NCJRS
National Criminal Justice Reference Service

Office of Justice Programs
BJA BJS NIJ OJJDP OVC SMART

The author(s) shown below used Federal funding provided by the U.S. Department of Justice to prepare the following resource:

Document Title:	**Poisson Processes and Randomly Acquired Characteristics: Are Wear Features on Footwear Randomly Distributed?**
Author(s):	**Jacqueline A. Speir**
Document Number:	**305756**
Date Received:	**January 2023**
Award Number:	**2018-MU-MU-0003**

This resource has not been published by the U.S. Department of Justice. This resource is being made publicly available through the Office of Justice Programs' National Criminal Justice Reference Service.

Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

EBERLY COLLEGE OF ARTS AND SCIENCES
FORENSIC & INVESTIGATIVE SCIENCE

Final Technical Report: 2018-MU-MU-0003
Project Title: Poisson Processes and Randomly Acquired Characteristics:
Are Wear Features on Footwear Randomly Distributed?

Principal Investigator: Jacqueline A. Speir, Associate Professor
West Virginia University (WVU), 208 Oglebay Hall, PO Box 6121, Morgantown,
WV 26506, `Jacqueline.Speir@mail.wvu.edu`, 304.293.9233

Recipient Organization: West Virginia University Research Corporation, 866 Chestnut
Ridge Road, PO Box 6845, Morgantown, WV 26506, **DUNS: EIN:**
`wvusponsoredprograms@mail.wvu.edu`, 304.293.3998
Submitting Official: Katie Schneller, Director of Office of Sponsored Research

Award Period & Amount: 01/01/2019 - 03/31/2022, $178,080
Reporting Period End Date: 03/31/2022

Keywords: randomly acquired characteristics, spatial randomness, Poisson processes,
generalized linear regression, Moran's I

This resource was prepared by the author(s) using Federal funds provided by the U.S.
Department of Justice. Opinions or points of view expressed are those of the author(s) and do not
necessarily reflect the official position or policies of the U.S. Department of Justice.

Contents

This resource was prepared by the author(s) using Federal funds provided by the U.S. Department of Justice. Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

1. Project Summary

Demonstrating a linkage between known footwear and a questioned impression left at the scene of a crime is a function of the observed agreement between class, subclass and wear features exhibited by the impressions being compared. If sufficient detail is present (beyond class and subclass characteristics), then the similarity, clarity, quantity and quality of what are termed '*randomly acquired characteristics*' (RACs) (such as nicks, tears, cuts, etc.) can form the basis of a source identification. When an examiner evaluates the possibility of a source identification using RACs, there is consideration of how likely the observed RAC agreement (in terms of position, size, shape, geometry, etc.) would be expected by chance alone when comparing unrelated outsoles. Within this approach is some consideration of the spatial distribution of RACs on outsoles. In other words, does the rarity of a RAC repeating by chance alone vary with its spatial position (*i.e.*, on the toe, heel, etc.), or are all RACs '*created equal*?'

Confining attention to RACs expected to reproduce in 2-dimensional test impressions, and those based on material-loss (nicks, tears, scratches, etc.) — rather than material acquisition (such as stones, nails, gum, etc.) — even if RACs are randomly '*acquired*,' they are not expected to be uniformly distributed across an outsole. First, when restricting attention to material-loss features, these features can only develop in locations where tread is in contact with the ground. Thus, an outsole's tread pattern dictates the possible distribution of RACs. Second, the attributes that lead to RAC development (interaction between the outsole and terrain) are not necessarily randomly distributed across an outsole. Instead, factors such as weight, gait, pronation/supination, etc. are all likely to contribute to the degree of interaction between the outsole and the terrain. As a consequence, these features are not expected to be uniformly distributed across an outsole since the factors that lead to their development (tread and degree of wear) are not uniformly distributed across an outsole.

The aim of this work was to investigate the distribution of RACs in an empirical dataset and compare it to an inhomogeneous Poisson point process modified by tread contact and wear. To achieve this goal, the RAC spatial frequency within an empirical dataset was compared against simulated and modeled data assuming a Poisson point process. Deviations in count between the empirical and simulated/modeled predictions were examined using a Poisson rate test and Moran's I. Results indicate that RAC frequency over 67% to 79% of an outsole can be reasonably well explained as a Poisson point process or by a Poisson generalized linear regression model (non-spatial GLM) with tread contact as a predictor. Moreover, if the predictor is extended to include both tread contact and wear, RAC count over 84% of the spatial locations on an outsole are well-explained (although autocorrelation in Pearson residuals persists). Overall, results indicate that RACs are not uniformly distributed in this dataset, most likely because the factors that dictate RAC development (friction, gait, etc.) are not uniformly distributed. Although this observation in no way negates the use of RACs in forming source associations, the value of a correspondence may require interpretation as a function of the feature's spatial location.

This resource was prepared by the author(s) using Federal funds provided by the U.S.
Department of Justice. Opinions or points of view expressed are those of the author(s) and do not
necessarily reflect the official position or policies of the U.S. Department of Justice.

1.1 Major Goal/Objective

The major goal of this research was to compare the spatial distribution of randomly acquired characteristics in an empirical opportunistic/convenience dataset to associated distributions with known attributes. The first comparison evaluated the observed data versus a simulated distribution that conformed to an inhomogeneous Poisson point process using tread in contact with the ground as a modifier. The second comparison sought to evaluate the empirical data versus a modeled distribution using contact and contact+wear as predictors. For details concerning the empirical dataset, the reader is referred to [1] (expanded from 1,000 outsoles to 1,300 outsoles, with a total of 72,306 randomly acquired characteristics at the time of use in addressing this research question).

