

Electronic Thesis and Dissertation Repository

---

8-22-2016 12:00 AM

## Automated Impact Crater Detection and Characterization Using Digital Elevation Data

Ian M. Pritchard, *The University of Western Ontario*

Supervisor: Dr. Jinfei Wang, *The University of Western Ontario*

Joint Supervisor: Dr. Phil Stooke, *The University of Western Ontario*

A thesis submitted in partial fulfillment of the requirements for the Master of Science degree in Planetary Science

© Ian M. Pritchard 2016

Follow this and additional works at: <https://ir.lib.uwo.ca/etd>



Part of the [Remote Sensing Commons](#)

---

### Recommended Citation

Pritchard, Ian M., "Automated Impact Crater Detection and Characterization Using Digital Elevation Data" (2016). *Electronic Thesis and Dissertation Repository*. 3982.

<https://ir.lib.uwo.ca/etd/3982>

This Dissertation/Thesis is brought to you for free and open access by Scholarship@Western. It has been accepted for inclusion in Electronic Thesis and Dissertation Repository by an authorized administrator of Scholarship@Western. For more information, please contact [wlsadmin@uwo.ca](mailto:wlsadmin@uwo.ca).

# Abstract

Impact craters are used as subjects for the remote study of a wide variety of surface and subsurface processes throughout the solar system. Their populations and shape characteristics are collected, often manually, and analysed by a large community of planetary scientists. This research investigates the application of automated methods for both the detection and characterization of impact craters on the Moon and Mars, using machine learning techniques and digital elevation data collected by orbital spacecraft. We begin by first assessing the effect of lunar terrain type variation on automated crater detection results. Next, we develop a novel automated crater degradation classification system for martian complex craters using polynomial profile approximation. This work identifies that surface age estimations and crater statistics acquired through automatic crater detection are influenced by terrain type, with unique detection error responses. Additionally, we demonstrate an objective system that can be used to automate the classification of crater degradation states, and identify some potential areas of improvement for such a system.

**Keywords:** automated crater detection, Chebyshev polynomials, degradation, digital elevation model, impact crater, machine learning, Mars, Moon, profile approximation, topography

# Acknowledgement

I would like to first acknowledge my supervisors Dr. Jinfei Wang and Dr. Phil Stooke for their guidance in forming my research topics, editing this work, and sharing their respective expertise with me. It has been a great pleasure learning from and working with them over the course of my MSc, and I am fortunate to have found them as supervisors. I would also like to thank Dr. Livio Tornabene, Dr. Gordon Osinski and Dr. Catherine Neish as well as the other members of CPSX for helping me learn more about impact cratering and remote sensing, and for providing unique opportunities to get involved in space exploration activities.

Additionally, I'd like to thank my GITA lab mates and CPSX colleagues for collectively helping me push forward, and providing support and stimulating conversation. I also owe a debt of gratitude to the planetary science community as a whole. Their tools and general philosophy of open science were instrumental parts of my progress.

Finally, I'd like to thank my family and loved ones, who did years and years of work to support me before I could even begin to write chapter one.

# Contents

<b>Abstract</b>	<b>ii</b>
<b>List of Figures</b>	<b>ix</b>
<b>List of Tables</b>	<b>xvi</b>
<b>List of Appendices</b>	<b>xviii</b>
<b>List of Abbreviations, Symbols, and Nomenclature</b>	<b>xix</b>
<b>1 Introduction</b>	<b>1</b>
1.1 Impact Cratering . . . . .	1
1.2 Crater Chronology . . . . .	4
1.3 Research Objectives . . . . .	7
1.4 Thesis Organization . . . . .	8
<b>Bibliography</b>	<b>9</b>
<b>References</b>	<b>10</b>
<b>2 Crater Detection and Terrain Type</b>	<b>13</b>
2.1 Introduction . . . . .	13



2.1.1	Automatic Crater Detection . . . . .	13
2.1.2	CDA Accuracy . . . . .	15
2.1.3	Purpose . . . . .	16
2.2	Methods . . . . .	17
2.2.1	Overview . . . . .	17
2.2.2	Study Areas and Data . . . . .	20
	Study Area 1 - Mare Serenitatis . . . . .	22
	Study Area 2 - Orientale Ejecta . . . . .	24
	Study Area 3 - Southern Highlands . . . . .	26
	Source Data . . . . .	28
2.2.3	Crater Detection . . . . .	30
	Crater Detection - <code>findcraters</code> . . . . .	30
	Stratified Sampling . . . . .	32
2.2.4	Crater Discrimination . . . . .	34
	Training Set Construction . . . . .	34
	Decision Tree . . . . .	36
2.2.5	Accuracy Assessment . . . . .	39
	Measurement Accuracy . . . . .	39
	Detection Efficiency . . . . .	40
2.3	Results . . . . .	42
2.3.1	Detection Performance . . . . .	45
	Measurement Accuracy . . . . .	45
	Decision Tree Accuracy . . . . .	48

Quality Measures . . . . .	50
2.3.2 Crater Statistics . . . . .	53
2.4 Discussion . . . . .	56
2.4.1 Comparison With Other Published Work . . . . .	56
AutoCrat for Martian Craters (Stepinski 2009) . . . . .	56
2.4.2 Scientific Integrity of Results . . . . .	58
Size-frequency Distribution and Crater Counting . . . . .	59
Depth-to-diameter Relationship . . . . .	60
Buried Craters in Orientale Ejecta . . . . .	62
2.5 Conclusion . . . . .	64
<b>Bibliography</b>	<b>65</b>
<b>References</b>	<b>66</b>
<b>3 Automated Crater Degradation Classification</b>	<b>73</b>
3.1 Introduction . . . . .	73
3.1.1 Crater Degradation . . . . .	73
Internal Crater Morphology . . . . .	76
Processes of Crater Modification . . . . .	79
3.1.2 Crater Profile Modelling . . . . .	80
Chebyshev Polynomials . . . . .	82
3.1.3 Purpose . . . . .	85
3.2 Methods . . . . .	86

3.2.1	Study Area and Data . . . . .	87
3.2.2	Pre-processing and Training Set Selection . . . . .	89
3.2.3	Profile Extraction and Approximation . . . . .	90
	Topographic Profile Extraction . . . . .	90
	Chebyshev Polynomial Fitting . . . . .	93
3.2.4	Crater Classification . . . . .	93
3.3	Results . . . . .	95
3.3.1	Degradation Class Model . . . . .	95
3.3.2	Interior Morphology Model . . . . .	96
3.4	Discussion . . . . .	98
3.4.1	Efficiency of Crater Profile Extraction . . . . .	98
3.4.2	Degradation Classification Performance . . . . .	100
3.4.3	Interior Morphology Classification Performance . . . . .	103
3.5	Conclusion . . . . .	106
	<b>Bibliography</b>	<b>108</b>
	<b>References</b>	<b>109</b>
<b>4</b>	<b>Conclusions</b>	<b>113</b>
4.1	Major Findings . . . . .	113
4.2	Motivation for Automated Planetary Image Processing . . . . .	115
4.3	Future Work . . . . .	116
	<b>Bibliography</b>	<b>117</b>

<b>References</b>	<b>118</b>
<b>A MATLAB Algorithms and Decision Trees</b>	<b>120</b>
<b>B Crater Degradation Classification</b>	<b>123</b>
<b>Curriculum Vitae</b>	<b>127</b>

# List of Figures

1.1	Cross-sectional diagrams of both a simple (a) and complex (c) crater (Osinski & Pierazzo, 2012). (b) shows a WAC mosaic image of Sarabhai crater, a 8 km diameter simple impact crater in Mare Serenitatis. Below (d) is Tycho, an 86 km diameter complex crater in the southern lunar highlands. Note the presence of the structural central uplift (SU) and terraced walls. . . . .	3
1.2	Plots of depth-to-diameter ratio of detected Martian craters as a function of southern latitude, from (Stepinski, Mendenhall, & Bue, 2009). Grey dots are ‘shallow’ craters, black are ‘deep’, and the larger circles represent binned measurements. A steep drop-off in $d/D$ is measured around $38^\circ$ S in all three areas; this is potentially correlated to a significant increase in the presence of ground ice. . . . .	5
1.3	(a) The lunar chronology function. (b) A sample size-frequency distribution (SFD), showing isochrons for 3.0, 3.6, 3.8 and 4.0 Ga. These plots were generated using CraterstatsII (Michael & Neukum, 2010). . . .	6

2.1	Flowchart depicting the major steps of this project. Boxes with dashed lines collect related steps. <b>a</b> : data collection and preprocessing <b>b</b> : crater detection and categorization <b>c</b> : machine learning and training set generation <b>d</b> : measurement accuracy and systematic error <b>e</b> : detection efficiency assessment . . . . .	19
2.2	Global lunar map in simple cylindrical projection. This map is a combination of 3 optical mosaics consisting of imagery from the Wide Angle Camera (WAC), in addition to colour elevation information from the LOLA instrument (red is high elevation, blue is low). Each study area, demarcated by the white boxes, is roughly 90,000 km <sup>2</sup> . . . . .	20
2.3	(a) Optical WAC mosaic of the Mare Serenitatis study area, at 100 m/pix. (b) Colourized elevation from the LOLA instrument. (c) Slope map, derived from the LOLA DTM. . . . .	22
2.4	(a) Optical WAC mosaic of the Orientale ejecta study area, at 100 m/pix. (b) Colourized elevation from the LOLA instrument. (c) Slope map, derived from the LOLA DTM. . . . .	24
2.5	(a) Optical WAC mosaic of the Southern highlands study area, at 100 m/pix. (b) Colourized elevation from the LOLA instrument. (c) Slope map, derived from the LOLA DTM. . . . .	26

2.6	A sample abbreviated image processing sequence for the Cratermatic system. Using an input DTM, the image is blurred at different spatial scales to separate nested depressions. Craters detected in each blurred landscape are recombined for the final catalog. Not all blurred landscapes are shown for the sake of simplicity. . . . .	31
2.7	Graphic representation of the training set generation process. Each basin that was selected via sampling is hand labeled as a crater or non-crater by overlaying the rim on both optical and elevation data in JMARS. . . . .	35
2.8	Decision tree built using the training set for the Southern Highlands study area. This tree was generated using the J48 machine learning algorithm in the WEKA software suite. Nodes are ellipses and leaves are rectangles. At each node, the numbers in parenthesis represent the number of instances reaching the leaf (green) /the number of misclassified instances at the leaf (red). . . . .	38
2.9	Automated crater detection results for study areas 1-3 (L to R, respectively). These maps show the complete crater catalogues after pruning. Background imagery is a 100 m/pix WAC mosaic. . . . .	43
2.10	Post-pruning cumulative distribution function (CDF) for the three study areas. . . . .	43
2.11	Small basins detected by Cratermatic. Note the vertically-linear arrangement. These basins do not survive the crater discrimination process, and are artefacts caused by the laser altimetry data type. . . . .	45

2.12	Assessment of crater detection in six sub-areas. Each sub-area is 100 x 100 km. Green crater profiles represent true positive detections, yellow are false positive, and red are false negative with $D > 600$ m. . . . .	51
2.13	The three performance metrics $DET$ (solid line), $Q$ (dashed line) and $B$ (dotted line) as a function of crater radius in pixels. Mare Serenitatis is represented in blue, Orientale in green, and the Southern highlands in red.	52
2.14	A false positive detection in Mare Serenitatis. (a) The Cratermatic detection result (b) The feature, as seen in LROC NAC optical imagery (c) The feature as seen in the raw LOLA DTM. . . . .	53
2.15	Histogram of the depth-to-diameter ratio for all pruned craters with $D > 1$ km. . . . .	54
2.16	SFD for each study area, plotted using CraterstatsII (Michael & Neukum, 2010). Three isochrons representing crater populations for 3.4, 3.8 and 4.0 Ga old surfaces are plotted as solid grey lines. Crater frequencies are binned in pseudo-logarithmic fashion (18 bins per decade). . . . .	55



3.1	<p>An example of a crater from each preservation class, from (Robbins &amp; Hynek, 2012b). All imagery is THEMIS Day IR. (a) A 23 km diameter crater in the pristine class (4), located at 16.50°N 11.49°E. Note the sharp rim, central pit, and pristine ejecta. (b) A 44 km diameter class 3 crater located at 14.79°N 9.63°E. This crater shows some infilling, as well as subsequent cratering on the ejecta blanket. (c) A 65 km diameter class 2 crater, located at 15.27°S 9.79°E. This crater shows substantial infilling, as well as minimal expression of the rim and ejecta. (d) A 38 km diameter class 1 crater, centred at 39.33°S 15.25°E. This crater has no raised rim, and is heavily infilled. Ejecta texture is no longer visible. . . . .</p>	77
3.2	<p>An example of a crater from each interior morphology class, from (Robbins &amp; Hynek, 2012b). All imagery is THEMIS Day IR. (a) A 30 km diameter crater exhibiting a central peak (CpxCPk), or localized topographic high near the crater center. This crater is located at 26.31°N 28.12°E. (b) A 38 km diameter central pit (CpxCPt) located at 14.68°N 20.67°E. (c) A 28 km diameter crater with a flat floor morphology (CpxFF), located at 8.81°N 35.82°E. (d) A 19 km diameter summit pit crater (CpxSuPt), centred at 15.24°N 16.45°E. The summit pit, located at the crater center, is differentiated from a central pit by the raised topography on which the pit is situated. . . . .</p>	78

3.3	(a) Plinius, a roughly 43 km diameter complex lunar impact crater located at 15.4°N, 3.7°E. The green line denotes the direction of the extracted profile in (b). (b) A scaled plot of the crater profile (black), alongside three Chebyshev profile approximations. As the order of approximation increases from $M = 4$ (red) to $M = 32$ (blue), the approximation error is reduced. . . . .	83
3.4	A general diagram of how crater profile reconstruction relates to basis functions and coefficients. Each basis function, shown on the left as $T_n(x)$ , has a weighted contribution to the final profile that is related to its coefficient $C_n(x)$ . . . . .	84
3.5	Flowchart of the major steps for this project. . . . .	86
3.6	Global map of Mars in simple cylindrical projection. Visualized data is MOLA Colourized Elevation. The study area is outlined by the white box. . . . .	88
3.7	Profile extraction and approximation process. (a) Crater is identified by catalogue values for location and radius. (b) Line of greatest slope (white/red dashed line) through crater center and length $4 \cdot R_{crater}$ is identified. (c) Topographic profile is extracted, with domain scaled to [-1,1]. (d) Profile is approximated using Chebyshev polynomial expansion. (e) Chebyshev coefficients are retrieved from the expansion. . . . .	91
3.8	(a) A THEMIS image of a 14 km crater, catalogued as degradation class 3 but misclassified as class 1. (b) The 2-D profile and Chebyshev reconstruction of the crater, showing a significant pre-impact slope. . . . .	102

3.9	(a) A THEMIS image of a 13 km crater, catalogued as degradation class 3 but misclassified as class 1. This crater is in the ejecta blanket of Cerulli, a 130 km diameter crater to the north. (b) The 2-D profile and Chebyshev reconstruction of the crater, showing a good reconstruction despite misclassification. . . . .	102
3.10	(a) A THEMIS image of a 20 km diameter crater with a central pit, located at 30.35°N, 18.60°E. The pit is visible in the crater center. (b) The extracted topographic profile (blue line) and reconstructed Chebyshev approximation (red dashed line). The central pit, noticeable as the dip in the elevation profile, is distinctly missing from the approximation. . . . .	104
3.11	(a) A THEMIS image of a 21 km diameter crater with a central peak, located at 22.09°N, 8.71°E. The peak is visible in the crater center. (b) The extracted topographic profile (blue line) and reconstructed Chebyshev approximation (red dashed line). This crater, listed in the source catalogue as unclassified (CpxUnc), exhibits a strong central uplift signature with a small summit pit when viewed as a 2D profile. . . . .	105
3.12	(a) A THEMIS image of a heavily degraded 42 km crater located at 18.95°N, 20.81°E, catalogued as CpxFF (flat floor) but given an ‘unclassified’ designation by the model. The south-west rim of the crater is superposed by a very pristine 11 km crater. (b) The 2-D profile and Chebyshev reconstruction of the crater, showing a rough interior topography. . . . .	106

# List of Tables

2.1	Details for the 3 selected study areas. . . . .	21
2.2	The six size-based crater classes. Craters are separated into classes based on their diametric size to perform a stratified sampling and build the training sets. . . . .	33
2.3	Numerical crater detection results for each study area. . . . .	44
2.4	Linear regression results for radius and depth in all three study areas. These fits were used to correct the values of the pruned crater catalogue. The true values are measured manually and are used to corrected the detected values. . . . .	46
2.5	Detailed accuracy statistics for ‘crater’ class, for each study area. . . . .	49
2.6	Detailed accuracy statistics for ‘non-crater’ class, for each study area. . . . .	50
2.7	Comparison of performance metrics with other published values. * The values listed for the three study areas here are for craters with $D \geq 4$ km. . . . .	57
3.1	Crater degradation classification criteria from (Robbins & Hynek, 2012a). . . . .	76
3.2	List of Chebyshev coefficients and coefficient combinations, with their interpreted topographical indications (Mahanti, Robinson, Humm, & Stopar, 2014). . . . .	85

3.3	Degradation class assignment results for the classification model. The number of instances listed is from the original catalogue. . . . .	96
3.4	Confusion matrix for the degradation class assessment. . . . .	96
3.5	Interior morphology assignment results for the classification model. The number of instances listed is from the original catalogue. CpxUnc is unclassified, CpxCpk is central peak, CpxCpt is central pit, CpxFF is flat-floored, and CpxSuPt is summit pit. . . . .	97
3.6	Confusion matrix for the interior morphology class assessment. . . . .	97
3.7	Degradation classification results for class combinations. The original 4 class results are listed, as well as the 3 class (degraded, moderately preserved/degraded, well preserved) and 2 class combination (degraded, preserved). . . . .	101

# List of Appendices

Appendix A MATLAB Algorithms and Decision Trees . . . . .	120
Appendix B Crater Degradation Classification . . . . .	123

# List of Abbreviations, Symbols, and Nomenclature

<b>ASCII</b>	American Standard Code for Information Interchange
<b>CDA</b>	Crater Detection Algorithm
<b>CDF</b>	Cumulative Distribution Function
$C_n$	$n$ -th Chebyshev Polynomial Coefficient
<b>CSV</b>	Comma Separated Values
<b>d/D</b>	Depth-to-Diameter
<b>DTM</b>	Digital Terrain Model
<b>FN</b>	False Negative
<b>FP</b>	False Positive
<b>GDR</b>	Gridded Data Record
<b>GLD</b>	Global Lunar DTM
<b>JMARS</b>	Java Mission-planning and Analysis for Remote Sensing
<b>lat</b>	Latitude
<b>LOLA</b>	Lunar Orbiter Laser Altimeter
<b>lon</b>	Longitude
<b>LRO</b>	Lunar Reconnaissance Orbiter
<b>LROC</b>	Lunar Reconnaissance Orbiter Camera
<b>MGS</b>	Mars Global Surveyor
<b>MOLA</b>	Mars Orbiter Laser Altimeter
<b>NAC</b>	Narrow Angle Camera
<b>PDS</b>	Planetary Data System
<b>pix</b>	pixel
<b>SFD</b>	Size-Frequency Distribution
<b>THEMIS</b>	Thermal Emission Imaging System
<b>TP</b>	True Positive
<b>WAC</b>	Wide Angle Camera
<b>WEKA</b>	Waikato Environment for Knowledge Analysis

# Chapter 1

## Introduction

### 1.1 Impact Cratering

Impact craters are basin-shaped features on a surface formed by a collision of an asteroid, comet or meteoroid on that surface. The structures are ubiquitous throughout the solar system, and are studied by a large community of planetary and earth scientists. Impact craters (hereafter just ‘craters’) exhibit different expressions and properties that depend on the pre-impact surface conditions, subsurface composition, and parameters of the impact event such as energy and impact angle. Additionally, they can be altered over time by surface processes which vary depending on celestial body. Craters can be divided into main groups: simple and complex. These two groups are separated by size; above a certain threshold, the crater transitions to a complex shape with a central peak. This transition size for a given body is related to its gravitational acceleration (Melosh, 2011). The anatomy of both a simple and complex crater are shown in Figure 1.1.



Simple craters are bowl-shaped, with an uplifted rim that can be altered by weathering or other modifying processes. The floor of the crater contains a mix of breccia (broken fragments of target material held together in a matrix) and possibly impact melt (target rock that was melted from the extreme heat of impact). Beyond the rim of the crater, debris known as ejecta is deposited on top of the surrounding terrain, with the thickness of the ejecta being inversely related to radial distance from the rim (McGetchin, Settle, & Head III, 1973). This ejecta, while relatively thin for smaller craters, can play a significant role in altering the landscape as the impact and thus amount of excavated material becomes larger. Significantly large enough impacts are known to form impact basins (hereafter referred to as ‘impact basins’ in full to identify them from regular topographic depressions). These structures, such as the one that formed the Orientale Basin, have diameters on the order of hundreds to thousands of km, and often form concentric ring systems. Impact basins have an age-resetting effect on the terrain surrounding them. Proximal craters will be entirely covered, while craters that are further out will be significantly in-filled so as to fundamentally change their expression. Craters which have been in-filled by either ejecta or lava, but still have a visible raised rim, are known as ghost craters. Ghost craters can affect crater counts and throw off age estimates.

The topographic expression of simple craters can be described to a lower degree as a local region of concavity. The morphology changes for craters that have a diameter larger than the transition size for its host body, which is around 15 km for the Moon, 7

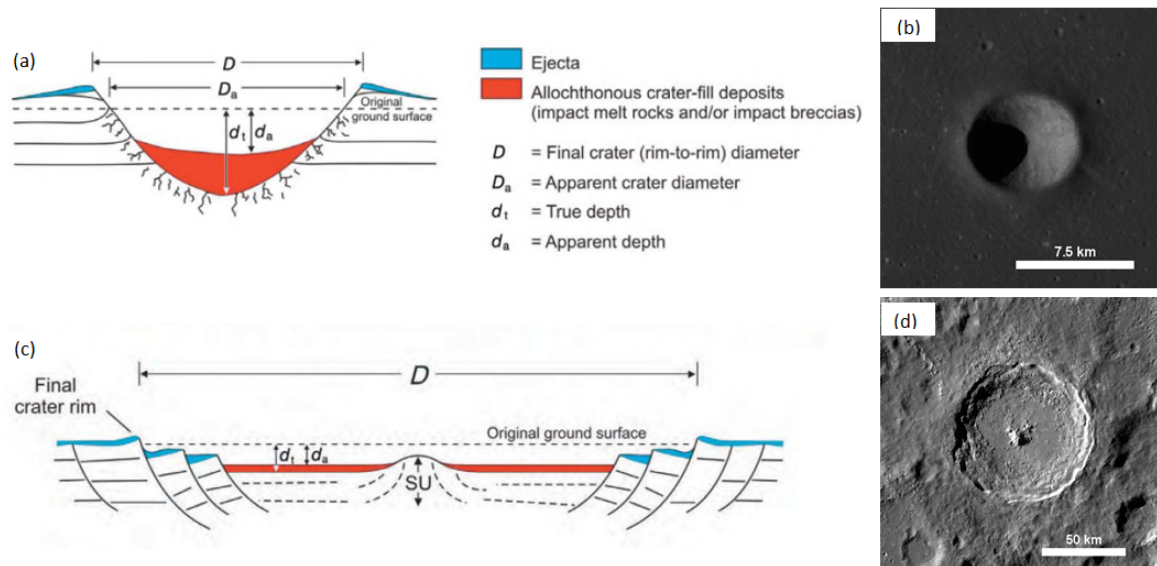


Figure 1.1: Cross-sectional diagrams of both a simple (a) and complex (c) crater (Osinski & Pierazzo, 2012). (b) shows a WAC mosaic image of Sarabhai crater, a 8 km diameter simple impact crater in Mare Serenitatis. Below (d) is Tycho, an 86 km diameter complex crater in the southern lunar highlands. Note the presence of the structural central uplift (SU) and terraced walls.

km for Mars and between 3 km and 5 km on Earth (Melosh & Ivanov, 1999). Complex craters, which have diameters above the transition size, are structurally different from their simple counterparts. In addition to the raised rim and ejecta features, complex craters exhibit a significant central peak, or local topographic high at the center of the crater. Additionally, complex craters can have terraced walls from collapse, which cause the rim to take on a more polygonal shape over time as the crater is modified. Quantities such as depth, circularity and continuity of the rim crest that describe the shape of a crater can be affected over time by various modifying processes. This has consequences for the automatic detection and measurement of these craters, as the topographic rules and assumptions about a craters expression become less valid.

The spatial and morphometric analysis of impact craters has proven over decades to be a useful tool for answering a variety of questions in planetary science. Early work established the use of impact crater studies in providing chronological information about the lunar surface, laying the foundation for a stratigraphic history of the Moon (Opik, 1960; Shoemaker, Hackman, & Eggleton, 1963; Hartmann, 1965). Additionally, impact crater populations have helped constrain models of impactor populations (Neukum, König, & Fechtig, 1975). Modern applications of impact crater morphometry have revealed information about crater formation processes (Xiao, Zeng, & Komatsu, 2014), as well as surface (Bart, 2014) and subsurface (Pathare, Paige, & Turtle, 2005) crustal compositions. An example of using crater morphometry to identify the presence of ice is shown in Figure 1.2. The study and detection of impact craters is also relevant to the field of space exploration and has been used to investigate autonomous spacecraft landing systems (Leroy, Medioni, Johnson, & Matthies, 2001; Johnson, Huertas, Werner, & Montgomery, 2008).

## 1.2 Crater Chronology

As mentioned previously, impact craters have been used as a tool to deduce the relative ages of surface units since the 1960's (Bland, 2003). Prior to the Apollo landings of the late 1960's and early 1970's, it was highly desirable to form a lunar geologic record so as to inform the landings and subsequent sample collection. Early work in lunar chronology employed crater counting as a means to deduce relative ages. Resting on the fundamental assumption that a pristine surface unit will have zero craters, this

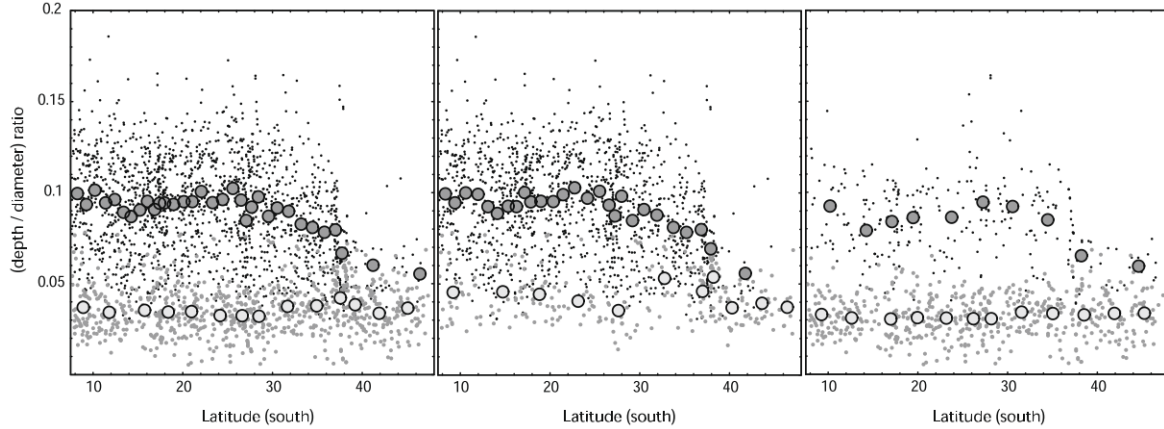


Figure 1.2: Plots of depth-to-diameter ratio of detected Martian craters as a function of southern latitude, from (Stepinski et al., 2009). Grey dots are ‘shallow’ craters, black are ‘deep’, and the larger circles represent binned measurements. A steep drop-off in  $d/D$  is measured around  $38^\circ$  S in all three areas; this is potentially correlated to a significant increase in the presence of ground ice.

technique involves counting the densities of craters of varying size that are found on the unit. By modelling and estimating the production rate, or rate at which craters of a given size are formed on the surface, it is possible to extract the age. It is then possible to calibrate these to absolute ages when compared against radiometrically dated samples from the Apollo missions. While this calibration can only be performed for bodies where physical samples have been obtained, more recent work has extended the crater production function from the Moon to other bodies such as Mars by examining impact rate and scaling law differences for the separate bodies. The crater density above a certain diameter has historically been expressed by a power law relationship, of the form

$$N_{cum}(D) = cD^b \quad (1.1)$$

where  $N_{cum}(D)$  is the cumulative number of craters with diameter larger than  $D$ ,  $b$  is a

value known as the size index, and  $c$  is a constant (Basaltic Volcanism Study Project, 1981). In (Neukum, 1983), the surface density of craters was related to the surface age by the chronology function

$$N(D > 1) = 5.44 \cdot 10^{-14}(e^{6.93T} - 1) + 8.38 \cdot 10^{-4}T \quad (1.2)$$

where  $N(D > 1)$  is the number of craters with a diameter larger than 1 km normalized to the counting area in sq. km, and  $T$  is the age of the surface in billions of years (Ga). For lunar surface calculations using the above equation and an automated crater detection system, it is imperative to understand the system's responsiveness for craters above 1 km as this will have an effect on the calculated age. The chronology function is shown in Figure 1.3.

