Utilizing Spatial Statistics In Crater Studies

J.D. Riggs, M.R. Kirchoff, S.J. Robbins

7^{th} Planetary Crater Consortium

August 17 - 19, 2016

Presentation Outline

- Introduction
- Impacts
- Secondary Impacts
- CSFD Method
- NND Method
- Grouping Methods
- Spatial Correlation Methods
- Optimal areal sampling
- Conclusions

Introduction

- Cratering process assumptions
  1. Craters form stochastically through time
  2. Craters form randomly (mostly) across a surface
  3. Impact cratering is an ongoing process
- Two cratering processes tied to stochastic assumptions
  1. Secondary cratering
  2. Crater saturation
- Spatial statistics to identify secondaries and saturation
  1. Familiar methods
  2. Spatial correlation methods

Context

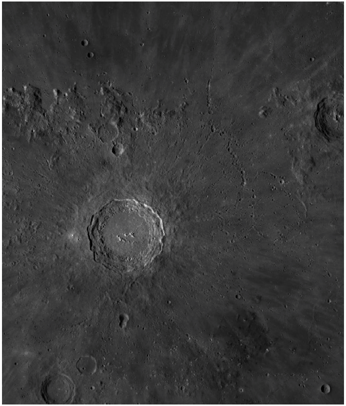
- Secondary cratering
  - Secondaries exhibit non-random patterns
    assumption 2 violation
  - Task is identification of secondaries from spatial statistics
- Crater saturation
  - New craters cannot form without erasing some number of older craters
  - Crater spatial population distributions static in time and space
    assumption 3 violation

Impacts

- Primary impacts
  - Complete spatial random (CSR) distributions (some deviation)
  - CSR follows a Poisson probability distribution function (pdf)
  - Poisson pdf is counts by area (univariate spatial density)
- Small crater impacts include
  - Primaries
  - Secondaries
- Primaries secondaries considered CSR
- Secondaries can cluster or appear as CSR

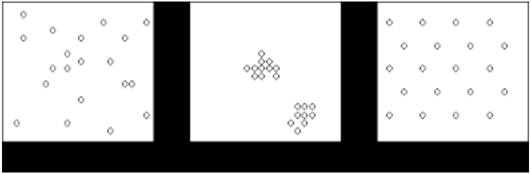
Secondary Impacts: Morphology

- Near relatively young primaries, secondaries tend to be unambiguous
- As craters age, its secondary field may not be as obvious.
- Far from primaries, secondaries more challenging to identify



Spatial Distributions

- Left panel: CSR
- Middle panel: clusters
- Right panel: regular (uniform)



CSFD Method

- Univariate spatial density comparisons among 2-D region subdivisions
- Sub-region differences suggest secondary contamination
- Similarities suggest no contamination or CSR ambiguity
- Steeper CSFD slopes (than primary SFDs) suggest secondaries

CSFD Method: Issues

- Dependent on region characteristics
- Sensitive to CSFD uncertainties
- Poor specificity (true positive rate) and selectivity (true negative rate)
- Spatial distribution identification challenging
- Spatial correlation ignored
- CSFD slope comparisons may improve using KDE and MLE

Distance Method

- Nearest Neighbor Distance (NND)
- Uses location (lat/lon), independent of region shape
- Under CSR, counts pdf transforms to univariate distance pdf

$$1 - \mathcal{P}(N = 0) = 1 - \frac{\rho\pi r^2}{0!} e^{-\rho\pi r^2}, \text{ NND} < r$$

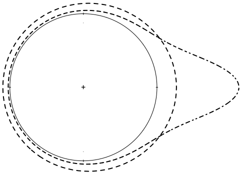
- Distances are Z-scores so standard deviations from CSR
  - $Z < 0 \Rightarrow$ clustering
  - $Z = 0 \Rightarrow$ CSR
  - $Z > 0 \Rightarrow$ uniformity
  - Location uncertainty quantifiable

Distance Method: Issues

- Spatial correlation information lost
- Only smallest scale information, no information on all scales
- Corrections often needed to accommodate crater diameters
- Only magnitude and direction of deviations from CSR, no coordinate-based variability among neighbors
- Affected by boundaries, though mitigation possible

Cluster Analysis Methods for Secondary Identification

- Known primary
  - Scan methods with primary as center
  - Jones-Pewsey circle distribution for scan segments
  - GAM, Kuldorff's, and Stone's tests for crater clustering
  - Best fit Jones-Pewsey distribution ties primary to clusters
- Unknown primary
  - Model-based cluster method
  - Clusters represent direction and distance vectors
  - Best Jones-Pewsey distribution fit to identify candidate primaries

