LUNAR CRATER POPULATION STATISTICS AND RELATED ACCURACIES FROM AN ELEVATION-BASED IMPACT CRATER DETECTION TECHNIQUE. I. M. Pritchard[1], Jinfei Wang[1], Philip J. Stooke[1], [1]Center for Planetary Science and Exploration (CPSX) / Department of Geography, Western University, London, ON Canada (ipritcha@uwo.ca).

Introduction: The spatial analysis of impact craters on a planetary body is a potent tool for characterizing the history and stratigraphy of its surface [1]. Additionally, crater identification has been used in autonomous landing systems [2] and is of great importance to geologic mapping [3]. As crater populations within a given region of interest can be quite large, the search for a fully robust and scale-invariant automated crater detection algorithm (CDA) has led to the creation of a variety of creative systems for such purpose. Early CDAs often worked with photographic imagery as their data source, employing complex and highly-specialized algorithms for detecting impact craters. These systems often suffered from inconsistencies in image quality and terrain complexity, and were not easily scaled to different size regimes [4]. Combinations of methods and data fusion, as employed in [5], helped to improve on the results. A newer set of approaches to this problem have used elevation data in place of imagery. There are some complex systems that have been developed to work with 2-D surfaces generated by interpolation, such as those in [6-7].

While these systems have returned varied levels of achievement, there has been considerably less effort placed into determining the conditions for success or failure of these systems, and the consequent viability of their results. This is necessary for improving the systems and interpreting their results. This work investigated the crater detection results of a particular system known as the AutoCrat system, proposed in [7]. Through a novel application of the system to three distinct lunar surface types using LOLA elevation data, this system was investigated to determine the relationship between surface type and detection conditions with resulting efficiency and accuracies.

Method: The AutoCrat system is a two-step crater detection process. The first part consists of an (ideally) exhaustive basin-detection routine, identifying basins of various sizes in a DEM and growing them to attempt to match the craters' rim borders. This first portion should output all basins in the landscape, without yet weeding out the non-craters. The second part of the system uses machine learning software to create a predictive model, and "prune" the results to remove non-craters using a provided training set. Ideally, one is left with a catalog of only true positive crater detections. This catalog provides the crater centroid location, radius, depth, area, and a few mathematical shape descriptors. For a thorough technical description of the algorithm, see [7].

This system was applied to 3 equal-area (~90,000 km^2) but topographically distinct areas of the moon: central Mare Serenitatis, the eastern ejecta blanket of Orientale Basin, and a region in the near-side southern highlands. Shown in Figure 1, these regions were chosen as they were collectively representative of much of the lunar surface, while being individually distinct from one another.

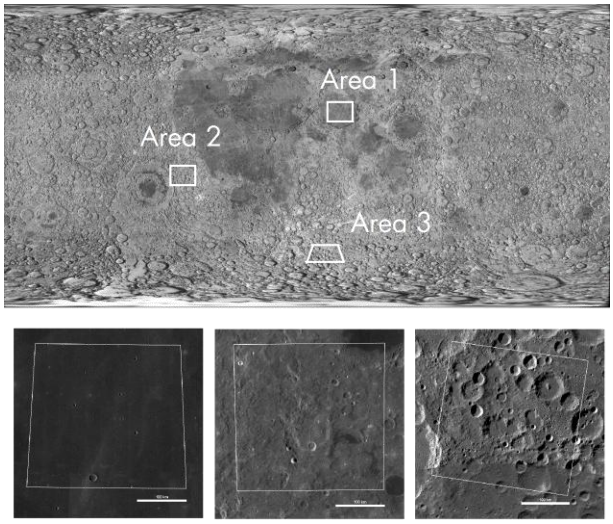

Figure 1: (*Above*) Global lunar map in cylindrical projection identifying the 3 study region boundaries. (Image: USGS Astrogeology) (*Below*) LROC WAC mosaics of areas 1-3 from left to right, zoomed to show detail. Scale bar in each represents 100 km.

The crater detection routine was then run on LOLA 512 ppd GDR elevation data for areas 1 and 2, and LOLA 100 m/pix GDR elevation data for the 3rd area. The training sets were generated by visual interpretation of the initial output and by manual labeling in JMARS [8] of a stratified sampling of roughly 140 craters for each study area, sampling from binned diameter ranges. Additionally, each crater in the stratified sample had its depth and radius manually measured.

Crater Discrimination: The training sets were then used to generate a decision tree using the J48 algorithm in WEKA, a free Java implementation of the C4.5 machine learning software first described in [9]. This decision tree analyzes the quantitative properties of each crater (including depth, diameter, d/D, and the

mathematical shape descriptors) and creates a predictive model to separate all crater candidates into a crater or non-crater class. This predictive model was then applied to all the originally detected craters, extracting the subset of craters from the set containing all basins. This subset of 'true' craters formed the fundamental dataset for extracting the population statistics.

Accuracy evaluation: To determine the algorithm's consistency in measuring the detected craters, depths and radii were collected manually for all craters used in the training sets. Relationships between detected and measured values were correlated using a standard linear regression to determine a systematic offset. To measure detection performance, two sub-areas accounting for roughly one-ninth of each study area were selected. Within these subareas, an exhaustive crater count was generated to be compared against our detection results.