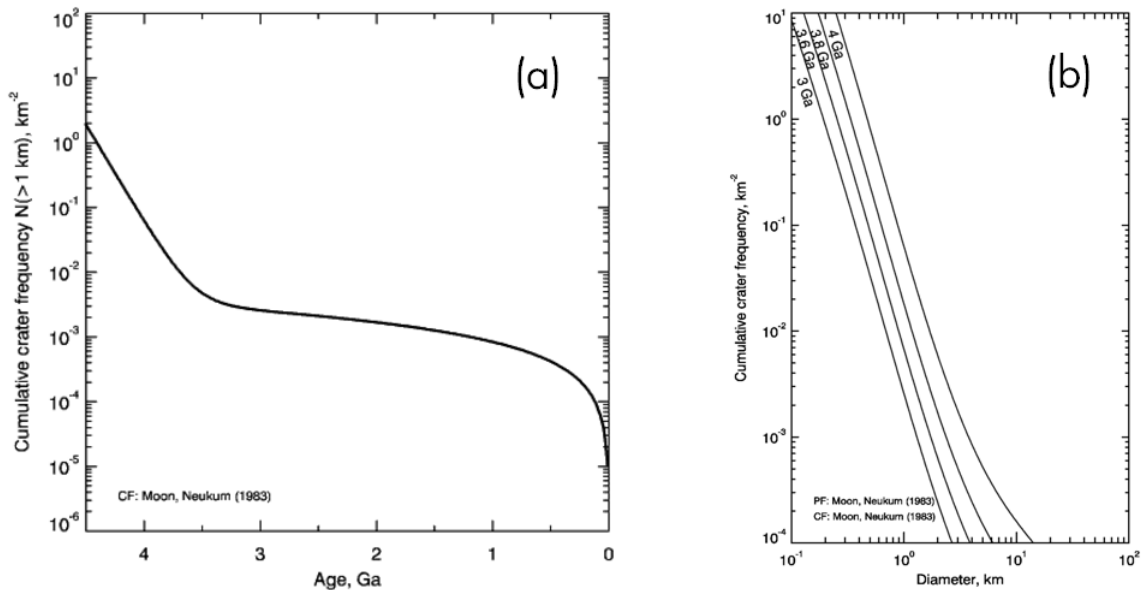


Figure 1.3: (a) The lunar chronology function. (b) A sample size-frequency distribution (SFD), showing isochrons for 3.0, 3.6, 3.8 and 4.0 Ga. These plots were generated using CraterstatsII (Michael & Neukum, 2010).

A common method of representing crater density in a context of surface age is to use

a size-frequency distribution. A SFD is a graphic representation of the density of craters contained within a certain diameter bin, against diameter in a log-log plot. Equation 1.1 shows that one should expect a power-law relationship on the graph, with slope  $b$ . However, various modifying processes affect crater densities, causing a deviation from the perfect power law relationship (Williams, Pathare, & Aharonson, 2014). This effect is strongest in the small- $D$  regime. In addition, SFDs can be used to plot isochrons, or lines of constant age on a SFD. For a given surface unit, an older age will mean a higher crater frequency for a given diameter bin. This will push older isochrons (or surface populations) to the right on the SFD plot. Comparison of these isochrons against the results of a crater count gives both an idea of the surface's age, as well as any modifying processes that cause the crater count to deviate from an ideal distribution.

### 1.3 Research Objectives

This thesis will examine the role of automated image processing methods for impact crater analysis. In doing so, it seeks to address the following questions:

1. Does lunar surface terrain type variation affect the results of automated crater detection?
2. What influences systematic measurement error using automated measurement systems?
3. How can the results of automated crater detection and measurement be applied to other topics in crater studies?

4. Can an objective crater degradation classifier be built using a polynomial profile approximation method?
5. How can existing crater catalogues be used to improve automated crater characterization methods?
6. Are automated methods alone currently sufficient for crater characterization?

The major research objectives are as follows:

1. To apply a topographic crater detection system to three distinct lunar surface types; to compare the results against visual inspection of the surface to determine detection efficiency as a function of crater size; to compare extracted radii and depths against measured values to identify systematic measurement error; and to build a crater chronology and bulk statistical description of the craters for each unit type and compare them against previous results.
2. To build an automated, quantitative method for the degradation and interior morphological classification of martian complex impact craters; to apply polynomial approximation methods to the craters' topographic profiles; to use the approximations as discriminators in a classification model; and to identify the classification accuracy of such an approach.

## 1.4 Thesis Organization

This thesis is presented in integrated article format. It contains two two papers that demonstrate the use of automated image processing methods in planetary science.

Chapter 1 introduces some background on impact craters and provides research topic context. Additionally, it outlines the research objectives. Chapters 2 and 3 are two individual but related studies. Chapter 2 presents a study on the effect of lunar terrain type variation on automated crater detection and measurement. This work was presented at the 47<sup>th</sup> Lunar and Planetary Science Conference in Houston, Texas. Chapter 3 describes the development of an automated crater degradation state classifier using polynomial profile approximations. Chapter 4 concludes the thesis, describing the primary findings of the work and discussing future applications of automated digital image processing in planetary science.



# References

- Bart, G. D. (2014). The quantitative relationship between small impact crater morphology and regolith depth. *Icarus*, *235*, 130–135.  
doi: 10.1016/j.icarus.2014.03.020
- Basaltic Volcanism Study Project. (1981). *Basaltic Volcanism on the Terrestrial Planets*.  
New York: Pergamon Press, Inc.
- Bland, P. (2003). Crater counting. *Astronomy and Geophysics*, *44*(4), 4.21–4.21.  
doi: 10.1046/j.1468-4004.2003.44421.x
- Hartmann, W. K. (1965). Terrestrial and Lunar Flux of Large Meteorites in the Last Two Billion Years. *Icarus*, *4*(2), 157–165.  
doi: 10.1016/0019-1035(65)90057-6
- Johnson, A. E., Huertas, A., Werner, R. A., & Montgomery, J. F. (2008). Analysis of On-Board Hazard Detection and Avoidance for Safe Lunar Landing. *Aerospace Conference, 2008 IEEE*, 1–9.  
doi: 10.1109/AERO.2008.4526301
- Leroy, B., Medioni, G., Johnson, E., & Matthies, L. (2001). Crater detection for autonomous landing on asteroids. *Image and Vision Computing*, *19*(11), 787–792.

- doi: 10.1016/S0262-8856(00)00111-6
- McGetchin, T. R., Settle, M., & Head III, J. W. (1973). Radial thickness variation in impact crater ejecta: implications for lunar basin deposits. *Earth and Planetary Science Letters*, *20*(2), 226–236. doi: 10.1016/0012-821X(73)90162-3
- Melosh, H. J. (2011). *Planetary Surface Processes* (1st ed.). Cambridge University Press.  
doi: 10.1017/CBO9780511977848
- Melosh, H. J., & Ivanov, B. A. (1999). Impact Crater Collapse. *Annual Review of Earth and Planetary Sciences*, *27*(1), 385–415.  
doi: 10.1146/annurev.earth.27.1.385
- Michael, G., & Neukum, G. (2010). Planetary surface dating from crater size-frequency distribution measurements: Partial resurfacing events and statistical age uncertainty. *Earth and Planetary Science Letters*, *294*(3-4), 223–229.  
doi: 10.1016/j.epsl.2009.12.041
- Neukum, G. (1983). Meteoritenbombardement und Datierung Planetarer Oberflaechen . *Habilitation Dissertation for Faculty Membership, Univ. of Munich*, 1–186.
- Neukum, G., König, B., & Fechtig, H. (1975). Cratering in the earth-moon system: Consequences for age determination by crater counting. *Proc. Lunar Planet. Sci. Conf. 6th*, 2597–2620.
- Opik, E. (1960). The Lunar Surface as an Impact Counter. *Oxford Journal*, *120*(5), 404–411.
- Osinski, G. R., & Pierazzo, E. (2012). *Impact Cratering: Processes and Products* (G. R. Osinski & E. Pierazzo, Eds.). Chichester, UK: John Wiley & Sons, Ltd.
- Pathare, A. V., Paige, D. A., & Turtle, E. (2005, apr). Viscous relaxation of craters

within the martian south polar layered deposits. *Icarus*, *174*(2), 396–418.

doi: 10.1016/j.icarus.2004.10.031

Shoemaker, E., Hackman, R., & Eggleton, R. (1963). Interplanetary Correlation of Geologic Time. *Advances in the Astronautical Sciences*, *8*, 70–89.

Stepinski, T. F., Mendenhall, M. P., & Bue, B. D. (2009). Machine cataloging of impact craters on Mars. *Icarus*, *203*(1), 77–87.

doi: 10.1016/j.icarus.2009.04.026

Williams, J. P., Pathare, A. V., & Aharonson, O. (2014). The production of small primary craters on Mars and the Moon. *Icarus*, *235*, 23–36.

doi: 10.1016/j.icarus.2014.03.011

Xiao, Z., Zeng, Z., & Komatsu, G. (2014). A global inventory of central pit craters on the Moon: Distribution, morphology, and geometry. *Icarus*, *227*, 195–201.

doi: 10.1016/j.icarus.2013.09.019

## **Chapter 2**

# **The Effect of Lunar Terrain Type**

## **Variation on the Results of**

# **Automated Crater Detection**

### **2.1 Introduction**

#### **2.1.1 Automatic Crater Detection**

Traditionally, impact craters were catalogued and measured manually, with visual inspection of stereoscopic images or utilizing shadow sizes and viewing geometry. This technique was hindered by low resolution imagery and inherent inconsistencies between experts. Around the year 2000, researchers began searching for a fully automated impact crater detection algorithm (CDA). An ideal CDA would be robust under varying surface conditions, independent of lighting conditions or viewing geometry

inherent to the data, and scale invariant. As planetary surface data became digitally available, the fields of image processing and computer science lent their developments to the task of crater detection. Early CDAs (Jahn, 1994; Honda et al., 2000; Alekseev, Pyatkin, & Salov, 1993) worked on photographic imagery, using edge detection and the Hough transform circular feature extraction tools in imagery. More complex systems (Kim et al., 2005; Sawabe, Matsunaga, & Rokugawa, 2006) included machine learning methods and image pipelines to improve their detection capabilities.

Newer systems have sought to improve the state of CDAs by using topographic data, primarily in the form of digital terrain models (DTMs). There are two primary arguments to be made for using elevation data: 1) impact craters manifest with a particularly strong topographic signature, and such a signature can be uniquely detected, and 2) this signature is independent of viewing conditions such as lighting or geometry. The Mars Orbiter Laser Altimeter (MOLA) instrument aboard Mars Global Surveyor (MGS) helped advance progress in the area of crater detection. Initially, 2-D line tracks of elevation were used alongside edge detection schemes to assist in crater identification (Kim, Muller, & Morley, 2000). Later systems (Michael, 2003; Kim, Muller, & Morley, 2004; Bue & Stepinski, 2007) were able to apply advanced image processing techniques to DTMs, with varying results in different study areas. One particular system known as the AutoCrat system (Stepinski, Mendenhall, & Bue, 2009) used a novel mixture of image processing and machine learning to detect craters on the martian surface, using a MOLA-derived DTM. The AutoCrat system combines topographic image processing with machine learning to identify impact craters in a topographic dataset. This system,

employed in this project, will be described in greater detail in the following chapters.

### 2.1.2 CDA Accuracy

The lack of a perfectly accurate and scale-invariant CDA has meant that each detection system contains some error in its detection efficiency, measurement, or both. Whether using a CDA for generating an age estimate or collecting bulk crater morphometric characteristics, these errors are likely to affect the scientific validity of the end product in some way. Many CDAs are compared directly against a set of craters that are manually-labelled by a human expert, often in the form of a catalogue. This process provides a good first order understanding of the CDA's inaccuracy. However, it is also important to control some of the testing variables in a manner to more broadly understand the system's accuracy and response to change. Some of these variables include geomorphic unit, crater size regime, and elevation data type. While there have been many published CDAs, achieving varied levels of success, there isn't yet a consistent framework for the objective evaluation of these systems (Salamunićcar & Lončarić, 2008). Different authors use different test areas, catalogues of ground-truthed craters, and different statistical measures of accuracy. A previous study using elevation data for Mars briefly examined the effect of automated crater detection on four different geomorphic units: volcanic, sea (planitia), and two plateaus (planum) (Yin, Xu, Li, & Liu, 2013). Significant differences in the true and false detection rates were identified based on unit type.

### 2.1.3 Purpose

This paper presents an in-depth evaluation of the AutoCrat system, using high-resolution elevation data for the lunar surface unavailable at the time the system was originally designed. As identified in previous work, there is a need for the objective evaluation and cross-comparison of CDAs (Salamunićcar & Lončarić, 2008). As the AutoCrat system is a particularly pure example of a topographic CDA (using only the topography, and no optical data or template matching), it is investigated using three different ‘characteristic’ lunar terrains to determine the robustness with varying surface type. Detection results can be compared against true images of the terrain to identify the strengths and weaknesses of a topographic approach to crater detection. As an additional objective, the population statistics of detected craters is investigated. By examining these statistics and their related accuracies, one can assess the validity of using the results for surface unit dating, or extraction of bulk morphometric properties. For the sake of comparison, this work does not involve a re-parametrization of the original system, nor does it change the machine learning algorithm used in discrimination. This will be further explained in Section 2.2.1.

We begin by explaining an overview of the project methodology. This includes a thorough discussion of the data type chosen, as well as the motivation for choice of study areas. Both the crater detection and discrimination systems will be described, in addition to the statistical methods chosen for assessment of the results. Finally, the crater detection and measurement results for the three study areas will be presented,

and compared with similar systems previously used in the literature. Closing with a discussion, we present themes on the dependency of terrain morphology on automated crater detection.

## 2.2 Methods

### 2.2.1 Overview

This study applied a set of processing steps on three digital elevation models representing topographically distinct regions of the lunar surface. These steps are grouped and identified with bold letters in the flowchart shown in Figure 2.1. For the sake of consistency, the processing and statistical analysis was replicated in all three areas. One of the three study areas used a DTM with a different resolution; this is further discussed in Section 2.2.2. The project is composed of five major steps: data collection, initial processing, machine learning discrimination, measurement accuracy, and detection efficiency assessment. Generally, each equal-area region was extracted from a larger DTM and preprocessed (box **a**). Next, a basin detection routine is run on the DTM, generating the preliminary crater candidates (box **b**). Using a discrimination algorithm generated by a machine learning method (**c**), the subset of impact craters is extracted from that set. Finally, corrections to the crater measurements are made (box **d**) and accuracy assessment is undertaken on the members of the final catalogue (box **e**).



The statistical analysis of the crater detection output is compared for the three study areas. Results such as detection efficiency and measurement accuracy are the primary focus. Additionally, we can comment on the viability of the results for use in both age estimation and population statistics. A combination of stratified sampling and linear regression are used in performing these assessments.

The AutoCrat system is a two-stage crater detection and discrimination system, using image processing operations on digital images representing elevation profiles of a landscape. This system approaches automated crater detection from a different angle than many other CDAs. In place of complicated systems that require constant re-parametrization for different data sets and on different spatial scales, the AutoCrat uses an impact crater's unique topographic signature as the basis of its modus operandi. In principle, a crater's topographic signature is independent of lighting conditions, viewing geometry or other instrument variation. Using some DTM image processing techniques, the AutoCrat system can find even nested depressions while operating under scale-invariance.

The primary crater finding algorithm, after operating on a DTM dataset, produces a catalog of crater candidate positions, depths and radii, and some mathematical shape descriptors. By seeking all nearly-circular depressions, the first portion of the system seeks completeness through performing an exhaustive search. This should minimize the number of false negative detections (missing craters). A second machine learning stage then uses these quantities to assign a crater or non-crater label to each candidate. This

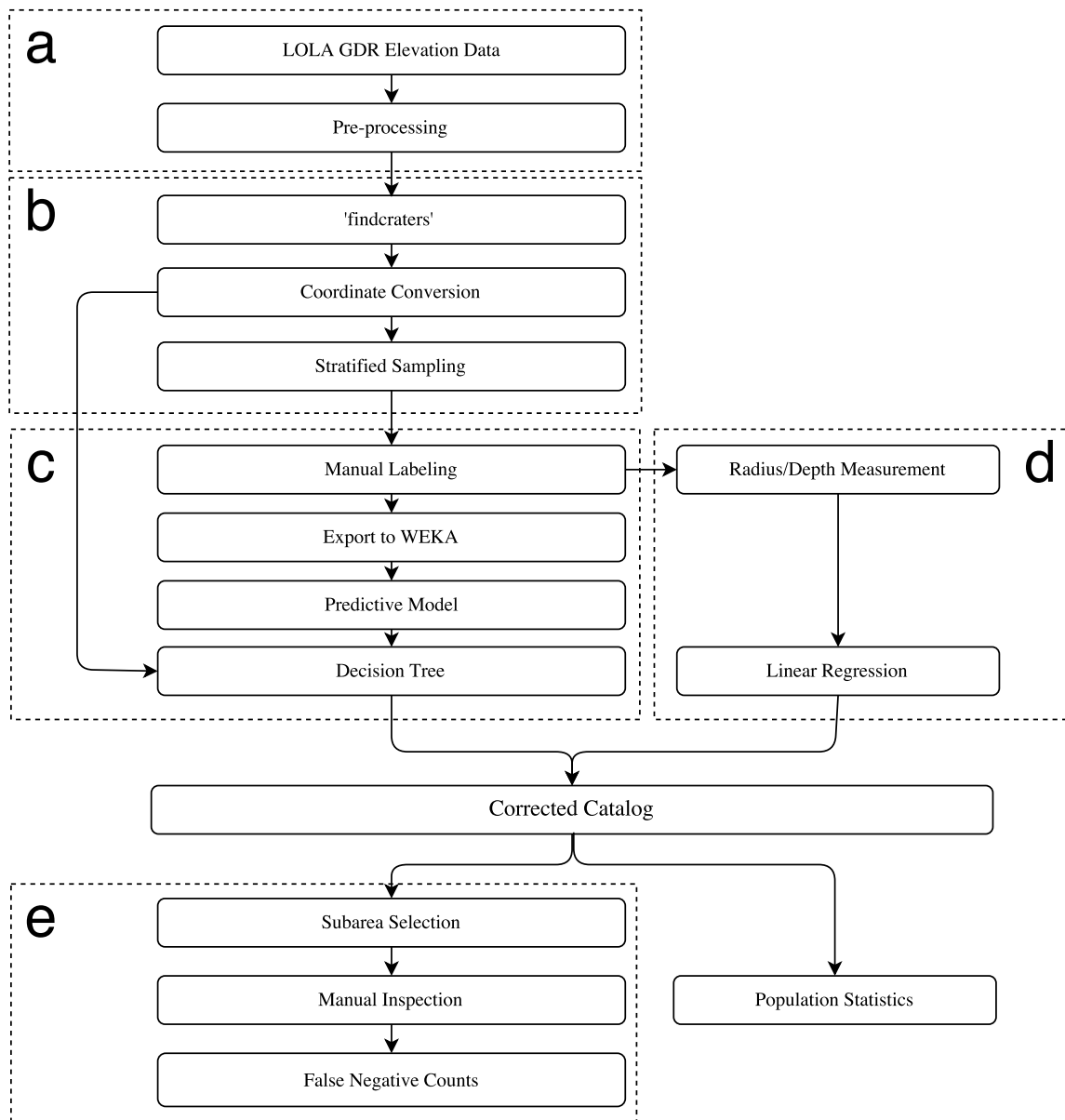


Figure 2.1: Flowchart depicting the major steps of this project. Boxes with dashed lines collect related steps. **a**: data collection and preprocessing **b**: crater detection and categorization **c**: machine learning and training set generation **d**: measurement accuracy and systematic error **e**: detection efficiency assessment

part provides accuracy by refining the preliminary catalogue of depressions, or minimization of false positives. A summarized description of these two steps will be given in the following subsections. For a more detailed description of the algorithm and its nuances, please consult (Stepinski et al., 2009). The two parts of the AutoCrat system are shown in boxes **b** and **c** in Figure 2.1.

### 2.2.2 Study Areas and Data

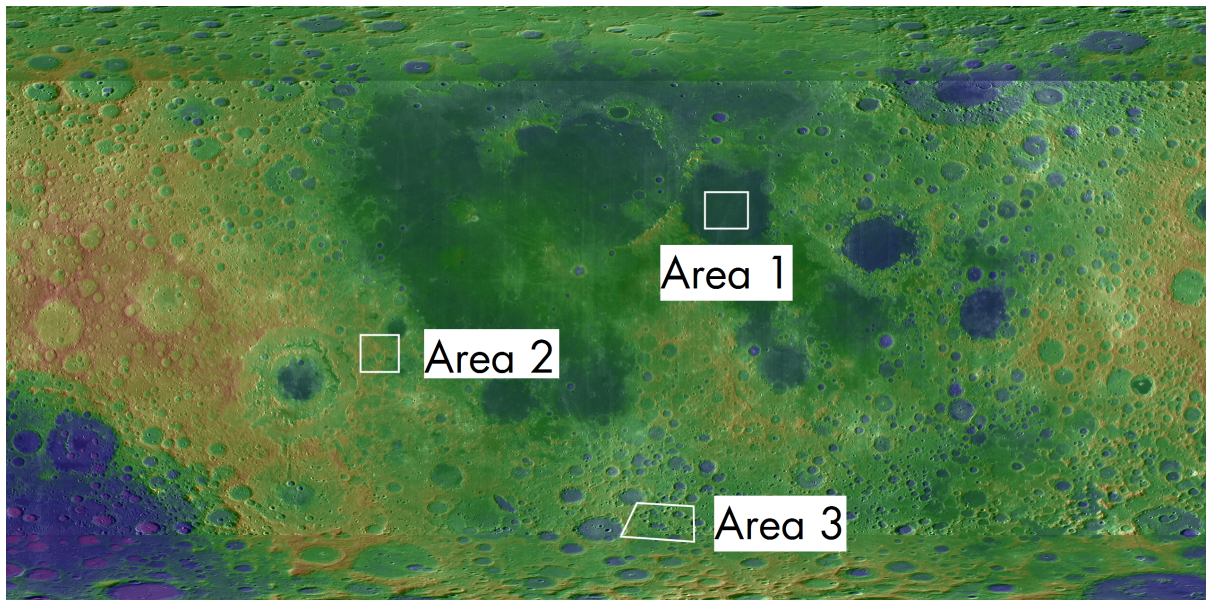


Figure 2.2: Global lunar map in simple cylindrical projection. This map is a combination of 3 optical mosaics consisting of imagery from the Wide Angle Camera (WAC), in addition to colour elevation information from the LOLA instrument (red is high elevation, blue is low). Each study area, demarcated by the white boxes, is roughly 90,000 km<sup>2</sup>.

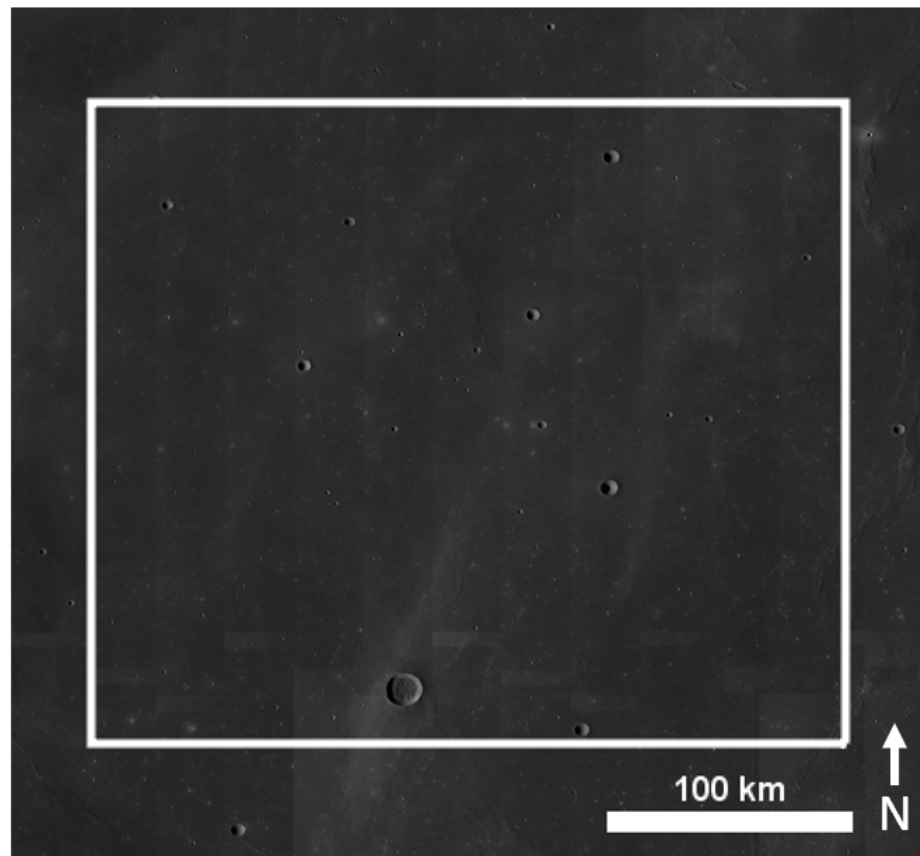
In examining the robustness of a CDA, broad-scale terrain morphology has been identified as a source of inconsistency in crater detection results (Li, Ling, Zhang, & Wu, 2015). As such, it is necessary to apply a CDA to different types of terrain for a more complete understanding of its performance. Additionally, for CDAs that also make measurements on the craters they detect, it is necessary to understand the

measurement response and variation with changing terrain. On the Moon, this includes the two major terrain types that comprise the lunar surface dichotomy: the light-toned, older, more heavily cratered highlands and the darker, younger maria. Additionally, there are the ejecta blankets of large basins, which deposit enormous amounts of material atop the surrounding terrain. These three terrain types collectively characterize much of the Moon's surface, while remaining distinct from each other individually. They can be thought of as a 'basis set' of terrain morphologies. By understanding the response of a CDA to these three units collectively, it can be employed on a majority of the lunar surface in a predictable fashion.

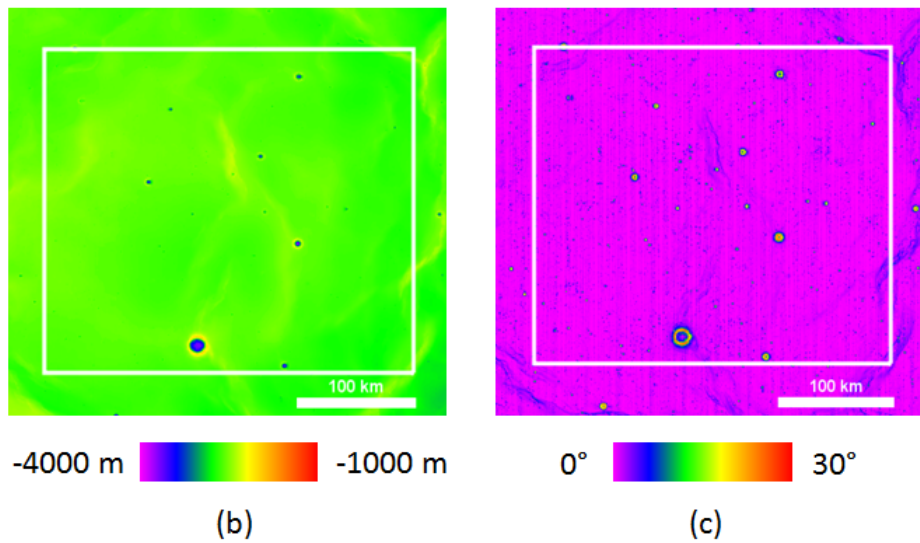
Table 2.1: Details for the 3 selected study areas.

