Gaussian blob detection algorithm implemented in the Python package scikit-learn (Pedregosa, et al., 2011), although other methods could had been employed (Zhen, et al., 2016).

2.3.1.1 Site 1 – Apple orchard in TN, USA

For ease of harvest and maintenance, orchards are typically planted in regular patterns that contrast with the disordered distribution of naturally occurring tree stands. The most common planting patterns are grid-based. The predictability of these patterns has allowed for some automated differentiation of orchards from natural tree stands on the basis the textural characteristics of orthoimagery (Aksoy, et al., 2012). However, alternative patterns such has single and double hedged, hexagonal, quincunx and topographically contoured planting systems are also used, limiting identification by standard methods (Khan, et al., 2017).

The IoD was applied to tree crown centroids extracted from a 7.5 km^2 area in a small commercial apple orchard in Tennessee, USA. The study area is rural and consists of two distinct planted areas of rectangularly gridded apple trees trending northwest to northeast. Mature natural forest is also present to the west and north of the orchard as well as sparse and isolated naturally occurring trees.

Trees crown centroids were manually classified using both the LiDAR-derived elevation models and contemporaneous aerial imagery. The IoD was calculated for all tree centroids in the study area, and trees were classified as "disordered" if their IoD was above an arbitrary IoD threshold and "ordered" if below the threshold. "Ordered" trees were assumed to correspond to orchard trees and a sensitivity analysis was conducted to characterize the effect that varying the scoring function's K_m and the neighborhood radius has on the classification quality. The cooperativity of the scoring function was held constant at 5 during the sensitivity analysis.

2.3.1.2 Site 2 – Reforested Area in Mooresville, NC

Site 2 is a 0.6 km^2 area centered on a partially reforested zone located in Mooresville, NC, USA. The study area is rural and consists of linear rows of young planted trees abutted by older natural mature forest. The trees comprising the replanted areas form roughly rectangular grid patterns of varying orientation that contrast with the disorder of the natural forest.

Like Site 1, trees were manually classified as being part of the replanted zone or not based on LiDAR-derived elevation models and contemporaneous aerial imagery. The IoD was calculated for all trees in the study area, and trees were classified as "disordered" if their IoD was above an arbitrary IoD threshold and "ordered" if below the threshold. "Ordered" trees were assumed to correspond to replanted zones and a sensitivity analysis was conducted to characterize the effect that varying the scoring function's K_m and the neighborhood radius has on the classification quality. The cooperativity of the scoring function was held constant at 3 during the sensitivity analysis.

2.3.1.3 Site 3 – Neighborhood in Nashville, TN

Urban planners have long noted the morphological peculiarities of unplanned or "organic" city growth as compared to those of planned growth. Planned portions of cities are marked by regularity of building placement and street network orientation, while organic growth does not (Nilsson & Gil, 2019; Boeing, 2019). On a smaller scale, residential communities often exhibit similar spacing patterns between homes, particularly when the area is one that is relatively planned. Auxiliary structures, such as sheds and garages, are more irregularly spaced and have "neighborhoods", in the sense of the IoD algorithm, that are offset from surrounding buildings.

Site 3 is a mixed-use neighborhood in Nashville, Tennessee with approximately 2 km^2 in size. It is densely developed compared to Sites 1 and 2. Main buildings (residences and businesses) are developed in relatively regular rows within neighborhood blocks, while auxiliary structures such as sheds and garages are sporadically placed with no consistent spacing or lot positioning (Figure 6). Building centroids were generated based on building footprints according to the City of Nashville, TN, USA.

In addition to spatial information, these datasets include additional information about buildings such as whether they are main or auxiliary structures. The IoD was then calculated for the extracted centroids, and buildings were classified as "disordered" if their IoD was above an arbitrary IoD threshold and "ordered" if below the threshold. "Ordered" buildings were assumed to correspond to "main" buildings, while disordered buildings were assumed to correspond to auxiliary structures. A sensitivity test was conducted to characterize the effect that varying the scoring function's K_m and the neighborhood radius has on the classification quality. The cooperativity of the scoring function was held constant at 5.

3 Results and Discussion

3.1 Theoretical Evaluation

For point patterns generated based on regular grids, the IoD differentiates greatly between the unperturbed points (within red circle) and perturbed points (outside red circle in Figure 5A-D). In more complex patterns, the IoD yields mild to moderate differentiation between the unperturbed and noisy patterns than the patterns based on regular grids. Nonetheless, in all cases the lowest IoD values are observed in the center of the figures, and a rapid increase in the IoD is observed as the pattern deviation approaches and exceeds the K_m of the scoring function, indicating that the IoD is capable of measuring relative levels of disorder and/or order within a dataset. Importantly, the IoD is agnostic to the general form of patterns, and thus is capable of detecting patterns with no knowledge of the exact form of the patterns.

Because pattern detection is scale-dependent, alternative input parameterization will result in varying levels of discrimination between ordered and disordered points. Optimal parameterization is generally simpler to achieve when patterns are also simple.

Realignment of neighborhoods during calculation of the IoD generally reduces both the unperturbed and noisy perturbed points by a similar amount. However, in certain cases, realignment can cause a slight increase in differentiation (*Figure* 8). In other words, realignment depresses the IoD of ordered points more than it depresses the IoD of disordered points. Discrete repeating patterns in these examples, sometimes referred to as “wallpaper groups” (Liu, et al., 2004), benefit from realignment because the effect of pattern offset (Figure *8*) becomes significant. Not realigning points will elevate the IoD even though pattern correspondence exists. Allowing neighborhood realignment can significantly increase computation time because point registration may be repeated multiple times per neighborhood rather than just once, so realignment should only be used when pattern offset is anticipated to occur and needs to be corrected for. In some cases, pattern offset may actually be of importance for feature identification, and so realignment is undesirable.

3.2 Natural Evaluations

3.2.1 Site 1

Sensitivity analysis was performed by varying IoD input parameters of neighborhood radius *r* and sigmoidal assigned threshold value K_m and comparing results with reference datasets (Table 3). The maximum kappa coefficient of agreement value of 0.81, interpreted as “substantial agreement” (Cohen, 1960), is achieved when the scoring function has a K_m of 5 and the neighborhood radius is 80 meters. The corresponding overall accuracy for this classification is 96%.

The mild planting pattern heterogeneity of the orchard increased the calculated IoD somewhat, but overall the orchard trees are largely differentiable from the surrounding forest due to the gridded nature of the orchard (Figure 9). The apparent heterogenic pattern of the orchard is likely as much due to inaccuracies in crown extraction from the DHM than it is due actual heterogeneity; crown extraction from LiDAR is itself highly parameter dependent. The high kappa value of this classification suggests that the IoD alone is sufficient to differentiate the orchard from surrounding trees without any other supporting data.

The classification quality of the IoD is highly sensitive to its parameterization (Table 3). Because patterns are a fundamentally scale-dependent phenomenon, it is not surprising that algorithms that quantify them must be parameterized appropriately. The neighborhood radius used to parameterize the IoD describes the scale of the anticipated patterns, while the K_m describes the level of expected deviation within the patterns. Though *r* is a parameter that will be present in any implementation of the IoD, K_m is technically a parameter of the sigmoidal function used for scoring and point assignment. If the classification quality is maximized and a strongly homogenous pattern is being differentiated from a highly disordered nonpattern, then *r* and K_m respectively characterize the actual pattern scale and deviation. Thus, the characteristic scale of the orchard in Site 1 is between 70 to 80 meters, and the threshold of the pattern deviation before points become disordered is approximately 5 meters.