1.2 Research Question & Impact

Achieving the aforementioned comparisons would serve to determine the degree of agreement between the empirical data and a random Poisson point process predicted using tread and tread + wear. Answering this question serves to increase the forensic footwear community's knowledge of the degree of '*uniformity*' expected in RAC distributions; if features are shown to deviate from a random distribution, then the examiner can incorporate the spatial position of a feature when forming an opinion about the chance of a characteristic of use repeating by chance in unrelated outsoles.

1.3 Research Methods

To answer the proposed research question, the tasks described in Table 1 were accomplished. First, the empirical dataset was evaluated for tread in contact with the ground, since this was a primary inhomogeneity factor that was expected to limit the distribution of RACs. Since the empirical dataset in question is an opportunistic or convenience sample of outsoles that vary in size, make and model, each had to be processed to create tread-contact binary maps. This was accomplished using an image processing procedure that included downsampling, median filtering using a 3 $\times$ 3 window, mode subtraction, thresholding using $-0.5 \times \sigma_i$ where σ_i is the image's grayscale standard deviation, adaptive histogram equalization, and finally, an edge detection step using a difference of Gaussian (DOG) with σ_{DOG} between 3.0 and 5.0. The result generated a black-and-white map for each outsole, where black reflects spatial locations with tread elements in contact with the ground, while white reflects spatial locations of raised areas. An example map is illustrated in Figure 1, which shows the outsole of a Nike® men's size 12 athletic shoe (A), its associated test impression (B) created using fingerprint powder and Handiprint sheets, and a binary (black-and-white) tread-contact image (C).

This resource was prepared by the author(s) using Federal funds provided by the U.S.
Department of Justice. Opinions or points of view expressed are those of the author(s) and do not
necessarily reflect the official position or policies of the U.S. Department of Justice.

Table 1: Tasks and associated descriptions associated with data generation, data processing and data analysis.

Task	Description
Data Processing	Create Contact-Area Modified Outsole Imagery
Data Generation	Simulate Inhomogeneous Poisson Point Processes
Data Processing	Characterize Wear
Data Generation	Non-Spatial Generalized Linear Model (GLM) using Contact and Contact + Wear as Predictors
Data Analysis	Evaluate Agreement in Distributions using a Poisson Rate Test and Local Moran's I

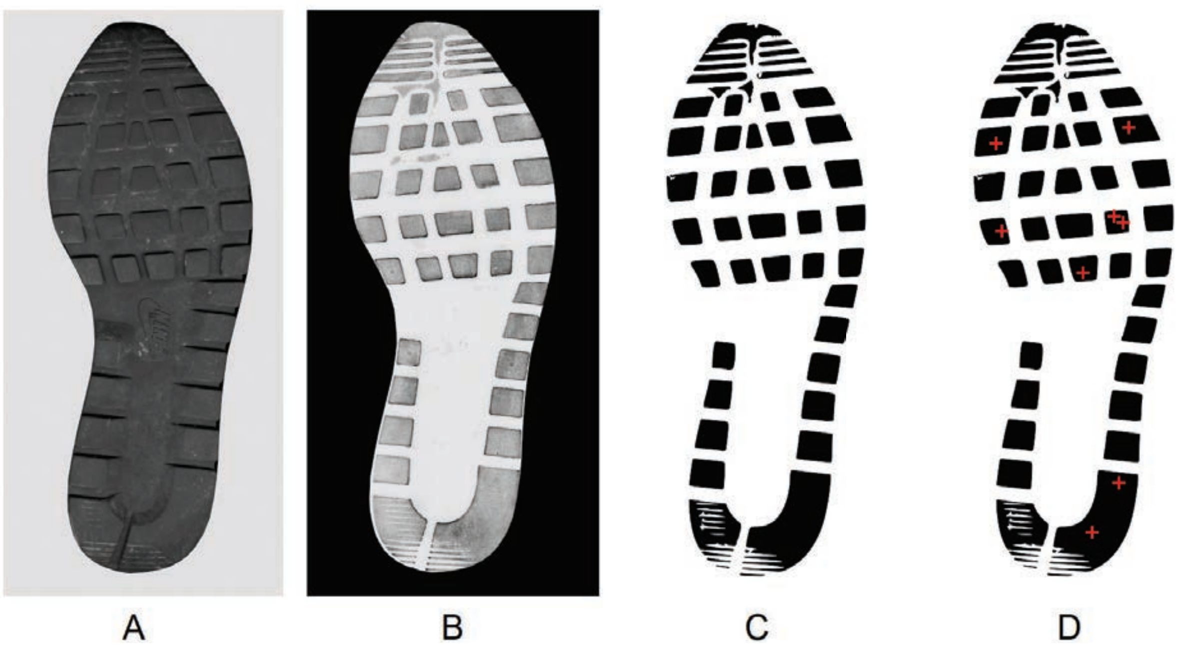

Figure 1: (A) Nike® men's size 12 outsole, (B) test impression, (C) contact-modified binary image, and (D) binary image with possible Poisson point process realizations illustrated using red-colored + symbols.