Area	Name	Central lon. (°E)	Central lat. (°N)	Area ( $10^4$ km <sup>2</sup> )	DTM Resolution (m/pix)
1	Mare Serentatis	18.9	27.0	9.00	52.8
2	Oriente Ejecta	287.1	-12.1	9.01	57.2
3	Southern Highlands	1.63	-56.9	9.02	100

## Study Area 1 - Mare Serenitatis



(a)



-4000 m  -1000 m

(b)

0°  30°

(c)

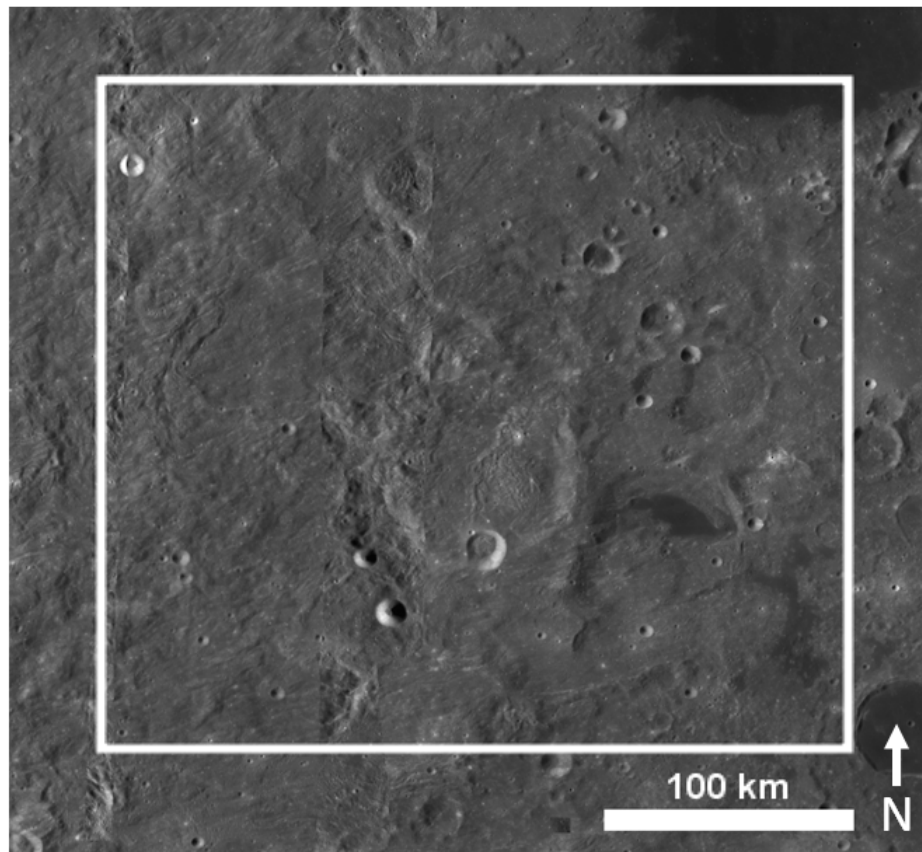
Figure 2.3: (a) Optical WAC mosaic of the Mare Serenitatis study area, at 100 m/pix. (b) Colourized elevation from the LOLA instrument. (c) Slope map, derived from the LOLA DTM.

Mare Serenitatis is a  $\sim 675$  km diameter impact basin on the lunar nearside. It contains a dark-toned surface unit, which together with other similar units on the moon are collectively called ‘maria’. The southeastern edge of Mare Serenitatis was visited by humans during the Apollo 17 mission in 1972. The maria, covering a total of about 17% of the lunar surface, are large basaltic plains that were emplaced by previous volcanic activity (Head III & Wilson, 1992). Typically, mare are relatively flat on large scales, with a more sparsely cratered surface. Previous mapping of Mare Serenitatis has subdivided the mare into many regions with varying ages. Surface unit dating and regional albedo differences imply that volcanism within the interior of the basin took place for hundreds of million years after the formation of the basin, leading to surface ages of roughly 3.4 - 3.8 Ga (Hiesinger, Jaumann, Neukum, & Head III, 2000).

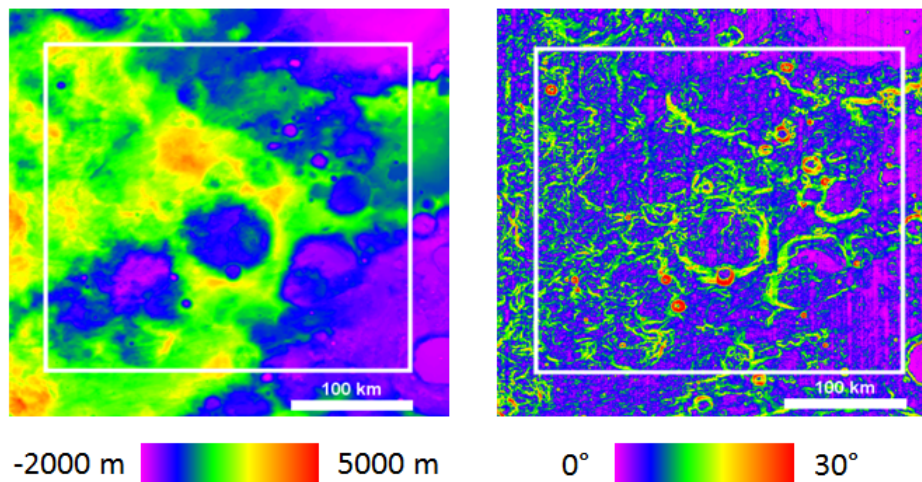
In addition to craters, the surface of the mare can exhibit wrinkle ridges and differences in tone. Few craters on the scale of tens of kilometres can be seen, as visible in the optical image from Figure 2.3. As the spatial scale under study is made more fine, both crater densities and local topographic variation increase. From a crater detection perspective, this unit type should be simplest. This is because of the flat pre-impact terrain and sparse cratering. In the chosen study area comprising of about one quarter of the total area of Mare Serenitatis (Whitford-Stark, 1982), there are no complex craters. Based on a study analysing the pulverization and overturning of the mare regolith, the mare surface should have saturation of craters up to roughly  $D = 200$  m (Hartmann & Gaskell, 1997). This process is what gives the mare its rugged surface on the decameter scale.



## Study Area 2 - Orientale Ejecta



(a)



-2000 m

5000 m

0°

30°

(b)

(c)

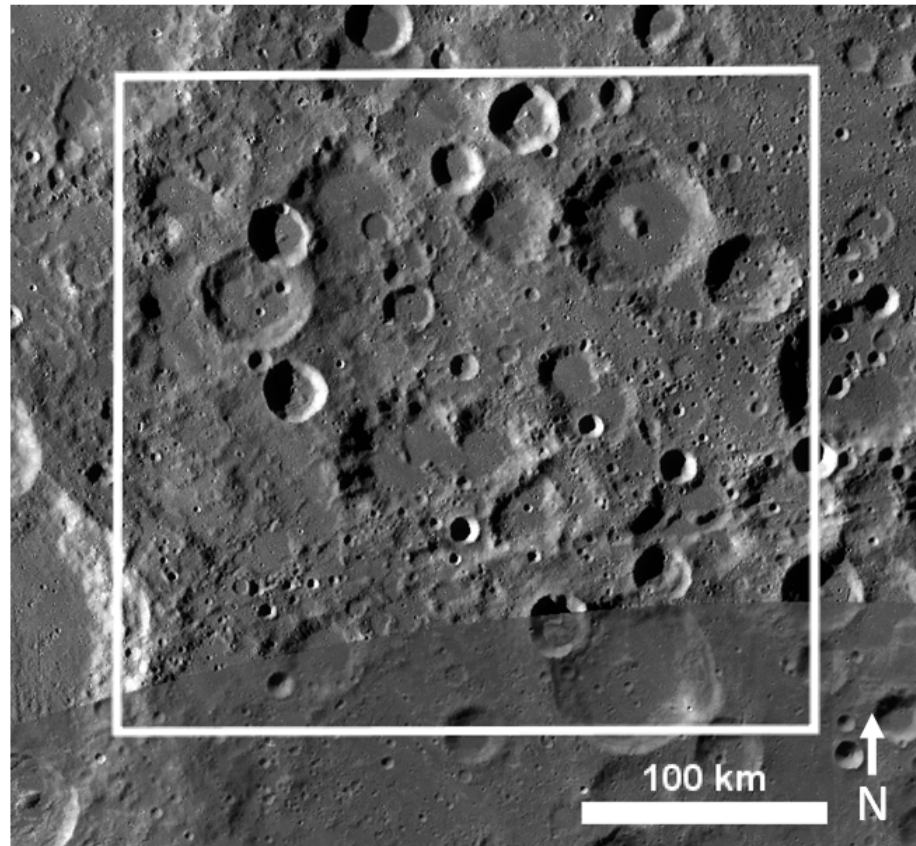
Figure 2.4: (a) Optical WAC mosaic of the Orientale ejecta study area, at 100 m/pix. (b) Colourized elevation from the LOLA instrument. (c) Slope map, derived from the LOLA DTM.

The Orientale Basin is a multi-ring impact structure on the western limb of the lunar nearside. A well-preserved structure, it is thought to have formed at the end of the Late Heavy Bombardment around 3.8 Ga ago (Spudis, Martin, & Kramer, 2014), a period of intense cratering in the early solar system. Measuring around 930 km in diameter, the impact basin's formation into underlying highland material affected much of the surrounding terrain, and lunar surface on the whole. The emplacement of ejecta on the surrounding terrain caused a proximal age resetting by modifying the prebasin population up to a distance of 2 basin radii from the center (Head III et al., 2010). Additionally, Orientale can be used to distinguish the presence of two populations of impactors, with the transition between the two occurring near the time the Orientale Basin formed.

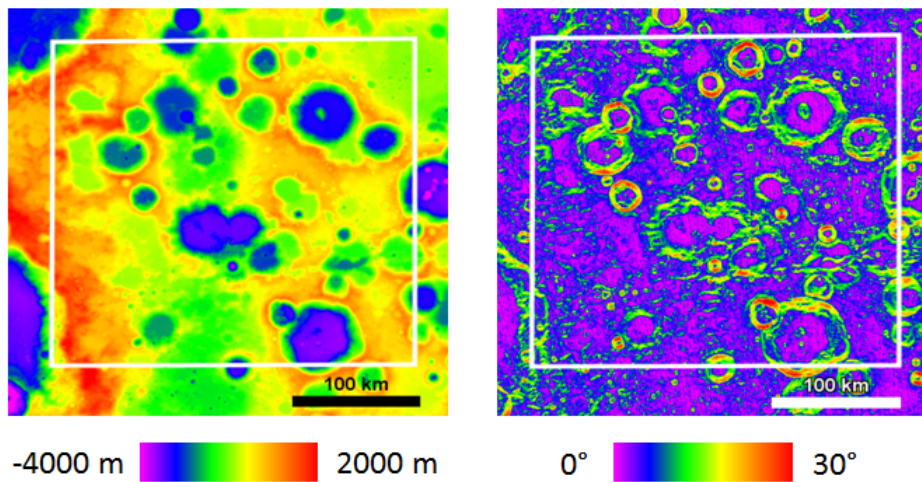
The study area for this project consists of the ejecta to the east and north of Orientale, just outside the Cordillera Mountain ring. This is within the area of age resetting, but since the study area extends significantly far in the radial direction from the center, there is expected to be some local variation in the terrain as the ejecta emplacement is non-uniform. As evident in the slope map in Figure 2.4(c), this terrain is more complex than the mare areas. Significant undulations caused by ejecta and the pre-basin terrain mean greater variation in slope on coarse scales. This will be a significant challenge for the CDA. Large craters that were on the pre-basin terrain may still present topographic expressions (as ghost craters), despite modification from the impact event.



## Study Area 3 - Southern Highlands



(a)



-4000 m

2000 m

0°

30°

(b)

(c)

Figure 2.5: (a) Optical WAC mosaic of the Southern highlands study area, at 100 m/pix. (b) Colourized elevation from the LOLA instrument. (c) Slope map, derived from the LOLA DTM.

The lunar highlands are the oldest terrains on the Moon. Lighter in colour than their mare counterparts, they are primarily composed of anorthosite, an intrusive igneous rock (Pieters, 1986). At higher elevation than the mare, they consist of the southern portions of the lunar nearside, and a significant percentage of the farside. The highlands have dense associations of impact craters, even in some areas achieving complete empirical saturation (Hiesinger et al., 2000) (the addition of any new crater is guaranteed to erase a pre-existing crater of comparable diameter). A state of empirical saturation is challenging for extracting surface ages from crater counts. This is because as new craters are formed, the total crater density for a given size range is unchanged, while the underlying crustal rock is obviously still getting older. Additionally, complex craters are most abundant in the highlands, with varying states of degradation. Many complex craters have been partially superposed by other craters, often eliminating significant portions of the rim.

The highlands study area is on the lunar nearside, towards the south pole. This study area poses a challenge for a CDA, as it has the highest requirement of scale-invariance. The presence of craters at a wide variety of diameters requires the system to be responsive to both small simple craters, as well as large complex craters. Additionally, the age of the surface indicates a larger array of crater degradation states - ranging from relatively young and intact to significantly destroyed craters.

## Source Data

The data chosen for this study are reduced and gridded DTMs generated by the LOLA instrument. These DTMs are freely available from the LOLA PDS Data Node at [imbrium.mit.edu](http://imbrium.mit.edu). LOLA, a laser altimeter and passive infrared radiometer on board the Lunar Reconnaissance Orbiter (LRO), is a time-of-flight laser altimeter which fires pulses of 5 infrared lasers onto the lunar surface at a frequency of 28 Hz (Neumann, 2011). By using 5 spots to sample the ground, the instrument can collect both along-track slope and across-track slope, in addition to measuring the spacecraft-to-ground distance (which can be used to calculate absolute elevation). Since the Moon does not have the notion of ‘sea level’, all elevations are measured relative to the selenoid, or a perfect sphere of radius 1,737 km.

As the spacecraft travels, measurements from the 5 spots are collected and used to construct a 2-D line of elevation along the spacecraft’s ground track. Over time, these tracks have collected over the lunar surface. Due to LRO’s arrangement in a polar orbit, the tracks are densest at the poles, and spaced most widely at the equator. From these tracks, a minimum-curvature interpolation is performed to fill in the space between tracks, providing a complete 3-D elevation model for the lunar surface. It is important to note the mathematical method of generating the 3D surface, as this will have an effect on the precision of the crater detection results. For any given pixel in the DTM that does not contain a direct LOLA elevation measurement, a mathematical rule is used to generate a value for that pixel that maintains a constant second surface

derivative, and minimizes overall curvature. While this provides a smooth surface, it makes assumptions about the surface that may not necessarily be true. In addition, craters whose diameters are significantly smaller than the cross-track width may be missed entirely.

To start, the appropriate DTM is selected from the repository. This project uses gridded data record (GDR) DTMs, which have undergone higher-level processing which includes binning, resampling and map projection. First, we located the appropriate parent DTMs which contain the study areas. For the equatorial study areas, DTMs in cylindrical map projection that span  $45^\circ$  of latitude and  $90^\circ$  of longitude were downloaded. The study area in the southern highlands is found in a south polar stereographic DTM. Next, the region of interest containing the study area was cropped out with a mask in ArcGIS. The digital values in the raster were then converted to elevations with respect to the selenoid by using a bulk raster calculation. Finally, the DTM is exported as an ASCII text file, which allows it to be read by the Cratermatic software. The DTM is now ready for crater detection.

For the sake of comparison with previous studies using DTMs of the martian surface generated in a similar fashion (from the MOLA instrument), only LOLA data were chosen for this study. Future work in automated crater detection can be expanded to DTMs generated by other means and instruments. Further discussion of the consequences of using laser altimetry data is discussed further on.

### 2.2.3 Crater Detection

This section details the process of identifying the crater candidates from a topographic landscape. Beginning with a DTM covering a region of interest, a series of image processing tasks are used to extract and compile the basins that form the initial list of candidates.

#### Crater Detection - findcraters

The first stage of the AutoCrat system uses a topographic basin finding routine called `findcraters`, which is within the Cratermatic topography analysis toolkit. This routine (box **b** in Figure 2.1) performs a series of image processing operations on a topography dataset whose pixel values represent the elevation for that pixel; either directly sampled from a LOLA track or via interpolation. Typically, identifying isolated basins in a topographic landscape can be performed with the use of a watershed algorithm (Meyer & Beucher, 1990), which segments the image into catchment basins. This approach is unsuccessful in the case of nested basins, since the watershed algorithm will simply combine the two into a single depression. The Cratermatic system uses an image convolution to address this issue. Shown in Figure 2.6 is the processing for a small sample region of the lunar surface.

The first step in crater detection is to apply an image convolution called a C-transform. This transform is similar to a Gaussian blur, and acts to smooth the DTM, removing depressions below a specific size  $\lambda$ , where  $\lambda_0 = 5$  pixels. This blurring

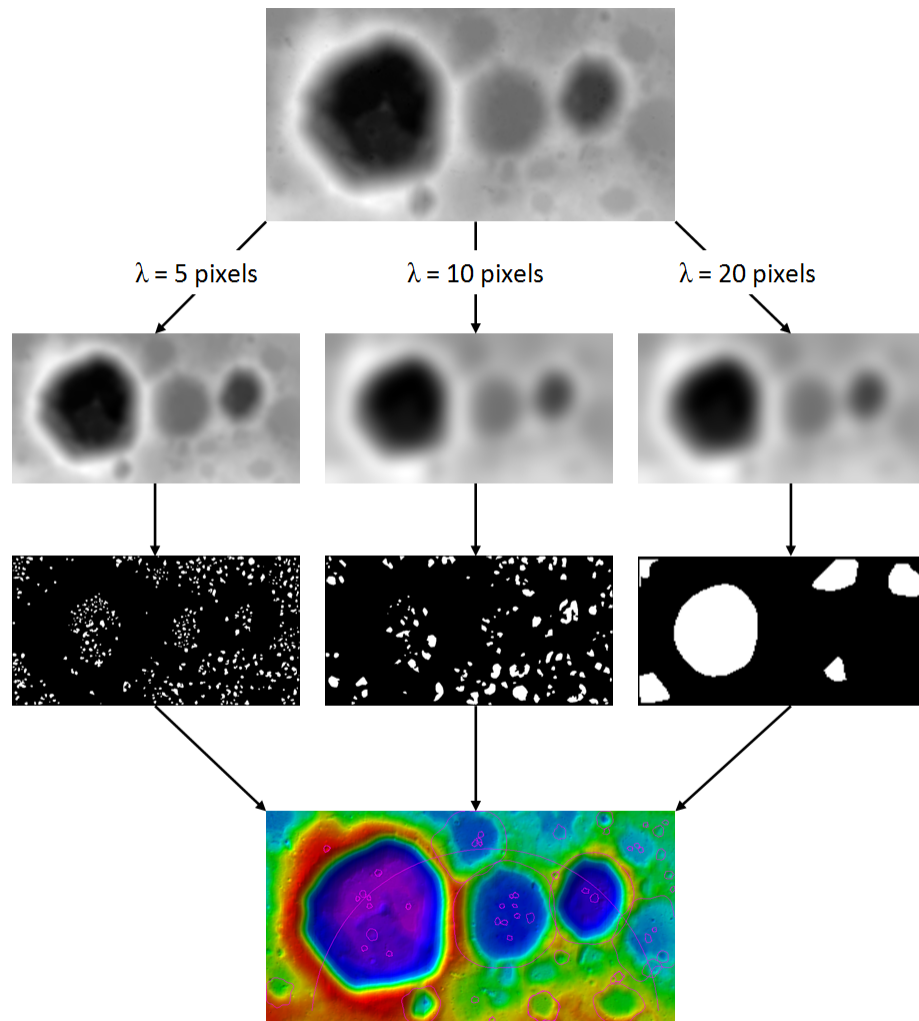


Figure 2.6: A sample abbreviated image processing sequence for the Cratermatic system. Using an input DTM, the image is blurred at different spatial scales to separate nested depressions. Craters detected in each blurred landscape are recombined for the final catalog. Not all blurred landscapes are shown for the sake of simplicity.

smooths the landscape emphasizing depressions on the order of size  $\lambda$ . This helps to identify nested depressions by ignoring depressions that are significantly larger or smaller. After finding an anchor pixel around which all points are concave up, the algorithm incorporates other pixels belonging to the crater region, stopping after a slope threshold. The centroid is identified as the center-of-mass of the crater (as the profile can be non-circular), and a polar function is used to represent the rim shape. A Fourier expansion is used to build the polar function. Coefficients derived from the Fourier expansion are used to describe the shape; a perfect circle would have all but the first coefficients equal to zero. Therefore we can use the coefficients to describe deviations from a perfect circle. Two particularly useful quantities derived from the coefficients are  $m_2$ , which describes the elongation of the crater shape, and  $m_3$ , which describes its ‘lumpiness’.

When the preliminary crater catalogue is produced, the crater locations are given in the image space. For example, the catalogue will list one depression in the set as an entry of the form  $\{\# \text{ ID, } x, y, r, \text{ area, depth, } m_2, m_3\}$ . Coordinate conversions are then performed on the catalogue to translate each centroid in  $x$  and  $y$  image space to a planetocentric latitude and longitude, and to convert the radius to meters.

### **Stratified Sampling**

The final step before categorizing the basins as crater or non-crater is to perform a size-based sampling on the preliminary results. The purpose of these sampling sets is

twofold and will be explained in greater detail in Section 2.2.5. Due to the uneven detection distribution as a function of crater size, special care must be taken to sample appropriately over the range of possible crater sizes in the landscape. To assess measurement accuracies as a function of crater size, the craters need to be split up into groups based on their diameters. We start by dividing craters into three major groups: small craters ( $D < 300$  m), intermediate craters ( $300 \text{ m} < D < 4$  km), and larger diameter craters ( $D > 4$  km)(Melosh, 2011). From there, the intermediate class of craters is split into three 1-kilometre sized groups between 1 and 4 km. The crater size categories are shown below in Table 2.2.

Table 2.2: The six size-based crater classes. Craters are separated into classes based on their diametric size to perform a stratified sampling and build the training sets.

Class	A	B	C	D	E	F
Diameter	$D < 300\text{m}$	$300\text{m} \leq D < 1\text{km}$	$1\text{km} \leq D < 2\text{km}$	$2\text{km} \leq D < 3\text{km}$	$3\text{km} \leq D < 4\text{km}$	$D \geq 4\text{km}$

The Cratermatic system works on iteratively larger radius scales, doubling the radial pixel size for the Gaussian convolution each round. After reaching a size limit which is dictated by the smaller dimension of the input image, the algorithm combines all the detected basins from each scale and outputs a catalogue identifying the basin location and quantities related to its size and shape. As mentioned previously, this is (ideally) an exhaustive list of *depressions*, a subset of which will consist of only the true craters. In the next section, we describe how machine learning can be used to extract this subset.



### 2.2.4 Crater Discrimination

The cadence of successful space missions over the past few decades has dramatically increased the wealth of data that is available for many solar system objects. This large amount of data has proven challenging to analyse in a time-efficient and thorough manner. Not only is there simply too much to look at, but often the nature of the analysis is quite complex and thus difficult for a human analyst to complete in significant enough amounts over reasonable timescales. Advancements and techniques in artificial intelligence and automated data analysis can be invaluable tools for such tasks. Previously, machine learning has been applied to problems in planetary science such as automated geologic mapping (Stepinski, Ghosh, & Vilalta, 2007) and surface unit annotation (Ghosh, Stepinski, & Vilalta, 2010). In the context of automated crater detection, machine learning can be used as a powerful tool for devising a set of rules to discriminate craters from non-crater depressions.

#### Training Set Construction

In order to later instruct the machine learning algorithm how to discriminate crater from non-crater, it is necessary to build training sets for each study area, consisting of examples of both craters and non-craters. The training sets also provide an opportunity to study the measurement accuracy of the system; this is elaborated on later in Section 2.2.5. For each study area, the list of crater candidates including their category assigned from Table 2.2 is generated. Next, 20 members from each available size class

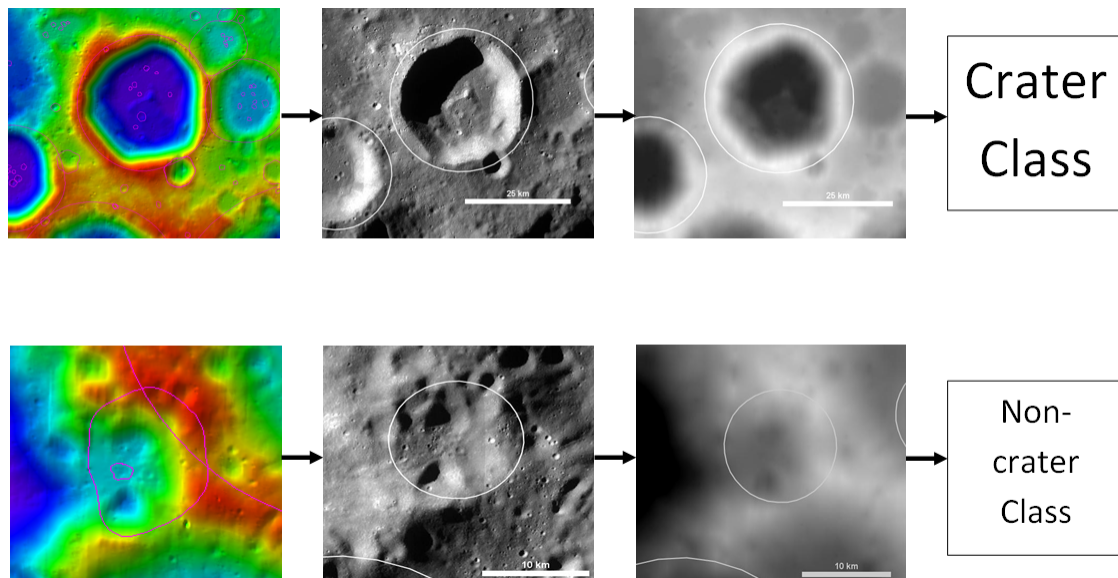


Figure 2.7: Graphic representation of the training set generation process. Each basin that was selected via sampling is hand labeled as a crater or non-crater by overlaying the rim on both optical and elevation data in JMARS.

are selected to begin constructing the training sets for each area. Each training set is then exported to JMARS for visualization. Starting at the top of the list, each crater is superimposed on optical data sets (primarily NAC image strips and WAC global mosaics) and inspected. For the high-resolution NAC images, the incidence angle search parameter is restricted to higher angles - typically between  $45^\circ$  and  $80^\circ$  to accentuate topography. The candidate is then hand labelled as either crater or non-crater. It is important to note here that individual craters of all types are labeled a crater, including secondaries (craters formed by the impact of ejecta from another crater) and ghost craters (craters heavily mantled by lava or other materials, leaving a faint rim). As long as it represents one single crater, it is counted. Differentiating between crater types is left for future work, and is also discussed further in Section 2.4.

After each category has had its 20 craters labelled, additional craters from the preliminary catalogue are added as necessary to provide equal numbers of true and false craters. This is necessary as the basin-detection algorithm typically finds more non-craters than craters (the degree to which this happens is a function of the category), and thus balancing of the training set is necessary to improve the robustness of the discrimination process. Exhaustive sampling of the larger crater categories (E and F) occurred in study area 1, where large diameter craters aren't as populous. After manual labelling and any additions to the list, the training sets for study areas 1-3 had 143, 133 and 128 members respectively.

### **Decision Tree**

As mentioned previously, machine learning in this context can generate a set of discrimination rules to separate crater from non-crater. These rules, when structured together, form the basis of a classification model. Fundamentally, this involves using a set of quantified characteristics for which an impact crater will have some set of identifying values. The pertinent values for this project are the quantities produced by the Cratermatic algorithm: depth, diameter, depth-to-diameter ratio, and two quantities derived from the Fourier coefficients described in Section 2.2.3 that are used to describe the rim shape. The AutoCrat system uses a specific type of machine learning algorithm, known as a decision tree classifier, to build the set of discriminatory rules. This algorithm, called J48, is an open source Java implementation of the C4.5 algorithm (Quinlan, 1993). The algorithm is available through the Waikato Environment for Knowledge Analysis (WEKA), an open source software suite

containing many tools for data mining (Hall et al., 2009).

In principle, the J48 algorithm works by finding attributes that most effectively split the data into different classes. Given  $n$  examples, the training set can be expressed as a set  $S = (s_1, s_2, \dots, s_n)$ . Each sample  $s_i$  in the training set contains  $m$  attributes, in the form  $(x_{1,i}, x_{2,i}, \dots, x_{m,i})$  with one of the attributes being the class to which the sample belongs. In this project, the set  $S$  is the training set with  $s_i$  being a single basin, either crater or non-crater. The  $x_j$  represent attributes for a single crater such as depth, diameter, and the shape descriptors, as well as class. The decision tree is constructed by generating nodes at each point where the data is effectively split by a single attribute. The best attribute is chosen by having the highest normalized information gain, which is a statistical quantifier of the level of purity in a group when split a particular way. After splitting on a node, the algorithm works iteratively, splitting again and again until all samples from the set have reached the end of the tree. These endpoints are called “leaves”. The decision tree for study area 3 is shown in Figure 2.8.

Each decision tree was implemented as a MATLAB algorithm to assign a crater or non-crater status to the original catalogue. This allows for the assignment of classes to thousands of crater candidates in a very short time frame, in a process hereafter called ‘pruning’. As with other remote sensing classification methods, it is possible to quantify the effectiveness of the classification model in doing its work. For example, confusion matrices and other performance metrics can be generated from the training sets. This is covered in Section 2.3.1.

