**A robust, generalizable model for canopy detection from generic LiDAR in the contiguous United States**

S. Jones1[[1]](#footnote-2)\*, R. El Kadiri1[[2]](#footnote-3) and H.G. Momm1b

1Department 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.

# Introduction

* Uses of LiDAR detection
* Existing canopy detection methods (both bespoke and general)
  + Closed source (ENVI, Arc, etc.)
  + Detection from point cloud vs. raster
  + Detection from aerial imagery (including photogrammetry)
* Need for generalized model

# Methods

## Model Design

* Limitations of general model (can’t use returns, intensity)
* Data products generated (DHM, textures, Laplace etc.)

## Study Areas

### 010500021301 – Penobscot, ME (trained)

### 030902040303 – Naples, FL (trained)

### 0070801050901 – Kelley, IA (trained)

### 080102040304 – Western TN (trained)

### 080902030201 – New Orleans, LA (naïve)

### 100301011309 – Helena, MT (naïve)

### 102901110304 – Freeburg, MO (naïve)

### 130202090102 – Alamo, NM (trained)

### 140801040103 – Central CO (trained)

### 180500020905 – San Francisco, CA (trained)

# Results and Discussion

## 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 Km 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 (*Figures 7 and 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.

## Natural Evaluations

### Site 1

Sensitivity analysis was performed by varying IoD input parameters of neighborhood radius *r* and sigmoidal assigned threshold value Km 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 Km 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 Km describes the level of expected deviation within the patterns. Though *r* is a parameter that will be present in any implementation of the IoD, Km 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 Km 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.

### 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 Km 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).

### 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 Km 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 Km 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 Km do not necessarily describe either pattern but rather represent, respectively, a discriminatory scale and a discriminatory deviance.

## Impact of Alternative Implementations of the IoD

### 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 (Km) imparts additional scale-awareness to the IoD; the Km describes the expected deviation, or “noise” within a pattern, and so displacements below Km are considered within the tolerance of the expected noise level and punished less. Without Km 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.

### 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.

# 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.

# 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.

# Tables and Figures

# Acknowledgements

We would like to thank the Natural Resources Conservation Service for providing funding for this project.

# Funding

This work was supported by the Natural Resources Conservation Service (agreement number NR194741XXXXC005).

# References

Potential works:

* Weinstein, B.G.; Marconi, S.; Bohlman, S.; Zare, A.; White, E. Individual Tree-Crown Detection in RGB Imagery Using Semi-Supervised Deep Learning Neural Networks. Remote Sens. 2019, 11, 1309 **AND** Geographic Generalization in Airborne RGB Deep Learning Tree Detection Ben Weinstein, Sergio Marconi, Stephanie Bohlman, Alina Zare, Ethan P White bioRxiv 790071; doi: <https://doi.org/10.1101/790071>
  + Related to <https://github.com/weecology/DeepForest>. Covers a generalizable RGB-based model for crown-detection. Essentially our work, except imagery based and goes a step further (detecting crowns). Maybe we can implement crown detection?
* <https://pypi.org/project/forestutils/>
  + Tree extraction directly from point clouds. I believe it needs full color point clouds, which come from either photogrammetry or full-color (i.e., expensive) LiDAR scans
* A Segmentation Method for Tree Crown Detection and Modelling from LiDAR Measurements - <https://link.springer.com/chapter/10.1007/978-3-642-31149-9_7>
  + 2012 paper on the inverted watershed (raster-based) method of crown detection
* Trends in Automatic Individual Tree Crown Detection and Delineation—Evolution of LiDAR Data
* <https://www.researchgate.net/publication/259128823_High-resolution_tree_canopy_mapping_for_New_York_City_using_LIDAR_and_object-based_image_analysis>]
* Characterizing urban surface cover and structure with airborne lidar technology
  + Contains schematic for image filter that remove linear structures (building edge detection)

1. Responsible for design, implementation and testing of model, collection and processing of data, and drafting of manuscript. [↑](#footnote-ref-2)
2. Responsible for supervision of model design, manuscript drafting and revision. [↑](#footnote-ref-3)