Results: As expected, the detection results were heavily dependent on the terrain type. Mare Serenitatis showed the 'best' results, with a high detection rate for even sub-km craters. The ejecta blanket of Orientale (Area 2) had the highest number of false negatives. Many missed craters were located on a significant surface gradient, and were not picked up. Example detection results are shown in Figure 2.

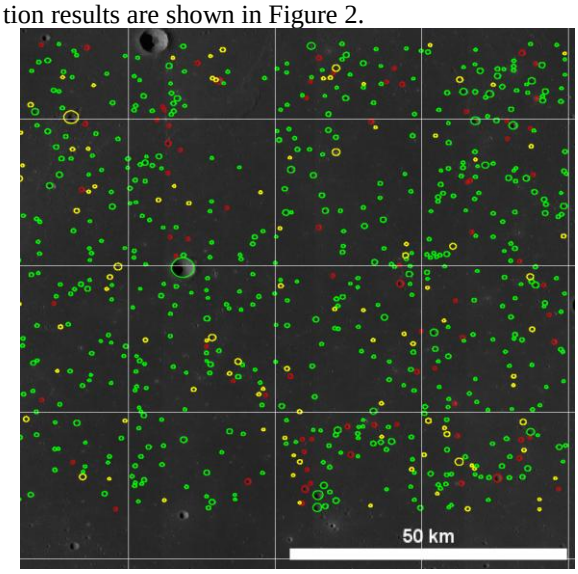

Figure 2: Detection results in a subarea of central Mare Serenitatis. Total detection area is ~8820 km^2. Green, yellow and red circles represent true positive, false positive, and false negative detections, respectively. Minimum FN diameter is 600m. Grid lines are 1 degree apart.

To compare with previous results in crater detection, we employ 3 main performance metrics, first proposed in [10]. The first is the detection percentage D, given by D = 100 * TP / (TP + FN), where TP is true positive (a correct detection), and FN is a false negative (a missed crater). Additionally, we use the quality percentage Q, expressed as Q = 100 * TP / (TP + FP + FN), and the branching factor B = FP/TP. Performance metrics are plotted together in Figure 3. A high detection percentage would be ideal for use in surface age estimation; a low branching factor would be ideal for collecting accurate crater size statistics. For all craters with D > 4km, we see suitable use for both applications.

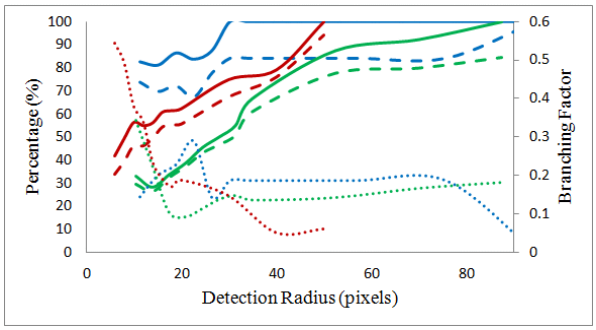

Figure 3: Detection percentage (solid line), quality percentage (dashed line), and branching factor (dotted line) for the 3 study areas. Mare Serenitatis is shown in blue, Orientale in green, and the highlands in red.

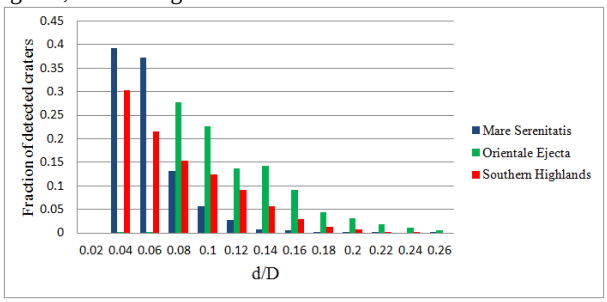

Figure 4: Histogram of depth-to-diameter ratio (d/D) values for craters in all three study areas. The results show generally deeper craters in the highlands, and an abscence of shallow crater detections in Orientale.

Acknowledgements: This research is partially funded by an NSERC Discovery Grant awarded to Dr. Wang. We would like to thank Michael Mendenhall, the author of the Cratermatic software for his generous assistance and help with technical issues. For software and tools, we thank NASA/ASU for the LROC Quickmap, as well as the JMARS team.

References: [1] Xie Y. et al. (2013) *IEEE Geosci. Remote Sens. Lett., 10,* 885-889. [2] Leroy B. et al. (2001) *Image Vision Comput., 19,* 787-792. [3] Kim J. R. et al. (2000) *Int. Arch. Photogramm. Remote Sens., 33,* 469-475. [4] Yin J. et al. (2015) *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens., 8,* 23-29. [5] Kim J. R. et al. (2005) *Photogramm. Eng. Remote Sens., 71,* 1205-1217. [6] Bue B. D. et al. (2007) *IEEE Trans. Geosci. Remote Sens., 45,* 265-274. [7] Stepinski T. et al. (2009) *Icarus, 203,* 77-87. [8] Christensen P.R. (2009) *AGU Fall Mtg,* #IN22A-06 [9] Quinlan J. (1993) *Programs for Machine Learning* [10] Shufelt J. and McKeown D.M. (1993) *Comput. Vision Image Understanding, 57,* 307-330.