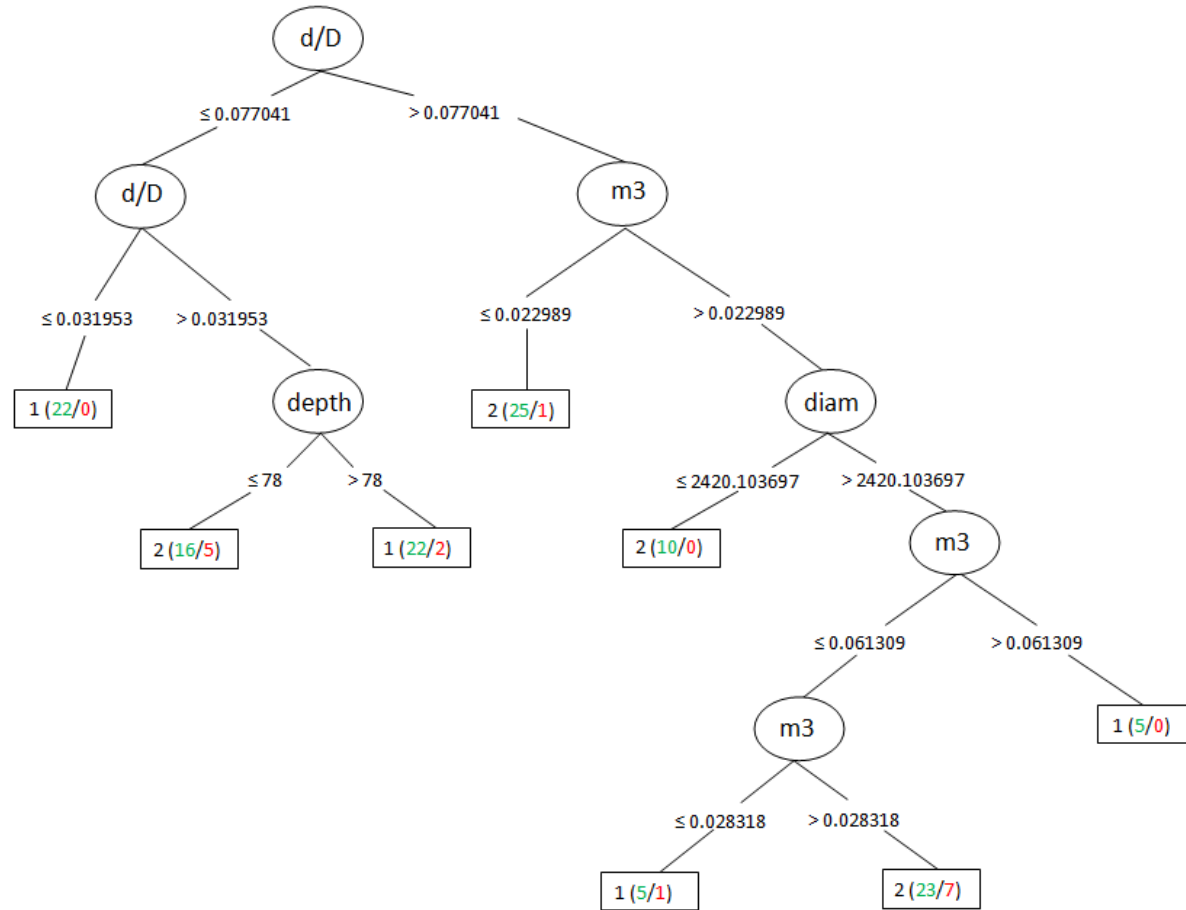


Figure 2.8: Decision tree built using the training set for the Southern Highlands study area. This tree was generated using the J48 machine learning algorithm in the WEKA software suite. Nodes are ellipses and leaves are rectangles. At each node, the numbers in parenthesis represent the number of instances reaching the leaf (green) /the number of misclassified instances at the leaf (red).

## 2.2.5 Accuracy Assessment

The AutoCrat system is a crater detection and measurement system. To understand the scientific conclusions drawn from the population statistics built using the measurements, it is useful to note the general measurement error and uncertainty as a function of terrain type and crater size. As mentioned previously, the training sets for each area serve two purposes: to build the classification model and to assess measurement error. In this section we break down how the training sets and pruned crater catalogue will be used to assess different kinds of accuracy.

### Measurement Accuracy

In Section 2.2.4, the process for assembling a training set is described. Each crater candidate in the training set is manually labeled a crater or non-crater. In addition, we can make measurements in the case where the candidate is a true crater. As two of the most common descriptors of an impact crater are its diameter and depth, it is important to compare the AutoCrat's measurements with 'true' values. For each crater in the training sets that was assigned to the crater class, its diameter and depth were measured manually. These measurements were made against the WAC GLD100 DTM available on the LROC Quickmap at <http://target.lroc.asu.edu/q3/>. Diameters were measured as an average of the N-S and E-W directions in cases where there existed obliquity or significant pre-impact surface gradients. Depths are measured as maximum elevation differences from rim crest to floor, in the same fashion as they are reported by the AutoCrat system. Additionally, any other notable and prominent properties of the

crater (such as being particularly inclined, or a ghost/buried crater) are recorded.

After collection of all manual measurements, it is possible to use simple linear regression to relate the measured values by hand (representing ‘true’ values) with the detected values provided by the CDA. This linear regression is used later to correct depths and diameters and provide a more accurate catalogue for use in analysing the bulk statistics. For the linear regression, the y-intercept is fixed to zero in each study area. This is to prevent unphysical measurements such as negative depths or radii from arising after correction. The results of the linear regression are described in Section 2.3.2.

### **Detection Efficiency**

The detection efficiency describes the sensitivity of the CDA to variations in crater size and terrain type. A perfectly efficient system will be invariant to changes in crater size, location, state of degradation, or other modification. Understanding the detection efficiency revolves around three primary measurements: the number of true positive detections (detection of a true crater), false positive detections (crater detections that do not correspond to an impact feature), or false negatives (missed true craters). To succinctly express the detection performance as a function of radius, metrics have been derived that combine the three quantities together (Shufelt & Mckeown, 1993). They are the detection percentage  $DET$ , the quality factor  $Q$ , and the branching factor  $B$ , shown below.

$$DET = 100 \cdot TP / (TP + FN) \quad (2.1)$$

$$Q = 100 \cdot TP / (TP + FP + FN) \quad (2.2)$$

$$B = FP / TP \quad (2.3)$$

The detection percentage,  $DET$ , has been renamed to prevent confusion with the diameter,  $D$ .  $DET$  gives an understanding of the CDAs capability to exhaustively find craters, neglecting false positives or ‘over-detection’. Including the FP rate allows us to calculate the quality factor  $Q$ . This value is increased by exhaustively finding the craters, but is reduced by the detection of non-craters. It gives a more complete picture of the CDA’s performance and can be used as a threshold for practical application of the results (Kim et al., 2005). The branching factor  $B$  expresses the tendency for over-detection in the system, and is a measure of the number of false craters detected for every true detection.

These quantities are associated with specific uses of the results of wide-scale crater detection and measurement. For example, if the results of the automated crater detection are being used to be able to discern secondaries or different types of craters, completeness is particularly important, so the detection percentage is valued. If crater densities are being measured to generate a relative age, completeness above a specific



diameter with minimal FPs is important, and so the quality factor is the best descriptor. Those collecting bulk crater shape parameters may be concerned with only a set of true craters (regardless of how complete the set is), and so a low branching factor is considered priority. A perfectly ideal system should have very high values for  $DET$  and  $Q$  with a low  $B$  value.

Since the total number of detections in each study area is high, to measure the values for TP, FP and FN, we subdivide each study area into nine equal-area portions, roughly 100 km x 100 km. Two of these study areas are chosen at random, and an exhaustive vetting of the crater detection results is performed within each. First, all detections whose centroids fall within that subarea area are loaded into JMARS. Next, each crater is manually inspected over WAC and NAC images, and is labelled with TP or FP. This is performed for all craters inside the box. Next, we count the FNs - craters that missed detection. As crater populations become increasingly dense as diameter is reduced, we set a minimum diameter of 600 m for false negatives. While this sets a lower limit for calculating  $DET$  and  $Q$ , it does not affect  $B$ . Additionally, it is below the diameter threshold for the age calculation for lunar surface units in (Neukum, 1983). Due to low populations of large craters in some study areas, all FNs with  $D > 5$  km are recorded.

## 2.3 Results

In all three areas, considerable numbers of impact craters were detected. As expected, their size distributions generally reflected the terrain type, but with varying response in

terms of sensitivity. Table 2.3 shows a total of over 34000 craters detected between the three areas. The crater discrimination process removed between one third to four fifths of the initial detected basins, depending on the area. Mare Serenitatis had the most detected craters, and the ejecta of Orientale the least.

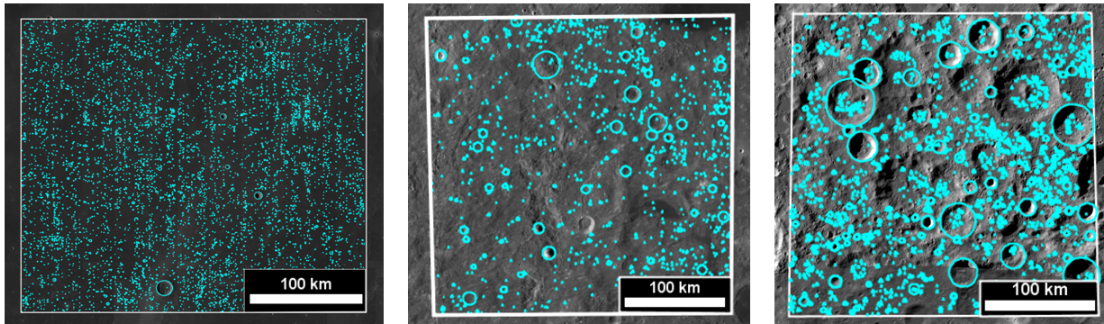


Figure 2.9: Automated crater detection results for study areas 1-3 (L to R, respectively). These maps show the complete crater catalogues after pruning. Background imagery is a 100 m/pix WAC mosaic.

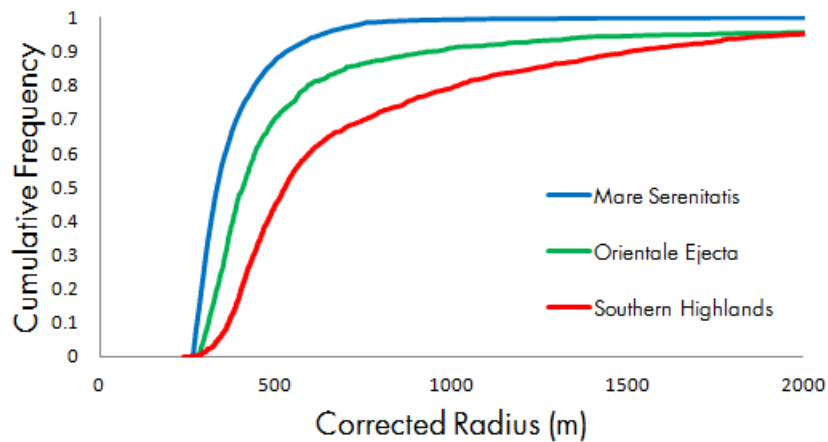


Figure 2.10: Post-pruning cumulative distribution function (CDF) for the three study areas.

The large numbers of crater detections in Mare Serenitatis is somewhat counter-intuitive at first. This is because the Mare areas are typically younger in age than the highlands, and should thus be less heavily cratered. This paradox is resolved when

looking at the size-frequency distribution of craters in the area. There is a correlation between small- $D$  detections and flat areas with minimal slope variation. This is evident in Figures 2.9 and 2.10, which show large populations of sub-kilometre diameter craters in study area 1. Visible as well are small-crater abundances in other flat areas, including the floors of larger craters in study area 3.

Table 2.3: Numerical crater detection results for each study area.

	Initial detections	Post-pruning	Post/initial (%)	Diameter range (km)	Depth range (m)	Crater area (% of total)
<b>Area 1</b>	29264	6237	21.3	0.530 - 15.255	12 - 1675	4.67
<b>Area 2</b>	2318	1031	44.5	0.559 - 26.957	36 - 3228	4.99
<b>Area 3</b>	2771	1831	66.1	0.474 - 44.259	16 - 3215	20.21

The smallest craters which can be detected reliably are on the order of  $D \approx 600$  m. While craters with depths in the tens of meters were detected, the smallest reliable depth measurements could only be made for craters with  $D > 1$  km. However, the local surface conditions and location of the crater with respect to LOLA ground tracks is also important in assessing the detectability of the crater. For all study areas, impact craters with  $D < 300$  m did not survive the crater discrimination process. Upon inspection, it was found that the only candidates at this size were small basins formed along the linear direction of the LOLA tracks (shown in Figure 2.11). As the ground tracks make directly-sampled measurements of the surface elevation, the resolution is much higher on the track, and lower in the between-track areas where interpolation is used. This causes small ‘basins’ to form where there is a small but significant region of surface concavity along the track direction. These are then classified as basins by the `findcraters` algorithm. This effect is strongest in flatter areas, which resulted in the

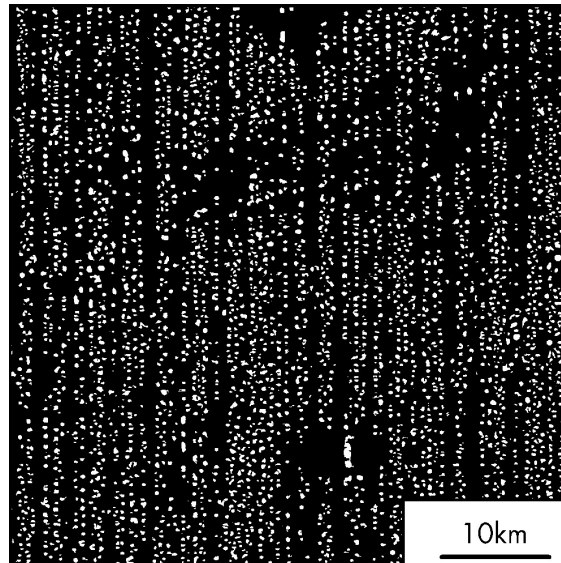


Figure 2.11: Small basins detected by Cratermatic. Note the vertically-linear arrangement. These basins do not survive the crater discrimination process, and are artefacts caused by the laser altimetry data type.

most detections in Mare Serenitatis. While these detections dominated the initial results, they did not survive the pruning process. This is because they did not satisfy the depth-to-diameter restrictions implemented by the decision tree.

### 2.3.1 Detection Performance

#### Measurement Accuracy

The results of the automated crater depth and radius measurement show a strong correlation to ‘true’ measured values from the training sets. The AutoCrat system’s ability to make accurate measurements was dependent on terrain and size. For a simple bowl crater on flat pre-impact terrain, there is little systematic error in measurement - in this case, it would be limited to the inherent error in the DTM. As crater size

decreases, or the pre-impact and surrounding terrain becomes more complex, the measurement error increases. Depth measurement accuracy suffers more as craters become smaller, since a crater a few hundred meters across will have a depth on the order of tens of meters or less, approaching the resolvable limit for z-axis variation.

Table 2.4: Linear regression results for radius and depth in all three study areas. These fits were used to correct the values of the pruned crater catalogue. The true values are measured manually and are used to corrected the detected values.

Area		Linear fit	R <sup>2</sup>
Mare Serenitatis	Radius	$r_{true} = 1.0013 \cdot r_{det}$	0.9739
	Depth	$d_{true} = 0.8906 \cdot d_{det}$	0.9618
Orientale Ejecta	Radius	$r_{true} = 0.9758 \cdot r_{det}$	0.9858
	Depth	$d_{true} = 0.9500 \cdot d_{det}$	0.9828
Southern Highlands	Radius	$r_{true} = 0.8775 \cdot r_{det}$	0.9970
	Depth	$d_{true} = 0.7416 \cdot d_{det}$	0.9220

Table 2.4 shows the results of the linear fitting and correction. These fits show a very strong correlation to true values, which allows for a bulk correction to be made to all values. In general, the system is over-estimating radii and depths. There is a slightly weaker correlation for depth measurements - this is most noticeable in the Southern Highlands. This is likely caused by the way the CDA defines crater profiles. The routine that is adding pixels to the crater adds pixels in a cellular automata-like fashion, as long as each added pixel satisfies some criteria related its local topography. In some areas, this process is causing an over-reach of the crater profile, especially in areas where there is a non-uniform surface gradient nearby, such as a steep hill near the rim of the crater. Thus, crater depths are being measured systematically deeper than they should be. This also affects the radius measurements; this is discussed further in

## Section 2.4.

An additional issue in automated crater measurement arises when trying to assign a radius value to a crater with non-perfect circularity, or one that is superposed on terrain with a significant pre-impact slope. While the fraction of lunar craters that form elliptical geometries is quite low (around 5%) (Bottke Jr., Love, Tytell, & Glotch, 2000), there is a significant fraction of craters that form on sloped terrain. This effect is particularly strong in the sub-km regime. As this deforms the standard circular expression of a crater, the radial distance from centroid to rim can change as a function of azimuthal angle. While the Cratermatic system uses a clever polar expansion to model the non-perfect circular rim, the returned radius is calculated using the total pixel area, as follows:

$$R_{crater} = \sqrt{\frac{A_{crat}(pix)}{\pi}} * res \quad (2.4)$$

where  $A_{crat}$  is the crater area in pixels and  $res$  is the DTM resolution. This gives the radius of an equal-area crater with perfect circularity. As mentioned in Section 2.2.5, the radii collected in the accuracy assessment stage were averages of the N-S and E-W radii for craters with any significant non-circularity or incline. Therefore, the goodness-of-fit of the linear regression for radius is affected by this dissonance in measurement technique. Standardization of measurements is a great benefit of automating processes, but it is important to examine the appropriateness of the measurement method for a variety of cases.

It was not possible to collect reliable depth measurements at all size scales. Sub-km diameter craters could not have their depth measurements compared reliably against the WAC GLD100 DTM, and thus the automatically detected depths were not useful for determining a systematic measurement error. For that reason, craters with  $D < 1$  km are not included in the analysis of bulk crater statistics such as depth-to-diameter ratio.

### **Decision Tree Accuracy**

The purpose of the decision tree is to analyse the quantitative characteristics of each detected basin, and decide whether the basin should be categorized as an impact crater. The basis for its judgement is a set of rules that it constructs after being provided a training set of examples. While an ideal decision tree working with a perfectly representative dataset would generate perfect classification results, there are factors that can cause less-than-ideal classification, and impurities in the results. Misclassification can have different root causes. There may be limitations in the size of training set or number of attributes provided. Additionally, the wide variety of crater forms (that also vary with size) may not be fully represented in the training set. Finally, there is the chance that some training set members are incorrect; this is more likely as the craters become very small and difficult to differentiate from other topographic expressions. Misclassification of craters ultimately affects the distribution of TP, FP and FN quantities.

Many measures that exist to quantify the performance of both supervised and unsupervised classification schemes in remote sensing image analysis can be applied to decision trees as well. Some simpler measures include the TP and FP rates, or the ratio of true or false classifications to the total number in a given class, respectively. We can use additional measures such as the precision, F-measure and Kappa statistics. The precision of a classification system describes the number of accurate decisions per classification (Buckland & Gey, 1994). Combining the precision and TP rate gives the F-measure, which is the harmonic mean of the two values. The harmonic mean is the reciprocal of the arithmetic mean of the reciprocals, which in this case is equivalent to  $2 \cdot \text{Precision} \cdot \text{TP Rate} / (\text{Precision} + \text{TP Rate})$ . Finally, for each decision tree, we can define the Kappa statistic. This number quantifies the agreement between the classifier and the training set, *minus* the probability that the agreement occurred by chance. For the sake of comparison, these metrics are divided into two tables representing all ‘crater’ class (Table 2.5) and ‘non-crater’ class (Table 2.6) assignments for the three study areas.

Table 2.5: Detailed accuracy statistics for ‘crater’ class, for each study area.

<b>Crater Class Statistics</b>					
<b>Study Area</b>	<b>TP Rate</b>	<b>FP Rate</b>	<b>Precision</b>	<b>F-Measure</b>	<b>Kappa</b>
<b>Mare Serenitatis</b>	0.891	0.023	0.961	0.925	0.880
<b>Oriente ejecta</b>	0.973	0.186	0.867	0.917	0.800
<b>Southern highlands</b>	0.953	0.203	0.824	0.884	0.750

Of primary concern is the ability of the system to label true craters. Table 2.5 shows trends in accurate crater labelling. Using the F-measure as a first order accuracy metric, we see that positive crater identification is most successful in the mare areas, and becomes



Table 2.6: Detailed accuracy statistics for ‘non-crater’ class, for each study area.

<b>Non-crater Class Statistics</b>					
<b>Study Area</b>	<b>TP Rate</b>	<b>FP Rate</b>	<b>Precision</b>	<b>F-measure</b>	<b>Kappa</b>
Mare Serenitatis	0.977	0.109	0.935	0.956	0.880
Oriente ejecta	0.814	0.027	0.960	0.881	0.800
Southern highlands	0.797	0.047	0.944	0.864	0.750

more challenging as the terrain becomes more complex, with worst performance in the highland areas. Upon deeper inspection, the TP rate for Mare Serenitatis is significantly lower than for either other study area, despite having a very low FP rate. The nature of the distribution of craters in Figure 2.10 can be invoked to explain this. As crater size decreases, more craters are detected for each study area. This increase is the steepest for Mare Serenitatis, and even while sub-km detections are the most reliable in this area out of the three study areas, the detection reliability is still diminished as crater size diminishes. However, the exceptionally low FP rate more than compensates for this in calculating the F-measure.

### **Quality Measures**

As described in Section 2.2.5, each study region was subdivided into nine equal-area subareas for a more complete study of the crater detection results. Within these subareas, every detection was labelled as a true or false positive. Additionally, all missed craters (false negatives) with a diameter greater than 600 m that could be identified using NAC imagery were labelled. These results could then be quantified using the quality metrics described previously. Images of the subarea analysis are shown in Figure 2.12.

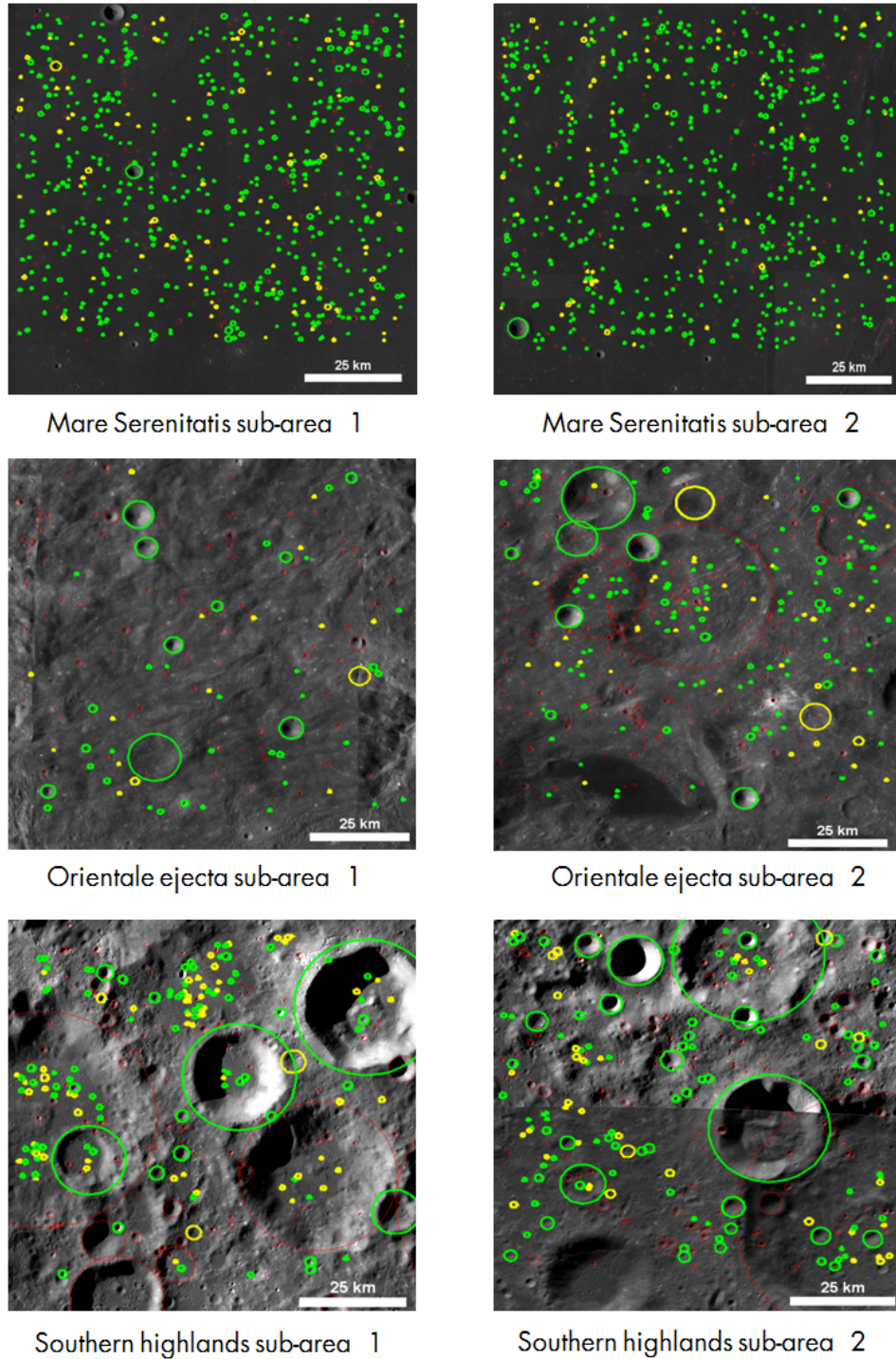


Figure 2.12: Assessment of crater detection in six sub-areas. Each sub-area is 100 x 100 km. Green crater profiles represent true positive detections, yellow are false positive, and red are false negative with  $D > 600$  m.

Starting at 600 m (the smallest diameter for which FNs were counted),  $DET$ ,  $Q$

and  $B$  are calculated using the tallied values for TP, FP and FN detections. These quantities were then re-calculated repeatedly with the minimum diameter increasing in 200 m increments, up to  $D = 5$  km. For  $D = 5$  km, all TP, FPs and FNs were counted for the study area to guarantee a sufficient sample size. These three quantities are then plotted in Figure 2.13. For practical applications, a critical value for  $Q$  has been identified as 80% (Kim et al., 2005). There is a general convergence towards this value in all three study areas around  $D = 50$  pix (roughly  $D = 2.9$  km for study areas 1 and 2, and  $D = 5$  km for area 3), with Mare Serenitatis achieving this value earlier, at  $D = 30$  pix. The branching factor, which ideally should be low, converges to below 0.2 at around  $D = 30$  pix.

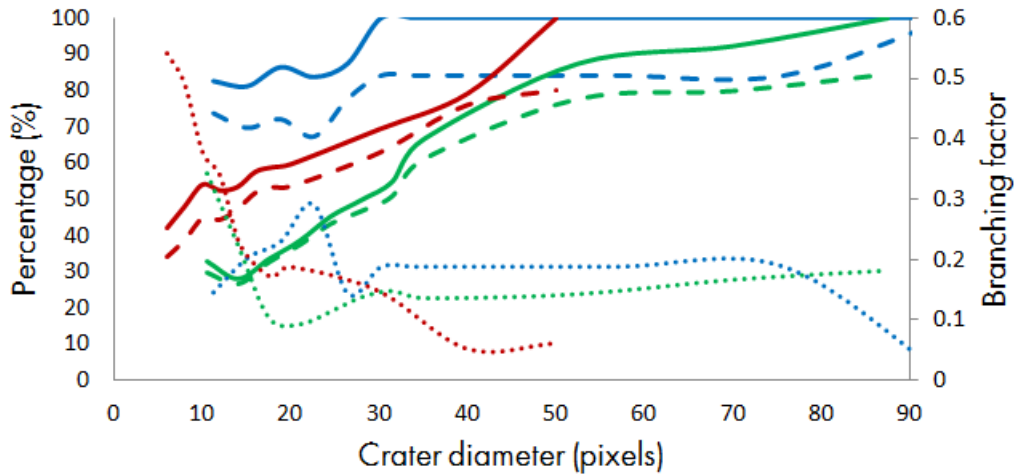


Figure 2.13: The three performance metrics  $DET$  (solid line),  $Q$  (dashed line) and  $B$  (dotted line) as a function of crater radius in pixels. Mare Serenitatis is represented in blue, Orientale in green, and the Southern highlands in red.

The crater detection efficiency demonstrated a differential response to size sensitivity. In Mare Serenitatis, a reasonably satisfactory quality factor  $Q$  was established, even when close to the detection limit. The system was able to consistently pick up sub-km

craters, and was primarily limited by the nature of the LOLA data. Multiple craters that superpose one another, especially on or very near a LOLA ground track, often fooled the system into thinking it was a singular basin. An example of this is shown in Figure 2.14. However, for crater sizes below 20 pixels, this area had a higher  $B$  value, indicating more false positive detections. An alternative way to analyse the contribution of FNs is by examining the vertical separation between  $DET$  and  $Q$  in Figure 2.13. A larger separation between the lines indicates a greater detracting to the quality by FNs.