Because of this scale dependence, the IoD is not appropriate when there is no prior knowledge of scale of pattern deviance or when trying to quantify patterns with multiple scales. In those cases, it is recommended using the IoD the characterize pattern scales on a subset of data before applying it more broadly.

3.2.2 Site 2

Similarly, sensitivity analysis of was performed for IoD calculations in Site 2 (Table 4). The maximum kappa value of 0.74, interpreted as “substantial agreement” is achieved when the scoring function has a K_m of 2 and *r* is 25 meters. The corresponding overall accuracy for this classification is 87%.

Like Site 1, the planted trees in this study area display a gridded structure that explains the lower IoD in the planted zones relative to the surrounding mature forest (Figure 10). The kappa coefficient of agreement suggests planted trees are differentiable from the surrounding trees based on the IoD alone. The sensitivity evaluation suggested a characteristic scale and pattern deviance for the planted trees, which are 25 meters and 2 meters respectively (Table 4).

3.2.3 Site 3

Performing sensitivity analysis, the maximum kappa value of 0.44, interpreted as “moderate agreement” (Cohen, 1960) and overall accuracy of 76% are achieved when the scoring function has a K_m of 6.5 and *r* is 19 meters (Table 5).

The classification agreement is lower at this site than previous sites. This is unsurprising due to the increased pattern complexity displayed by the building centroids (Figure 11). While the planted trees at the previous sites displayed relatively simple grids patterns that contrasted with the highly disordered positioning of the naturally occurring trees, the buildings in this site are arranged in gridded blocks that vary in scale, orientation, and deviance. Though the main buildings generally adhere to straight lines within these blocks, the exact position of their centroids can vary along this line and occasionally there is no discernible pattern to their placement at all. Conversely, auxiliary buildings overall do not display the level of pattern adherence that the main buildings do—not every main building has an auxiliary building, and when an auxiliary building is present its placement on the property is relatively varied—but their placement is not perfectly random. Occasionally a block will have enough auxiliary buildings that a pattern similar to that of the main buildings emerges.

These phenomena together elevate the IoD of the main buildings and reduce the IoD of the auxiliary buildings, reducing the ability of the IoD alone to differentiate these building types. However, the IoD has some level of classification power even in complex systems and so can be used to improve the accuracy of classification schemes based on machine learning; the IoD can be calculated and added to a dataset, increasing its dimensionality.

Higher classification results were achieved with a r of 19 meters and a K_m of 6.5 meters. The respective interpretation of these values as the characteristic pattern scale and pattern deviation is not necessarily as clear is it is for Sites 1 and 2, which consist of strongly patterned and nonpatterned points. The placement of building centroids in Site 3, in contrast, are moderately patterned (main buildings) or weakly patterned (auxiliary buildings). Because of this, the ideal r and K_m do not necessarily describe either pattern but rather represent, respectively, a discriminatory scale and a discriminatory deviance.

3.3 Impact of Alternative Implementations of the IoD

3.3.1 Scoring and Point Assignment Functions

Swapping any monotonic increasing function with another for the purposes of scoring will not change the relative ranking of the disorder of the points, and thus the choice of scoring function is ultimately an aesthetic choice. However, it is convenient to use the same function used to calculate assignment costs to calculate the IoD to simplify interpretation of the results. The assignment cost function *does* have an impact on what points are assigned to one another, and thus may have an impact on the relative ranking of IoD scores for points in a set (Figure *4*). Using the Euclidean distance between points as the assignment function results in many suboptimal pairings; many points are assigned to a point for which there is no obvious correspondence but results in an overall minimization of the assignment cost. Using a sigmoidal function, however, allows for more intuitive assignments for most points by reducing penalties for assignments with large Euclidean displacements, in turn allowing assignment of points very close in space to one another.

Other functions, such as the square root of the Euclidean distance, similarly reduce the penalty for large assignment displacements and generally improve assignment. An advantage of the sigmoidal function over other options is that the midpoint (K_m) imparts additional scale-awareness to the IoD; the K_m describes the expected deviation, or “noise” within a pattern, and so displacements below K_m are considered within the tolerance of the expected noise level and punished less. Without K_m or a similar metric, the IoD is aware of characteristic pattern scales (via the neighborhood radius) but will be unable to account for intra-pattern noise.

3.3.2 Realignment Function

There are many alternative methods for realignment, typically called point set registration or point matching algorithms in the field of computer vision. Discussion of their differences and similarities is beyond the scope of this paper, but it is worth noting that all registration methods must either calculate or be provided with the correspondence between the point sets being aligned.

Additionally, realignment algorithms are not guaranteed to “correctly” align the neighborhoods, and realigning neighborhoods may also lead to spurious reductions in the IoD even when no pattern similarity exists. If this decrease exceeds reduction in IoD when true pattern correspondence exists, then realignment will actually reduce the ability of the IoD to differentiate ordered and disordered point sets (*Figure 7*). Thus, the use of realignment may or may not be appropriate depending on the intent of the study. Realignment is also computationally expensive as it requires repeated recalculation of neighborhood point assignments.

4 Conclusions

Existing methods to quantify disorder have relied on either raster data or non-geospatial algorithms that quantify the disorder of point data based on the assumption that order is grid-like, an assumption that is often violated in geospatial contexts. Thus, the Index of Disorder algorithm provides a new way to quantify spatial disorder of individual points in a set, which is achieved by quantifying the similarity of a point’s “neighborhood” to the neighborhoods of its neighbors.

Datasets evaluated in this study indicate that the IoD alone is sufficient to differentiate planted stands of trees, which tend to be planted in curvilinear grids, from mature forest, which displays no pattern in the positioning of trees. On this principle the IoD can be used to estimate reforestation extent or identify orchards. The IoD can also be used to enrich datasets for classification in systems where spatial patterns alone may not be sufficient to make a classification, such as when classifying building types in complex urban systems.

Because spatial patterns are inherently scale-dependent phenomena, the IoD requires parameterization in order to satisfactorily quantify disorder. Thus, its utility may be limited in systems where the scale of the patterns being analyzed is poorly understood or if the pattern scale is variable across the study area. However, this limitation allows the IoD to be used backwards: if there is a priori knowledge of the classification labels of points in a system, then then an optimization algorithm can be applied to the IoD in order to estimate the scale of the pattern and magnitude of the pattern deviation.

Because the measure is quantitative (though relative) it can also be used as an additional dimension of analysis for problems that benefit from data enrichment, such as machine learning classification. Further work is planned to explore in more detail the effects of alternative implementations of the IoD, as well as its utility in classification problems beyond the scope of what is presented here. In particular, the IoD may be of use in quantifying patterns present in 3-dimensional point sets.

5 Data and Materials Availability

All code used in the analysis for this paper is publicly available at https://github.com/rsjones94/point-disorder. The digital height models used to generate the tree crown datasets and Nashville building centroids are available on request.

6 Tables and Figures

Table 1. Explanation of input parameters used in the implementation of the IoD demonstrated in this study.

