SPATIAL POINT PROCESS MODELS FOR THE CLUSTERING BEHAVIOR OF NORTHERN PLAINS BOULDERS D. R. Thompson[1] and D. Wettergreen[1] {drt,dsw}@ri.cmu.edu [1]The Robotics Institute, Carnegie Mellon University, Pittsburgh PA 15213. USA.

Introduction: Spatial point processes describe the arrangement of discrete objects on the plane [1,2]. Here we use them to model clustering patterns for approximately 500,000 boulders in a HiRISE orbital image. Summary statistics suggest varying degrees of aggregation at different image locations. Nevertheless, the overall clustering behavior exhibits considerable structure and local continuity. The space of observed summary statistics has a low intrinsic dimensionality, and the observed clustering is well-described by a small handful of archetypes.

Image and Preprocessing: The HiRISE instrument aboard the Mars Reconnaissance Orbiter captures high-resolution visible-light images. Here we examine the orthorectified image TRA_000828_2495 containing 31,000x16,000 pixels at a resolution of about 0.31m/pixel [3]. It centers on a point in the Northern Plains at 69.3 degrees latitude, 130.2 degrees East longitude. This original image appears in figure 1. It shows a plain strewn with boulders up to several meters in size. We aim to model the spatial distribution of these boulders, their clustering behavior and the differences across image locations.

The large number of boulders makes comprehensive manual labeling impractical. Instead we identify rocks automatically. In the general case rock detection is a difficult image analysis problem [4]. Fortunately, the strong directed lighting in this image causes each rock to cast a dark shadow. One can identify these shadows by looking for contiguous blobs of pixels below a constant intensity threshold. We reject single-pixel shadows to reduce noise, and use the centroids of remaining shadows as a proxy for rocks' locations. This initial processing gives the image pixel locations of 498,582 rocks. Figure 2 shows a kernel-based estimate of the number of rocks per pixel at each location in the image.

Method: We use point processes to describe the spatial distribution. Point processes models can describe gaps and aggregation in spatial data. Previous research has used them to model geophysical features like rootless cone groups [5] and Venusian craters [6]. Reseearchers commonly characterize spatial point processes with diagnostic *summary statistics.* Møller and Waagepetersen provide extensive discussion [2].

Here we consider two summary statistics related to nearest-neighbor properties. The empirical nearest-neighbor function $G(r)$ gives the probability of an observed point's nearest neighbor appearing at any given distance r. It describes the degree of aggregation or regularity in the point process. The second statistic is the empirical empty space function $F(r)$. This gives the probability of a random empty location having a nearest neighbor at a given distance r. This characterizes the gaps between clusters.

We divide the image into a grid of 500x500 pixel subwindows and calculate summary statistics independently for each subwindow. A Kaplan-Meier estimator [2] corrects for edge effects. The result associates each subwindow with two curves representing a particular instantiation of the summary functions. These curves define two vector-valued attributes totaling over 140 dimensions.

Results: The eigenvalues of the two statistics' covariance matrices indicate the relative importance of each principal component. In this case the variance between subwindows is nearly unidimensional; for each statistic the first covariance matrix eigenvalue is larger than the second by an order of magnitude. Thus we can describe the curves using their first principal components while preserving the vast majority of inter-subwindow variance. After projecting the summary statistics and normalizing for contrast we

