UNDERSTANDING AND REDUCING CRATER COUNTING ERRORS. P. D. Tar[1], N.A. Thacker[1], [1]The university of Manchester, UK

Introduction: The ability to make reliable measurements forms a cornerstone upon which the empirical sciences are built. In order for a scientific measurement to be meaningful it must be accompanied by an assessment of its stability. Statistical and systematic uncertainties can be summarised using error bars which typically indicate +/- 1 standard deviation, meaning that in approximately 68% of cases the "true" value sought will lie within the error bar's range.

Crater counts are fundamental measurements in crater studies. The standard counting model for craters assumes Possion statistics, which predicts sqrt(N) standard deviation errors, where N is the number of craters counted within some size range. This error prediction is testable. Crater counts can only be considered valid if their error bars are honest reflections of actual levels of uncertainty. Scientific conclusions can not be made with any level of confidence unless this is the case. However, high levels of variability are observed in counts from expert, non-expert and automated crater counting approaches. These % errors are larger than those predicted using the Poisson assumption, as can be seen in Figure 1 using data take from [1].

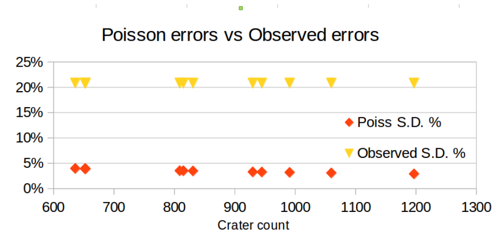

Figure 1: One Standard Deviation % errors

We propose an alternative model for crater counting incorporating the efficiencies with which counters identify false positive and false negative craters. We show that certain terms in this counting model can be estimated using Binomial statistics from repeatability data. More importantly, we show how a semi-automated approach can be applied to mitigate against false positive sources of uncertainty, potentially making the statistical error on empirical crater counts much smaller than is possible using raw counts alone. This reduction in error is achieved using Linear Poisson Models [2]. We make use of MoonZoo [3] citizen science crater data, undergraduate and expert counts from the Apollo 17 site to test the proposed methods.

Crater counting model: The estimated number of craters within a region can be summarised using

$$N_D = N_T P_T + N_F P_F$$

where N_D is an estimated count; P_T is the efficiency (0.0 to 1.0) with which "true" craters are counted; N_T is the "true" unknown number of craters; P_F is the efficiency (0.0 to 1.0) with which "false" craters are counted; and N_F is the number of potentially ambiguous "false" craters which might be counted accidentally. The N terms can reasonably be considered Poisson for independent surfaces, but variability in the efficiency P terms, due to different crater counter's abilities and personal biases, can explain the larger than Poisson errors observed in practice. In cases where P_T is 1.0 and P_F is 0.0 then the traditional sqrt(N_D) error is valid, however, subjectivity in humans and ambiguity in automated methods prevents this best case from being achieved.

Estimating efficiencies: The efficiency term for "true" craters can be estimated by repeated annotation. The ratio of craters annotated once, F_1, to those annotated twice, F_2, after two attempts at marking all craters can be used to give

$$P = \frac{2F_2}{F_1 + 2F_2}$$

This efficiency was computed for 8 regions within NAC images M104311715LE and M10431171RE for a group of undergraduate students and an expert crater counter (see acknowledgments). Results are presented in Figure 2.

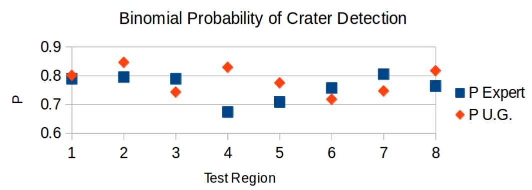

Figure 2: Expert vs Undergraduate Efficiencies

Estimating false positives: Where false positives are reported as a problem, such as in automated methods and citizen science data, the number of false positives, $P_F N_F$, can be estimated by applying Linear Poisson Models (LPM). For MoonZoo data, all candidate craters can be compared to a "true" crater template (see Figure 3). A dot product can be used to compare pixels of a template crater with candidate craters which may or may not be real. The distribution of template match scores for true and false craters can be linearly modeled from training data. The match scores from false craters are distinctly different from those of true craters (see Figure 4).