Parameter	Meaning
r	The neighborhood radius
K_m	The assignment deviation at which the sigmoidal scoring function assigns a score of 0.5
n	The degree of cooperativity of the sigmoidal scoring function. Higher values of n make the function increasingly sigmoidal. Values of 1 or lower result in a hyperbolic function.
Punishment Level	The assignment score given to unpaired points
Punishment Type	Whether to assign the punishment score to unpaired points. If True, punish all unpaired points. If False, punish only those that fall within the convex hull of assigned points.
Assignment Type	Whether to use a euclidean or sigmoidal assignment cost function. If True, use a the euclidean distance as the assignment cost. If False, use the sigmoidal scoring function.
Reorientation Boolean	Whether to apply ICP reorientation during the calculation of the IoD. If a numerical value, the ICP is applied until 20 iterations have elapsed or the deviation is less than the value supplied, whichever comes first. If False, no reorientation is applied.

Table 2. Summary of classification results for each study area. Classification quality was assessed using Cohen's kappa coefficient (κ) and accuracy. Differentiation of natural and planted trees using only the IoD displays moderate to high classification agreement, indicating that the IoD alone is sufficient to differentiate planted and naturally occurring tree stands. Differentiation of building types shows weaker agreement. For this purpose, the IoD alone may not be sufficient for acceptable classification quality but could be used to enrich more comprehensive classification methods.

Study Area	Disordered Group	Ordered Group	Peak κ	Corresponding Accuracy
Orchard in Crab Orchard, TN	Natural forest	Orchard	0.81	0.96
Reforested area in Mooresville, NC	Natural forest	Reforested zone	0.74	0.87
Neighborhood in Nashville, TN	Auxillary structures (sheds, detached garages)	Major buildings (homes and commerical buildings)	0.44	0.76

Table 3. Sensitivity evaluation of the IoD for Site 1 (orchard in TN) using a threshold of 0.8 and varying neighborhood radius (r) and midpoint of sigmoidal function (Km).

Cohen's κ		K_m (m)										
		1	3	5	7	9	11	13	15	17	19	21
Radius (m)	30	0.05	0.38	0.31	0.19	0.12	0.09	0.07	0.06	0.05	0.05	0.04
	40	0.02	0.42	0.47	0.19	0.10	0.06	0.05	0.04	0.03	0.03	0.03
	50	0.00	0.29	0.65	0.21	0.07	0.04	0.03	0.03	0.02	0.02	0.02
	60	0.00	0.24	0.78	0.26	0.08	0.04	0.03	0.02	0.02	0.02	0.01
	70	0.00	0.06	0.81	0.31	0.09	0.05	0.03	0.02	0.02	0.02	0.01
	80	0.00	0.02	0.81	0.35	0.09	0.05	0.03	0.03	0.02	0.01	0.01
	90	0.00	0.00	0.78	0.39	0.11	0.06	0.03	0.02	0.02	0.01	0.01
	100	0.00	0.00	0.73	0.43	0.11	0.06	0.04	0.02	0.02	0.01	0.01
	110	0.00	0.00	0.65	0.44	0.12	0.07	0.04	0.02	0.02	0.01	0.01
	120	0.00	0.00	0.57	0.45	0.12	0.07	0.04	0.03	0.02	0.01	0.01
	130	0.00	0.00	0.51	0.47	0.13	0.07	0.04	0.03	0.02	0.01	0.01
	140	0.00	0.00	0.43	0.44	0.13	0.07	0.05	0.03	0.02	0.01	0.01
	150	0.00	0.00	0.33	0.45	0.14	0.07	0.05	0.03	0.02	0.01	0.01

Table 4. Sensitivity test for Site 2 (replanted forest in NC) with an IoD threshold of 0.75.

Cohen's κ		K_m (m)											
		0.25	0.5	0.75	1	1.25	1.5	1.75	2	2.25	2.5	2.75	3
Radius (m)	5	0.00	0.02	0.05	0.05	0.06	0.06	0.06	0.08	0.07	0.07	0.07	0.08
	10	0.00	0.01	0.10	0.25	0.39	0.48	0.51	0.52	0.50	0.45	0.41	0.36
	15	0.00	0.00	0.01	0.15	0.37	0.55	0.64	0.62	0.50	0.36	0.25	0.18
	20	0.00	0.00	0.00	0.06	0.29	0.51	0.67	0.72	0.55	0.36	0.22	0.15
	25	0.00	0.00	0.00	0.03	0.22	0.44	0.62	0.74	0.59	0.36	0.21	0.14
	30	0.00	0.00	0.00	0.01	0.16	0.38	0.58	0.73	0.59	0.36	0.21	0.15
	35	0.00	0.00	0.00	0.00	0.11	0.30	0.52	0.69	0.60	0.36	0.21	0.15
	40	0.00	0.00	0.00	0.00	0.06	0.24	0.45	0.65	0.59	0.35	0.22	0.15

Table 5. Sensitivity test for Site 3 (building centroids in a neighborhood in Nashville, TN) with an IoD threshold of 0.70.

Cohen's κ		K_m (m)											
		4.5	5	5.5	6	6.5	7	7.5	8	8.5	9	9.5	10
Radius (m)	15	0.10	0.09	0.08	0.08	0.09	0.09	0.07	0.05	0.06	0.04	0.03	0.02
	17	0.31	0.33	0.34	0.37	0.37	0.37	0.37	0.36	0.35	0.34	0.33	0.31
	19	0.34	0.37	0.40	0.43	0.44	0.42	0.40	0.39	0.38	0.38	0.37	0.37
	21	0.33	0.36	0.39	0.41	0.43	0.41	0.39	0.39	0.38	0.37	0.37	0.36
	23	0.32	0.33	0.35	0.37	0.37	0.37	0.35	0.33	0.32	0.30	0.29	0.28
	25	0.28	0.31	0.33	0.34	0.34	0.33	0.32	0.30	0.27	0.25	0.23	0.20
	27	0.25	0.28	0.30	0.31	0.31	0.29	0.28	0.24	0.20	0.17	0.15	0.13
	29	0.24	0.26	0.28	0.28	0.27	0.25	0.24	0.21	0.19	0.17	0.14	0.12
	31	0.27	0.28	0.31	0.31	0.31	0.30	0.25	0.23	0.20	0.19	0.15	0.12