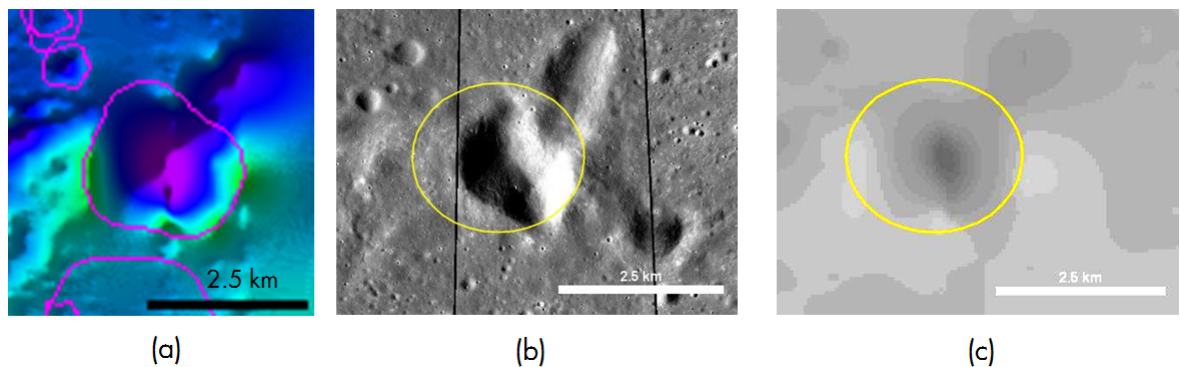


Figure 2.14: A false positive detection in Mare Serenitatis. (a) The Cratermatic detection result (b) The feature, as seen in LROC NAC optical imagery (c) The feature as seen in the raw LOLA DTM.

### 2.3.2 Crater Statistics

Automation allows us to perform a relatively simple (and sometimes complex) task much more quickly and consistently than a human analyst. Measuring the depths and diameters of a dozen craters is likely better done manually, but if the emphasis is on collecting the same information for thousands of craters, an automated approach is preferable. In addition to speed, automation also provides a consistency that human experts can lack; the subjective expertise of a manual count can be detrimental to the

results, especially when comparing counts between different experts.

In addition to identifying craters in a topographic landscape, this project also collected bulk measurements. These measurements consist of a crater's centroid position on the landscape, the radius and depth, and a few rim shape descriptors. Of particular importance here are the radius and depth, which are used in calculating the depth-to-diameter ( $d/D$ ) ratio for craters. This quantity was first examined in the mid 1960's as a method of understanding crater scaling laws - how depth and diameter relate as crater size changes (Baldwin, 1965). For each study area, all pruned craters with diameters above 1 km were corrected using the regression corrections outlined in Table 2.4, and then included in a histogram. The minimum diameter restriction is due to the inability to accurately measure depths for craters in classes A and B using the GLD100 DTM. This histogram is shown in Figure 2.15.

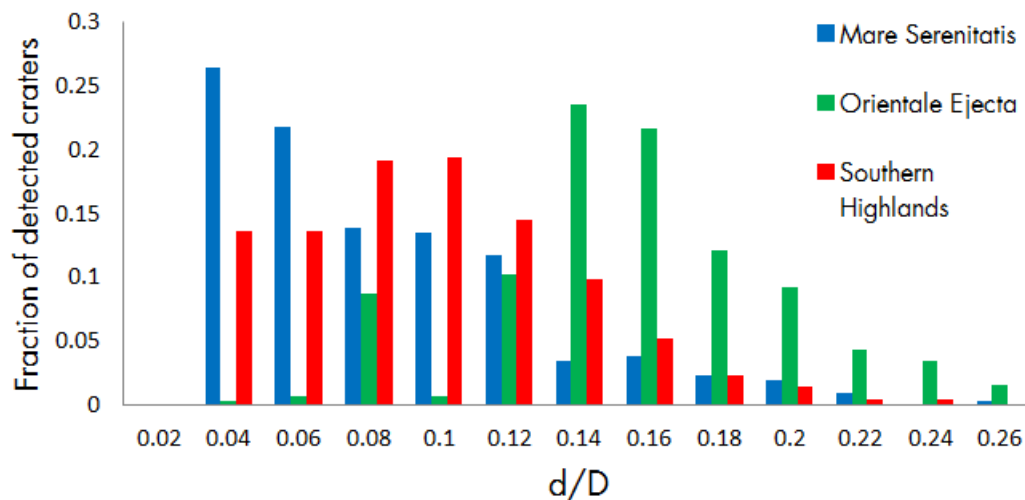


Figure 2.15: Histogram of the depth-to-diameter ratio for all pruned craters with  $D > 1$  km.

As mentioned previously, crater counting has an important role in deducing age differences between different surface units. Using the crater detection results, a SFD (as shown in Figure 1.3) was plotted for the 3 study areas. This plot is shown in Figure 2.16.

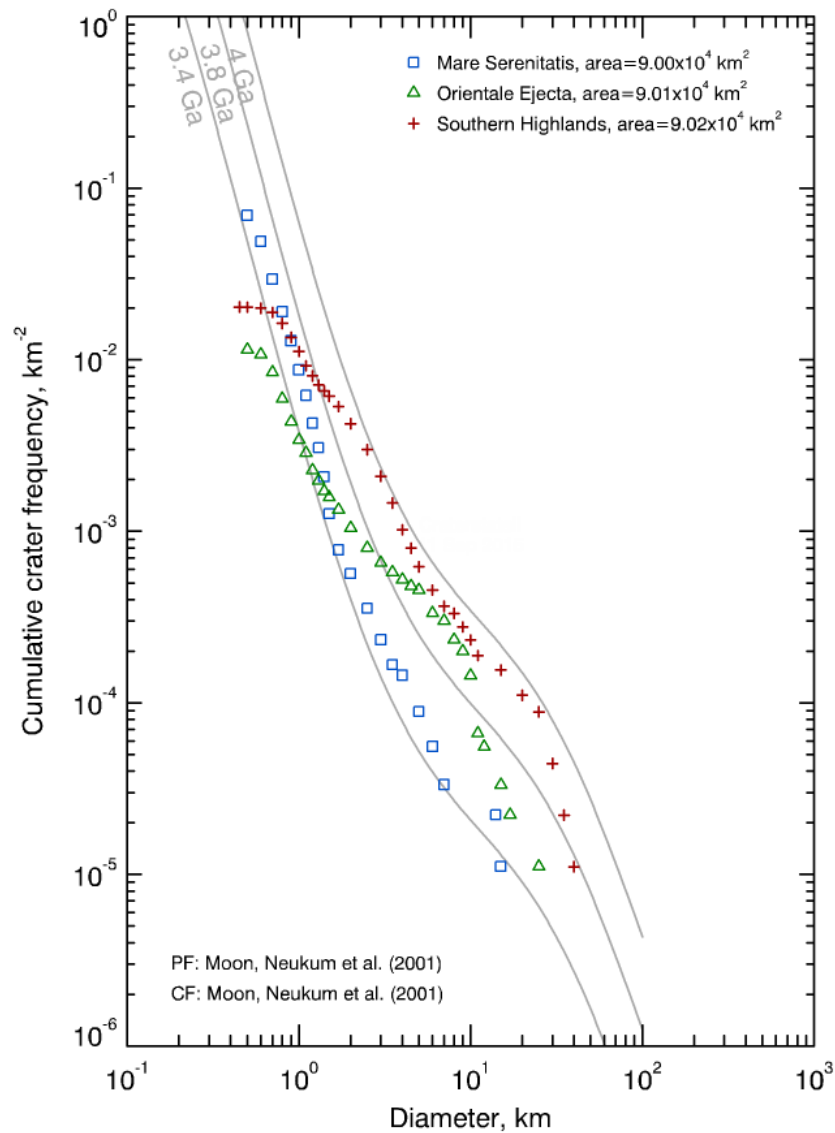


Figure 2.16: SFD for each study area, plotted using CraterstatsII (Michael & Neukum, 2010). Three isochrons representing crater populations for 3.4, 3.8 and 4.0 Ga old surfaces are plotted as solid grey lines. Crater frequencies are binned in pseudo-logarithmic fashion (18 bins per decade).

## 2.4 Discussion

### 2.4.1 Comparison With Other Published Work

Automated CDAs using a variety of methods and data sources have been investigated prior to this study. Their results have often been published using similar performance metrics as those used in this project. For the purpose of putting these results into context, it is important to directly compare the results (both in measurement quality and detection efficiency) with some of these other systems.

#### **AutoCrat for Martian Craters (Stepinski 2009)**

The project described in this chapter applied the AutoCrat system to high-resolution LOLA laser altimetry data for the Moon. At the time this system was created, laser altimetry data was only readily available for Mars (LRO's orbital insertion was one month after the publication of the AutoCrat paper). We can thus do a direct comparison between the results of Stepinski et al., 2009 (henceforth referred to as Step09) and this study.

The primary data source used in Step09 is MOLA altimetry data, collected and processed in much the same way as LOLA data. The resolution of the dataset is 128 ppd, which equates to 463 meters per pixel at the equator. This is roughly eight times coarser than the LOLA data used in study areas 1 and 2, and 4.6 times the resolution of that used in study area 3. Using this coarse elevation data, the author's result was a minimum detectable crater size of  $D \geq 2.5$  km, with reliable depth measurements for

craters with  $D \geq 5$  km. Additionally, Step09 lists future application of the system by stating “[u]sing a DTM with resolution of 100 m, the AutoCrat is expected to catalog craters with  $D \geq 500$  m and to obtain a reliable values [sic] of depths for craters with  $D \geq 1$  km”. This is remarkably close to the limits identified in this study. In the most successful unit type (mare), craters with  $D \geq 600$  m were reliably detected, and depths were reliable for craters  $D \geq 1$  km. However, for the more complex areas like the ejecta of Orientale, or heavily cratered units like the highlands, these limits were not so applicable. Similarly to the results of this study, the Martian craters detected in Step09 found a systematic overestimation of both radius and depth, and could successfully account for these overestimations by applying a linear fit to the values. From Table 2.4, the average overestimation of the value for diameter and depth is 10% and 14%, compared with 15% and 30% from Step09.

Table 2.7: Comparison of performance metrics with other published values. \* The values listed for the three study areas here are for craters with  $D \geq 4$  km.

Study	DET (%)	Q (%)	B	Reference
Mare Serenitatis*	100	100	0	This study
Orientale Ejecta*	92	80	0.17	This study
Southern Highlands*	79	76	0.05	This study
Bue	74	61	0.29	(Bue & Stepinski, 2007)
Barlow	75	75	N/A	(Barlow, 1988)
Barata	64	31	1.65	(Barata, Alves, Saraiva, & Pina, 2004)
Kim	88	78	0.15	(Kim et al., 2005)
Li	87	82	0.07	(Hui Li, Jihao Yin, & Zetong Gu, 2014)

It is also possible to directly compare the detection performance metrics listed in Section 2.2.5 with other published values. Table 2.7 displays the values for  $DET$ ,  $Q$  and  $B$ . The values listed for each study area apply to craters with a diameter of 4 km



or larger, corresponding to class F craters. The choice of 4 km for these values is to provide the quantities after they have stabilized, as they change rapidly approaching the detectable size limit. The system performs in a manner that is often superior to previous systems, but with variation and shortcomings associated with unit type. The high branching factor  $B$  in Orientale indicates the systems inability to discriminate crater from non-crater basin in this area. The highlands region does not suffer as much from this problem, but the lower  $Q$  value indicates that missing craters are driving up the FN value, so completeness is an issue in this region. The system excels in mare areas for larger craters, and even for craters of much smaller size, as can be seen tracking the metrics in Figure 2.13.

### 2.4.2 Scientific Integrity of Results

As mentioned previously, CDAs are used to generate crater counts and crater morphometry statistics on large scales, collecting information on a scale that would be time-inefficient for a human analyst. Systematic inaccuracies in an automated CDA can have an affect on the scientific integrity of the results. For example, knowledge of the total detection efficiency would be crucial for performing an automated crater count of very small craters, perhaps in trying to date a small unit such as an impact melt sheet or crater ejecta blanket. The same applies with making size measurements, as inherent error has to be corrected before comparing values.

### Size-frequency Distribution and Crater Counting

To determine the age of a surface unit, or compare the ages of two separate units, crater counts can be used to build a size-frequency distribution (described in detail in Chapter 1). The SFD, a plot of the cumulative crater frequency per square kilometer against crater diameter, is usually displayed in a log-log graph. SFDs for the three study areas are shown in Figure 2.16. Isochrons, representing distributions that correspond to particular unit ages, are plotted as well. An ideally perfect crater count, including only primary craters on a limited surface in question, should adhere perfectly to the isochron representing the unit's age. Since CDAs are (at present) imperfect systems, there are deviations from the isochron that represent inconsistencies in the system. Also, large enough count areas will eventually cover units of different ages. This effect is particularly important in Mare Serenitatis; this is discussed in Section 2.2.

Mare Serenitatis, represented with blue boxes in Figure 2.16, closely follows the shape of its nearest isochrons. With ages ranging from 3.4 to 3.8 Ga, the SFD falls in the 3.5 - 3.6 Ga region of the plot. With no significantly large deviations, it begins to slowly round off as the minimum diameter is reached. This turnover effect is caused by the inability to resolve small craters, and is an observational bias (Basaltic Volcanism Study Project, 1981). As a SFD moves from the lower left to the upper right of the plot, it increases in age. As expected, the next SFD is for study area 2, the ejecta of Orientale. As this impact occurred roughly 3.8 Ga ago, it would be expected that the ejecta blanked, forming without any superposed craters, should show a crater distribution that reflects this age.

The SFD for study area 2 is heavily modified from an expected distribution. Below 3-4 km, there is a paucity of craters in the distribution. Analysis of the detection efficiency in previous sections has shown that this is just a shortcoming of the detection system, resulting in false negatives. Around  $D = 8$  km, a kink is visible in the distribution. This is likely a manifestation of buried craters being detected. Even with ejecta infilling the underlying craters, their topographic expression was still significant enough so as to be classified. Further evidence for the detected presence of buried craters is provided in the following subsections. Finally, the distribution again deviates towards lower densities as the crater diameter regime is increased. The Southern Highlands, as the oldest unit, is the right-most SFD on the plot. Building SFDs of very heavily cratered units for the purpose of age dating is tricky, as the unit will eventually adopt a state of cratering equilibrium. In this case, as cratering continues, the destruction of other craters will leave the relative distribution unchanged. For this reason, any age determined from a SFD in an area that has reached crater density equilibrium is to be taken as a lower limit to the age of the unit (Richardson, 2009). The SFD for the highlands can be most closely associated with an age of 3.9 - 4.0 Ga. This SFD also experiences a negative deviation from expected densities at larger diameters. Larger complex craters, especially those with partially destroyed rims, were not adequately detected and thus under-represented in the plot.

### **Depth-to-diameter Relationship**

Figure 2.15 shows the results of the bulk collection of depth-to-diameter values for all craters with  $D > 1$  km. This lower size limit was dictated by the ability to accurately

measure depths in the WAC GLD100 DTM. All depths and diameters had already been corrected following the linear regression procedure outlined in Section 2.2.5 and the values given in Table 2.4. To first order, the histogram shows a relationship between  $d/D$  and unit type - craters formed on the highlands are generally deeper than mare craters. The modal bins for the Southern Highlands and Mare Serenitatis were 0.10 and 0.04, respectively. This relative result is in agreement with previous work on the  $d/D$  relationship for lunar craters (Kalynn, Johnson, Osinski, & Barnouin, 2013). While not entirely understood, this is likely to be an effect related to target material strength (Senft & Stewart, 2007). As surface materials in the highlands are composed of overlapping layers of regolith (loose material ejected from impacts) collected over time, they form a less cohesive target than would a thinner regolith layer over top stronger mare basalts. Additionally, the plot shows a sharp peak for Mare Serenitatis in the low  $d/D$  regime. This result is likely a convolution of detection bias as well as the physical nature of the craters. As can be seen in CDF shown in Figure 2.10, the sensitivity of the system results in around 90% of detections in this study area with  $D < 1$  km. At small scales, craters are more substantially modified by infilling processes which reduce their total depth. Also, many smaller craters are likely to be secondaries, which form shallower craters as a result of their slower impact velocity (which is by necessity less than the escape velocity for the body). These factors affect the ability to measure the depth accurately at small sizes.

The Orientale Ejecta study area also manifests interestingly on the  $d/D$  histogram. The only bimodal distribution, it contains peaks at  $d/D = 0.08, 0.14$ . The distribution

is heavily skewed towards the deeper side of the histogram. Unlike the skewed distribution of measurements in the mare area, this is likely caused by systematic measurement effects. The impact that formed the Orientale Basin impacted a surface type that would have been similar to the highland, throwing large amounts of heterogeneous ejecta on the surrounding landscape. This material should thus have a consistency that is not too dissimilar to the highland material. By a basic argument of the target material consistency, we would expect a similar  $d/D$  distribution. Figure 2.4 shows the highly undulating terrain in the Orientale study area, with great topographic variation at many length scales. Craters forming on these slopes form inclined basins, where the downslope rim is at a significantly different elevation than the upslope rim. As the crater-region growing process adds pixels, it searches for pixels with a positive gradient from the anchor point. This causes an overreach in the diameter measurement, which has a lowering effect on the  $d/D$  value. However, in regions where the crater is on a severely sloped pre-impact surface, the overreach also significantly increases the depth of the crater, more than offsetting the diameter overreach. The net result is an inflated value for  $d/D$ .

### **Buried Craters in Orientale Ejecta**

The SFD for the Orientale ejecta study area is discussed in Section 2.4.2. This discussion details the presence of a kink in the SFD, showing an inflated presence of craters in the area of  $D = 8$  km. This surplus is thought to be a result of buried craters, which were significantly large as to survive partial burial by the ejecta

emplaced during the impact that formed the Orientale basin. It is also possible that the first mode at  $d/D = 0.08$  in the distribution is related to this kink. A crater with a diameter of 8 km and  $d/D = 0.15$  (near the larger mode) should have a depth of 1200 m. The same crater, if it were part of the  $d/D = 0.08$  population, would have a depth of 640 m. The discrepancy between these two depths is 560 m.

Previous work has quantified the volume of ejecta generated by the impact. The study area in this project is contained within a region between the Cordillera ring (the outermost circular ridge formed by the basin) with radius  $R_{CR} = 465$  km, and  $3R_{CR}$ . The volume of ejecta within this region has been given as  $2.9 \cdot 10^6$  km<sup>3</sup> (Fassett, Head III, Smith, Zuber, & Neumann, 2011). While the thickness scales with radial distance through a power law, we can use a crude approximation to determine the average thickness over the area. Treating the total area as an annulus bounded by the concentric circles formed by  $R_{CR}$  and  $3R_{CR}$ , the total area can be calculated as  $8\pi R_{CR}^2 = 5.43 \cdot 10^6$  km<sup>2</sup>. Dividing the volume of the ejecta by the area yields an average ejecta thickness  $t_{avg} = 534$  m. This value is quite close to the depth discrepancy identified in Figure 2.15, and provides evidence for the significant detection of buried craters in this region. Detection and characterization of buried craters using automated methods can have useful application in verifying quantitative models of ejecta emplacement.

## 2.5 Conclusion

Automated crater detection systems are tasked with the challenge of identifying and measuring impact craters under a wide variety of conditions. Surface unit type, crater morphology (related to size, degradation/modification state), crater density and data type all play a role in determining the effectiveness of these solutions. By applying an elevation-based CDA to high resolution data for different lunar surface types, we find:

1. Different lunar surface units have an effect on the results of automated crater detection. These differences include detection efficiencies, minimum detectable crater size, and accuracy of crater measurements. Mare units provide the best performance, with quality decreasing as crater saturation equilibrium is approached.
2. The AutoCrat system can be used to quickly and accurately generate SFDs that reflect *relative* ages between units. Absolute ages can be approached as detection quality improves. However, the balance of false positives and false negatives tends to underestimate absolute unit ages.
3. Bulk crater statistics can be collected efficiently with an incorporation of systematic measurement error. The results of bulk statistics collection is a convolution between the population quantities in question, and inherent detection biases.
4. The results of bulk statistics collection can be used to verify models of crater formation, ejecta emplacement, and other cratering processes.

While imperfect, progress by many researchers has been made to improve these systems and begin to more broadly introduce their use in crater studies. The choice of using an automated system versus a manual count must be made by weighing factors such as the counting area, minimum crater size desired in the count, or the specificity of measurements to be made. However, automated systems will become more useful moving forward as new datasets become available. The CaSSIS instrument on board ExoMars Trace Gas Orbiter (TGO), which will insert into Mars orbit in October 2016, will greatly expand the wealth of terrain information we have about the planet. The imaging system will be able to generate stereo pair images of the surface at high quality at a resolution of 6 meters. This will allow for many crater counts and characterization at more fine scales.

There is still work to be done in improving CDAs. Aside from simply improving detection capabilities, their measurement capabilities will need expansion. As there are many ways to quantify the morphology of a crater, it will be important to create standards for measurement that can be adhered to, so that measurements made by different systems can be directly compared. Comparative studies of impact crater detection on terrestrial worlds and icy satellites are also important for working towards a more complete solution. Ultimately, the task of crater detection and characterization can be incorporated with geologic mapping and unit annotation to improve automated methods in planetary science and exploration.



# References

- Alekseev, A. S., Pyatkin, V. P., & Salov, G. I. (1993). Crater detection in aero-space imagery using simple nonparametric statistical tests. In *Proceedings of the 5th international conference on computer analysis of images and patterns* (pp. 793–799).  
doi: 10.1007/3-540-57233-3\_110
- Baldwin, R. B. (1965). The Crater Diameter-Depth Relationship from Ranger VII Photographs. *The Astronomical Journal*, 70(8), 545–547. doi: 10.1086/109778
- Barata, T., Alves, E. I., Saraiva, J., & Pina, P. (2004). Automatic recognition of impact craters on the surface of Mars. *Image Analysis and Recognition*, 489–496.
- Barlow, N. G. (1988). Crater size-frequency distributions and a revised Martian relative chronology. *Icarus*, 75(2), 285–305. doi: 10.1016/0019-1035(88)90006-1
- Basaltic Volcanism Study Project. (1981). *Basaltic Volcanism on the Terrestrial Planets*. New York: Pergamon Press, Inc.
- Bottke Jr., W. F., Love, S. G., Tytell, D., & Glotch, T. (2000). Interpreting the Elliptical Crater Populations on Mars, Venus, and the Moon. *Icarus*, 145(1), 108–121.  
doi: 10.1006/icar.1999.6323

- Buckland, M., & Gey, F. (1994). The relationship between Recall and Precision. *Journal of the American Society for Information Science*, *45*(1), 12–19. doi: 10.1002/(SICI)1097-4571(199401)45:1(12::AID-ASI2)3.0.CO;2-L
- Bue, B. D., & Stepinski, T. F. (2007). Machine detection of Martian impact craters from digital topography data. *IEEE Transactions on Geoscience and Remote Sensing*, *45*(1), 265–274. doi: 10.1109/TGRS.2006.885402
- Fassett, C. I., Head III, J. W., Smith, D. E., Zuber, M. T., & Neumann, G. A. (2011). Thickness of proximal ejecta from the Orientale Basin from Lunar Orbiter Laser Altimeter (LOLA) data: Implications for multi-ring basin formation. *Geophysical Research Letters*, *38*(17), 1–5. doi: 10.1029/2011GL048502
- Ghosh, S., Stepinski, T. F., & Vilalta, R. (2010). Automatic annotation of planetary surfaces with geomorphic labels. *IEEE Transactions on Geoscience and Remote Sensing*, *48*(1), 175–185. doi: 10.1109/TGRS.2009.2027113
- Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., & Witten, I. H. (2009). The WEKA Data Mining Software: An Update. *ACM SIGKDD Explorations*, *11*(1), 10–18.  
doi: 10.1145/1656274.1656278
- Hartmann, W. K., & Gaskell, R. W. (1997). Planetary cratering 2: Studies of saturation equilibrium. *Meteoritics & Planetary Science*, *32*(1), 109–121.  
doi: 10.1111/j.1945-5100.1997.tb01246.x
- Head III, J. W., Fassett, C. I., Kadish, S. J., Smith, D. E., Zuber, M. T., Neumann, G. A., & Mazarico, E. (2010). Global distribution of large lunar craters: implications for resurfacing and impactor populations. *Science*, *329*(5998), 1504–7.

doi: 10.1126/science.1195050

Head III, J. W., & Wilson, L. (1992). Lunar mare volcanism: Stratigraphy, eruption conditions, and the evolution of secondary crusts. *Geochimica et Cosmochimica Acta*, 56, 2155–2175.

Hiesinger, H., Jaumann, R., Neukum, G., & Head III, J. W. (2000, dec). Ages of mare basalts on the lunar nearside. *Journal of Geophysical Research: Planets*, 105(E12), 29239–29275.

doi: 10.1029/2000JE001244

Honda, R., Konishi, O., Azuma, R., Yokogawa, H., Yamanaka, S., & Iijima, Y. (2000). Data Mining System for Planetary Images - Crater Detection and Categorization -. In *Proceedings of the international workshop on machine learning of spatial knowledge in conjunction with icml* (pp. 103–108). Stanford, Ca.

doi: 10.1.1.28.1172

Hui Li, Jihao Yin, & Zetong Gu. (2014, jul). Crater detection based on local non-negative matrix factorization. In *2014 IEEE Geoscience and Remote Sensing Symposium* (pp. 521–524). IEEE.

doi: 10.1109/IGARSS.2014.6946474

Jahn, H. (1994, aug). Crater detection by linear filters representing the Hough Transform. In H. Ebner, C. Heipke, & K. Eder (Eds.), *Spie 2357, isprs commission iii symposium: Spatial information from digital photogrammetry and computer vision* (pp. 427–431). Munich.

doi: 10.1117/12.182789

Kalynn, J., Johnson, C. L., Osinski, G. R., & Barnouin, O. (2013). Topographic

- characterization of lunar complex craters. *Geophysical Research Letters*, 40(1), 38–42. doi: 10.1029/2012GL053608
- Kim, J. R., Muller, J.-P., & Morley, J. G. (2000). Towards Automated MOLA track registration in MOC and Viking Images and it's application to the establishment of new 3-D control point network on Mars. In *International archives of photogrammetry and remote sensing* (Vol. XXXIII, pp. 469–475). Amsterdam.
- Kim, J. R., Muller, J.-P., & Morley, J. G. (2004). Quantitative assessment of automated crater detection on Mars. *Proceedings of the XXth ISPRS*.
- Kim, J. R., Muller, J.-P., van Gasselt, S., Morley, J. G., Neukum, G., & Team, H. (2005). Automated crater detection, a new tool for Mars cartography and chronology. *Photogrammetric Engineering & Remote Sensing*, 71(10), 1205–1217.
- Li, B., Ling, Z., Zhang, J., & Wu, Z. (2015). Automatic Detection and Boundary Extraction of Lunar Craters Based on LOLA DEM Data. *Earth, Moon, and Planets*, 115(1-4), 59–69.  
doi: 10.1007/s11038-015-9467-9
- Melosh, H. J. (2011). *Planetary Surface Processes* (1st ed.). Cambridge University Press.  
doi: 10.1017/CBO9780511977848
- Meyer, F., & Beucher, S. (1990). Morphological segmentation. *Journal of Visual Communication and Image Representation*, 1(1), 21–46. doi: 10.1016/1047-3203(90)90014-M
- Michael, G. (2003). Coordinate registration by automated crater recognition. *Planetary and Space Science*, 51, 563.  
doi: 10.1016/S0032-0633(03)00074-6

- Michael, G., & Neukum, G. (2010). Planetary surface dating from crater size-frequency distribution measurements: Partial resurfacing events and statistical age uncertainty. *Earth and Planetary Science Letters*, *294*(3-4), 223–229.  
doi: 10.1016/j.epsl.2009.12.041
- Neukum, G. (1983). Meteoritenbombardement und Datierung Planetarer Oberflaechen . *Habilitation Dissertation for Faculty Membership, Univ. of Munich*, 1–186.
- Neumann, G. A. (2011). *Lunar Reconnaissance Orbiter Lunar Orbiter Laser Altimeter Reduced Data Record and Derived Products Software Interface Specification* (Tech. Rep.).
- Pieters, C. M. (1986). Composition of the Lunar Highland Crust From Near-Infrared Spectroscopy. *Reviews of Geophysics*, *24*(3), 557–578. doi: 10.1029/RG024i003p00557
- Quinlan, J. R. (1993). Programs for Machine Learning. *Machine Learning*, *240*, 302.  
doi: 10.1016/S0019-9958(62)90649-6
- Richardson, J. E. (2009). Cratering saturation and equilibrium: A new model looks at an old problem. *Icarus*, *204*(2), 697–715.  
doi: 10.1016/j.icarus.2009.07.029
- Salamunićar, G., & Lončarić, S. (2008). Open framework for objective evaluation of crater detection algorithms with first test-field subsystem based on MOLA data. *Advances in Space Research*, *42*(1), 6–19.  
doi: 10.1016/j.asr.2007.04.028
- Sawabe, Y., Matsunaga, T., & Rokugawa, S. (2006). Automated detection and classification of lunar craters using multiple approaches. *Advances in Space*

- Research*, 37(1), 21–27. doi: 10.1016/j.asr.2005.08.022
- Senft, L. E., & Stewart, S. T. (2007). Modeling impact cratering in layered surfaces. *Journal of Geophysical Research E: Planets*, 112(11), 1–18. doi: 10.1029/2007JE002894
- Shufelt, J., & Mckeown, D. (1993, may). Fusion of Monocular Cues to Detect Man-Made Structures in Aerial Imagery. *CVGIP: Image Understanding*, 57(3), 307–330. doi: 10.1006/ciun.1993.1021
- Spudis, P. D., Martin, D. J. P., & Kramer, G. (2014). Geology and composition of the orientale basin impact melt sheet. *Journal of Geophysical Research E: Planets*, 119(1), 19–29. doi: 10.1002/2013JE004521
- Stepinski, T. F., Ghosh, S., & Vilalta, R. (2007). Machine Learning for Automatic Mapping of Planetary Surfaces. In *Proceedings of the national conference on artificial intelligence* (pp. 1807–1812).
- Stepinski, T. F., Mendenhall, M. P., & Bue, B. D. (2009). Machine cataloging of impact craters on Mars. *Icarus*, 203(1), 77–87. doi: 10.1016/j.icarus.2009.04.026
- Whitford-Stark, J. L. (1982, may). A preliminary analysis of lunar extra-mare basalts: Distribution, compositions, ages, volumes, and eruption styles. *The Moon and the Planets*, 26(3), 323–338. doi: 10.1007/BF00928015
- Yin, J., Xu, Y., Li, H., & Liu, Y. (2013, jul). A novel method of crater detection on digital elevation models. In *2013 ieee international geoscience and remote sensing symposium - igarss* (pp. 2509–2512). IEEE.

doi: 10.1109/IGARSS.2013.6723331

# Chapter 3

## Automated Martian Crater

## Degradation Classification Using

## Chebyshev Polynomial

## Approximations

### 3.1 Introduction

#### 3.1.1 Crater Degradation

Impact craters are a ubiquitous surface feature in the solar system with a wide variety of scientific applications. The identification and measurement of impact craters can be used for estimating surface unit ages, and has been the primary tool for such a task for decades. Their presence on many different bodies exhibiting a spectrum of different



ambient surface conditions and processes allows for the study of impact crater morphology over time. Modification of the crater's morphology can act to affect the quantities that are used to represent craters, in addition to the precision with which those measurements can be made. For highly degraded terrestrial impact structures, even a quantity as fundamental as the diameter can vary between authors by a factor of two (Grieve, 1998).