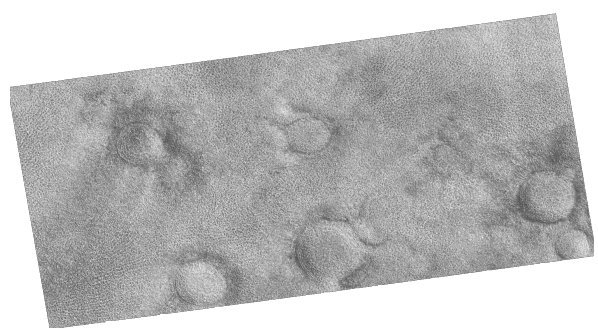

Figure 1: Original Image

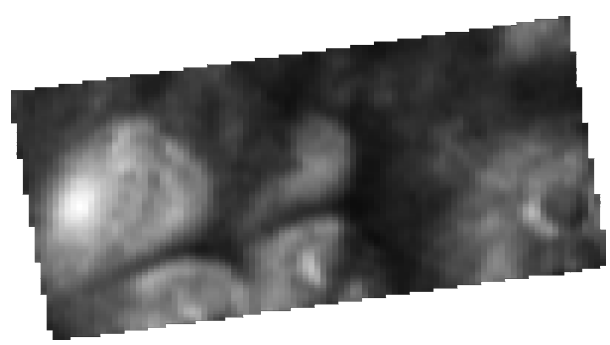

Figure 2: Empirical density of boulders

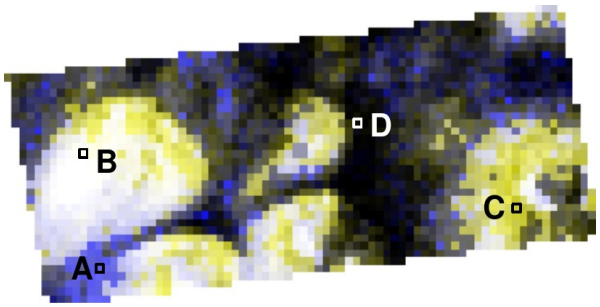

Figure 3: Visualization of empirical F and G functions

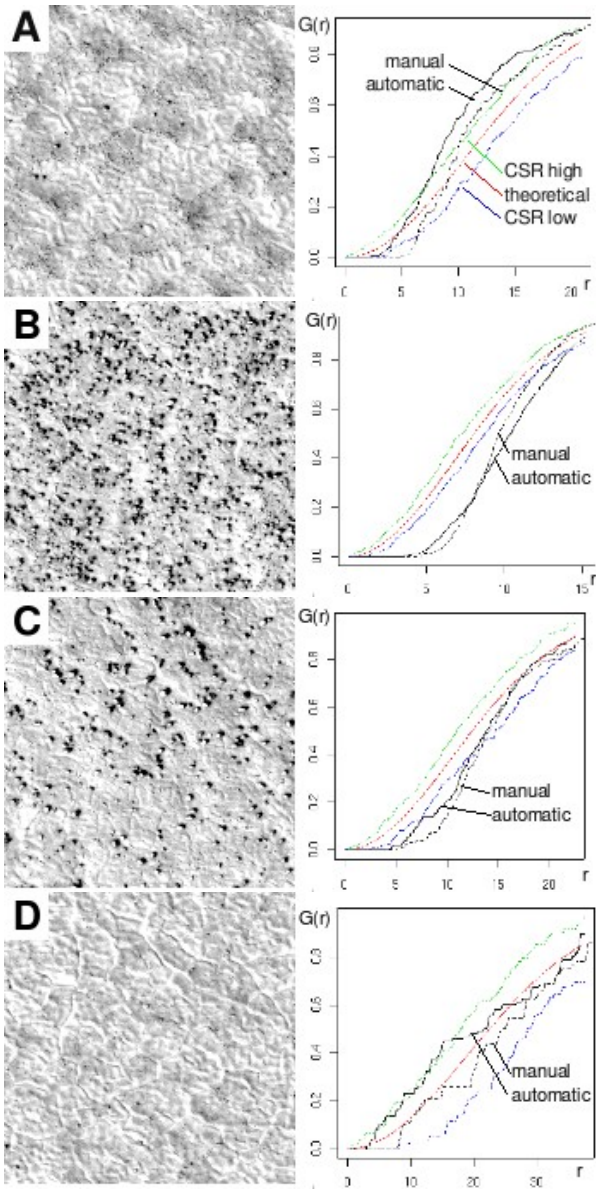

Figure 4: G function estimates for four subwindows. These plots use different horizontal scales for distance r to compensate for rock densities in each subwindow. Nearest-neighbors closer than five pixels are rare due to finite rock diameters.

map each to a color intensity: yellow for the *G* function and blue for the *F* function. This produces a color representation of the boulders' clustering patterns as they vary across the image (Figure 3).

We study four archetypal subwindows labeled A, B, C, and D. Figure 4 shows a larger view of each subwindow. The associated plots show the observed *G* functions alongside the theoretical *G* function generated by complete spatial randomness (CSR). The complete spatial randomness curve is diagnostic of a process with no underlying structure. Subwindow A has a *G* function value above this curve, suggesting aggregation relative to a random process. On the other hand, subwindows B and C show atypical regularity as evidenced by *G* function values below the theoretical curve for complete spatial randomness.

The plots of figure 4 also indicate lower and upper envelopes for 100 actual simulated trials of a completely random process. The empirical G functions for subwindows A, B, and C fall outside this envelope, motivating us to reject complete randomness as a model for these patterns. We cannot reject randomness in subwindow D based on its *G* function behavior alone. However, a more powerful test might still reject CSR on other grounds.

The plots in figure 4 also show the *G* function estimate resulting from a hand-labeling of the rocks in each subwindow. These manual analyses follow the general behavior of the automated estimates; the largest errors occurs in subwindows A and B that contain small rocks with weak shadows.

Discussion: This work shows how diagnostic techniques for spatial point processes can reveal hidden structure in the distribution of features from high-resolution orbital data. Automatic feature detection permits efficient analysis of hundreds of thousands of rocks spread over many square kilometers. A more complete analysis might consider boulders' size or a broader range of summary statistics.

Acknowledgments: We used Spatstat, an R package for spatial statistics [7]. This research was supported by NASA grant NNG0-4GB66G.

References: [1] Cressie, N. (1991) Statistics for Spatial Data. [2] Møller, J. and Waagepetersen, R. P. (2004) Statistical Inference and Simulation for Spatial Point Processes. [3] Thompson, D. R. and Castaño, R. *IEEE Aerospace* 2007. [4] [5] Bruno, B.C. et al., (2004) *JGR* 109: E07009. [6] Hauck et al. (1998), *JGR* 103:E6 p. 13,635-42. [7] Bradley, A. and Turner, R. (2005) *Journal of Statistical Software 12.*