DETERMINING STATISTICALLY SIGNIFICANT DEVIATIONS FROM A MODEL CRATER PRODUCTION FUNCTION FOR ESTIMATING RESURFACING EVENTS. B.P. Weaver[1], J.M. Hilbe[2], S.J. Robbins[3], C.S. Plesko[4], J.D. Riggs[5]. [1]Statistical Sciences, CCS-6, Los Alamos National Laboratory; [2]Arizona State University, Tempe, AZ 85287; [3]Southwest Research Institute, 1050 Walnut Street, Suite 300, Boulder, CO 80302; [4]Applied Physics, Theoretical Design, XTD-NTA, Los Alamos National Laboratory; [5]Northwestern University. theguz@lanl.gov

Introduction: Crater population data are often compared with model crater populations ("production functions" or "PF") to determine differences. These differences from the model PFs are then used to assume that an event other than stochastic impacts. The most common assumption is that a resurfacing event took place (if the crater population deflects to less than the PF), an unexcluded population of secondary impact craters is present (if the crater population deflects to greater than the PF), or the population has reached an equilibrium where no new craters can form without an equal area of other craters being removed. At issue for purposes of this abstract is at what point a deviation from a model PF is considered significant enough to be interpreted – and with what confidence – as a deviation event. A practical application of this problem is illustrated in Fig. **1**. Whether the model PF is "correct" and differences between published PFs is a separate issue beyond the scope of this investigation.

Complicating Factor— Researcher Variation and Bias: It is often assumed – typically because of simplicity or practicality – that when an individual crater analyst identifies and measures a crater population, that measured population is "the answer." In other words, it is a repeatable measurement, and if the investigator is honest, they will consider it to be repeatable to at least within the standard-assigned Poisson $N^{1/2}$ counting uncertainty (however, even this is not always the case, for there are examples in the literature where a researcher will ignore even these uncertainties in lieu of considering every small deflection to be a geophysical event).

Other researchers have recognized that this is not the case, that crater measurements are usually objectively repeatable, but only as an ensemble: *I.e.*, there will *always* be variation, not only between different crater analysts, but even from one hour to the next for a single individual. Among spacecraft mission teams, and within research groups, certain calibrations are often run internally to verify that individuals are qualitatively "on the same page." When variations are found, they are noted, perhaps recorded, but then individuals will often go on to do their own crater counts and ignore this variation in most analyses and conclusions.

Robbins *et al*. [1] published the first-ever analysis of multiple crater analysts with a range of experience who independently, with their own tools and techniques, identified impact craters on the same two lunar images. (They also examined how lay, minimally trained individuals varied among each other and from professionals, but that discussion is not relevant to this abstract.) They had ten enumerated conclusions, but the most important to this discussion were that they found the *number* of craters found by the different researchers for the same images varied up to a factor of ~2 (min to max) with a 1σ envelope of 15–45% depending on terrain type and crater morphology. While the absolute numbers varied considerably, the *population* (*i.e.*, the "shape" of the crater size-frequency distribution) of craters found by the different researchers was relatively consistent (determined by a Kolmogorov-Smirnov test). However, they emphasized that there were numerous small deviations over narrow diameter ranges between each researcher that, individually, could have been interpreted as a resurfacing event. [Note: This work is also being submitted as Robbins *et al*., this volume, and will be discussed at the May workshop.]

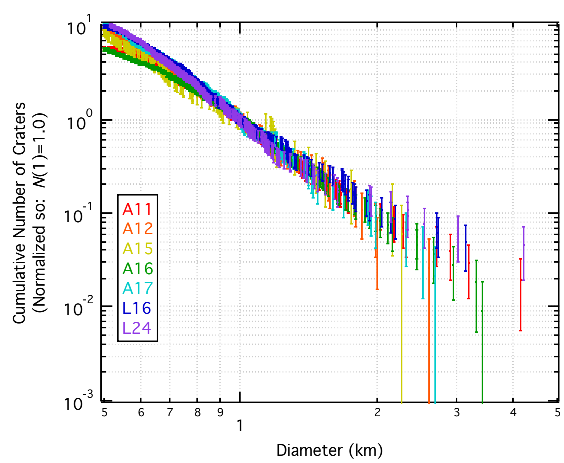

Figure 1: Example set of cumulative crater size-frequency distributions, from [2]. These show the data for seven lunar sampling sites, all normalized such that $N(1) = 1$. Uncertainty is $N^{1/2}$ of the cumulative counts. While these populations overlap within their collective uncertainties, *Apollo 15* literally sticks out at $D \approx 1.2$ km, with a value nearly twice as large as the mean. The question therefore is: Is this a statistically significant deviation? Is it likely meaningful, indicating a process has acted to enhance those craters, or is it within the expected variation? Similarly, at small diameters, the deviations are likely due to incomplete counts, but if one were to assume that the *Luna 16* and *24* site counts were complete for $D \geq 0.5$ km, then one could ask at what diameter do the other crater populations – such as *Apollo 16* – deviate from that population, and to what confidence? That deviation diameter could then be used to infer a process has

acted to remove all craters shallower than the corresponding depth of those craters, *e.g.*, 70 m of surface removal.

Discussion at Workshop: For the statistically significant deviations problem, we will review the current literature in statistical model assessment. This is a broad area of statistics that seeks to understand what features of the observed data are not adequately explained by the model. The procedures we discuss utilize Monte Carlo simulations to also quantify and visualize the variability specified by the model. In particular, we will be discussing the broad tool of "bootstrapping" as a potential solution.

In addition, we will look into the issue of researcher variation and bias. For example, there are statistical methods that account for sources of variability, such as effects induced by observers or measurement devices. In industrial engineering, much research has been done on what is called "Gauge R&R" studies which seek to understand what are all "repeatable" (in some sense) sources of variability or error and how much of the total variability of a series of measurements can be explained by these repeatable sources of error. (See [3] for more on an overview of Gauge R&R studies.) Modeling variability may be a necessary input to being able to place realistic confidence levels on when statistically significant deviations in the data are present, relative to a model production function.

References: [1] Robbins, S.J. *et al*. (2014). *Icarus*, doi: 10.1016/j.icarus.2014.02.022. [2] Robbins, S.J. (2014). *EPSL*, doi: 10.1016/j.epsl.2014.06.038. [3] Weaver, B.P. *et al*. (2012). *Quality Engineering*.