Next, null distributions were generated. The contact-modified dataset forms the basis for creating null distributions, where a null distribution was defined as one that conformed to a random or Poisson point process but with spatial locations restricted to tread contact areas. Using the observed number of RACs in the empirical dataset [1] as the number of points to simulate and the contact-area modified binary maps as a '*window*' to describe the positions where these points could fall [2], a minimum of 1,000 Poisson random realizations were created per shoe using the `rpoisspp()` function from the `spatstat` package for `R` [2]. An example is illustrated in Figure 1 (D) where the red + symbols indicate possible realizations. After creation of all shoe-specific simulations, 10,000 random heatmap realizations were sampled, where each heatmap is based on 1,300 outsoles and includes 72,306 simulated '*RACs*' localized and binned in the same manner as the empirical dataset [1, 3].

This resource was prepared by the author(s) using Federal funds provided by the U.S. Department of Justice. Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

Since the empirical dataset contains a variety of shoes with different class and subclass characteristics, these shoes likewise exhibit considerable variation in outsole condition, specifically regarding the locations and degree of wear. In order to characterize wear, the criteria described in Table 2 were defined. Factors such as degradation of texture, changes in element appearance and spacing, erosion, and balding were used to create an ordinal scale of light, moderate and heavy wear.

Table 2: Factors and associated criteria to define wear as light, moderate or heavy.

Factor	Light Wear	Moderate Wear	High Wear
Outsole texture degradation	Minimal degradation of outsole texture/minimal areas of balding	Moderate degradation of outsole texture (with some areas of texture persisting)/observable areas of balding	Advanced/complete degradation of outsole texture/complete balding
Outsole material degradation	Little to no erosion of outsole material	Some erosion of outsole material (without evidence of tearing/shearing)	Severe erosion of outsole material, likely resulting in tearing/shearing
Tread element appearance	Minimal degradation of tread elements (crisp edges with lack of discernable height/size difference)	Moderate degradation of tread elements (smoothing of edges with some observable changes in height/size)	Advanced degradation of tread elements (missing or broken, significant observable changes in height/size)
Tread element spacing	Tread element spacing relatively unaffected	Some reduction in tread element spacing (elements do not completely run together)	Tread element spacing greatly impacted (elements may run entirely together or be flush with the base layer of outsole material)
Other	N/A	N/A	Areas of significant wear causing holes and/or visualization of the upper outsole structure layer (*e.g.,* honeycomb, lattice)

To maximize the number of RACs under consideration, all 1,300 shoes were sorted based upon the number of RACs they exhibited, and the top 10% (130 shoes) were selected for wear evaluation, yielding 27,933 total accidentals available for assessment (39% of all RACs included in the full database). Based on the criteria outlined in Table 2, each shoe was inspected for degree of wear and 130 shoe-specific "wear maps" were created as illustrated in Figure 2.

These maps were created using Adobe® Photoshop® Elements 10 via annotation of a semi-transparent mask layer. Each test impression was opened and the layer's grayscale value altered in order to annotate the location and degree of wear across the entire frame of the shoe. More specifically, three different grayscale values were attributed to pixels in order to generate a coarse scaling of wear, such that a grayscale of 50 was used to identify high wear, a grayscale of 255 to identify moderate wear, and a grayscale of 175 to identify areas of light wear.

This resource was prepared by the author(s) using Federal funds provided by the U.S. Department of Justice. Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

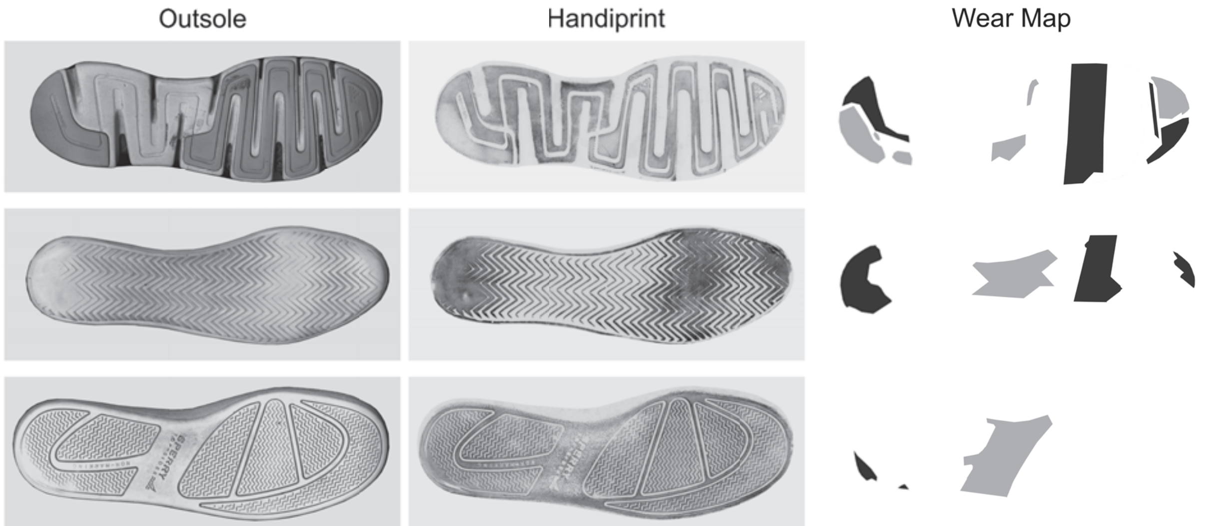

Figure 2: Shoe-specific wear maps (right) for outsoles (left) and associated test-impressions (middle). Image reproduced from [4].

