APPLICATION OF SUPPORT VECTOR REGRESSION TO DERIVE CRATER DEPTH/DIAMETER FROM SATELLITE IMAGES.

L. R. Chin[1], W. A. Watters[1], E. T. Chickles[2], and C. I. Fassett[3], [1]Dept. of Astronomy, Whitin Observatory, Wellesley College, Wellesley, MA 02481 (lc2@wellesley.edu), [2]Dept. of Physics, Massachusetts Institute of Technology, MA 02139, [3]NASA Marshall Space Flight Center, Huntsville, AL 35805.

Introduction: Through the study of impact crater shapes, one can draw important conclusions about the nature and evolution of planetary surfaces [e.g., 1-4]. In particular, studying the depth (d) to diameter (D) ratio (d/D) of a population of impact craters, in combination with crater count statistics, can yield valuable insights regarding rates of erosion and burial [5].

Motivated by the great abundance of available planetary surface image data, the goal of this project is to develop an efficient way to estimate d/D from satellite images of impact craters for which stereo information is not available [6]. We set out to develop and train a machine learning algorithm to extract d/D from a dataset of synthetic impact crater images for which model d/D is known.

The applications of machine learning to planetary science are numerous and diverse [7], including automatic planetary surface mapping [8] and the detection of impact craters [9]. Our algorithm makes use of Support Vector Regression (SVR), which is a type of Support Vector Machine (SVM) [10, 11]. SVMs are a branch of supervised machine learning valued for their straightforward implementation and versatility in solving both classification and regression problems. In regression analysis, an SVR algorithm produces a hyperplane function to fit the training data points, as well as an ε-tube that surrounds the hyperplane. Tunable hyperparameters include the width of the ε-tube (ε) and the amount an algorithm is penalized for points which fall outside the ε-tube.

Methods: We developed synthetic datasets using Python and POV-Ray (a ray-tracing scene-renderer) [12]. Our goal was to simulate the variation in illumination of impact craters for a proof of concept study. The synthetic crater images account for factors that may affect actual cropped images of craters, such as crater-center offset (mapping error), background noise, and lighting and shadows determined by the position of the Sun relative to the planet surface. We generated two datasets of 100,000 synthetic craters each, holding the solar elevation angle (α) constant at 30º for Dataset 1 and allowing it to range from 0-90º for Dataset 2. Figure 1 shows a sample of synthetic craters.

We have trained and tested SVR algorithms on three subsets of each dataset. In each case we split the data in half to create a training subset and a testing subset. In our first attempt, we used only ~4,000 synthetic craters due to memory limitations and in order to conduct a grid search to optimize hyperparameters. Our second and third attempts made use of ~10,000 and 25,000 synthetic craters, respectively.

Results: We measured the SVR algorithm's effectiveness by comparing actual d/D to the algorithm's predicted d/D for each synthetic crater in the testing subset, obtaining the correlation coefficient r^2. Figure 2 shows the algorithm's actual vs predicted d/D for the N = 25,000 subsets of Dataset 1 and Dataset 2. For the three sample sizes (~4,000, ~10,000, and 25,000) r^2 increased from 0.865 to 0.940 and then 0.963 for Dataset 1 and from 0.339 to 0.650 and then 0.71 for Dataset 2, respectively (see Fig. 2).

Discussion: Comparing the results from these three sample sizes indicates that using a larger dataset improves the algorithm's r^2 score significantly until the training set size exceeds about 10,000. Going forward, we intend to probe the limits of this improvement by training and testing the SVR algorithm on even larger subsets of Datasets 1 and 2, as well as by attempting the same analysis on real image data.

In our synthetic datasets, solar elevation angle (α) determines the illumination pattern on crater floors, including the size and placement of shadows. For the same training set size, the SVR algorithm performed significantly better when α was held constant at 30º. We may infer that the variation at high values of α leads to less variation in the illumination pattern, and is therefore less discriminating between values of d/D. This implies that our final model, trained on real image data, may be the most useful (or only useful at all) in the case of relatively shallow elevation angles.

An important next step is to test algorithm performance on real data rather than synthetic data. We intend to train and test the SVR algorithm on a dataset of ~20,000 images of Martian craters for which d/D is known (from stereo-derived measurements) and where D ranges from 1-5 km. In the event of positive results, our final step would be to use the algorithm to estimate d/D for a dataset of ~300,000 non-stereo MRO-CTX images of Martian craters for which d/D is not known. Based on our results to date using synthetic data, we believe our algorithm will at least be able to distinguish between shallow (d/D < 0.075), moderately filled (d/D = 0.07-0.013), and deep (d/D > 0.013) craters. Given that our synthetic craters lack realism that introduces significant variation, including ejecta blankets, non-circular crater shapes, and variations in tone and texture, our results are likely to fall short of the accuracy demonstrated using synthetic images. The range of α values in this dataset may also pose a challenge to the SVR algorithm, but our results suggest that the large size of each dataset will at least allow us to train the algorithm with a sufficient number of images.