Early work established the use of studying the 'freshness' of defining morphologic characteristics for lunar craters as metrics for crater modification and relative age determination (Pohn & Offield, 1970; Head III, 1975; Wood, Head III, & Cintala, 1977). This idea works from an assumption that a newly formed crater will exhibit morphologic characteristics that are either less present or completely subdued in a crater that has been considerably modified over time. By studying the degree to which these characteristics are present, some authors began to devise ordinal systems of crater classification (McGill & Wise, 1972; Arvidson, 1974). These systems divided craters into classes based on morphology expressions that were noticeable in the datasets available at the time, which were primarily optical in nature. Thus, the examination of these expressions were prone to inconsistencies such as resolution limitations, viewing geometry and lighting conditions. Barlow (2004) describes a semi-quantitative classification scheme incorporating multiple datasets to distinguish crater preservation based on characteristics such as relative depth and the morphology of the rim, ejecta, and crater interior. This classification amalgamated optical, topographic and thermal information in its decision-making process.

More recent work has provided large-N statistics to crater preservation information. A recent global database of martian impact craters, complete to  $D \geq 1$  km, also addressed the problem of preservation state classification (Robbins & Hynek, 2012b). This catalogue, consisting of over 384,000 craters, identifies positions, interior and exterior morphologic information, and modification state with identifying morphologic expressions such as channels or gullies. Additionally, each entry is assigned a confidence factor, which expresses the analysts assessment of the feature's likelihood of being an impact crater and not another type of circular depression. Measurements were made with both optical data from the global THERmal EMission Imaging System (THEMIS) Daytime IR mosaics (Christensen et al., 2004) and Mars Orbiter Laser Altimeter (MOLA) topographic datasets (Zuber et al., 1992), with both circular and elliptical rim-shape fitting. The crater preservation class was assigned similarly to previous work - by analysing the 'freshness' of the crater's morphology. The judgement criteria are shown in Table 3.1.

As can be seen in Table 3.1, there are four primary criteria used in the catalogue for assessing crater degradation. These are relative depth, rim, ejecta, and interior state. The author chooses relative depth as a way to account for pristine crater morphology variations at different latitudes. Instead of comparing the depth to a global average, the author uses equatorial or polar averages for comparison, depending on the latitude of the crater (Robbins & Hynek, 2012b).

Table 3.1: Crater degradation classification criteria from (Robbins &amp; Hynek, 2012a).

Relative Depth	Rim	Ejecta	Interior	Rank <sup>d</sup>	Class
<1/4 (1)	Rimless (1)	None (1)	Mostly Infilled/ Highly Modified <sup>b</sup> (1)	4-6 (3-4)	1
1/4 - 1/2 (2)	Slightly Elevated (2)	None (1)	Significant Deposits/ Modification <sup>b</sup> (2)	7-9 (5-6)	2
1/2 - 3/4 (3)	Some Degradation/ Modification <sup>b</sup> (3)	Some Erosion/ Modification (2)	Some Infilling/ Modification <sup>b</sup> (3)	10-13 (7-9)	3
>3/4 (4)	Sharp <sup>c</sup> (4)	Pristine (3)	Pristine (4)	14-16 (10-11)	4

<sup>a</sup>Craters are classified with three morphologic characteristics and the relative depth from topography (if available). The corresponding rank is converted to a preservation class. The majority of craters in a given class will have characteristics from that row, but that is not always the case. It is possible - if highly unlikely - for a crater to have, for example, a “Sharp” rim while having no ejecta and being mostly infilled.

<sup>b</sup>Modification includes: Gullying/dissection, fracturing, lava flows, ice flows, mass wasting (e.g., from the rim), superimposed cratering, etc.

<sup>c</sup>Does not necessarily mean “pristine” (i.e., can have a small crater superimposed or a very small bit of modification)

<sup>d</sup>Parenthetical values are if depth information is not present.

## Internal Crater Morphology

In addition to providing the degradation class, the catalogue also provides a description of the internal crater morphology. These morphologic identifiers, originally described in (Barlow & Bradley, 1990), describe the presence of distinct features within the crater rim. They are indicated by shorthand abbreviations, starting with ‘Cpx’ for complex craters, and then appending a morphology descriptor. These features include flat floors (CpxFF), central peaks (CpxCPk), central pits (CpxCPt), summit pits (CpxSuPt), and unclassifiable (CpxUnc). An example of each of these morphologies is shown in Figure 3.2. These morphologies are fundamentally related to, and affected by crater modification. Flat floor craters, for example, can be formed by the infilling of a material that is exogenous to the impact crater’s formation such as wind-blown material or lava.

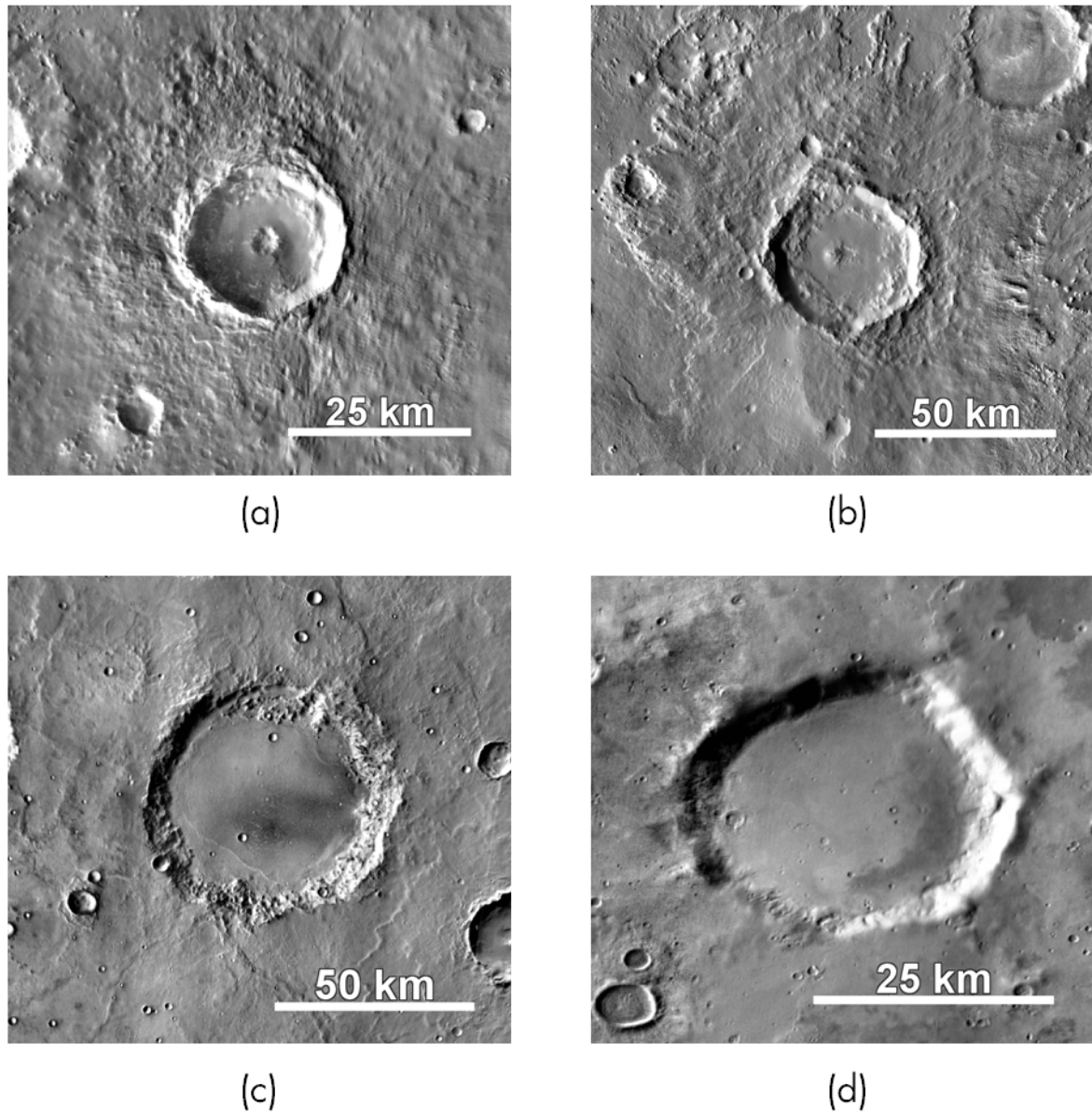


Figure 3.1: An example of a crater from each preservation class, from (Robbins & Hynek, 2012b). All imagery is THEMIS Day IR. (a) A 23 km diameter crater in the pristine class (4), located at  $16.50^{\circ}\text{N } 11.49^{\circ}\text{E}$ . Note the sharp rim, central pit, and pristine ejecta. (b) A 44 km diameter class 3 crater located at  $14.79^{\circ}\text{N } 9.63^{\circ}\text{E}$ . This crater shows some infilling, as well as subsequent cratering on the ejecta blanket. (c) A 65 km diameter class 2 crater, located at  $15.27^{\circ}\text{S } 9.79^{\circ}\text{E}$ . This crater shows substantial infilling, as well as minimal expression of the rim and ejecta. (d) A 38 km diameter class 1 crater, centred at  $39.33^{\circ}\text{S } 15.25^{\circ}\text{E}$ . This crater has no raised rim, and is heavily infilled. Ejecta texture is no longer visible.

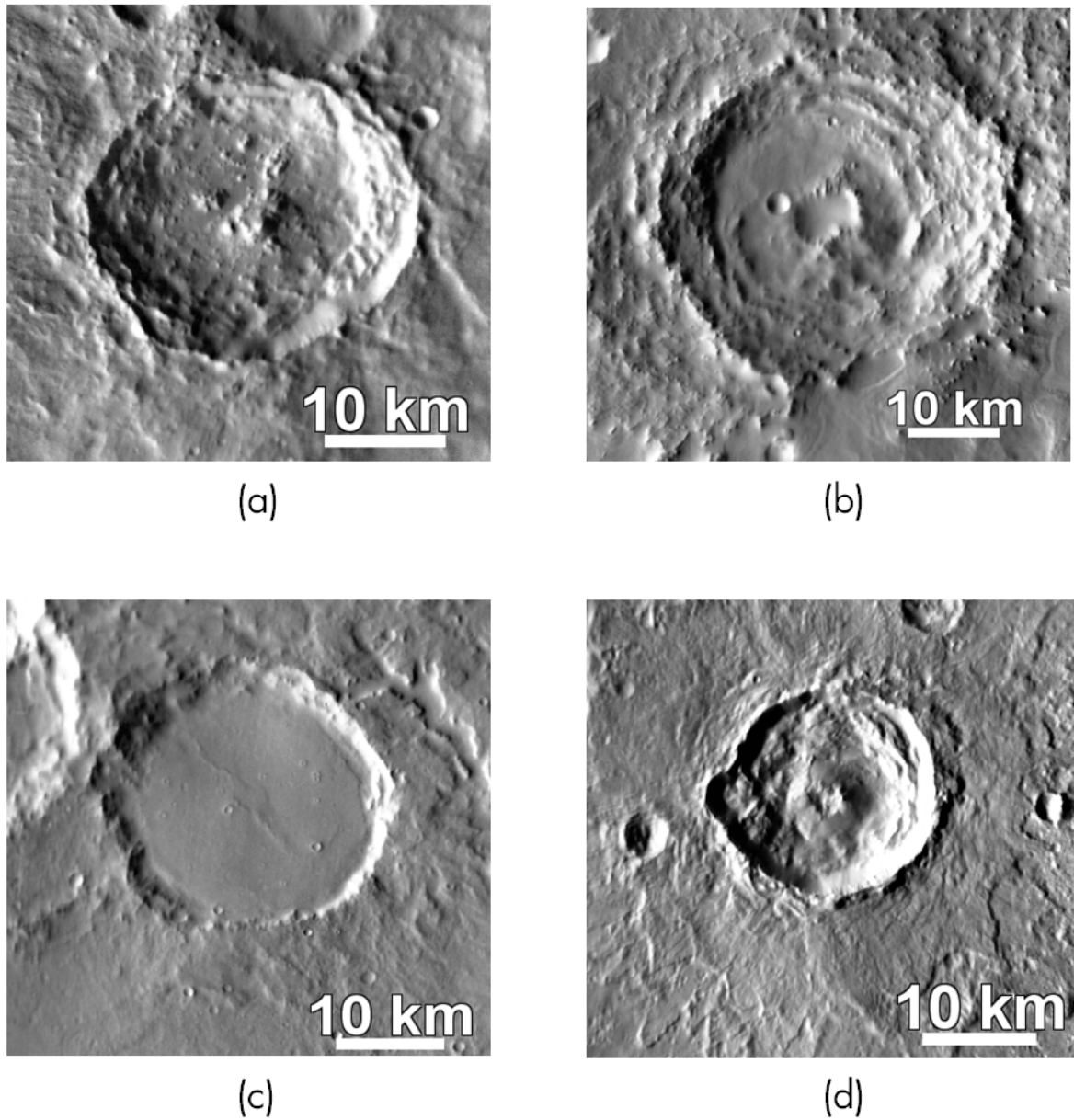


Figure 3.2: An example of a crater from each interior morphology class, from (Robbins & Hynek, 2012b). All imagery is THEMIS Day IR. (a) A 30 km diameter crater exhibiting a central peak (CpxCPk), or localized topographic high near the crater center. This crater is located at  $26.31^{\circ}\text{N } 28.12^{\circ}\text{E}$ . (b) A 38 km diameter central pit (CpxCPT) located at  $14.68^{\circ}\text{N } 20.67^{\circ}\text{E}$ . (c) A 28 km diameter crater with a flat floor morphology (CpxFF), located at  $8.81^{\circ}\text{N } 35.82^{\circ}\text{E}$ . (d) A 19 km diameter summit pit crater (CpxSuPt), centred at  $15.24^{\circ}\text{N } 16.45^{\circ}\text{E}$ . The summit pit, located at the crater center, is differentiated from a central pit by the raised topography on which the pit is situated.

### Processes of Crater Modification

The processes which modify an impact crater are dependent on the parent body on which it is located. By extension, the variety of modified crater morphologies is related to the number of degrees of freedom in the surface and subsurface environment of the body. For example, a body that has no atmosphere will not possess craters that exhibit aeolian (wind-driven) modification. However, the same body might experience more rapid degradation from subsequent small impacts and even smaller scale micrometeorite ablation due to a lack of protection from the atmosphere. This subsection details some of the primary methods that modify a crater's topographic signature.

One of the most common processes of crater modification is from bombardment. This is described as the destruction of a crater profile by the superposition of other craters over top of it. Additionally, bombardment can cause infilling by ejecta emplacement inside a neighbouring crater, which reduces its depth. In more extreme cases, the ejecta of a significantly large nearby impact can cause severe modification by depositional blanketing. Generally, meteoritic bombardment has been shown to soften topography and reduce the pristine profile of craters (Ross, 1968). This process is dominant on surfaces not protected by any considerable atmosphere; on Mars, the effect would be less important than on the Moon or another airless world. A second consequence of bombardment is the seismic shaking of material caused by nearby impacts. By shaking loose material from less stable slopes, this would have a net effect of marginally increasing the crater diameter, while also reducing its depth through

infilling.

On Mars, and in environments with a significant enough atmosphere for wind to be present, aeolian erosion can modify a crater's signature. For arid, cold desert environments like those on Mars, aeolian activity dominates the environment (Kumar, Head, & Kring, 2010). Aeolian modification rounds out sharp features such as the crater rim. Additionally, it deposits material delivered by wind into the crater, reducing the depth profile. Other modification processes active on Mars are from periglacial and volcanic sources. Modification of craters occurs at higher latitudes and polar regions, where the presence of ground ice causes a gradual shallowing of the crater profile over time (Pathare, Paige, & Turtle, 2005). For volcanic modification, extrusive volcanism can infill, partially destroy or completely bury impact structures. This is applicable to craters near the volcanic provinces. Fluvial degradation of impact craters in the martian highlands has also been proposed to explain morphometry related to crater infilling (Forsberg-Taylor, Howard, & Craddock, 2004).

### 3.1.2 Crater Profile Modelling

The topographic analysis of impact crater profiles is a useful tool for advancing our knowledge on their morphologies at all scales and crater states. Previously, work has been done to ascribe mathematical representations to crater profiles. A straightforward example of this would be using a mathematical function to model the crater surface height  $z(r)$  as a function of radial distance  $r$  from the crater center. One of the earlier

applications of using a mathematical model was to describe the scaling relationship between crater depth and diameter (Pike, 1977). For modelling the crater profile, previous work had used various polynomial expansions or hyperbolic representations to model the crater shape (Craddock & Howard, 2000; de Vet & de Bruyn, 2007). A typical polynomial expansion of one variable can be represented as a linear sum:

$$z(r) \approx f(x) = \sum_{n=0}^M C_n T_n(x) \quad (3.1)$$

where  $T_n(x)$  is the chosen polynomial of degree  $n$  (also called the basis function),  $C_n$  is the  $n^{\text{th}}$  coefficient, and  $M$  is the total approximation order. Choosing a basic polynomial expansion, where the basis functions are powers of  $x$ , and expanding out this sum to degree  $M$  results in the full polynomial:

$$z(r) \approx f(x) = a_0 + a_1x + a_2x^2 + \dots + a_Mx^M \quad (3.2)$$

Representations using this ordinary choice of polynomial suffer as they are not directly comparable across studies, or even across individual craters. This is because the basis functions in this representation are not orthogonal. Orthogonal basis functions are those which have zero correlation between one another. This means that modifying one of the terms, or changing the degree of expansion ( $M$ ) does not affect the other terms. For this reason, a different set of polynomials must be used to create a standardized and internally consistent system for representing crater topographic profiles.



### Chebyshev Polynomials

Recent work on standardizing the quantitative representation of craters has improved our ability to represent crater profiles. The novel application of a different set of mathematical objects, called Chebyshev polynomials of the first kind (Chebyshev, 1854) (furthermore just ‘Chebyshev polynomials’), has made significant progress in crater profile representation (Mahanti et al., 2014). Chebyshev polynomials are a sequence of polynomials that are defined recursively through the following relationship:

$$T_{n+1}(x) = 2xT_n(x) - T_{n-1}(x) \quad (3.3)$$

with  $T_0 = 1$  and  $T_1 = x$ . The first 9 polynomials, corresponding to an approximation order  $M = 8$ , are listed in Appendix B. These polynomials are particularly well suited to the problem of crater profile representation for few reasons. First, the basis functions are orthogonal, allowing for direct comparison between functions for different craters, even with different expansion orders. This also results in a reduced minimum mean squared error in the approximation (Mahanti et al., 2014). Also, when scaling the x-axis to be limited to the domain  $x \in [-1, 1]$ , all Chebyshev polynomials have a maximum value of +1 and a minimum value of -1 in this interval, simplifying the computation.

Since it is not computationally possible to have a truly infinite sum of Chebyshev terms, it is necessary to select an order of approximation. The topographic information that cannot be accurately represented due to the limited approximation order is known as the truncation error. The previous study using Chebyshev expansions for lunar

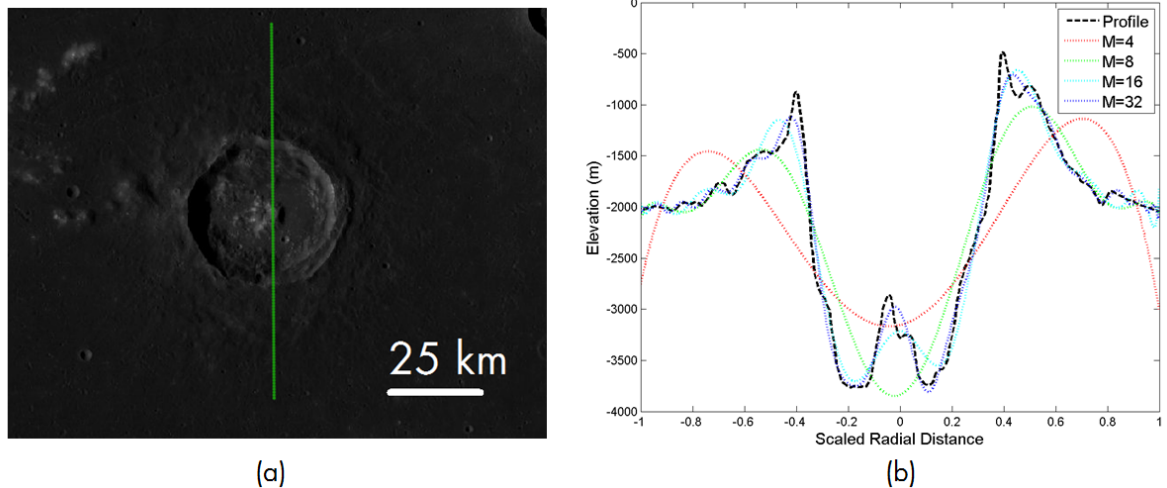


Figure 3.3: (a) Plinius, a roughly 43 km diameter complex lunar impact crater located at  $15.4^{\circ}\text{N}$ ,  $3.7^{\circ}\text{E}$ . The green line denotes the direction of the extracted profile in (b). (b) A scaled plot of the crater profile (black), alongside three Chebyshev profile approximations. As the order of approximation increases from  $M = 4$  (red) to  $M = 32$  (blue), the approximation error is reduced.

crater profile reconstruction found that choosing an approximation order of  $M = 16$  was sufficient for representing, characterizing and classifying craters (Mahanti et al., 2014).

By scaling the distance axis to  $x \in [-1, 1]$ , we can also directly compare coefficients across craters. The coefficients correspond to weighting factors that represent the amount of each contribution from their respective basis function. A larger coefficient indicates a larger contribution from that coefficient's basis function. This process and how it is used in crater profile reconstruction is shown in Figure 3.4.

The coefficients derived from a Chebyshev polynomial expansion of crater topography have been shown to directly relate to crater morphologic characteristics such as depth-to-diameter relationship, local surface gradient, crater profile asymmetry,

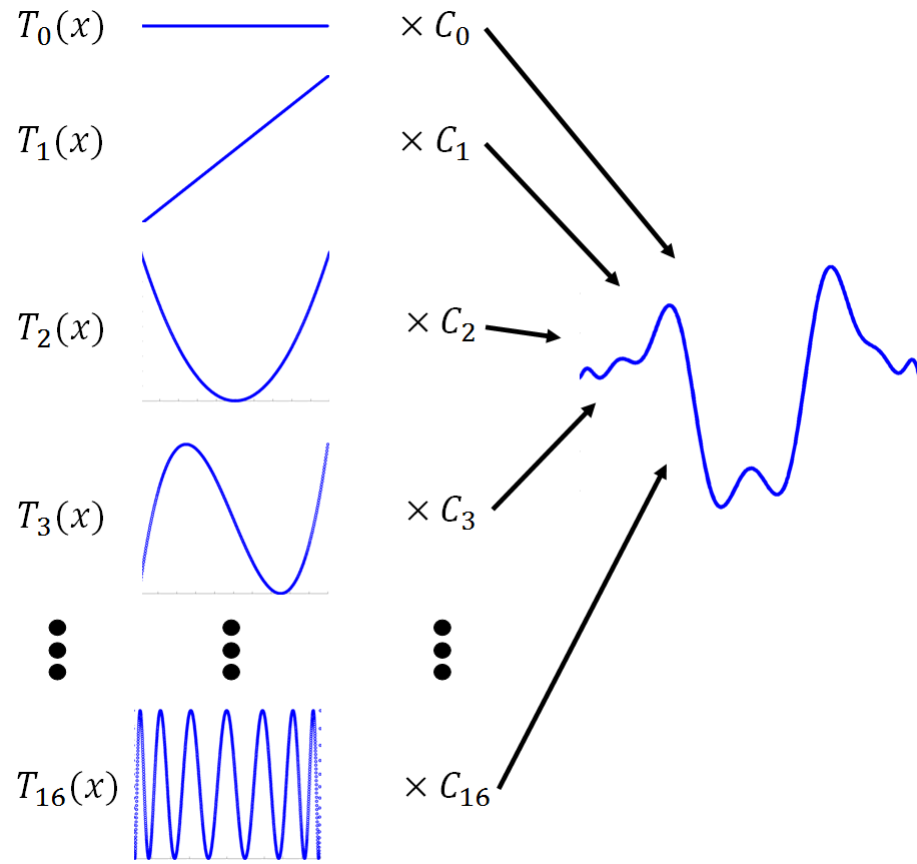


Figure 3.4: A general diagram of how crater profile reconstruction relates to basis functions and coefficients. Each basis function, shown on the left as  $T_n(x)$ , has a weighted contribution to the final profile that is related to its coefficient  $C_n(x)$ .

and central peak height (Mahanti et al., 2014). These relationships are a direct result of the symmetry and extrema properties of Chebyshev polynomials. For example, Chebyshev polynomials alternate as even (symmetric) and odd (antisymmetric) functions. Since only odd functions pass through the origin (i.e. have a value of zero there), we can use the combination of even Chebyshev coefficients to learn about the crater profile at  $x = 0$ . Since these characteristics are modified by the processes that cause crater degradation, it should therefore be possible to use the extracted Chebyshev expansion coefficients to quantify crater degradation.

Table 3.2: List of Chebyshev coefficients and coefficient combinations, with their interpreted topographical indications (Mahanti et al., 2014).

<b>Coefficient</b>	<b>Equation (M=16)</b>	<b>Interpretation</b>
$C_0$	-	Avg. crater profile
$C_1$	-	Local topographic gradient
$C_2$	-	Depth (1st order approx)
$C_3$	-	Asymmetry in rim shape
$C_4$	-	Central peak height (1st order)
$I_1$	$C_0 - C_2 + C_4 - C_6 + \dots$	Elevation at $x = 0$
$I_2$	$C_0 + C_4 + C_8 + C_{12} + C_{16}$	Central peak height (2nd order)
$I_3$	$C_0 - C_1 + C_2 - C_3 + \dots$	Crater elev. at left-hand extreme
$I_4$	$C_0 + C_1 + C_2 + C_3 + C_4 + \dots$	Crater elev. at right-hand extreme
$I_5$	$2 \cdot (C_2 + C_6 + C_{10} + C_{14})$	Depth (2nd order)
$\tilde{I}_5$	$I_5/D$	Depth-to-diameter ratio

### 3.1.3 Purpose

The purpose of this paper is to propose a standardized and quantitative method for the automated preservation classification of martian complex impact craters. Using MOLA topographic data for the martian surface, a standardized method will be described for extracting and modelling the 2-D profiles of complex impact craters on Mars. The results of this numerical modelling will be used to build a classification model for assigning a preservation state quantity to the craters. This process will be automated using a catalogue of input craters, whose locations are provided from Robbins and Hynek (2012b). This system should provide an objective and efficient method for assigning a preservation class, which has historically been only semi-quantitative at best. Additionally, predicting the general interior morphology (flat-floored, central pit, or other descriptors) by the same means will be investigated. This work should provide a stepping-off point for the development of a more complete and accurate automated crater classification system.

