UTILIZING SPATIAL STATISTICS IN CRATER STUDIES. J.D. Riggs[1] , M.R. Kirchoff[2], S.J. Robbins[2]. [1]Northwestern University; [2]Southwest Research Institute; jamie.riggs@northwestern.edu, kirchoff@boulder.swri.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. Spatial statistics have been used to analyze two processes that undermine these assumptions [e.g., 1, 2]: 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) [e.g., 3]. 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., 4]. 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. Here, we outline spatial statistic approaches used to potentially identify these two processes.

Secondary Crater Identification: There are numerous morphologic techniques to identify secondary impact craters [e.g., 5], but they are not entirely reliable, for secondaries will often look like primaries and hence cannot be distinguished based on morphologic criteria alone. Therefore, a variety of statistical methods are also frequently employed [e.g., 5], such as comparing the spatial density of craters and/or the crater size-frequency distributions to a model production function for primaries. In this abstract we focus on how spatial statistics can help.

In most work to date the Z-statistic, which is within the class of distance measurement for nearest neighbors statistics ("NND"), is the spatial statistic that has been used [e.g., 1]. This can be computed overall for the region being studied or for subregions. 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. 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.

The Z-statistic has some advantages in that 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 this method 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 this method is susceptible to fixed scales over which analyses are conducted and thus lose variable scale information.

The formation of secondaries in clusters and in an annular pattern around a primary suggests three particular spatial statitical analyses that could work better than the Z-statistic: two point correlation function (TPCF), Ripley's K (and related) functions, and circular statistics using Jones-Pewsey probability distributions. The TPCF was introduced by [6, 7] to describe galaxy clustering. The technique counts the number of potential non-primaries in a series of annuli around a selected point (e.g., the primary 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.

Bartlett [8] first proposed a second-order (spatial) correlation function which Ripley [9, 10] 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. 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. 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 that the Ripley's K function is not trivial to interpret. However, Besag's L* function transformation [11] produces a plot that is intuitive, and hence interpretation is more straightforward. Bierhaus [1] used K-functions as a method to identify clustering in Europa's small-crater population.

The circular Jones-Pewsey [12] family of probability distribution functions can be used to form scan statistics [13, 14] once the Ripley's K function has indicated a clustered, circular region around a primary crater. For this there are two statistical approaches to using spatial methods with secondary crater identification in conjunction with morphological considerations. The first set of statistical methods assumes the primary crater has a known location. The second assumes secondaries are identified, at least in part, but the primary location is not known. In the case for the known primary coordinates, the identification task is locating the associated secondaries. The scan method uses the circular probability distribution function to define a search window, and a local test for clustering is then employed. Commonly used tests for localized clustering are Openshaw's [15] Geographical Analysis Machine, Kulldorff's [16] statistic, and Stone's [17] test. Upon identifying the secondaries, a circular distribution from the Jones-Pewsey family then describes the spatial distribution.

The second set of statistical methods, used when secondaries are identified without knowledge of the location of the primary, use a combination of a spatial variance-covariance matrix and a model-based cluster analysis. Model-based cluster methods [18-20] assume the population of secondaries are from n subpopulations, each corresponding to a cluster associated with N primary craters, and each secondary distribution follows a probability density function, such as one from the Jones-Pewsey family of circular distributions. The model is parameterized with a direction and distance vector, and the resulting clusters are identified with these vectors. The vectors then may be used to find intersections and the potential primary craters.

Saturated Crater Population Identification: Again, multiple techniques have been developed to potentially determine if crater distributions are in saturation equilibrium. One technique ascertains if the density of a crater distribution has reached a proposed maximum attainable by crater populations at which they become saturated [e.g., 4, 21]. Another technique examines the crater SFD slopes, as some populations attain a cumulative SFD slope of –2 when they reach equilibrium, effectively shallowing [e.g., 4, 22].

Spatial statistical studies of crater saturation have so far only utilized the Z-statistic described above [2, 23, 24]. In this case, however, the crater population is expected to be more uniform than random. Lissauer [2] and Squyres et al. [23] have used the Z-statistic to show that the dense cratered terrains of Callisto and Rhea are likely saturated. Kirchoff [24] has explored more terrains to show that densely cratered surfaces of many inner and outer solar system objects are likely saturated. However, because the Z-statistic has the issues discussed above, TPCF and Ripley's K (and related) functions could be better tools to assess uniformity (regularity) and study crater saturation.

Conclusion: In general, spatial statistics can be helpful in analyzing secondary or saturated crater populations. TPCF, Ripley's K (and related) functions, and circular statistics using Jones-Pewsey probability distributions can provide more robust spatial methods than the traditionally used Z-statistic. They 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 these methods operate over a scale range suitable for the study objectives. Due to these advantages, we will discuss if these techniques can be applied to other issues in crater studies.

References: [1] Bierhaus, E.B. (2004). Ph.D. Thesis. [2] Lissauer, J.J. et al. (1988). JGR. [3] Shoemaker, E.M. (1962). in Physics & Astronomy of the Moon. [4] Gault, D.E. (1970). Radio Sci. [5] McEwen, A.S. & E.B. Bierhaus (2006). Annu. Rev. Earth Planet. Sci. [6] Totsuji, H. & Kihara, T. (1969). Pub. Astron. Soc. Japan. [7] Peebles, P.J.E. (1973). ApJ. [8] Bartlett, M.S (1964). Biometrika. [9] Ripley, B.D. (1976). J. App. Prob. [10] Ripley, B.D. (1977). J. of Royal Stat. Soc. [11] Besag, J.E. (1977). J. of Royal Stat. Soc. [12] Jones, M.C. & Pewsey, A. (2005) J. Am. Statist. Assoc. [13] Hjalmars, U. et al., (1996) Statist. Medic. [14] Lawson, A. et al. (2006) Statist. Medic. [15] Openshaw, S. et al. (1987) Inter. J. Geo. Inform. Sys. [16] Kulldorff, M. & Nagarwalla, N. (1995) Statist. Medic. [17] Stone, R.A. (1988) Statis. Medic. [18] Scott, A.J. & Symons, M.J. (1971) Biometrics. [19] Banfield, J. D. & Raftery, A.E. (1993) Biometrics. [20] Fraley, C. & Raftery, A.E. (2002) J. Am. Statist. Assoc. [21] Hartmann, W.K. & R.W. Gaskell (1997). MAPS. [22] Chapman, C.R. & McKinnon W.B. (1986) Satellites. [23] Squyres, S.W. et al. (1997). Icarus. [24] Kirchoff, M.R. (2016) MAPS.