UNDERSTANDING SPATIAL STATISTICS FOR PURPOSES OF IDENTIFYING NON-PRIMARY AND SATURATED IMPACT CRATER POPULATIONS. J.D. Riggs[1], S.J. Robbins[2], M.R. Kirchoff[2], E.B. Bierhaus[3], B.P. Weaver[4]. [1]Northwestern University; [2]Southwest Research Institute, 1050 Walnut Street, Suite 300, Boulder, CO 80302; [3]Lockheed Martin Space Systems Company, PO Box 179, Mail Stop S8110, Denver, CO 80201; [3]Statistical Sciences, CCS-6, Los Alamos National Laboratory. jamie.riggs@northwestern.edu

Introduction and Background: Three of the key assumptions that drive most interpretations of crater populations are: (1) Craters form stochastically around a derivable time-dependent function, (2) craters form randomly across a surface, and (3) the population has not reached equilibrium. Two main types of process undermine these assumptions: secondary cratering and crater saturation. Secondary cratering occurs when the ejecta blocks launched from a primary impact event return to strike the planetary surface with enough energy to create their own craters ("secondary craters" because the event is secondary to the initial, or "primary," impact event). These occur in a geologic instant and are not distributed randomly across the surface, in direct conflict with the first two assumptions. Crater saturation occurs when so many craters have formed that no new craters can form without an equal fraction of old ones being erased [*e.g.*, 1]. This results in a crater population with a distribution that does not change its characteristics in time and space, in direct conflict with the third assumption.

There are significant ongoing discussions as to the magnitude of the effect of secondary cratering and crater saturation on these three important assumptions, though that discussion is a separate issue from this abstract. In this abstract, we outline the spatial statistic approaches used to identify these two different processes affecting populations of craters and in our talk, we will discuss potential improvements on these and the use of other techniques.

Secondary Crater Identification from Spatial Statistics: There are numerous morphologic techniques to identify secondary impact craters [*e.g.*, 2-6], but they are not entirely reliable, for secondaries will often look like primaries and hence cannot be distinguished based on morphologic criteria alone. Therefore, statistical methods are also frequently employed.

There are three methods typically used in crater studies to examine the likelihood of contamination by non-morphologically-obvious secondary craters, some more rigorous than others. First is a simple spatial density comparison: A region that is thought to represent a single, random population is sub-divided into 2 or more parts and the crater populations are compared. If the size-frequency distributions (SFDs) of each are similar to within the specified uncertainties, they are assumed to, at worst, have the same level of secondary crater contamination and, at best, no contamination. This is simple to implement but is poor in specificity.

The second method is to examine the crater SFD structure itself. Secondary craters typically have a steeper SFD than primary craters; while primaries often have slopes of –2 to –3, secondaries can have slopes up to –8 [*e.g.*, 4-5]. If the SFD of the measured population deflects to greater than a model production function, then it is likely to be contaminated by unidentified secondary craters.

The third method is the *Z*-statistic, which is within the class of distance measurement for nearest neighbors statistics ("NND"). This can be computed overall for the region being studied or for sub-regions. The *Z*-statistic is the number of standard deviations from a Poisson distribution (which primaries should follow) due to random impacts. Because secondary craters tend to form more clustered than random, and the *Z*-statistic indicates if a studied distribution is more clustered, this method potentially finds crater populations that are affected by secondaries. This has been used to examine crater spatial distributions looking for secondaries in the past [*e.g.*, 7]. The specific interpretation of the *Z*-statistic and to what certainty the null hypothesis (that the craters are spatially random) is not rejected is subject to variation amongst individual researchers.

Saturated Crater Population Identification: Again, multiple techniques have been developed to potentially determine if crater distributions are saturated and in equilibrium. One technique ascertains if the spatial density of the crater distributions has reached a proposed maximum density attainable by crater populations before they become saturated [1,8]. Another technique examines the crater SFD slopes, as some populations attain a cumulative SFD slope of –2 when they reach equilibrium, effectively shallowing [1]. The one we detail here uses the *Z*-statistic described above. In this case, however, the population is expected to be more uniform than random [9-11]. Therefore, the *Z*-statistic is used to look for dense cratered distributions that are also more uniform than random (as opposed to more clustered than random). Lissauer and Squyres *et al.* [9,10] have used this technique to show that the dense cratered terrains of Callisto and Rhea are likely saturated. Kirchoff (this volume) has explored more terrains to show that densely cratered surfaces of many inner and outer solar system objects are likely saturated.

Problems with These Approaches and Introduction of Suggested Alternatives: The above-described identification methods have advantages and issues. In general, they are commonly used, easy to calculate, and have some error structure that may be utilized.

Some advantages of the SFD comparison and analysis are that it is in common use and therefor reasonably well understood, there is (generally) consistent agreement on slope interpretation, and it is based on independent model crater populations. Some difficulties in comparing these shapes is that it is inherently a comparative technique that can only give "as good as" or "worse than" comparisons, and that it can give a false negative when secondary crater contamination is significantly smaller than the overall population.