In addition to SVR, we intend to explore the application of other branches of machine learning. A logical next step would be to develop a Support Vector Classifier (SVC), which is an SVM used for classification rather than regression [10]. Discretely binning d/D could allow us to use an SVC algorithm to classify crater images into different d/D ranges. We could then decrease the step size between bins to improve the accuracy of the algorithm's predictions. Convolutional Neural Networks (CNNs) are another promising avenue that could make use of either classification or regression [13].

Acknowledgments: This project was made possible by funding from the Massachusetts Space Grant as well as the Annie Jump Cannon Fellowship.

References: [1] Craddock, R. A. et al. (1997) *JGR Planets, 102, 13321–13340*; [2] Forsberg-Taylor, N. K. et al. (2004) *JGR, 109, 10.1029/2004JE002242*; [3] Golombek, M.P. et al. (2014) *JGR, 119, 10.1002/2014JE004658*; [4] Sweeney, J. et al. (2018) *JGR, 123, 10.1029/ 2018JE005618*; [5] Watters, W. A. et al. (2020) *LPS LI*, Abstract #3061; [6] Robbins, S. J. et al. (2018) *Meteoritics & Planet. Sci., 53, 583-626*; [7] Azari, A. R. et al. (2020) *arXiv preprint arXiv:2007.15129*; [8] Stepinski, T. F. et al. (2007) *AAAI, 1807-1812*; [9] Stepinski, T. F. et al. (2012) *Intelligent Data Analysis for Real-Life Applications: Theory and Practice, 146-159*; [10] Vapnik, V. (1997) *Advances in Neural Information Processing Systems, 281-287*; [11] Drucker, H. et al. (1997) *Advances in Neural Information Processing Systems, 9, 155-161*; [12] Buck, D. K. et al. (2004) *Computer software available from http://www.povray.org*; [13] O'Shea, K. et al. (2015) *arXiv preprint arXiv:1511.08458*

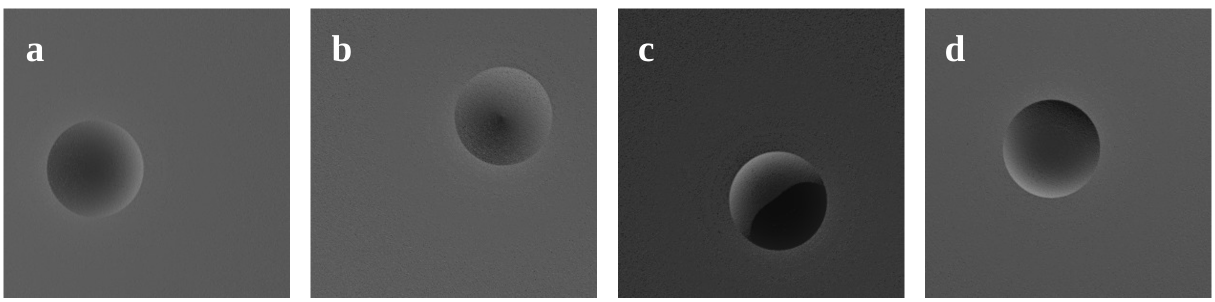

Fig. 1. (a,b) Synthetic craters from Dataset 1 (solar altitude α = 30°). (c,d) Examples from Dataset 2 (α = 0-90°)

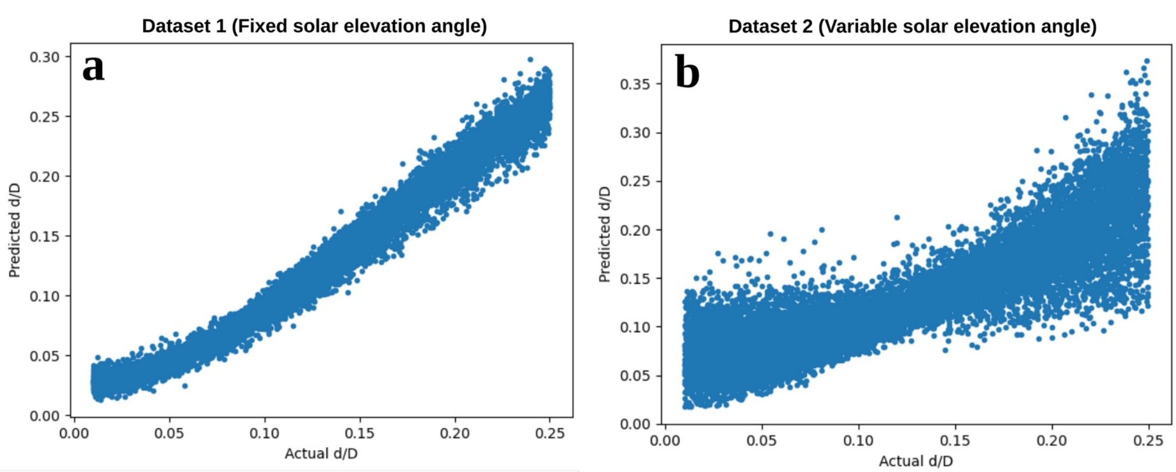

Fig. 2. (a) Actual vs Predicted d/D for Dataset 1 (N 25,000, r^2=0.963). (b) Actual vs Predicted d/D for Dataset 2 (N = 25,000, r^2=0.711).