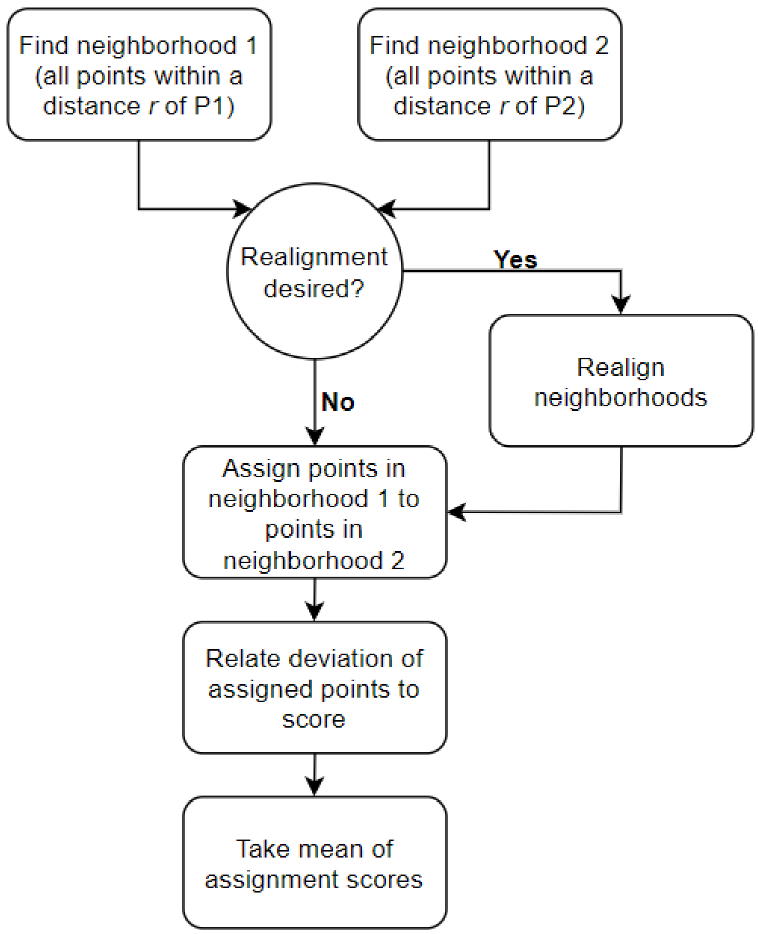

Figure 1. Generalized process for calculating the IoD sub-score of two points P_1, P_2. Due to the heuristic nature of the algorithm, point set realignment, point pair assignment, and deviation scoring techniques can vary between implementations of the IoD.

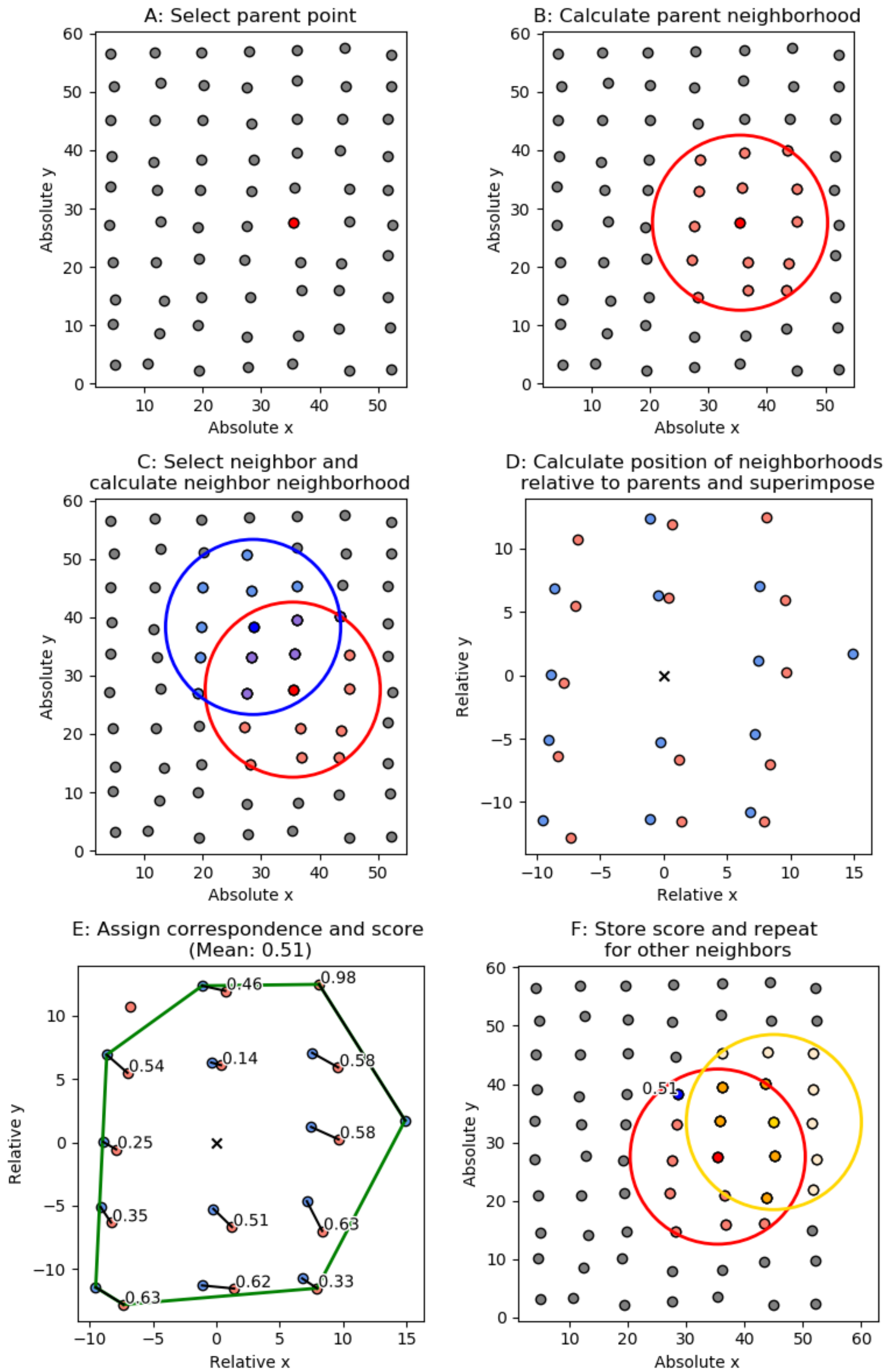

Figure 2. Illustration of the calculation of the IoD sub-score between two points. Given a parent point (A), its neighbors are selected (B). Similarly, given a child point its neighbors are selected (C). Both parent and child neighborhoods are converted from absolute to relative coordinate system (D). Each point in the parent set is assigned to a point in the child set (E). The process then repeated by selecting another child point set (F). The IoD sub-score between two neighborhoods is the mean of the assignment scores between a parent point and every neighbor.