## 3.2 Methods

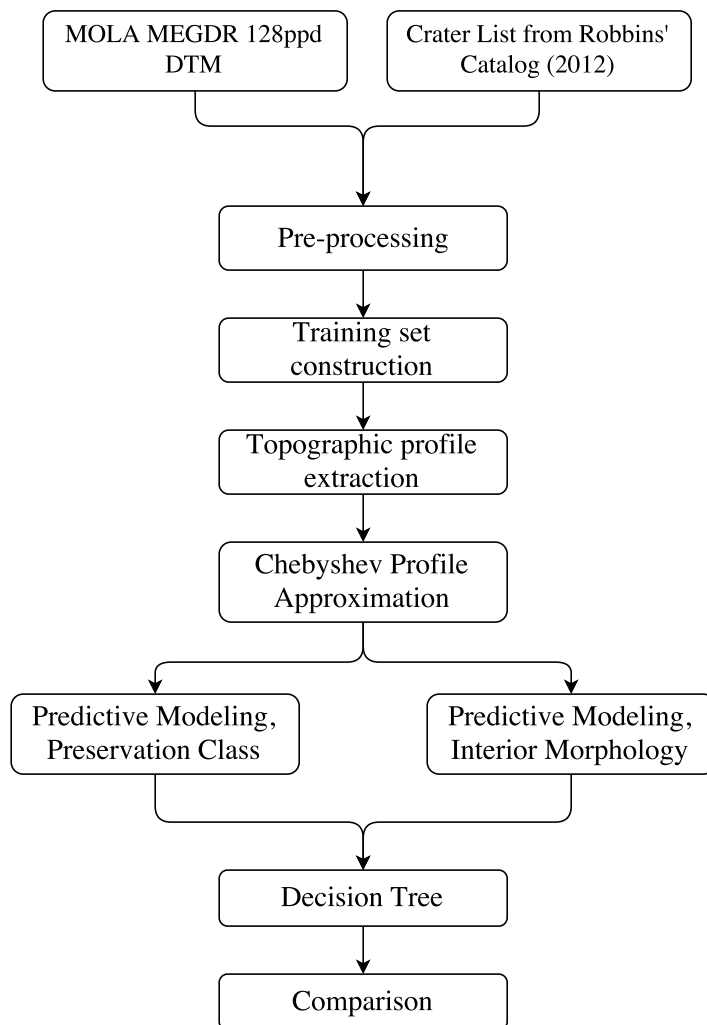


Figure 3.5: Flowchart of the major steps for this project.

This section will describe the four primary elements of the project methodology:

1. Data collection.
2. Pre-processing and training set construction.
3. Profile extraction and polynomial approximation.
4. Building the classification models and evaluation of results.

### 3.2.1 Study Area and Data

As mentioned previously, the global crater catalogue in (Robbins & Hynek, 2012b) is complete for craters with  $D \geq 1$  km. The topographic measurements made for craters in this catalogue were made using the highest resolution global DTMs available for Mars. These are the MOLA Mission Experiment Gridded Data Records (MEGDR) raster topographic maps of the martian surface, at 128 pixels per degree. The locations and size measurements provided in the catalogue are similarly made using this dataset. This choice dictates our use of the MOLA 128 ppd topography dataset for this project. To compare our polynomial approximations with the degradation classes assigned by the original authors, it is important to use the same source data. The Chebyshev profile approximation described in (Mahanti et al., 2014) was applied to lunar craters with diameters in the range  $100 \text{ m} \leq D \leq 145 \text{ km}$ , using two different elevation data sources. Since this study uses lower resolution data for the martian surface, we choose to set the minimum crater size to  $D \approx 10 \text{ km}$ . This is because MOLA data has been shown to reproduce crater morphology down to this size with satisfactory fidelity (Robbins & Hynek, 2013).

The authors of the catalogue used in this study have provided a very convenient online database searching tool. The search interface (<http://craters.sjrdesign.net/>) has a selectable list of crater data that can be exported to a CSV file. The primary selections used for the craters in this study were a minimum diameter of  $D \approx 10 \text{ km}$ , and for all measurements (diameter, location) to be made against the MOLA topography data set.

Additionally, all information relating to crater morphology was selected for inclusion in exported catalogue.

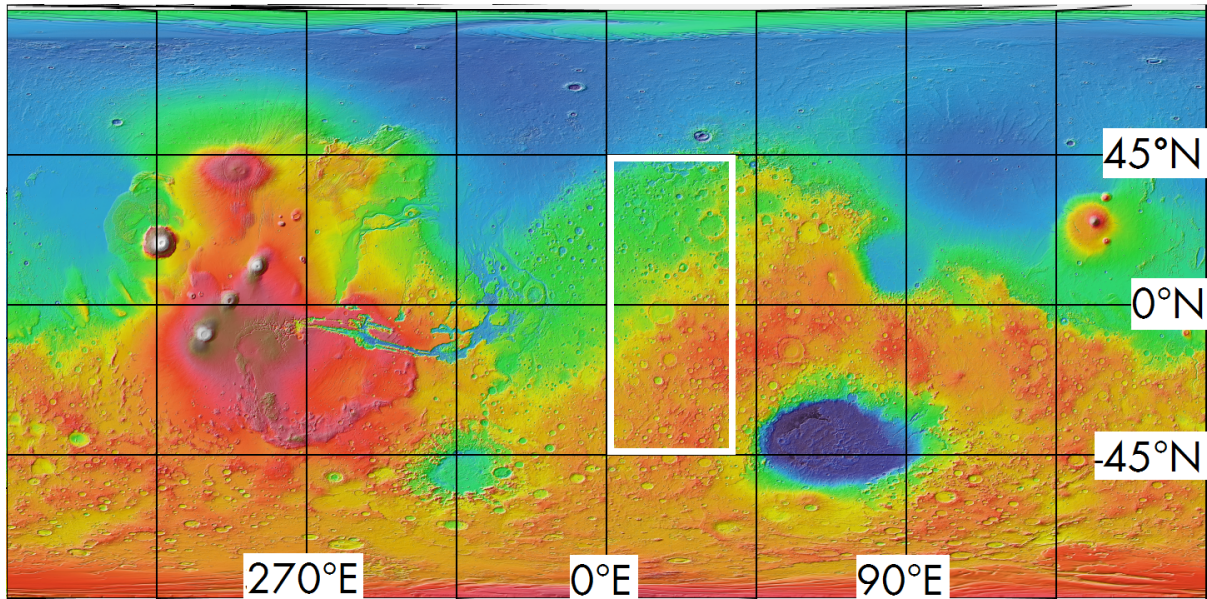


Figure 3.6: Global map of Mars in simple cylindrical projection. Visualized data is MOLA Coloured Elevation. The study area is outlined by the white box.

The chosen study area for this project encompasses around  $1.27 \cdot 10^7$  km<sup>2</sup> of the martian surface (roughly 9% of Mars' surface area), over equatorial and mid-latitude regions. This area, shown in Figure 3.6, encompasses three primary areas: Arabia Terra in the north, Terra Sabaea to the east, and Noachis Terra to the south. These areas consist primarily of rugged, high-relief terrain that is dated to the early and middle Noachian period of Mars' history (Tanaka et al., 2014). Noachian units are thought to be some of the oldest units on Mars. These units are interpreted as being undifferentiated materials of impact, volcanic, fluvial, and basin origin that are moderately to heavily degraded. However, this unit is also interspersed with more recent impacts that superpose the surface. These impacts are interpreted as being upturned and ejecta target rocks, which may be modified over time by fluvial-lacustrine and aeolian processes. Robbins' catalogue

lists 3,274 craters with  $D > 10$  km within the study area, some of which will belong to this unit, but most will be much older with varying levels of degradation.

### 3.2.2 Pre-processing and Training Set Selection

After acquiring the full crater catalogue, some pre-processing steps are applied to improve the quality of the results. First, the set is split into two halves: one corresponding to the northern section of the study area (Arabia Terra) and the other to the southern section (Noachis Terra/Terra Sabaea). This is to ensure ultimately that an even number of craters from each region go into building the training sets. Next, as mentioned in Section 3.1.1, we examine the confidence factor assigned to each crater entry. Each crater was assigned a confidence factor by the authors, from 1 to 4, with 1 being non-confidence and 4 being complete confidence in the structure being a true impact crater. For this study, we remove all craters from the list that have a confidence factor  $< 3$ . We also remove any craters that do not have a degradation state assigned, or that were missing topographic data. The final check is to remove craters whose elevation profile lines are going to extend beyond the extents of the DTM - this is explained further in Section 3.2.3. To prepare the DTM, we crop our region of interest from the larger tile DTMs, and export them as ASCII raster files.

After pre-processing the craters down to a list that is appropriate for this study, we next select our training set. This list is a sub-selection of craters that will be used to train the classification model to recognize craters in the different degradation classes. To



acquire a near-random sampling of craters in each of the four degradation classes, we employ a stratified sampling technique. The classification model works best when each crater class is represented in as even proportions as possible, so we are limited by the class with the smallest number of craters - this is class 4, the ‘pristine’ class. Twenty members of each class from each area are selected using the stratified sampling. A second stratified sampling is then done to build a training set for the interior morphology descriptor. After exporting the collective training set (including the members from both areas), we then move onto the profile extraction and approximation.

### 3.2.3 Profile Extraction and Approximation

The classification model used to determine degradation class values is built by using Chebyshev polynomial coefficients as discriminators. To extract these coefficients, we develop a two-step process in MATLAB (code shown in Appendix B). The two steps include profile extraction, and then polynomial fitting. The main steps of this method are shown in Figure 3.7.

#### Topographic Profile Extraction

The Chebyshev polynomial expansion is used to model the true topographic profile using reconstructed basis functions. The first necessary step is to define a consistent way to extract a 2-D profile for the crater. Craters can exhibit a variety of morphologies, often possessing asymmetry in some form. This means that crater profiles of a given radial length can be quite different as the azimuthal angle is varied. To be consistent with the method described in Mahanti (2014), we constrain all profiles

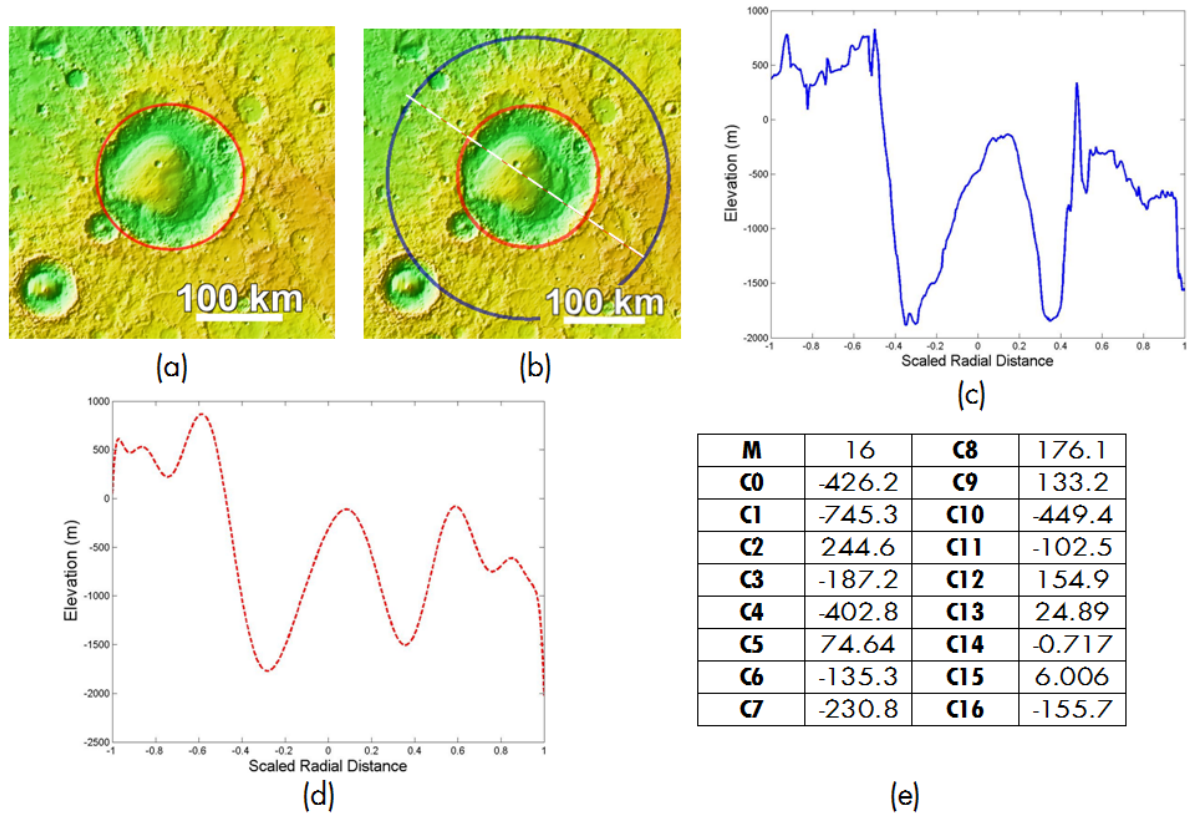


Figure 3.7: Profile extraction and approximation process. (a) Crater is identified by catalogue values for location and radius. (b) Line of greatest slope (white/red dashed line) through crater center and length  $4 \cdot R_{crater}$  is identified. (c) Topographic profile is extracted, with domain scaled to  $[-1,1]$ . (d) Profile is approximated using Chebyshev polynomial expansion. (e) Chebyshev coefficients are retrieved from the expansion.

to be in the direction of greatest surface gradient outside the crater, with a profile length of  $2 \cdot R_{crater}$  from the crater center. This results in a profile of total length  $4 \cdot R_{crater}$ .

Once all craters are loaded into the landscape, a circle of points with radius  $2 \cdot R_{crater}$  is defined around the crater centroid. Each point and its opposing member (diametrically opposite) are then used to calculate the slope along the line that connects them. This is done iteratively over all 100 lines (lines are separated by  $\theta = \frac{\pi}{50}$ ) to identify the line with greatest slope. Once the point pair has been identified, it is necessary to generate a line between the two points to create the profile. To build a line in a discretized coordinate system (such as a raster map), we employ Bresenham's line algorithm (Bresenham, 1963). We then save the line points for each crater profile.

Generating the profile involves combining the line points with the DTM landscape. At each point along a crater's profile line, the elevation value is extracted from the DTM. As mentioned in Section 3.2.2, it is important that all profile points lie within the DTM - this justifies the removal of craters near edges or with significantly long profile lines. Finally, the distance scale is normalized so the range lies between -1 and 1, for the reasons described in Section 3.1.2. We now have our scaled topographic profiles for the craters, and can perform the polynomial approximation.

### Chebyshev Polynomial Fitting

Once the elevation profile is extracted and scaled, we can apply a polynomial approximation to the function that represents the height of the surface,  $z(r)$ , as a function of radial distance  $r$ . This is done with the help of the Chebfun toolbox (<http://www.chebfun.org/>), an open-source MATLAB package for numerical function computation (Driscoll, Hale, & Trefethen, 2014). The Chebfun toolbox includes a polynomial fitting tool that can be used to overload MATLAB's default tool, replacing the ordinary polynomial choice with Chebyshev polynomials. The polynomial fitting works through a least-squares approximation method. This is done by minimizing the square of the difference between the polynomial values and the values corresponding to our crater profile,  $z(r)$ . The coefficients  $C_n$  (where  $0 \leq n \leq M$ ) are varied to minimize this value. The set of coefficients which satisfy this condition are chosen for the polynomial approximation. After running the polynomial fit, we store the coefficients for each crater as well as plotting the approximated profile.

### 3.2.4 Crater Classification

With the coefficients of each crater retrieved from the polynomial fit, the final step is to create a classification model for the degradation state using the coefficients as discrimination attributes and the classes from Robbins' catalogue as the training set values. As mentioned in Section 3.1.2, the reconstructed Chebyshev coefficients have been shown to be directly correlated to crater topographic characteristics. Additionally, careful combination of related coefficients can be employed to expand the number and

accuracy of morphology descriptors. Therefore, we should be able to use these coefficients as a quantitative means for describing crater topography and assigning a classification scheme.

The J48 statistical classifier constructs the decision tree using the attributes for each crater. This algorithm identifies which attributes best sort the training set (and at which value) by varying these values and examining the ‘purity’ of the result. This tree assigns a preservation class value to each crater by comparing its attribute values (coefficients) against these values. The resulting set of rules is then applied to the entire catalogue in the form of a decision tree. To run the statistical classifier, we import the training sets (one for degradation and one for interior morphology) into WEKA, an open source suite of machine learning tools (Hall et al., 2009). The training sets consist of a list of each crater with the attributes to be used as discriminators. The first attributes are standard size descriptors from the catalogue: depth  $d$ , diameter  $D$ , and depth-to-diameter ratio  $d/D$ . In addition, we import the 17 Chebyshev coefficients retrieved from the polynomial approximation:  $C_0, C_1 \dots C_{16}$ . Additionally included are the coefficient combinations that have been also been shown to relate to morphology:  $I_1, I_2, \dots, I_5, \tilde{I}_5$ . Finally, the value to be predicted (degradation class or interior morphology indicator) is included. The J48 classifier is then run on each of the training sets.

## 3.3 Results

Topographic information retrieved from the Chebyshev polynomial coefficients served as a discrimination framework for assessing both crater degradation class, as well as major interior morphologic characteristics. The models were compared against values assigned by expert analysts and published in a global martian crater catalogue. Two separate classification models were built; the classification results of each are discussed separately in this section.

### 3.3.1 Degradation Class Model

The primary goal of this project is to demonstrate a standardized, quantitative method for automatically classifying the degradation state of complex martian impact craters. Using a training set of 160 craters, representing 40 craters from each class over the study area, a classification model was built to be able to automatically assign a crater degradation class using a list of Chebyshev polynomial coefficients as input. The training set size was limited by the class with the smallest number of representatives - this was the pristine class. The model produced by the J48 statistical classifier was applied to all craters in the study area, with the resulting degradation class compared against the expert-provided class.

The overall accuracy for the degradation classification system was 70.1%. At first glance, we see a variation in the system's ability to discriminate crater degradation. Pristine craters are very well matched, and the most degraded are classified with

Table 3.3: Degradation class assignment results for the classification model. The number of instances listed is from the original catalogue.

Degradation Class	# of Instances	Diameter Range (km)	Accuracy (%)
1	2009	8.7-425.4	75.6
2	463	9.7-156.7	65.5
3	507	9.8-112.6	50.7
4	36	10.1-31.8	96.8

reasonable accuracy of around 76%. For the two intermediate classes, the detection accuracy is not as high, with class 3 having the lowest accuracy at around 51%. Interestingly enough, by separating the confusion matrix (shown in Table 3.4) into two halves for the predicted class (the halves can be thought to represent two broad classes - ‘degraded’ and ‘well-preserved’), the total error is greatly reduced. Classes 1 and 2 have many false classifications that are in class 2 or 1, respectively. Similarly, classes 3 and 4 tend to have the highest number of errors within them as well.

Table 3.4: Confusion matrix for the degradation class assessment.

		Predicted Class				Total
		1	2	3	4	
Reference (Catalog) Class	1	1518	469	21	1	2009
	2	92	304	54	13	463
	3	16	82	257	152	507
	4	0	0	1	35	36
	Total	1626	855	333	201	

### 3.3.2 Interior Morphology Model

In addition to crater degradation state, Robbins’ catalogue also includes information on the morphology of the craters, including their interiors, ejecta (if present), and other features. This extra information was used to try and build a similar model for some

interior morphologic phenomena including central peaks, central and summit pits, and flat floors.

Table 3.5: Interior morphology assignment results for the classification model. The number of instances listed is from the original catalogue. CpxUnc is unclassified, CpxCpk is central peak, CpxCpt is central pit, CpxFF is flat-floored, and CpxSuPt is summit pit.

Interior Morphology	# of Instances	Diameter Range (km)	Accuracy (%)
<b>CpxUnc</b>	684	9.0 - 182.5	51.8
<b>CpxCPk</b>	214	9.6 - 112.6	35.0
<b>CpxCPt</b>	182	10.0 - 177.5	59.3
<b>CpxFF</b>	1859	8.7 - 425.4	40.0
<b>CpxSuPt</b>	76	10.2 - 65.8	80.2

The overall accuracy of the interior morphology classifier with  $M = 16$  was 44.5%. This is significantly lower than the degradation classifier. Once again, there is variation within the different classes. The summit pit class had a significantly higher accuracy ( $\sim 80\%$ ) when compared against the other four classes. Most other accuracies are spread between 35% and 60%. The central pit and unclassified classes were higher than the central peak and flat floor classes.

Table 3.6: Confusion matrix for the interior morphology class assessment.

		Predicted Class					Total
		CpxUnc	CpxCPk	CpxCPt	CpxFF	CpxSuPt	
Reference (Catalog) Class	CpxUnc	354	78	67	157	28	684
	CpxCPk	26	75	33	13	67	214
	CpxCPt	7	25	108	5	37	182
	CpxFF	572	289	159	743	96	1859
	CpxSuPt	0	11	3	1	61	76
	Total	959	478	370	919	289	

The confusion matrix for the morphology classifier, shown in Table 3.6, generally depicts a lack of consistency in identifying morphology types. Other than the summit pit craters (whose classification performance will be discussed in Section 3.4), there seems



to also be a trend in confusion between unclassified and flat-floored craters, as well as central peak and summit pit craters.

## 3.4 Discussion

This system was designed to use Chebyshev coefficients from a polynomial approximation as quantitative discriminators for classification models. These models were used to assign both a crater degradation classification which quantified the preservation state of the crater, as well as describe any features that exist interior to the crater. There was a very notable difference in the performance of this system in performing the two classifications. Some reasons for this will be described in this section. In addition, we discuss the crater profile extraction methodology.

### 3.4.1 Efficiency of Crater Profile Extraction

The method described in Section 3.2.3 was created as an efficient means for extracting and approximating crater profiles. This process, written in MATLAB, can be found in Appendix A. The process was capable of generating a set of crater profiles in orders of magnitude less time than would be possible with manual collection. There are, however, a few elements of this process that are worth mentioning, as they have an effect on the resulting crater profile. The first pertains to how the profiles are generated. To identify the line of greatest gradient, a circle consisting of one hundred points is generated around the crater centroid. The opposing point pairs are used to

identify slopes, with the greatest being picked. Since the points are separated by the same angular distance ( $\frac{\pi}{50}$ ), the spacing between points changes as a function of the crater radius. This creates an inconsistency identifying the line of greatest slope when scaling crater size, which is likely to affect the standardized nature of such a method.

By choosing only the line associated with the greatest external surface gradient, we also neglect information provided by the asymmetry of profiles selected in different directions. Sometimes, asymmetry in a single-line profile can be useful for assessing degradation (for example, if one section of rim is mantled by lava) but it is more likely that we are losing more than gaining. It would thus be desirable going forward to devise a standardized method for considering or amalgamating crater profiles in many directions. This could include averaging a second profile that is orthogonal to the direction of the one already chosen based on gradient. A more comprehensive system might average a large number of profiles, drawn at even angular separations. Whichever method is chosen would need to maintain a standardized approach.

The order of approximation for the Chebyshev polynomial fit plays a significant role in the process. As with other polynomial expansions (or Fourier expansions), higher degree terms correspond to high-frequency components in the reconstructed profile. For an impact crater, this would correspond to sharp changes in topography such as a rim, or terracing inside the crater. As mentioned previously, an order of approximation  $M = 16$  was chosen as previous work has shown it to be sufficient in reproducing crater topography while minimizing reconstruction error (differences in the true crater profile

shape and the reconstructed shape). While this may have been sufficient for a degradation classification system, the high-frequency components are necessary for determining internal morphology. Computational efficiency must also be considered when choosing the reconstruction order. The polynomial fitting function has  $\mathcal{O}(n(\log n)^2)$  complexity. If we were to double the reconstruction order from 16 to 32, this would result in over triple the runtime. For large crater catalogues, consideration of this choice is necessary.

### 3.4.2 Degradation Classification Performance

This project demonstrated that Chebyshev polynomial coefficients can be used to quantitatively assess crater degradation states on large scales in an efficient manner. However, the results were not perfect in an absolute nor internally consistent sense. The primary problem pertains to a difference in accuracy based on degradation class. Table 3.4 identifies a trend in terms of misclassified craters. Classes 1 and 2 share most misclassifications within themselves, and the same applies with classes 3 and 4. It may therefore be useful at this approximation order to reduce the four classes into two by combination, considering the new classes as ‘degraded’ and ‘well-preserved’. This would improve the accuracies of those two classes to 96.4% and 82.0% respectively. Merging together just classes 2 and 3 into a ‘moderately preserved/degraded’ class would result in a new accuracy of 71.8%, so that no class would have an accuracy below 70%. These results are shown in Table 3.7. It can be reasoned that the number of discrete degradation classes that can accurately be resolved is a function of the approximation

order.

Table 3.7: Degradation classification results for class combinations. The original 4 class results are listed, as well as the 3 class (degraded, moderately preserved/degraded, well preserved) and 2 class combination (degraded, preserved).

Number of Classes	Description	Average Accuracy (%)
4 classes	1,2,3,4	72.2
3 classes	1,2+3,4	81.4
2 classes	1+2,3+4	89.2

Inspection of the results reveals a few trends related to crater misclassification. For craters that were catalogued as being in class 3 (lightly degraded), but ended up being classified as classes 1 or 2, it seems that pre-impact local topography plays a role in confusing the model. Craters with a highly-significant pre-impact slope were often classified one or two classes below their assigned value. Additionally, newer craters that formed in the ejecta blankets of much larger craters were also subject to this effect. Examples of both cases are shown in Figures 3.8 and 3.9, respectively.

Section 3.4.1 discusses that higher-ordered coefficients are related to higher-frequency spatial information. The crater degradation processes that area active in this region, described in Section 3.1.1, tend to subdue sharp features and flatten the crater interior. These processes emphasize low spatial frequency components over time. This may explain the system's advantage in identifying degraded craters more accurately than pristine craters (on the whole).

The unusually high detection accuracy for class 4 (pristine) craters is likely a mixture

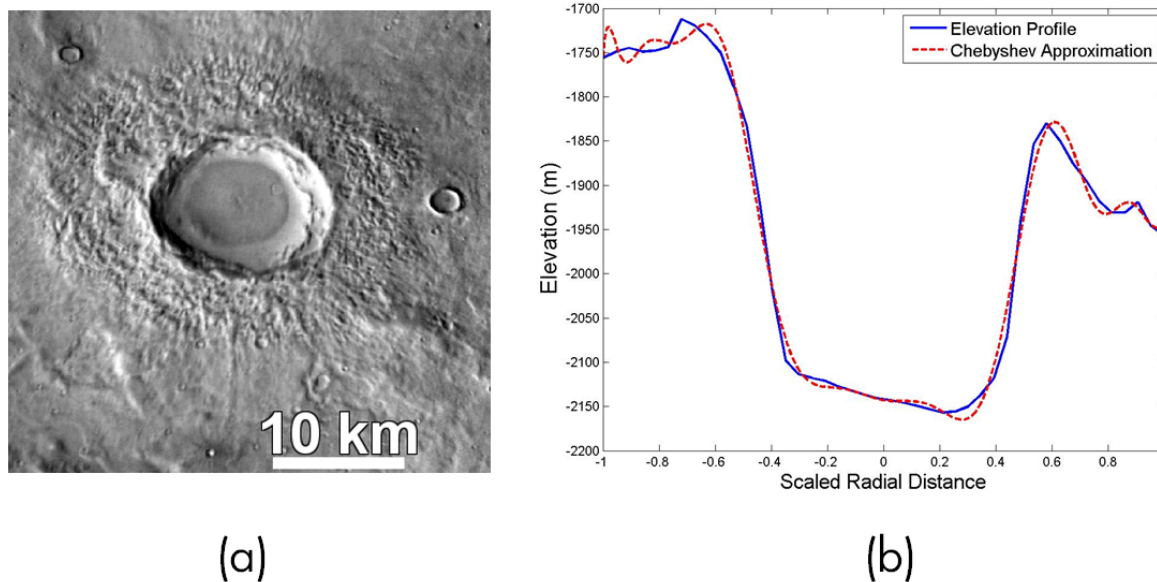


Figure 3.8: (a) A THEMIS image of a 14 km crater, catalogued as degradation class 3 but misclassified as class 1. (b) The 2-D profile and Chebyshev reconstruction of the crater, showing a significant pre-impact slope.

