EPSC Abstracts
Vol. 9, EPSC2014-756, 2014
European Planetary Science Congress 2014

Quantification of false positives within Moon Zoo crater annotations

P.D. Tar(1), N.A. Thacker (1) and The Moon Zoo Team (2)

(1) Imaging Science, Dept. of Life Sciences, The University of Manchester, UK (2) www.moonzoo.org/team

Abstract

The Moon Zoo citizen science project [1] allows members of the public to annotate lunar images, providing researchers with a wealth of location and size information regarding the population of small craters on the Moon. To date, approximately 4 million images have been inspected. Here, we show how a quantitative pattern recognition system can be used to estimate the quantity of contamination in Moon Zoo data from erroneous annotations. The proposed method produces not only estimates of true verses false crater annotations, but also a full error covariance, with additional conformity checks, which is essential for the meaningful interpretation of measurements, e.g. for plotting error bars.

1. Introduction

The analysis of impact craters plays an important role in chronological studies of planetary bodies, yet annotating the location and size of impact craters is a time consuming and subjective activity. To mitigate against this, the Moon Zoo project brings together large numbers of volunteers to identify lunar craters. Users are presented with images, via a web-based interface (www.moonzoo.org), and asked to place markers around visible craters. However, mistakes are made, which introduces false positive annotations where no craters actually exist. The amount of contamination from these false positives must be quantified in order to calibrate any Size-Frequency Distributions derived from Moon Zoo data.

The amount of contamination can be estimated by analysing the match scores returned when comparing a template crater image to each annotation. It was shown in the associated abstract 'Coalescence and refinement of Moon Zoo crater annotations' that true and false annotations had distinctive match score distributions. Linear Poisson Models [2] can be applied to learn these distributions, then fit them to future data to perform the estimation.

40,000+ Moon Zoo annotations from around the Apollo 17 site (NASA's Lunar Reconnaissance Orbiter images M104311715LE and M104311715RE) are used to test the method.

2. Methodology

The false positive quantification process involves: learning, through example, the distribution of match scores for true and false annotations; a detailed error theory which computes measurement covariances via the application of error propagation; and a model conformity check using a chi-squared per degree of freedom test.

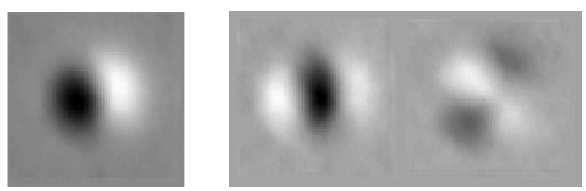

Figure 1: Left, mean grey level crater template (Grey); right, derivative (x and y) template (Grad).

2.1 False positive quantification

Two alternative crater templates and two match scores are investigated. The templates include a grey scale template (Grey) and a derivative template (Grad). Examples of these can be seen in Figure 1. The match scores include a dot-product (DP) and a mean square error (MSE) function. For all combinations, the distribution of match scores for example true and false annotations are sampled into histograms. Linear models of the resulting distributions are then trained using an Independent Component Analysis, based upon Likelihood, which is optimised using an Expectation Maximisation (EM) algorithm.

After training, the linear models can be fitted, using EM, to new unseen data. The weighting parameters returned from the fit are proportional to the amount of each class present in the data, and thus give a measurement of false verses true annotations.

2.2 Measurement covariances

The stability of the estimates are computed by considering how noise in training data and noise in incoming data affects model weighting parameters. This is done via error propagation [3] which approximates measurement perturbations by computing the derivatives of model weights with respect to the sources of error. The sources of error are assumed to be independent Poisson sampling noise in training and testing histogram bins.

2.3 Conformity check

Problems with an analysis (e.g. outliers in data or images with atypical craters) can be spotted using a chi-squared per degree of freedom function [3]. The residuals between modelled and observed histograms, which are Poisson, can be made approximately Gaussian with a square-root transform, then a standard chi-squared per degree of freedom test can be applied.

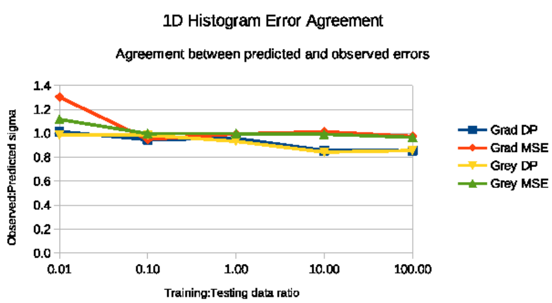

Figure 2: Agreement between predicted measurement accuracies and those achieved in practice.

3. Results

The 40,000+ Moon Zoo annotations were used to test the method, with repeated sampling with replacement used to confirm that the repeatability of measurements was correctly predicted by the error covariances. A known quantity of true and false annotations were included within each trial, allowing estimates to be compared against ground truth. On each trial a different quantity of training and testing data was used, and the empirical spread of measurements was compared against those predicted by the error covariance estimates. Figure 2 shows the ratio of the observed to predicted measurement errors (which should always be unity), corroborating the validity of the method. Figure 3 shows the level of accuracy attainable in estimating the quantity of true and false positive annotations.

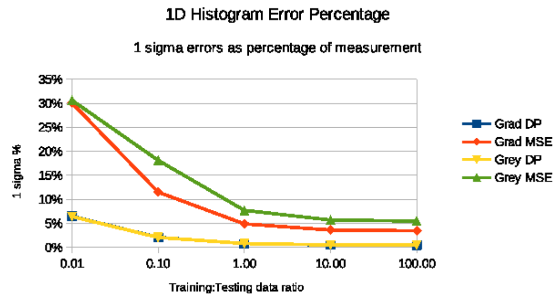

Figure 3: Measurement accuracies attainable when quantifying true and false annotations.

4. Summary and Conclusions

As seen in Figure 2, all combinations of template and match score allowed repeated measurements to be made with accuracies correctly predicted by the error theory, i.e. ratio of unity. Figure 3 shows that the best estimates of true verses false positives are achieved using a dot product type match score, with measurement errors typically better than 5%.

Acknowledgements

Thanks to K. Joy, R. Bugiolacchi, I. Crawford, the Moon Zoo team and many citizen scientists.

References

[1] K Joy et al.; Moon Zoo: Citizen science in lunar exploration; Astronomy & Geophysics vol 52 issue 2; 2011
[2] P.D. Tar, N.A. Thacker, Linear Poisson Models: A Pattern Recognition Solution to the Histogram Composition Problem, Annals of the BMVA 2014, Vol. 1, 2014 [3] R.J. Barlow; Statistics: A guide to the use of statistical methods in the physical sciences; John Wiley and Sons, UK; 1989