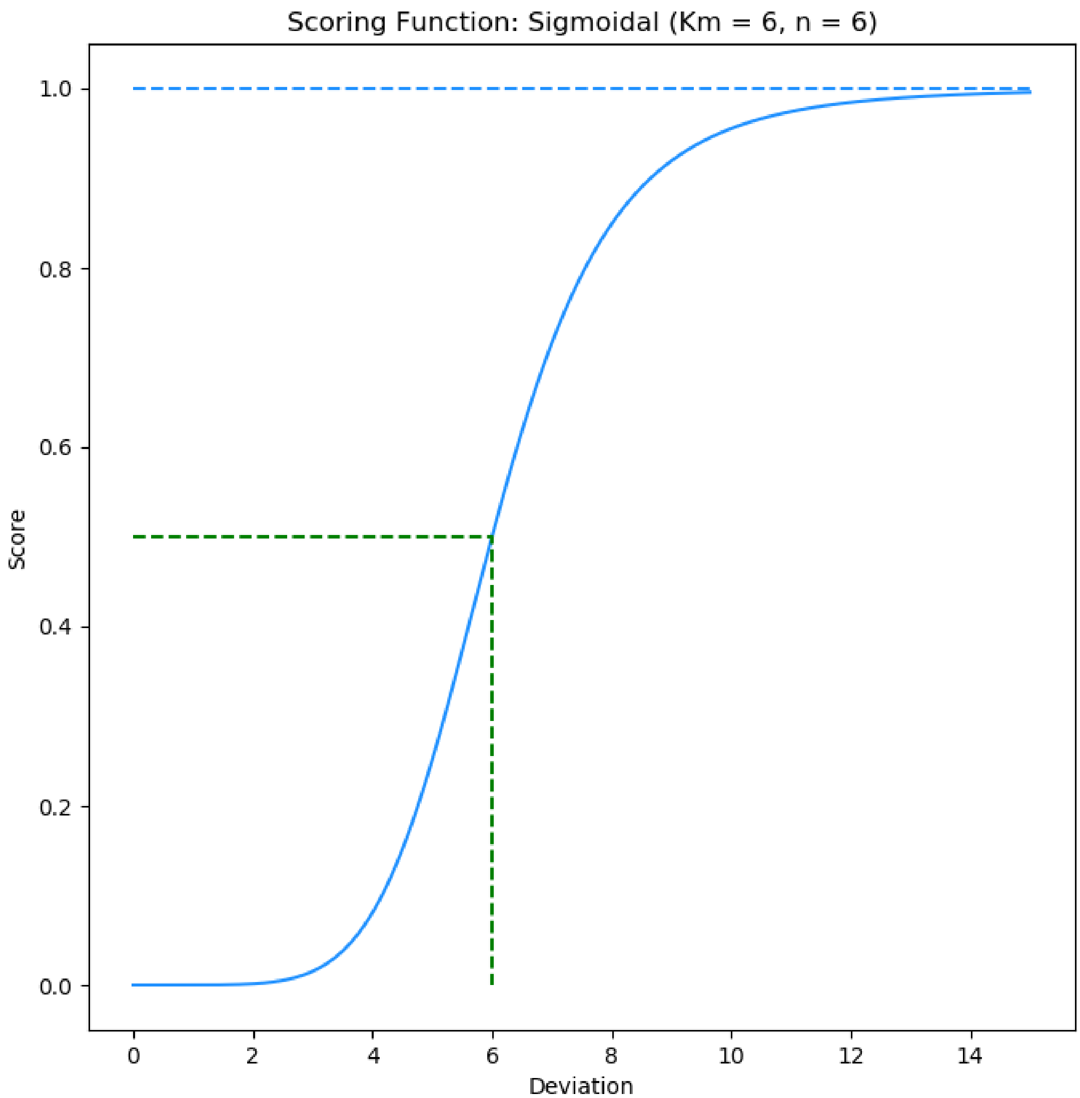

Figure 3. An example of a sigmoidal scoring function, $s = \frac{d^n}{{K_m}^n + d^n}$, where d is the point-pair deviation, n is the function cooperativity and K_m is the deviation midpoint. An n of 1 will create a hyperbolic function, while values greater than 1 will make the function increasingly sigmoidal. The value of K_m is the value of d at which the score is 0.5; this is represented by the dashed green lines.

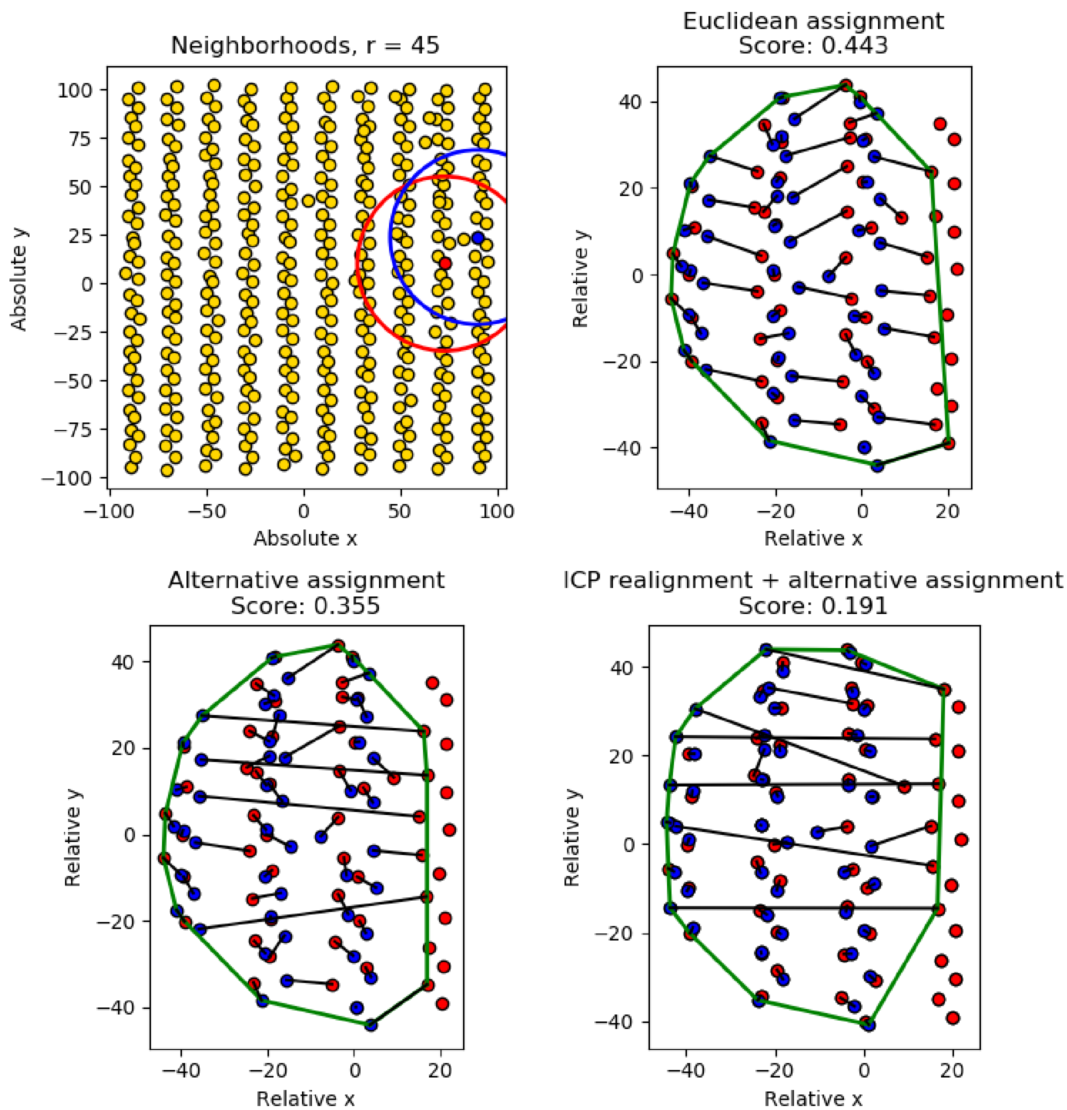

Figure 4. Visualization of the effects of using different cost functions on point assignment and IoD sub-score. The neighborhood radius is shown for both the red (parent) and blue (child) points in the upper left figure. The remaining three figures explore various point assignment methods and the calculated IoD.