Equipped with contact information and wear maps, the next phase was to model the distribution of RACs using a non-spatial generalized linear regression (or GLM). This is a global model that does not consider spatial relationships between observations and therefore contains an underlying assumption of independence. The general expression is shown in Equation 1 where Y is the response or RAC count, X are predictors, β are modeled coefficients for each predictor, and is residual error.

$$\begin{aligned} Y &\sim Poisson(\lambda) \\ log(\lambda) &= \beta_o + X\beta + \end{aligned} \tag{1}$$

Using this model, Equation 2 describes RAC count as a function of tread contact, and bin partiality. Contact is defined as the average degree of tread in contact with the ground, while bin partiality is an offset variable used to prevent confounding as a function of unequal cell size for any cell on the shoe's perimeter (full cells are 5mm x 5mm or 14,161 pixels).

$$\begin{aligned} log(\lambda) &= \beta_o + \beta_1 C + \beta_2 log(B) \\ C &= \frac{\sum_1^{1,300} \#\text{ pixels with contact}}{1,300 \text{ shoes}} \\ B &= \frac{\#\text{ pixels on frame}}{14,161 \text{ pixels}} \end{aligned} \tag{2}$$

This resource was prepared by the author(s) using Federal funds provided by the U.S. Department of Justice. Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

Again by extension, Equation 3 describes contact-modified wear using the notation L_c for average contact-localized light wear, M_c for average contact-localized moderate wear, and H_c for average contact-localized high wear.

$$log(\lambda) \quad = \quad \beta_o + \beta_1 L_c + \beta_2 M_c + \beta_3 H_c + \beta_4 log(B) \tag{3}$$

1.4 Data Analysis

The simulated distributions (conforming to the assumption of an inhomogeneous Poisson point process) and the non-spatial GLMs were compared to the empirical data in order to test for deviations. This comparison was comprised of two metrics. The first was a Poisson rate test. The propensity for a RAC to be ascribed to a specific cell was considered an outcome x_i modeled as a Poisson process with parameter λ_i and Poisson rate $\lambda_i = s_i \gamma_i$ (*i.e.*, $x_i \sim Poisson(\lambda_i)$ where s_i is the sampling frame or window) [5]. In testing for the equality of rates, the hypotheses $H_0 : \gamma_e = \gamma_{s/m}$ versus $H_1 : \gamma_e = \gamma_{s/m}$ was assessed, where e denotes the empirical dataset and s and m denote a simulated and/or modeled dataset, respectively. The test statistic is provided in Equation 4 for a constant sampling window/frame, and W_5 is distributed as a standard normal [5], such that we reject H_0 when $p < \alpha$ when $\alpha = 0.05$.

$$W_5 \quad = \quad \frac{2}{\sqrt{2}} \left(\sqrt{x_e + 3/8} - \sqrt{x_{s/m} + 3/8} \right) \tag{4}$$

In addition to the Poisson rate test, local spatial autocorrelation was also evaluated. This metric computes the degree to which RAC counts (or residuals) in a spatial area are similar to each other, such that positive spatial correlation indicates that similar features cluster, while negative correlation indicates that dissimilar features cluster. This was performed on the Pearson residuals of the simulated (or modeled) versus empirical datasets in order to determine if differences in residuals cluster together.

More specifically, local Moran's I (I_i) (assuming total randomization) was computed for each of the N = 987 cells across the RAC heatmap according to Equation 5 [6, 7], where i represents a heatmap cell of interest, j represents a second heatmap cell against which i is being compared, and x represents the attribute of interest (which are Pearson residuals of RAC count for empirical versus simulated/modeled datasets). In Equation 5 the variable w_{ij} defines a spatial weight value, meant to describe the physical relationship between cells i and j under consideration. In this regard, a simple row-normalized inverse Euclidean distance weight matrix was used for any cell inter-distance less than a predefined threshold, and a weight of zero for any cell more distant than a series of predefined neighborhoods of interest

This resource was prepared by the author(s) using Federal funds provided by the U.S. Department of Justice. Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

(400px = 16.7mm (the average diameter of a lug or tread element in this dataset), 800 px = 33.3mm (upper limit of lug diameter), 1,000px = 41.7mm, 1,500px = 62.5mm, 1,750px = 72.9mm and 2,500px = 104mm (the widest width of the toe on the '*standardized*' shoe for the empirical dataset [1])).

$$
\begin{aligned}
s_i^2 &= \frac{\sum_{j=1}^{N}(x_j - \bar{x})^2}{N-1} \\
I_i &= \frac{x_i - \bar{x}}{s_i^2} \sum_{j=1}^{N} w_{ij}(x_j - \bar{x})
\end{aligned} \tag{5}
$$

1.5 Expected Applicability

This research is applicable to the forensic footwear community. It serves to evaluate the spatial distribution of wear features in an opportunistic/convenience dataset consisting of 1,300 outsoles and 72,306 randomly acquired characteristics. However, since the dataset is based on characteristics of use identified in high-quality test impressions, the same spatial distribution may not be observed in lower quality questioned impressions collected from crime scenes. Instead, the results, and any deviation of uniformity in distribution, should be used as a baseline for what is *possible*, within the confines of the assumptions and limitations described below.

---

2. Changes in Approach

The original/funded project intended to evaluate tread contact as a predictor of RAC spatial frequency. This was extended to include both tread contact and wear as a predictor, thereby expanding the original scope or the proposed research. In addition, measures of complete spatial randomness (CSR) were originally proposed as a means to evaluate the simulated and empirical datasets. However, since the simulated data were created using an inhomogeneous Poisson point process, traditional CSR measures were ineffective for evaluating the spatial distribution. Instead, several models were employed, including a non-spatial generalized linear model, a spatial auto-Poisson model, a spatial Poisson Durbin regression, and a spatially geographically-weighted Poisson model. The reader is referred to [4] for additional details.