Advantages of NND are: It is conceptually straightforward, is not dependent on arbitrary region sizes as it uses precise mapping locations, and it is easily computed. However, NND loses large-scale information for it is dependent on nearest neighbor proximity, individual event information is lost, it gives only the direction of the deviation from complete spatial randomness (CSR), and the statistical properties are not well understood with large departures from CSR.

Additional drawbacks to using these methods are: Loss of information as a result of reducing at least two-dimensional mapping to a one-dimensional summarization, the error structures are dependent on researcher bias and method implementation such as area determination, and these methods are susceptible to fixed scales over which analyses are conducted and thus lose variable scale information.

Two spatial point process statistical methods are presented here for identification of non-primary or saturated crater populations; these account for the issues when using one-dimensional measures. They are the Two Point Correlation Function (TPCF) and Ripley's K function. Both of these spatial methods operate on as many spatial dimensions as are required to manage the research question of interest, they each have specific error structures that allow for precise error assignment, and the two methods operate over a scale range suitable for the study objectives.

Two Point Correlation Function: The TPCF was introduced by [12 and 13] to describe galaxy clustering. The technique counts the number of potential non-primaries or saturated craters in a series of annuli around a selected point (*e.g.*, the primary, a possibly saturated surface, or secondary cluster), or it counts the features as part of the background. This is unlike the NND which analyzes the statistics around each specific crater. The TPCF derives from the joint probability that two secondaries, for example, lie in infinitesimally small area annuli around the two vector locations of the two secondaries; then, the TPCF is a function of the vectorial distance between these locations. Thus, given a putative secondary crater location, the TPCF is a function of the probability of finding, at a specified distance, another secondary. The larger the value of the TPCF, the more clustered and hence nonrandom are the secondaries at the specified distance. Secondaries' clustering can also be used to suggest the originating primary when that primary is the center of the annuli. Note that the annulus size must be predetermined, which is a disadvantage. However, the TPCF is a function of the differential of Ripley's K function, which does not suffer this difficulty.

Ripley's K and Related Functions: Bartlett [14] first proposed a second-order (spatial) correlation function which Ripley [15,16] developed into a widely-used spatial point statistic that captures the spatial dependence between different regions of a point process, such as mapped locations of impact craters. Bierhaus (2004) [7] used K-functions as a method to identify clustering in Europa's small-crater population. Others have offered transformations of Ripley's K function. Besag's L* transformation [17] is one that will be described here.

Ripley's K function is defined to be the expected number of non-primary craters (more generally, any designated event) within a specified distance of some additionally selected other crater, weighted by the region crater density (the intensity function). Under CSR, the K function value is the area of a circular region with the specified distance as the radius. Values of the K function larger than this circle's area suggest clustering of events, and K function values less than this area suggest uniform distribution of events (regularity).

Advantages of Ripley's K function include independence from region shape, corrections for region boundary biases, retention of spatial information on crater distributions at all scales of interest, and use of precise spatial locations of the events in the estimations. A disadvantage is the Ripley's K function is that it is not trivial to interpret. However, Besag's L* function transformation produces a plot that is intuitive, and hence interpretation is uncomplicated. We will demonstrate this at the May workshop.

References: [1] Gault, D.E. (1970). *Radio Sci.*, doi: 10.1029/RS005i002p00273. [2] Shoemaker, E.M. (1962). in *Physics & Astronomy of the Moon*. [3] Oberbeck, V.R. & R.H. Morrison (1974). *Moon*, doi: 10.1007/BF00562581. [4] McEwen, A.S. & E.B. Bierhaus (2006). *Annu. Rev. Earth Planet. Sci.*, doi: 10.1146/annurev.earth.34.031405.125018. [5] Robbins, S.J. & B.M. Hynek (2011). *JGR*, doi: 10.1029/2011JE003820. [6] Robbins, S.J. & B.M. Hynek (2014). *EPSL*, doi: 10.1016/j.epsl.2014.05.005. [7] Bierhaus, E.B. (2004). Ph.D. Thesis. [8] Hartmann, W.K. & R.W. Gaskell (1997). *MAPS*, doi: 10.1111/j.1945-5100.1997.tb01246.x. [9] Lissauer, J.J. *et al.* (1988). *JGR*, doi: 10.1029/JB093iB11p13776. [10] Squyres, S.W. *et al*. (1997). *Icarus*, doi: 10.1006/icar.1996.5560. [11] Kirchoff, M.R. (2013). *PCC*, #1311. [12] Totsuji, H. & Kihara, T. (1969). *Pub. Astron. Soc. Japan*. [13] Peebles, P.J.E. (1973). *ApJ*. [14] Bartlett, M.S (1964). *Biometrika*. [15] Ripley, B.D. (1976). *J. App. Prob*. [16] Ripley, B.D. (1977). *J. of Royal Stat. Soc*. [17] Besag, J.E. (1977). *J. of Royal Stat. Soc*.