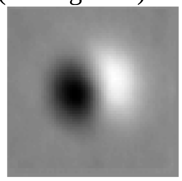

Figure 3: MoonZoo crater template, generated by taking an average of "true" craters

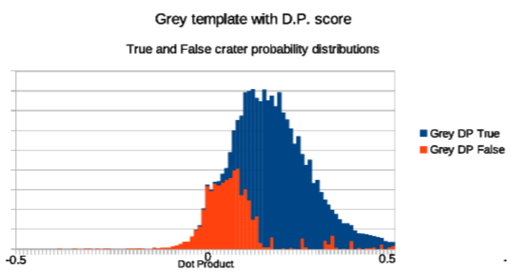

Figure 4: Distribution of dot product match scores, when dot product is applied to compare template to true (blue) and false (red) craters in MoonZoo dataset

The relative quantities of these different distributions can be estimated using Likelihood. If X is a template match score and k is a class of crater (e.g. a false or true component) then

$$\ln \mathcal{L} = \sum_X \ln \left[\sum_k P(X|k) \mathbf{Q}_k \right] \mathbf{H}_X - \sum_k \mathbf{Q}_k$$

where the P(X|k) terms model the probability distribution of the crater template match scores, the Q_k terms are the relative quantities of the different components, and H_X is a histogram of template matches from which false positives are to be estimated. The Expectation Maximisation algorithm is used to estimate the Q terms, then error propagation is used to assess the effects of Poisson noise in training and testing data on final measured counts

$$\mathbf{C}_{ij(data)} = \sum_X \left[\left(\frac{\partial \mathbf{Q}_i}{\partial \mathbf{H}_X} \right) \left(\frac{\partial \mathbf{Q}_j}{\partial \mathbf{H}_X} \right) \sigma^2_{\mathbf{H}_X} \right]$$

$$\mathbf{C}_{ij(model)} = \sum_X \left[\sum_k \left(\frac{\partial \mathbf{Q}_i}{\partial \mathbf{H}_{X|k}} \right) \left(\frac{\partial \mathbf{Q}_j}{\partial \mathbf{H}_{X|k}} \right) \sigma^2_{\mathbf{H}_{X|k}} \right]$$

where C is a covariance matrix, (data) is a statistical error from incoming data, (model) is an error from training data, and the sigmas are assumed Poisson noise in match score histograms.

Testing: Bootstrap re-sampling from true and false MoonZoo craters was used to test that predicted quantities of false positives were able to be estimated to within errors predicted using error propagation. Figures 5 and 6 show comparisons of predicted to observed counting errors when differing quantities of training and testing data are used within the LPM.

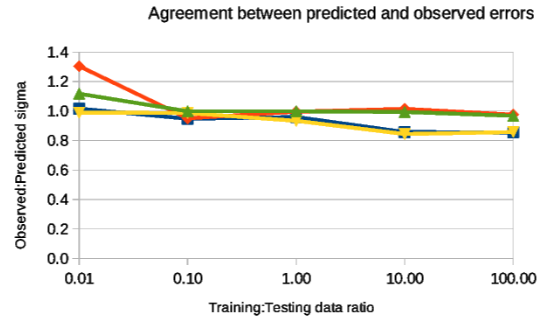

Figure 5: Agreement between predicted to observed errors when estimating false positives from MoonZoo data using Linear Poisson Models with different quantities of training data.

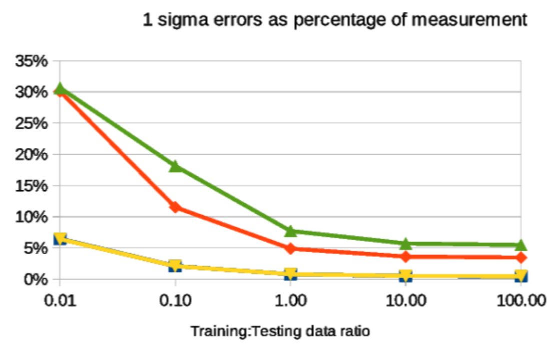

Figure 6: % predicted error on crater counts from LPM with different quantities of testing data

Results: Repeatability studies of expert and undergraduate crater counts show mark-up efficiencies of between 70% to 90%, with expert and non-expert efficiencies varying across the 8 test regions. This moves towards explaining larger than Poisson errors in real crater counts. Linear Poisson Models, applied to crater template matches, successfully estimates false positive quantities giving counts with predictable errors (Figure 5 showing a ratio of close to unity for predicted to observed re-sampled data). Correcting for false negatives, however, is a subject of future work.

Acknowledgements: We thank Sean Corrigan, Alex Griffiths, Tim Gregory, Hazel Blake, Dayl Martin, Maggie Sliz, Joe Scaife and Pavel Kamenov. Also expert crater counter *Roberto* Bugiolacchi. Leverhulme Trust (RPG-2014-019).

References: [1] Robbins S.J. et al., (2014) Icarus, 234:109-131. [2] Tar P.D. Thacker N.A. (2014) Annals of BMVA 1-22. [3] Joy K. at al., (2001) Astronomy & Geophysics 52 2.10-2.12