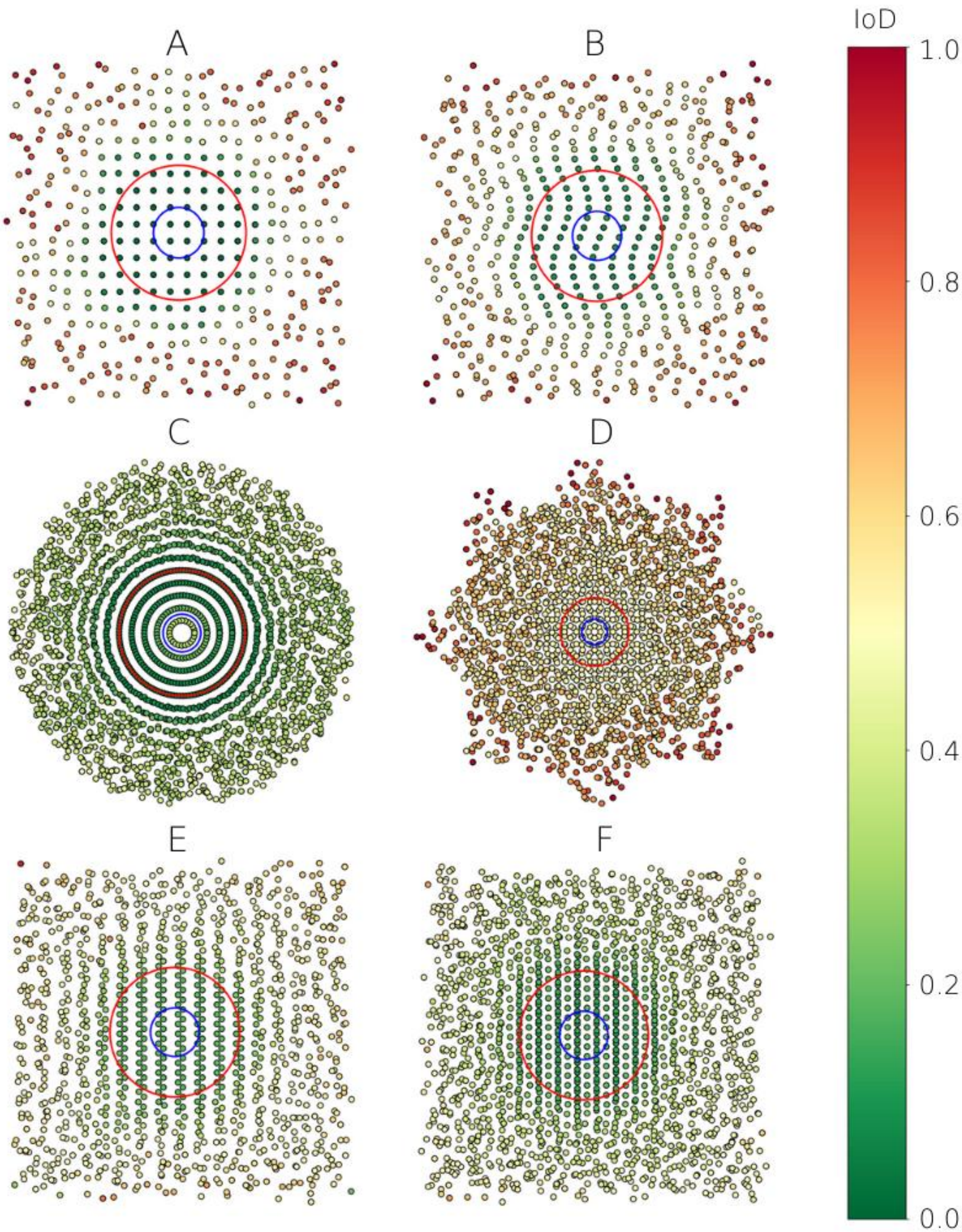

Figure 5. Results of theoretical evaluation of the IoD. The pattern is perturbed with random noise that increases in strength as a function of distance beyond the red circle, and the blue circle represents the neighborhood size. Multiple patterns were investigated: square grid (A), rectangular grid modified with a sinusoidal function (B), pattern formed by overlaying concentric circles with equal linear point densities (C), pattern formed by overlapping a square grid with a copy of the same grid rotated 45 degrees (D), pattern formed by overlapping two offset rectangular grids (E), pattern formed by overlapping three offset rectangular grids (F). No reorientation was applied.

Figure 6. A map of building types in a neighborhood in Nashville, TN.

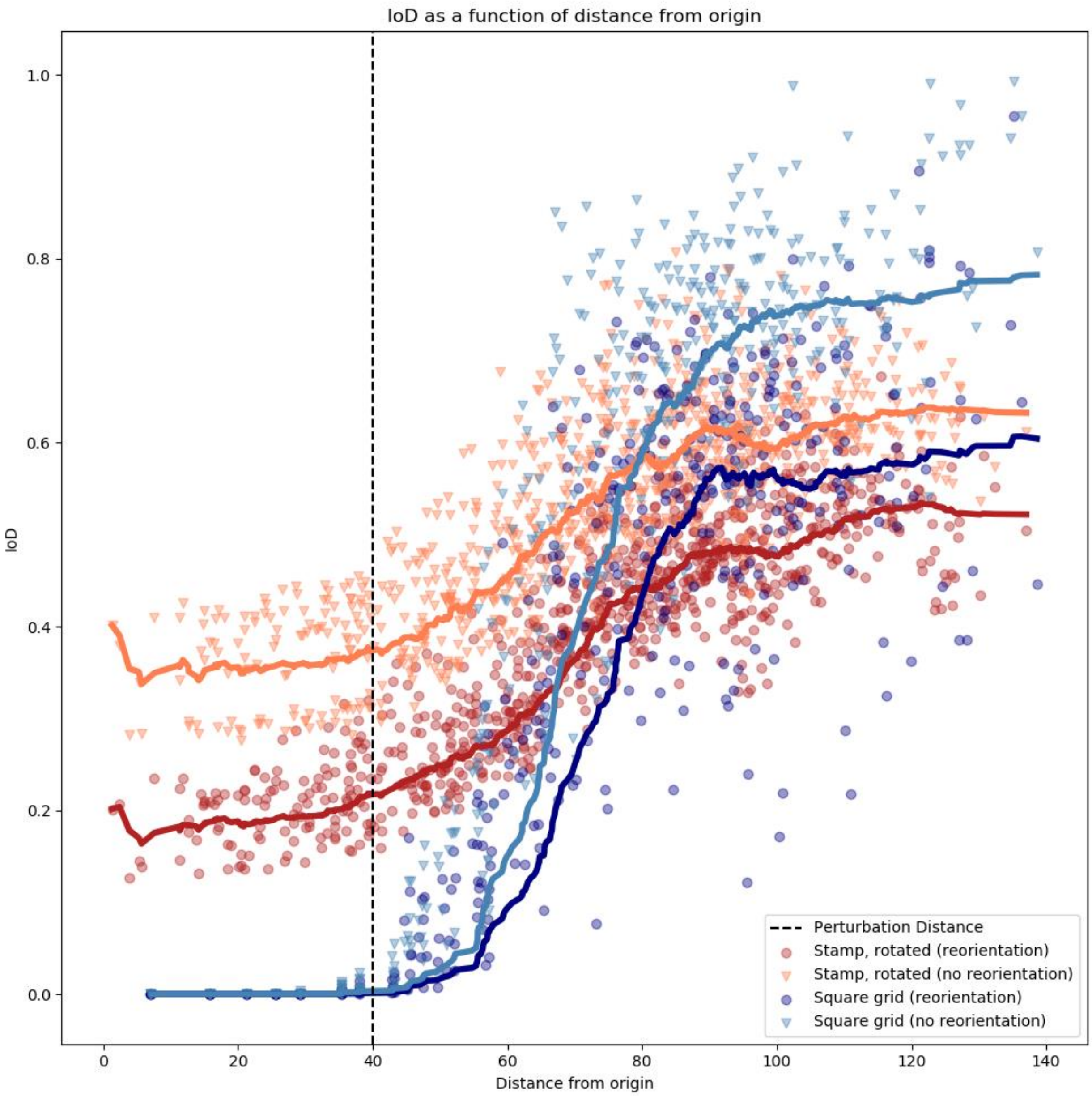

Figure 7. Illustration of index of disorder in theoretical evaluation with increasing noise beyond the perturbation distance. Point sets with a defined pattern were created and were then perturbed with noise of increasing magnitude beyond the perturbation distance. Some types of patterns, such as repeated stamps, show slightly increased discrimination between ordered and disordered points when reorientation is applied (realignment depresses the IoD of unperturbed points slightly more than it depresses the IoD of points beyond the perturbation distance). More homogenous patterns, such as a square grid, show reduced discrimination when reorientation is applied.