This resource was prepared by the author(s) using Federal funds provided by the U.S. Department of Justice. Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

3. Outcomes

3.1 Simulation Results: Tread Contact

The simulated heatmaps, based on a tread-modified Poisson process, were compared against the empirical heatmap by examining the ratio of cell-specific Poisson rates. To accomplish this, the cell-specific rate equality of the simulated and empirical datasets was examined, assuming a null hypothesis of unity (*i.e.*, that the simulated and observed rates per cell were equal between the empirical and the Poisson simulated heatmaps). Several test statistics are available for this computation, of which W_5 is proposed as having the greatest power [5] (or the greatest ability to reject the null hypothesis of equality when it is false). Since 10,000 simulated heatmaps were prepared, this test was conducted 10,000 times per cell, and the proportion of significant results is illustrated in Figure 3.

For 33% of the spatial locations on the heatmap, the rate test failed to detect (or rarely detected; less than 1% of the time) a significant difference between the empirical and simulated results (*i.e.*, detected a difference in a bin in fewer than 100 simulations out of the 10,000 possible simulations). Likewise, 51% of the cell locations revealed a significant difference in no more than 10% of the comparisons (*i.e.*, detected a difference in fewer than 1,000 simulated heatmaps). At the other extreme, nearly a quarter (24%) of all cell locations exhibited a significant rate difference between the empirical and the simulated data 75% of the time or more often (*i.e.*, detected a difference in greater than 7,500 simulated heatmaps).

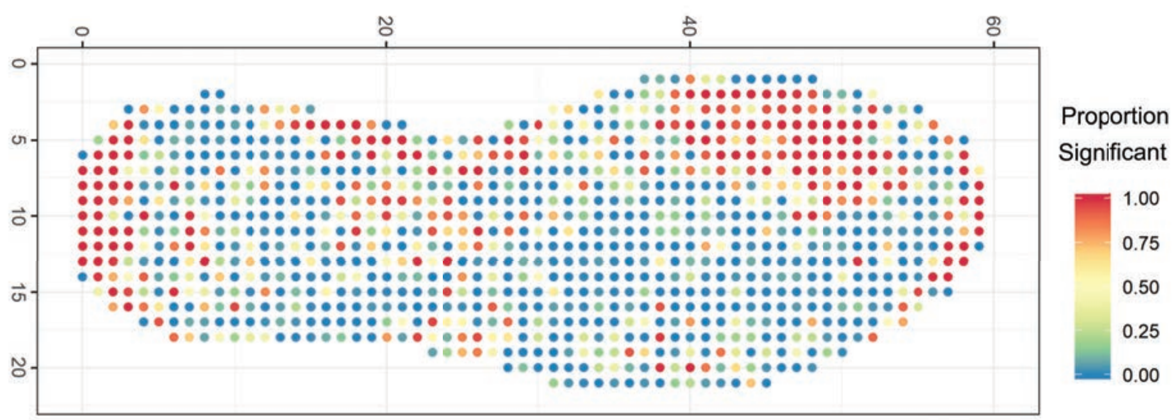

Figure 3: Results of Poisson ratio test for empirical versus 10,000 Poisson simulations where color indicates the proportion of times the rate in a specific cell yielded a significant difference.

Considering all 10,000 simulations, the cumulative results indicate that the Poisson simulations well predict empirical RAC counts across an average of 67% of bin locations (*i.e.*, the rate test failed to detect significant differences between the empirical and simulated RAC

This resource was prepared by the author(s) using Federal funds provided by the U.S. Department of Justice. Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

frequencies in about two-thirds of locations on average, across all 10,000 simulations), while the remaining 33% are poorly predicted and require additional investigation. These are illustrated as red-orange colored points in Figure 3. Note that these significant differences cluster in the medial ball of the toe, the heel, and to a lesser degree, the instep or arch of the shoe. If this comparison is repeated using a chi-square test where the Poisson simulated results are considered the '*expected*' RAC counts, the same basic pattern and rough percentage of significant locations is found.

When the difference in RAC count per cell between the empirical and simulated heatmaps is computed, the resulting residuals were evaluated for spatial autocorrelation using local Moran's I. The proportion of Pearson residuals exhibiting significant correlation is illustrated in Figure 4. Regardless of neighborhood cut-off, a consistent pattern of correlation in results emerged, clustering in the medial ball of the toe, the heel, and to a lesser degree, the instep. On average, the RAC Pearson residual for frequency in 70% of all bin locations does not exhibit spatial correlation, at the mean diameter of a lug (800px; 33.3mm). Conversely, spatial correlation is detected in approximately 30% of the bin locations across the standardized outsole.