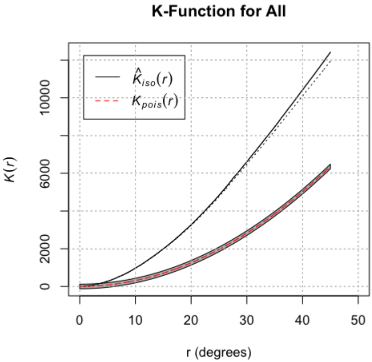


Spatial Point Methods

- Saturation and secondary identification
- Intensity function measure of counts / unit area: $\lambda = \frac{N}{|A|}$, (counts, N, for normalized area, $\mid A \mid$)
- Implementations using λ (specified) or $\hat{\lambda}$ (estimated)
  - Ripley's K-function
  - Besag's L-function
  - Two point correlation function
- These methods identify
  - CSR: primaries
  - Regularity: saturation
  - Clustering: secondaries

Ripley's K-Function: $K(r) \equiv \lambda^{-1}\mathcal{E}$ (# nonrandom events $0 \leqslant r \leqslant h$)

- Captures spatial areal dependence (correlation)
- Red dashed curve is CSR with 0.05 uncertainty region $K(r) = \pi r^2, \; r \geqslant 0, \; r =$ distance
- Black solid curve shows clustering at scales $> 5^\circ$ $K(r) > \pi r^2$ (dashed curve has boundary correction)
- If black curve below CSR region, distribution regular $K(r) < \pi r^2$

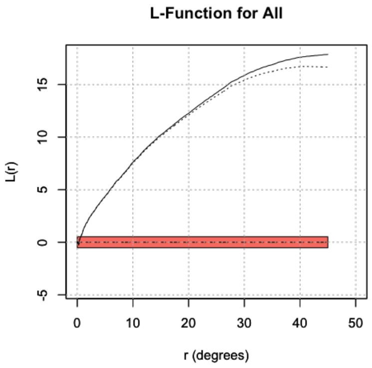


Ripley's K-Function

- K independent of region shape
- K has region boundary accountability
- Information exists on all scales
- Accounts for lat/lon uncertainties
- Accounts for spatial correlation

Besag's L-Function: $L(r) = \sqrt{K(r)/\pi} - r, r \geqslant 0$

- Function of Ripley's K, has same information
- Plot easier to interpret than K-function
- Zero-line and uncertainty represents CSR
- Excursions above zero-line imply clustering $> 5^\circ$
- Excursions below zero-line imply dispersion
- Constant L-value below zero-line implies saturation



Besag's L-Function: $L^*_{12}(r; n_1, n_2)$

- Adapted to address clustering interactions
  - Two or more populations, e.g., same-sized craters of primaries and secondaries
  - May be used with GAM, Kuldorff's, and Stone's tests

- Counts of 1 population assessed around counts of another population
  - > 0 segregated
  - < 0 aggregated
  - $= 0$ no interaction

- Can help identify types in mixes of same-size primaries and secondaries

Spatial Statistics for Saturation Identification

- NND (small scale only)
- Ripley's K-function scale dependent distance below CSR curve
- Besag's L-function scale dependent distance below the zero-line
- Spatial correlation (variance-covariance) matrix has constant correlation for all location pairings
- Lattice spatial methods
  - Independent of lat/lon
  - Utilizes 2-D NND techniques that include spatial correlation
  - Untried? for saturation (or secondary) crater identification

Spatial Sampling

- Optimal Sample Region Size
  - Spatial variance-covariance (vcov) structure (scale dependent)
  - Features under study
  - Study purpose
- vcov includes location (lat/lon) and markers (e.g, mark = craters with central peaks)
- Population vcov estimated from
  - Catalogs
  - Production function
  - Simulation
- Sample vcov $\stackrel{?}{=}$ population vcov
  - Canonical correlation
  - $\Sigma_{sple}\Sigma_{pop}^{-1} = I$ (if nonsingular)

Sampling Methods for Spatial Analysis

- For markers
  - Stratified systematic most efficient (e.g., random pages in phone book, every 10*th* name on page
  - Stratified with line transect with decreasing exponential detection function (e.g., random page in phone book, draw diagonal line on page, every last name crossing line)
- For saturation and secondary identification
  - Adaptive cluster strip sampling (e.g., search region until marker found, search surrounding area (Jones-Pewsey)
  - Adaptive cluster systematic sampling (e.g,, as previous but with checkerboard)
- Must have homogeneous vcov within strata
- Population variances added as strata sampled independently
- Post-stratification inflates estimates, but calculable

Limited (Flyby) Strip Samples

- vcov may be unknown
- Production function may be unknown
- Select regions with largest variability if possible
- Capture-Recapture sampling is two independent samples of the same region. Features of interest should have the same proportion in the two samples (e.g., 2 independent researchers)
- Multistage sampling is selecting primary regions, then selecting sub-regions within the primaries (within an area, randomly select disjoint regions, then sample sub-regions of regions)

Conclusions

- Need studies of sims and NND to compare with proposed spatial methods
- Currently applying spatial methods to Mimas region thought to be saturated
- Need compare sims and NND with spatial methods for secondaries
- Spatial-temporal methods may be helpful for geologically active areas
- Other potential issues to address with spatial statistics?