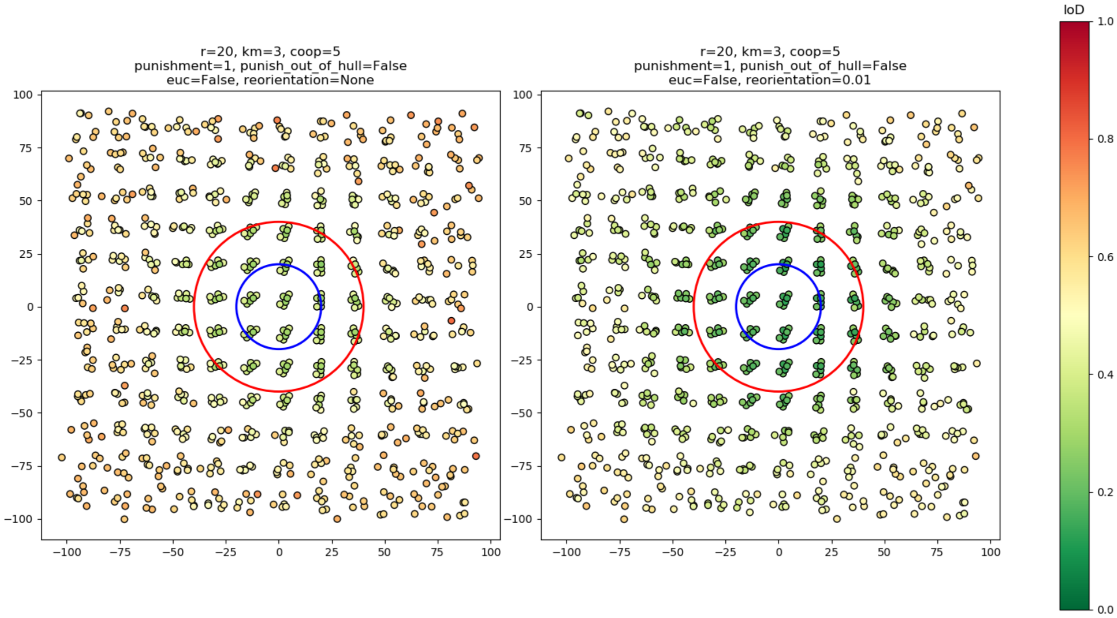

Figure 8. Results from applying two versions of the IoD algorithm without and with realignment (left and right respectively) to a pattern that consists of a repeating wallpaper group (a “stamp”) that is continuously rotated. The pattern is perturbed with random noise that increases in strength as a function of distance beyond the red circle, and the blue circle represents the neighborhood size.

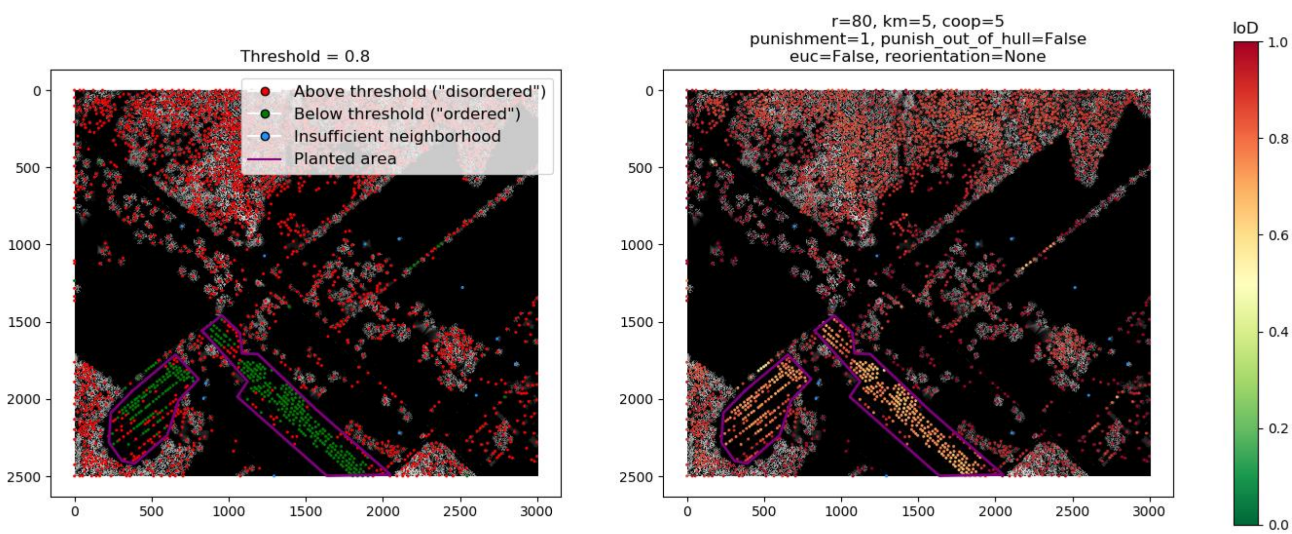

Figure 9. Results from applying the IoD to trees extracted from a DHM of an orchard near Crab Orchard, TN. Each point represents a tree crown. Axis units are meters. Trees with an IoD value above the threshold are classified as “disordered”, interpreted to be non-orchard trees. Trees with an IoD value below the threshold are classified as “ordered”, interpreted to be orchard trees.

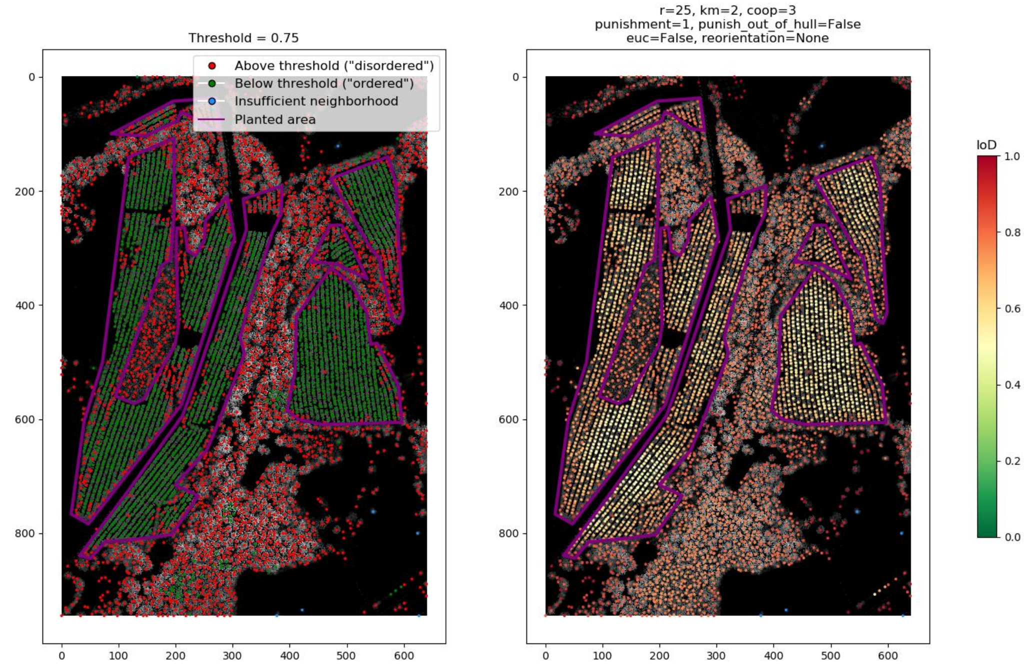

Figure 10. Results from applying the IoD to trees extracted from a DHM of a mixed planted and natural forest stand near Mooresville, NC. Each point represents a tree crown. Axis units are in meters. Trees with an IoD value above the threshold are classified as “disordered”, interpreted to be naturally occurring trees. Trees with an IoD value below the threshold are classified as “ordered”, interpreted to be planted trees.