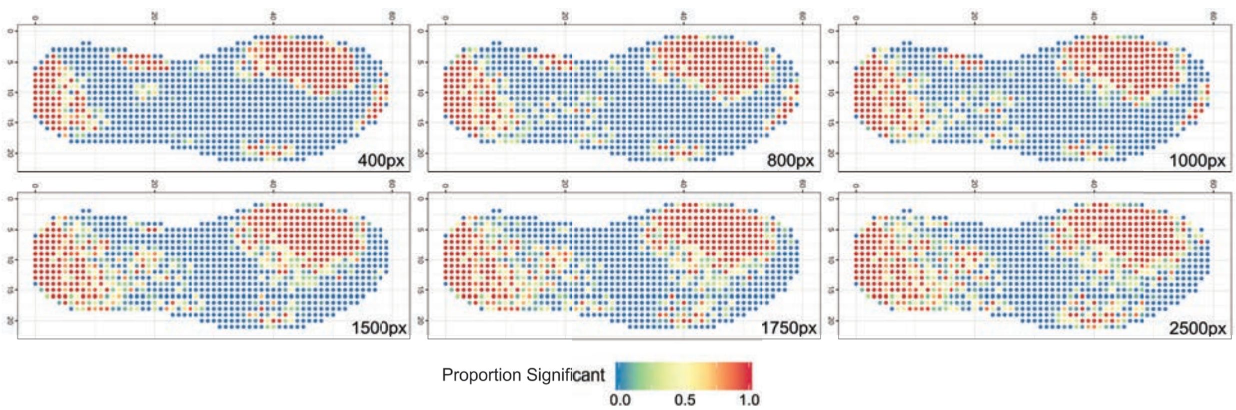

Figure 4: Results of local Moran's I for empirical RAC counts versus 10,000 Poisson simulations where color indicates the proportion of times the residuals in a specific bin yielded a significant difference at six difference neighborhood cut-off values (400px = 16.7mm, 800 px = 33.3mm, 1,000px = 41.7mm, 1,500px = 62.5mm, 1,750px = 72.9mm and 2,500px = 104mm).

3.2 Modeled Results: Tread Contact

The model coefficients for the Poisson generalized linear regression (or a non-spatial generalized linear model (GLM)), to predict RAC count (Y) as a function of tread contact and cell % offset are detailed in Table 3.

Table 3: Model coefficients for the non-spatial Poisson generalized linear model to predict RAC frequency for the full dataset of 72,306 RACs on 1,300 outsoles, and the subset of 27,933 RACs on 130 outsoles.

Model	**Dataset**	**Intercept (β_0)**	**Contact (β_1)**	**Cell % Offset (β_2)**
Non-Spatial GLM	Full (1,300 shoes)	2.957	2.2e-04	0.5944
Non-Spatial GLM	Subset (130 shoes)	2.376	1.6e-04	0.9433

This resource was prepared by the author(s) using Federal funds provided by the U.S. Department of Justice. Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

The Poisson rate test was used to compare the non-spatial GLM modeled predictions versus empirical heatmap for the full 1,300 and the subset of 130 shoes (where the subset contributed 39% of the RACs in the full dataset). The results are illustrated in Figure 5 where gray-colored points denote failure to detect a significant difference, while aqua-colored points indicate significance differences in RAC count ($p < 0.05$) which accounts for 31% of cell locations in the full dataset, and 21% of locations in the subset.

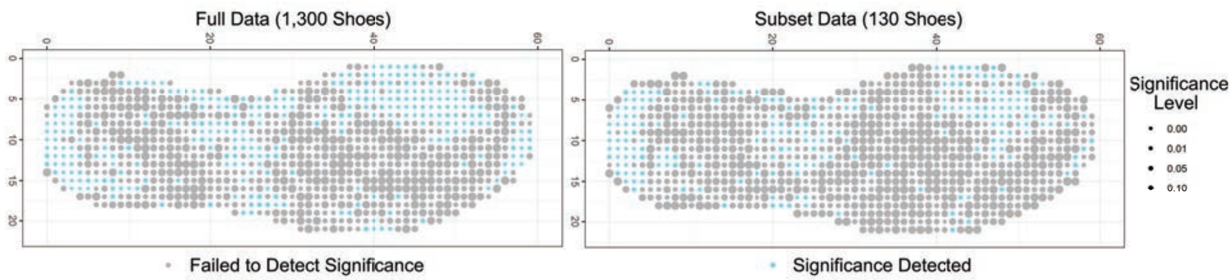

Figure 5: Spatial variation of the Poisson rate test p-value (significance level) for bins with significant differences (aqua) in RAC count for the empirical data versus a non-spatial GLM model prediction based on tread contact for the full data (31% significant) and the subset (21% significant).

Figure 6 plots the residuals (point size) and local Moran's I spatial autocorrelation (color) for the non-spatial GLM predictions for both the subset and full dataset as compared to the empirical data. Note that the results are visually consistent, and numerically, 40% of the bins in the subset (and 40% for the full dataset) were found to have significant correlation in residuals when comparing predictions to ground truth. Of the bins exhibiting correlation, the vast majority are positive (as indicated by the red points) and localized to the areas highlighted in Figure 4 (the medial ball of the toe, the arch, and the edge of the heel). Moreover, negative residuals (over-prediction of RAC frequency) are observed in the arch and heel (as indicated by the smaller point sizes), while positive residuals (under-prediction of RAC frequencies) are observed in the toe (as indicated by the larger point sizes).

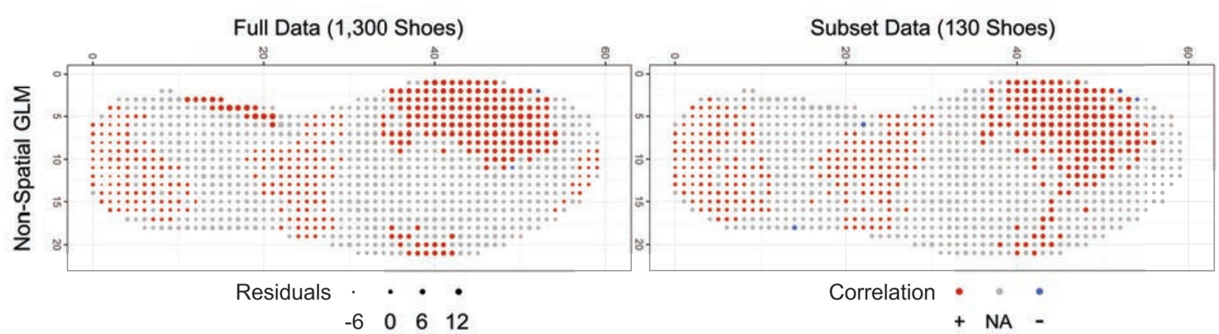

Figure 6: Spatial correlation (color, where positive = red, negative = blue and none = gray) and associated Pearson residuals (point size) for RAC counts predicated as a function of contact area and bin percentage offset using a non-spatial Poisson GLM with cut-off of 800px (33.3mm) for the full database versus the subset sampled for wear. Image reproduced from [4].

This resource was prepared by the author(s) using Federal funds provided by the U.S. Department of Justice. Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

3.3 Modeled Results: Tread Contact + Wear

In order to evaluate the hypothesis that both tread contact and wear can better explain RAC spatial distributions (as opposed to contact alone), the degree of wear in the 130 shoe subset was evaluated. Results indicate that the heel exhibits a greater degree of moderate-to-high wear, the arch exhibits mostly light-to-moderate wear, and the ball of the toe exhibits mostly moderate wear, with the lateral top edge predominately light-to-moderate and the medial edge moderate-to-high.

Next, the GLM predictors were updated to include contact-modified light, moderate and high wear. These inputs were determined on a per-cell basis as the average wear, obtained by dividing the total number of contact pixels containing each level of wear by the number of shoes. This same process was completed for each cell and each category of wear to obtain the set of predictors used in the non-spatial GLM. Table 4 reports the resulting model coefficients.

Table 4: Model coefficients for the non-spatial Poisson generalized linear model to predict RAC frequency for the partial dataset of 27,933 RACs on 130 outsoles.

Model	**Intercept (β_0)**	**Light Wear (β_1)**	**Moderate Wear (β_2)**	**High Wear (β_3)**	**Cell % Offset (β_4)**
Non-Spatial GLM	2.523	-2.4e-04	2.5e-04	-1.0e-04	1.162

Finally, the Poisson rate test was used to determine deviations. Figure 7 reveals that 16% of the bins (aqua-colored) exhibit empirical RAC counts that differ ($p < 0.05$) from counts expected based upon modeling via a non-spatial GLM using contact + wear as predictors.

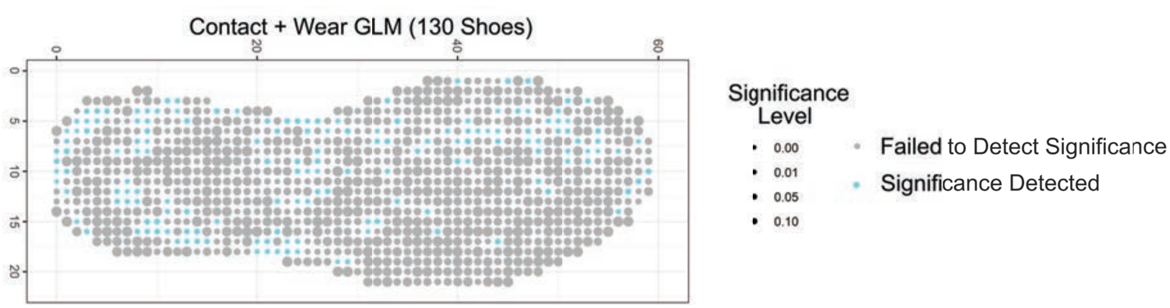

Figure 7: Spatial variation of the Poisson rate test p-value (significance level) for cells with significant differences (aqua) in RAC count for the empirical data versus a contact + wear based GLM for the subset of 130 shoes.

3.4 Discussion

Wear characteristics on shoe outsoles are acquired via interactions between the wearer, the activity, the terrain, and the material comprising the outsole. Although there is no evidence to suggest that one feature promotes or deters the acquisition of another, RAC distributions do not appear to be uniformly distributed across an outsole.

This resource was prepared by the author(s) using Federal funds provided by the U.S. Department of Justice. Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

If contact alone were sufficient to describe RAC frequency, then inhomogeneous Poisson simulations and GLMs modified by contact information would be expected to show reasonably high agreement with empirical data. Results indicate that the RAC frequency in approximately 60-70% of the spatial locations on an outsole can be reasonably well described under this assumption. However, RAC frequency over approximately 30-40% of the outsole is not well explained by tread contact alone, with underprediction of counts in the medial ball of the toe, and overprediction of counts in the instep and heel.

Given that tread contact alone does not account for interaction between the wearer and terrain, it is reasonable to suggest that RAC count in areas poorly modeled by a Poisson process modified by tread contact can be better explained by the combination of tread and localized wear. Under this assumption, results improved such that all but one-sixth of the cell locations on the outsole (16%) were well explained using a non-spatial GLM model when coefficients were informed by contact-localized wear. The cumulative results suggest that in the absence of confounding factors, contact-localized wear reasonably predicts much of the RAC frequency across the empirical dataset of shoes.

3.5 Limitations

The results presented here should be cautiously interpreted based on the following known limitations.