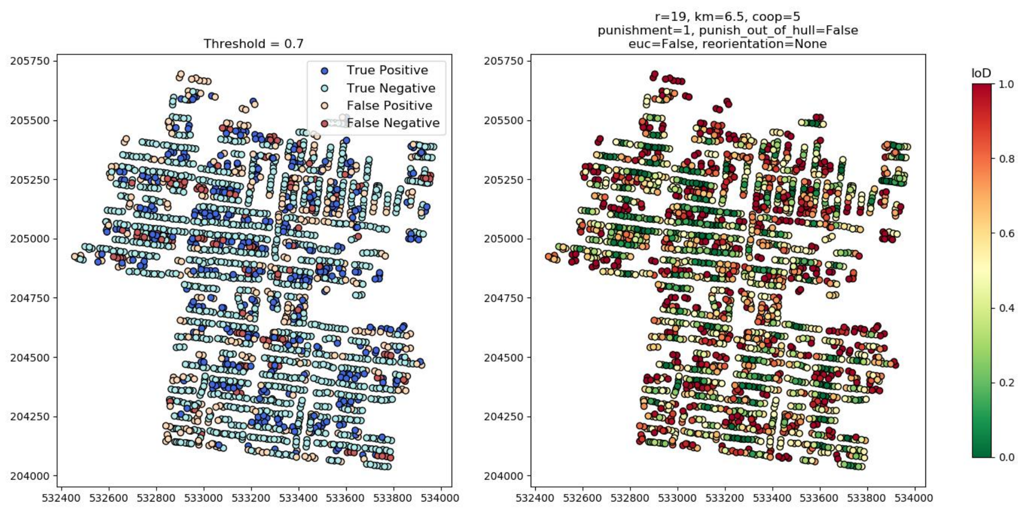

Figure 11. Building centroids in Nashville, TN. On the right, the IoD is shown. On the left, classification results are shown for a threshold of 0.7. Buildings with an IoD below the threshold are classified as “major buildings”, such as houses (a negative response) and those above the threshold are classified as “other”, which is comprised of auxiliary buildings such as sheds and detached garages (a positive response). Axis units are in meters and correspond to the TN State Plane coordinate system.

7 Acknowledgements

We thank Dr. Racha El Kadiri for her input and guidance while writing this manuscript.

8 Funding

This work was partially supported by the Natural Resources Conservation Service (agreement number NR194741XXXXC005) and a Middle Tennessee State University, Undergraduate Research Experience and Creative Activity (URECA) grant.

9 References

Aksoy, S., Yalniz, I. Z. & Tasdemir, K., 2012. Automatic Detection and Segmentation of Orchards Using Very High Resolution Imagery. *IEEE Transactions on Geoscience and Remote Sensing,* 50(8), pp. 3117-3131.

Antuono, M., Bouscasse, B., Colagrossi, A. & Marrone, S., 2014. A measure of spatial disorder in particle methods. *Computer Physics Communications,* 185(10).

Besl, P. & McKay, N., 1992. A Method for Registration of 3-D Shapes. *IEEE Transactions on Pattern Analysis and Machine Intelligence,* 14(2), p. 239–256.

Boeing, G., 2019. Urban Spatial Order: Street Network Orientation, Configuration, and Entropy. *Applied Network Science.*

Cohen, J., 1960. A coefficient of agreement for nominal scales. Educational and Psychological Measurement. *Educational and Psychological Measurement,* 20(1), pp. 37-46.

Dong, P., 2000. Lacunarity for Spatial Heterogeneity Measurement in GIS. *Annals of GIS,* 6(1), pp. 20-26.

Haralick, R. M., Shanmugam, K. & Dinstein, I., 1973. Textural Features for Image Classification. *IEEE Transactions on Systems, Man, and Cybernetics,* SMC-3(6), pp. 610 - 621.

Isenburg, M., 2019. *LAStools - Efficient Tools for LiDAR Processing.* s.l.:rapidlasso GmbH.

Khan, M. M., Al-Yahyai, R. & Al-Said, F., 2017. *The Lime: Botany, Production and Uses.* s.l.:CABI.

Kuhn, H. W., 1955. The Hungarian method for the assignment problem. *Naval Research Logistics Quarterly,* pp. 83-97.

Liu, Y., Collins, R. & Tsin, Y., 2004. A computational model for periodic pattern perception based on frieze and wallpaper groups. *IEEE Transactions on Pattern Analysis and Machine Intelligence,* 26(3), pp. 354-371.

Lou, S. et al., 2016. Fusion of Airborne Discrete-Return LiDAR and Hyperspectral Data for Land Cover Classification. *Remote Sensing,* 8(1).

Momm, H., Easson, G. & Kuszmaul, J., 2009. Evaluation of the use of spectral and textural information by an evolutionary algorithm for multi-spectral imagery classification. *Computers, Environment and Urban Systems,* 33(6), pp. 463-471.

Nilsson, L. & Gil, J., 2019. *The Mathematics of Urban Morphology.* s.l.:Springer.

Oliphant, T. E., 2006. *A guide to NumPy.* s.l.:Trelgol Publishing.

Pauly, M. et al., 2008. Discovering structural regularity in 3D geometry. *ACM Transactions on Graphics,* 27(3).

Pedregosa, F. et al., 2011. Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research,* Volume 12, pp. 2825-2830.

Phiri, D. & Morgenroth, J., 2017. Developments in Landsat Land Cover Classification Methods: A Review. *Remote Sensing,* 9(9).

Remmel, T. K. & Csillag, F., 2003. When are two landscape pattern indices significantly different?. *Journal of Geographical Systems,* Volume 5, pp. 331-351.

Shekhar, S., Evans, M., Kang, J. & Mohan, P., 2011. Identifying patterns in spatial information: A survey of methods. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery,* 1(3), pp. 193-214.

Swain, P. H. & Stephen B. Vardman, J. C. T., 1981. Contextual classification of multispectral image data. *Pattern Recognition,* 13(6), pp. 429-441.

Trias-Sanz, R., 2006. Texture Orientation and Period Estimator for Discriminating Between Forests, Orchards, Vineyards, and Tilled Fields. *IEEE Transactions on Geoscience and Remote Sensing,* 44(1020).

Zhen, Z., Quackenbush, L. J. & Zhang, L., 2016. Trends in Automatic Individual Tree Crown Detection and Delineation—Evolution of LiDAR Data. *Remote Sensing,* 8(4), p. 333.