- No consideration was given to variations in outsole material type or chemical composition.
- The dataset does not include a single shoe make, model or size. RAC count was normalized to a single reference shoe in order to allow for comparisons (see [1] for additional details).
- Wear was coarsely (and subjectively) characterized as light, moderate and high, and a more detailed description of this feature may very well allow for improved predictions.
- It was challenging to identify an appropriate metric to evaluate statistical differences. Testing the equality of two Poisson rates using the W_5 metric, and/or the chi-square metric using the Poisson process as the expected, and the empirical data as the observed, are equally problematic. The total number of RACs (72,306) in the empirical heatmap are distributed among 987 cells, which suggests a level of dependency in the computations, despite the fact that each cell (empirical versus modeled or simulated) are being individually (independently) compared. In addition, no Bonferroni-type correction for multiple comparisons was conducted, nor any Yates-type correction for small counts which exist in some bins (approximately 3%).
- Confounding factors may exist, and if present, limit our ability to understand the physical predictors that lead to RAC count and therefore limit our understanding of RAC acquisition on outsoles.

This resource was prepared by the author(s) using Federal funds provided by the U.S. Department of Justice. Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

- The RAC spatial frequency examined in this dataset was extracted from high-quality exemplar impressions. This implies an extremely high level of transfer of RACs of all sizes, shapes, and geometries. In reality, many of the features present on an outsole may not transfer to crime scene impressions. An example are small pin-pricks or lines not well transferred/confirmed in a crime scene impression deposited in blood on a mottled surface such as a ceramic tile. Under these conditions, one could surmise that only large features are likely to be observed (or not obliterated by the interaction of media, substrate, and activity). It can also be implied that, at least sometimes, the larger the feature the more complex its geometry. However, the results presented here do not reflect the distribution of necessarily large or complex RACs, but all '*possible*' RACs and therefore may not mirror the distribution observed in casework impressions.

3.6 Conclusion

Wear characteristics on shoe outsoles are acquired via interactions between the wearer, the activity, the terrain, and the material comprising the outsole. Without inferring that one feature promotes or deters the acquisition of another, RAC distributions do not appear to be uniformly distributed across an outsole. Thus, a feature's value in determining a source association may be a function of not only its quality, clarity, and complexity, but also its physical location. By analogy, this is similar to pattern-force in fingerprint evaluation; minutiae near the delta develop due to pattern forces that lead to ridge endings in this region, and therefore may be more readily duplicated in non-mated pairs versus minutiae in other areas of the print. Finally, although there is no evidence to suggest that this observation would change in crime scene impressions, until empirically confirmed or refuted, the conclusions reported here should be used to inform understanding and prompt additional inquiries, rather than being used when interpreting casework.

This resource was prepared by the author(s) using Federal funds provided by the U.S. Department of Justice. Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

4. Dissemination: Publications & Presentations

Richetelli, N., Speir J.A. (June 2022). Spatial Frequency of Randomly Acquired Characteristics on Outsoles. Journal of Forensic Sciences. *Under Re-review - Minor Revisions.*

Speir, J.A, Richetelli N. (March 2022). Evaluating the Spatial Distribution of Randomly Acquired Characteristics on Outsoles. National Institute of Justice (NIJ) Forensic Science Research and Development (R&D) Symposium. Impression, Pattern & Trace Evidence Session.

Richetelli, N. (2020). Statistical Evaluation of Randomly Acquired Characteristics on Outsoles with Implications Regarding Chance Co-Occurrence and Spatial Randomness. West Virginia University, PhD Dissertation. Chapters 3-5, pp.148-169.

---

5. Collaborators

All work was conducted at West Virginia University and by West Virginia University students and faculty. PI-Speir and PhD candidate Nicole Richetelli were responsible for the majority of the work on this project. Additional contributions such as wear-map generation and data verification were provided by various students, including Tabitha DeBat, Nathan Weston and Swathi Murali. The primary collaborator information (at the time the research was conducted) are listed below.

Jacqueline A. Speir, Associate Professor
Principal Investigator
West Virginia University
208 Oglebay Hall, PO Box 6121
Morgantown, WV 26506-6121
P: 304.293.9233, F: 304.293.2663
E: Jacqueline.Speir@mail.wvu.edu

Nicole Richetelli, PhD Candidate
Graduate Research Assistant
West Virginia University
208 Oglebay Hall, PO Box 6121
Morgantown, WV 26506-6121
P: 304.293.4982
E: nrichete@mix.wvu.edu

This resource was prepared by the author(s) using Federal funds provided by the U.S. Department of Justice. Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

Bibliography

[1] Speir J, Richetelli N, Fagert M, Hite M, Bodziak WJ. Technical Note: Quantifying randomly acquired characteristics on outsoles in terms of shape and position. Forensic Science International. 2016;266:399-411.

[2] Baddeley A, Rubak E, Turner R. Spatial Point Patterns: Methodology and Applications with R. Apple Academic Press Inc.; 2015.

[3] Richetelli N, Bodziak WJ, Speir JA. Empirically observed and predicted estimates of chance association: Estimating the chance association of randomly acquired characteristics in footwear comparisons. Forensic Science International. 2019;302:109833.

[4] Richetelli N. Statistical Evaluation of Randomly Acquired Characteristics on Outsoles with Implications Regarding Chance Co-Occurrence and Spatial Randomness; 2020. West Virginia University, PhD Dissertation.

[5] Gu K, Ng HK, Tang ML, Schucany WR. Testing the ratio of two poisson rates. Biometrical Journal. 2008;50(2):283-98.

[6] Schabenberger O, Gotway C. Statistical Methods for Spatial Data Analysis. Boca Raton, FL: Chapman & Hall/CRC; 2005.

[7] Anselin L. Local Indicators of Spatial Association-LISA. Geographical Analysis. 2010;27(2):93-115.

This resource was prepared by the author(s) using Federal funds provided by the U.S. Department of Justice. Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.