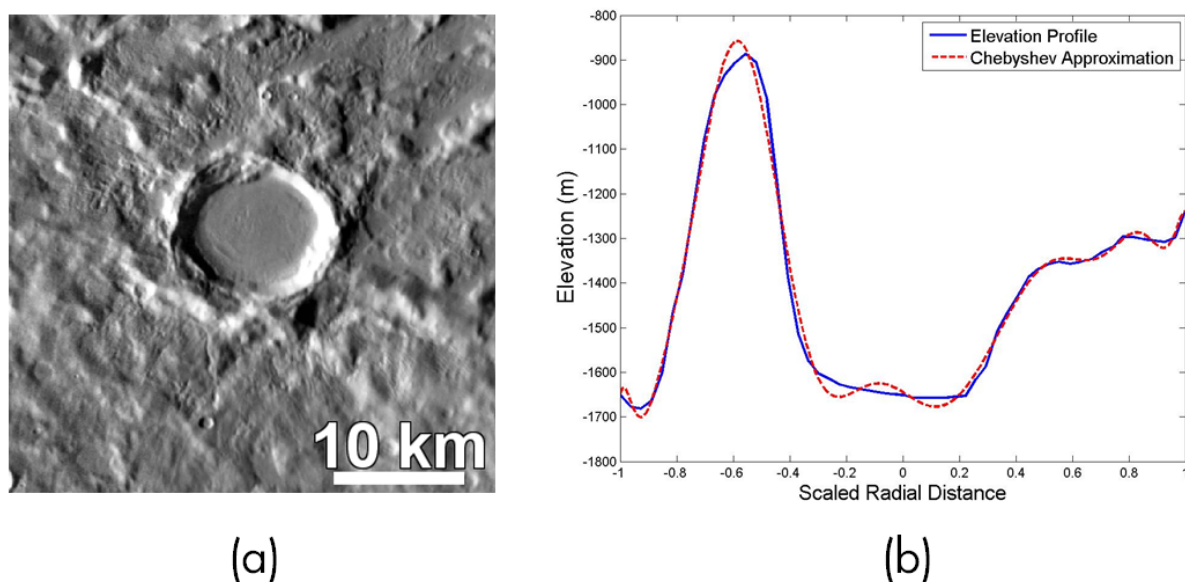


Figure 3.9: (a) A THEMIS image of a 13 km crater, catalogued as degradation class 3 but misclassified as class 1. This crater is in the ejecta blanket of Cerulli, a 130 km diameter crater to the north. (b) The 2-D profile and Chebyshev reconstruction of the crater, showing a good reconstruction despite misclassification.

of purity in cataloguing and an ‘artefact’ of the machine learning process. Even with large study areas such as the one chosen in this project, there are not many craters that

are assigned with a pristine designation. This results in nearly-exhaustive sampling of these craters in the training set. The training set is then perfectly representative of the population on the whole, which returns a high accuracy. However, strict uniformity in assigning craters as ‘pristine’ results in a very reliable crater set, which in theory should result in a set of discrimination attributes with minimum variation.

### 3.4.3 Interior Morphology Classification Performance

Assessment and classification of interior crater morphology is a task that should be feasible through the application of an automated method. This project attempted to assign a morphology label to crater candidates using the first 17 Chebyshev coefficients and combinations of the coefficients that represent the crater’s shape. The results shown in Tables 3.5 and 3.6 show relatively weak performance with identifying crater morphologies. This is for some features likely related to the approximation order. Choosing an order of  $M = 16$  does not sufficiently reproduce some of the features, like the central pit. This is represented in Figure 3.10.

This case is likely to explain misclassified instances of summit pits as central peaks as well. Summit pit features, being a special case for central peak craters, are not being fully resolved. Smoothing out the summit pit, but keeping lower frequency spatial information, would cause the crater to resemble a central peak crater. However, this effect should not be the case for flat-floored crater misclassifications, since they are low frequency features. Therefore, we can look to the coefficients as culprits. It is likely

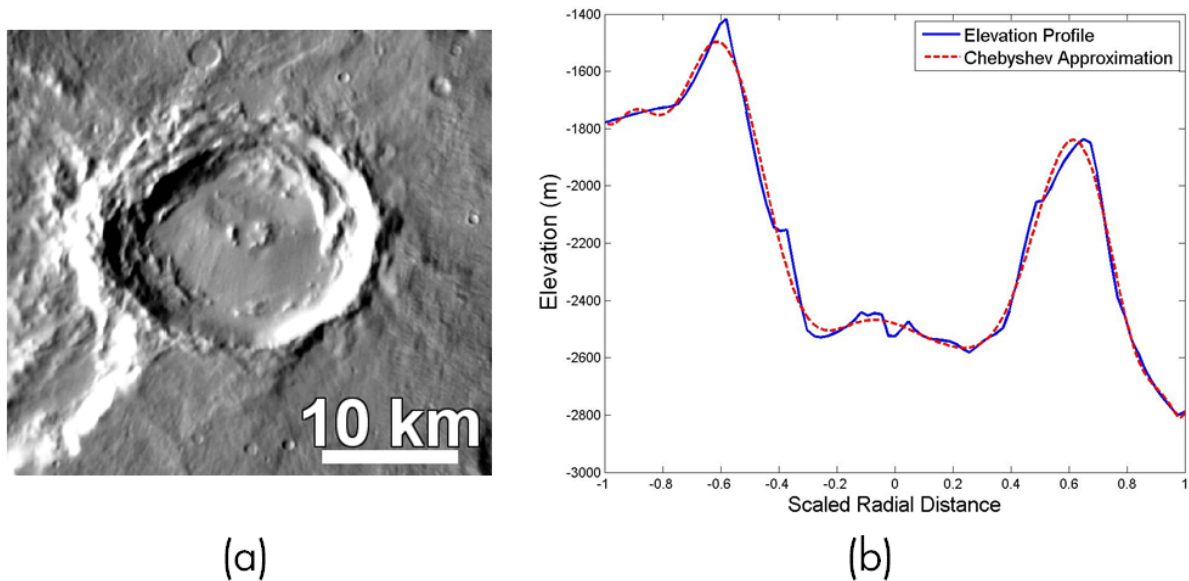


Figure 3.10: (a) A THEMIS image of a 20 km diameter crater with a central pit, located at  $30.35^{\circ}\text{N}$ ,  $18.60^{\circ}\text{E}$ . The pit is visible in the crater center. (b) The extracted topographic profile (blue line) and reconstructed Chebyshev approximation (red dashed line). The central pit, noticeable as the dip in the elevation profile, is distinctly missing from the approximation.

that to describe some of these features, it is necessary to define some new coefficient combinations that better represent some of these features. This, in addition to more robust profile collection which was previously described, is likely to improve the accuracy of the results.

Certain processes that significantly affect the craters were cause for misclassification as well. For example, the presence of a nearby crater that formed later and caused infilling of ejecta affected the classification results. Shown in Figure 3.12, a heavily degraded crater was misclassified as a result of a nearby, more recent impact. This crater, given a flat floor designation in the catalogue, has had a large portion of its flat floor mantled by the ejecta of a nearby crater that impacted its rim. This ejecta has a

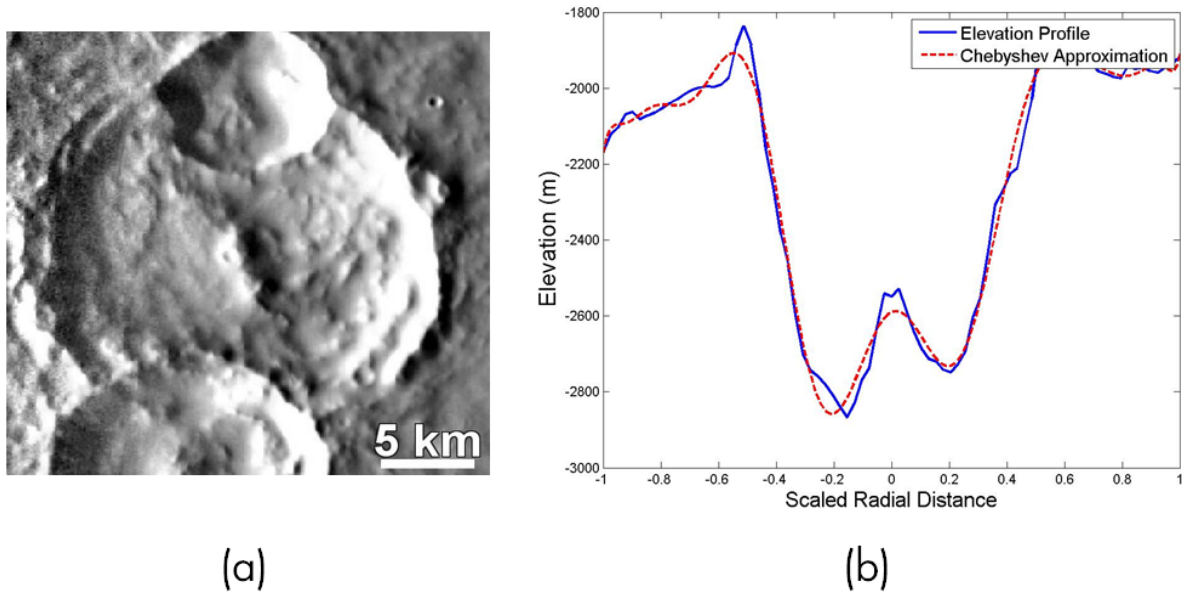


Figure 3.11: (a) A THEMIS image of a 21 km diameter crater with a central peak, located at  $22.09^{\circ}\text{N}$ ,  $8.71^{\circ}\text{E}$ . The peak is visible in the crater center. (b) The extracted topographic profile (blue line) and reconstructed Chebyshev approximation (red dashed line). This crater, listed in the source catalogue as unclassified (CpxUnc), exhibits a strong central uplift signature with a small summit pit when viewed as a 2D profile.

rougher topographic signature than the smooth flat floor. When examining the profile in Figure 3.12(b), we see the interior of the crater is rough, caused by the overlying ejecta. The rim of the newer crater is also included in the profile, causing a sharp peak around  $R_{scaled} = -0.9$ . This resulted in an ‘unclassified’ designation from the model.

Potential catalogued misclassifications are also likely responsible for some confusion in creating an accurate classification model. This is not necessarily the fault of the expert analysts; the set of possible crater morphologies is very large and there can be some ambiguity in assigning a class. An example of this is shown in Figure 3.11. Here, we have a degraded crater with multiple superposed impacts on its rim. This crater, originally given an unclassified designation, shows a strong central peak signature, especially when



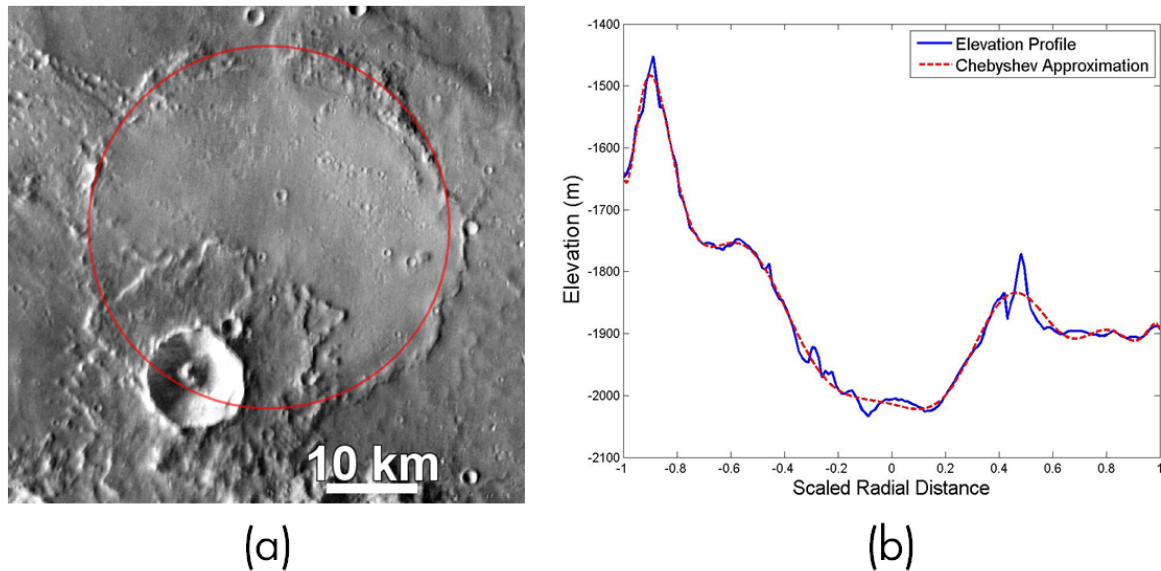


Figure 3.12: (a) A THEMIS image of a heavily degraded 42 km crater located at  $18.95^{\circ}\text{N}$ ,  $20.81^{\circ}\text{E}$ , catalogued as CpxFF (flat floor) but given an ‘unclassified’ designation by the model. The south-west rim of the crater is superposed by a very pristine 11 km crater. (b) The 2-D profile and Chebyshev reconstruction of the crater, showing a rough interior topography.

viewed as a linear elevation profile. These instances, which may be incorrectly labelled in the catalogue, affect the purity of the training set and reduce the accuracy of the model.

### 3.5 Conclusion

The current state of crater degradation state classification is dominated by the manual process of human experts assigning class values using a variety of datasets. The subjectivity introduced by an expert’s scrutiny can also incorporate inconsistency to the classification process. In addition, manual classification is tremendously time-inefficient. As new datasets are generated for other planetary bodies, a standardized and automated method for classifying crater morphology would become quite useful. This project applied an automated crater topography extraction and modelling process to

the problem of crater degradation assessment. The conclusions of this work include:

1. Polynomial reconstructions of crater topography can be efficiently extracted from topographic datasets in a standardized fashion.
2. Automated crater characterization systems can use the breadth and scope of existing crater catalogues to improve and assess their accuracy.
3. Using a Chebyshev polynomial expansion of order  $M = 16$ , craters can be divided into three classes ‘pristine’, ‘moderately preserved/degraded’, and ‘heavily degraded’ with accuracies of over 70% and as high as 96% if using a 2-class split.
4. Accurate description of crater interior morphology using the same method is not achievable at the same order, and requires polynomial approximation of a higher order to represent fine spatial detail with high fidelity.
5. Further refinement of standardized crater profile definition, as well as shape descriptors, is necessary for a more accurate automated classifier.

Expert assessment of crater degradation is not a process that relies solely on topography information. As such, it is believed that a more complete system could be developed that fuses multiple datasets together. This system would assign a preservation state based not only on topography, but perhaps using optical and thermal information as well. These more complete systems will improve the output from machine learning tools, and provide for a more robust system. Ultimately, the results will help researchers who are interested in extracting a specific class of craters, as they

seek to build more accurate models of crater modification and the surface processes that cause it.

# References

- Arvidson, R. E. (1974). Morphologic classification of Martian craters and some implications. *Icarus*, *22*(3), 264–271. doi: 10.1016/0019-1035(74)90176-6
- Barlow, N. G., & Bradley, T. L. (1990, sep). Martian impact craters: Correlations of ejecta and interior morphologies with diameter, latitude, and terrain. *Icarus*, *87*(1), 156–179.  
doi: 10.1016/0019-1035(90)90026-6
- Bresenham, J. (1963). An incremental algorithm for digital plotting. In *Acm nat. conf.*
- Chebyshev, P. L. (1854). Théorie des mécanismes connus sous le nom de parallélogrammes. *Mém. Acad. Sci. Pétersb.*, *7*, 539–568.
- Christensen, P. R., Jakosky, B. M., Kieffer, H. H., Malin, M. C., McSween, Jr., H. Y., Neelson, K., ... Ravine, M. (2004). The Thermal Emission Imaging System (THEMIS) for the Mars 2001 Odyssey Mission. *Space Science Reviews*, *110*(1/2), 85–130.  
doi: 10.1023/B:SPAC.0000021008.16305.94
- Craddock, R. A., & Howard, A. D. (2000, aug). Simulated degradation of lunar impact craters and a new method for age dating farside mare deposits. *Journal of*

- Geophysical Research: Planets*, 105(E8), 20387–20401.  
doi: 10.1029/1999JE001099
- de Vet, S. J., & de Bruyn, J. R. (2007, oct). Shape of impact craters in granular media. *Physical Review E*, 76(4), 041306.  
doi: 10.1103/PhysRevE.76.041306
- Driscoll, T. A., Hale, N., & Trefethen, L. N. (2014). *Chebfun Guide*. Oxford: Pafnuty Publications.
- Forsberg-Taylor, N. K., Howard, A. D., & Craddock, R. A. (2004). Crater degradation in the Martian highlands: Morphometric analysis of the Sinus Sabaeus region and simulation modeling suggest fluvial processes. *Journal of Geophysical Research*, 109(E5), E05002.  
doi: 10.1029/2004JE002242
- Grieve, R. A. F. (1998, jan). Extraterrestrial impacts on earth: the evidence and the consequences. *Geological Society, London, Special Publications*, 140(1), 105–131.  
doi: 10.1144/GSL.SP.1998.140.01.10
- Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., & Witten, I. H. (2009). The WEKA Data Mining Software: An Update. *ACM SIGKDD Explorations*, 11(1), 10–18.  
doi: 10.1145/1656274.1656278
- Head III, J. W. (1975). Process of Lunar Crater Degradation: Changes in Style with Geologic Time. *The Moon*, 12(3), 299–329. doi: 10.1007/bf02629699
- Kumar, P. S., Head, J. W., & Kring, D. A. (2010). Erosional modification and gully formation at Meteor Crater, Arizona: Insights into crater degradation processes on

- Mars. *Icarus*, 208(2), 608–620.  
doi: 10.1016/j.icarus.2010.03.032
- Mahanti, P., Robinson, M. S., Humm, D. C., & Stopar, J. D. (2014). A standardized approach for quantitative characterization of impact crater topography. *Icarus*, 241, 114–129.  
doi: 10.1016/j.icarus.2014.06.023
- McGill, G. E., & Wise, D. U. (1972). Regional variations in degradation and density of Martian craters. *Journal of Geophysical Research*, 77(14), 2433–2441.  
doi: 10.1029/JB077i014p02433
- Pathare, A. V., Paige, D. A., & Turtle, E. (2005, apr). Viscous relaxation of craters within the martian south polar layered deposits. *Icarus*, 174(2), 396–418.  
doi: 10.1016/j.icarus.2004.10.031
- Pike, R. J. (1977). Apparent depth/apparent diameter relation for lunar craters. *Lunar and Planetary Institute Science Conference Abstracts*, 3, 3427–3436.
- Pohn, H. A., & Offield, T. W. (1970). Lunar crater morphology and relative-age determination of lunar geologic units - Part 1. Classification. *U.S. Geological Survey Professional Paper*, 153–162.
- Robbins, S. J., & Hynek, B. M. (2012a). A new global database of Mars impact craters  $\geq 1$  km: 1. Database creation, properties, and parameters. *Journal of Geophysical Research E: Planets*, 117(5), 1–18. doi: 10.1029/2011JE003966
- Robbins, S. J., & Hynek, B. M. (2012b). A new global database of Mars impact craters  $\geq 1$  km: 2. Global crater properties and regional variations of the simple-to-complex transition diameter. *Journal of Geophysical Research E: Planets*, 117(6),

- 1–21. doi: 10.1029/2011JE003967
- Robbins, S. J., & Hynes, B. M. (2013). Utility of laser altimeter and stereoscopic terrain models: Application to Martian craters. *Planetary and Space Science*, *86*, 57–65.  
doi: 10.1016/j.pss.2013.06.019
- Ross, H. P. (1968). A simplified mathematical model for lunar crater erosion. *Journal of Geophysical Research*, *73*(4), 1343–1354.  
doi: 10.1029/JB073i004p01343
- Tanaka, K. L., Skinner, J. A., Dohm, J. M., Irwin III, R. P., Kolb, E. J., Fortezzo, C. M., ... Hare, T. M. (2014). *Geologic Map of Mars* (Tech. Rep.). U.S. Geological Survey.  
doi: 10.3133/sim3292
- Wood, C. A., Head III, J. W., & Cintala, M. J. (1977). Crater degradation on Mercury and the Moon: Clues to surface evolution. In *Proc. lunar sci. conf. 8th* (pp. 3503–3520).
- Zuber, M. T., Smith, D. E., Solomon, S. C., Muhleman, D. O., Head III, J. W., Garvin, J. B., ... Bufton, J. L. (1992). The Mars Observer laser altimeter investigation. *Journal of Geophysical Research*, *97*(E5), 7781–7797. doi: 10.1029/92JE00341

# Chapter 4

## Conclusions

### 4.1 Major Findings

This project sought to apply automated image processing methods to the study of impact craters and their bulk morphological characteristics for the Moon and Mars. This first study (Chapter 2) analysed how varying lunar terrain types affects the results of bulk crater detection and measurement, building from previous work and recently developed tools (Bue & Stepinski, 2007; Stepinski, Mendenhall, & Bue, 2009; Yin, Xu, Li, & Liu, 2013; Salamunićcar & Lončarić, 2008). The second (Chapter 3) developed the first fully-automated quantitative crater degradation classifier and applied it to martian complex craters. This system used polynomial approximations to model crater topographic profiles, and built a classification model using the coefficients of this approximation and expert-assigned degradation classes as a training set. This project drew from a large body of work on crater cataloguing, topographic analysis and degradation state classification (Robbins & Hynek, 2012; Barlow, 1988, 1995; Mahanti,



Robinson, Humm, & Stopar, 2014). Both projects seek to further the use of automated and quantitative methods for common processes in planetary science studies.

The major findings from the projects are as follows:

1. Automated systems that use topographic data to detect and measure craters are affected both in detection efficiency and measurement accuracy by the type of surface unit to which they are applied. Understanding surface type variation dependency is necessary for interpreting the integrity of the detection results.
2. Bulk impact crater statistics (such as morphometry) collected through automated means will be composed of a mixture of true population values, inherent detection biases and systematic measurement errors.
3. Despite these variations in robustness, automated collection of crater statistics can provide valuable information on crater populations, and the results can be used to inform models of crater scaling laws and ejecta emplacement.
4. Polynomial approximations of 2-D crater topographic profiles can be used to build an automated, objective crater degradation classifier.
5. Automated systems can be trained and guided by catalogues and other databases of expert-assigned crater descriptors.
6. Until more complete and robust systems for the automated detection and characterization of craters are developed, a semi-autonomous approach to image processing and analysis in planetary science remains the optimal approach.

## 4.2 Motivation for Automated Planetary Image Processing

The two papers contained in this work are fundamentally linked by their application of the concept of ‘automation’ to orbital data for planetary bodies. Automation of image processing tasks has not been restricted to orbital imagery, as work has been done for decision-making autonomy for landed assets such as rovers (Francis, 2014). In addition to building complicated image correction procedures (Chien & Mortensen, 1996), the automation of some analytical tasks in planetary science such as feature identification, measurement, annotation, or classification provides some important benefits. These include:

1. *Efficiency* - Digital image processing techniques can perform many tasks in rapid sequential order through batch processing
2. *Standardization* - Automated systems rely on a set of hard-wired rules for operating, and are thus not subject to the inconsistencies that would arise from a human analyst
3. *Repeatability* - Repeating the analysis using an automated method will yield identical results

For crater counting and measurement, points 1 and 3 are the most applicable. Crater densities become high as crater size decreases or count area increases, and thus the time required to perform a manual count also drastically increases. Also, craters could be

measured in an imprecise fashion or missed altogether when identified manually. The second point is most applicable to automated classification of crater degradation. While a manual crater degradation assessment can integrate information from different data sets that current automated processes cannot, it is also prone to inconsistencies that arise from different analysts performing the classification. An automated process relying entirely on consistently-extracted quantitative image information would resolve this issue.

### 4.3 Future Work

The last major finding enumerated in Section 4.1 posits that semi-autonomous approaches to image analysis in planetary science are the best analytical approach at the moment. This statement should not be interpreted as capitulation of the pursuit of full autonomy. Autonomous approaches suffer from inherent biases related to their particular approach; this often results in some form of dependence on the subject area, data type, or operating scale. Careful study of these biases can be used to improve autonomous systems that are used for many different tasks. Some potential future applications of the previously described work would be to expand the systems to other bodies. For example, it would be valuable to develop crater detection and characterization systems that work for icy satellites, in addition to terrestrial bodies. Developing more complete systems by integrating data fusion methods could also greatly increase their accuracy and scope.

One of the most valuable returns from using autonomous approaches is the

standardization and objectivity that they provide. Looking forward, it will be necessary when refining older methods or defining new approaches to carefully describe a standardized methodology for whichever task is being automated. By starting with simpler tasks such as data processing and handling, autonomy can be scaled up to more complex tasks. Eventually, machine learning and autonomy could be applied to all stages of the scientific method, from hypothesis formulation to experimentation and modelling (Mjolsness & DeCoste, 2001).

The continued cadence of robotic space exploration has begun a new age of solar system science, where tremendous amounts of data are being generated and returned to Earth every day. This wealth of data provides an opportunity for developing automated methods and systems that can be used going forward to increase both the efficiency of data analysis and science return. Although contemporary applications are relatively humble, further development could lead to a future where fully autonomous systems become the primary agents of exploration, engaging in the study of phenomena around the solar system and rapidly accelerating our understanding of the forces that shape our cosmic neighbourhood.

# References

- Barlow, N. G. (1988). Crater size-frequency distributions and a revised Martian relative chronology. *Icarus*, *75*(2), 285–305. doi: 10.1016/0019-1035(88)90006-1
- Barlow, N. G. (1995). The degradation of impact craters in Maja Valles and Arabia, Mars. *Journal of Geophysical Research*, *100*(E11), 23307.  
doi: 10.1029/95JE02492
- Bue, B. D., & Stepinski, T. F. (2007). Machine detection of Martian impact craters from digital topography data. *IEEE Transactions on Geoscience and Remote Sensing*, *45*(1), 265–274. doi: 10.1109/TGRS.2006.885402
- Chien, S., & Mortensen, H. (1996). Automating image processing for scientific data analysis of a large image database. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *18*(8), 854–859.  
doi: 10.1109/34.531806
- Francis, R. (2014). *Automated Image Interpretation for Science Autonomy in Robotic Planetary Exploration* (Ph.D., Western University).
- Mahanti, P., Robinson, M. S., Humm, D. C., & Stopar, J. D. (2014). A standardized approach for quantitative characterization of impact crater topography. *Icarus*,

241, 114–129.

doi: 10.1016/j.icarus.2014.06.023

Mjolsness, E., & DeCoste, D. (2001). Machine Learning for Science: State of the Art and Future Prospects. *Science*, 293(5537), 2051–2055.

doi: 10.1126/science.293.5537.2051

Robbins, S. J., & Hynek, B. M. (2012). A new global database of Mars impact craters  $\geq 1$  km: 1. Database creation, properties, and parameters. *Journal of Geophysical Research E: Planets*, 117(5), 1–18. doi: 10.1029/2011JE003966

Salamunićar, G., & Lončarić, S. (2008). Open framework for objective evaluation of crater detection algorithms with first test-field subsystem based on MOLA data. *Advances in Space Research*, 42(1), 6–19.

doi: 10.1016/j.asr.2007.04.028

Stepinski, T. F., Mendenhall, M. P., & Bue, B. D. (2009). Machine cataloging of impact craters on Mars. *Icarus*, 203(1), 77–87.

doi: 10.1016/j.icarus.2009.04.026

Yin, J., Xu, Y., Li, H., & Liu, Y. (2013, jul). A novel method of crater detection on digital elevation models. In *2013 IEEE International Geoscience and Remote Sensing Symposium - IGARSS* (pp. 2509–2512). IEEE.

doi: 10.1109/IGARSS.2013.6723331

# Appendix A

## MATLAB Algorithms and Decision Trees

### Mare Serenitatis Decision Tree

```
1 copyfile('fullseren.dat','fullseren_prune.dat')
2 catalog = 'fullseren_prune.dat';
3 cat = csvread(catalog);
4
5 for i = 1:size(cat,1);
6     if cat(i,3) > 0.087524,
7         cat(i,6) = 2;
8     else
9         if cat(i,3) <= 0.023431,
10            cat(i,6) = 1;
11        else
12            if cat(i,1) > 1511.930804,
13                cat(i,6) = 1;
14            else
15                if cat(i,3) > 0.035994,
16                    cat(i,6) = 2;
17                else
18                    if cat(i,4) > 0.066479,
19                        cat(i,6) = 2;
20                    else
21                        cat(i,6) = 1;
22                    end
23                end
24            end
25        end
26    end
27 end
```

```

28
29 dlmwrite('fullseren_prune.dat',cat,'delimiter',' ','precision'
, 16)

```

## Oriente Ejecta Decision Tree

```

1 copyfile('fullorient.dat','fullorient_prune.dat')
2 catalog = 'fullorient_prune.dat';
3 cat = csvread(catalog);
4
5
6 for i = 1:size(cat,1);
7     if cat(i,3) > 0.113858,
8         if cat(i,4) > 0.066541,
9             cat(i,6) = 1;
10        else
11            cat(i,6) = 2;
12        end
13    else
14        if cat(i,2) <= 85,
15            if cat(i,3) > 0.064734,
16                cat(i,6) = 2;
17            else
18                cat(i,6) = 1;
19            end
20        else
21            if cat(i,1) < 8385.661148,
22                cat(i,6) = 1;
23            else
24                if cat(i,1) > 11066.868596,
25                    cat(i,6) = 1;
26                else
27                    cat(i,6) = 2;
28                end
29            end
30        end
31    end
32 end
33
34 dlmwrite('fullorient_prune.dat',cat,'delimiter',' ','precision'
, 16)

```

## Southern Highlands Decision Tree

```

1 copyfile('fullhighland.dat','fullhighland_prune.dat')

```



```
2 catalog = 'fullhighland_prune.dat';
3 cat = csvread(catalog);
4
5
6 for i = 1:size(cat,1);
7     if cat(i,3) <= 0.077041,
8         if cat(i,3) <= 0.031953,
9             cat(i,6) = 1;
10        else
11            if cat(i,2) <= 78,
12                cat(i,6) = 2;
13            else
14                cat(i,6) = 1;
15            end
16        end
17    else
18        if cat(i,5) <= 0.022989,
19            cat(i,6) = 2;
20        else
21            if cat(i,1) <= 2420.103697,
22                cat(i,6) = 2;
23            else
24                if cat(i,5) > 0.061309,
25                    cat(i,6) = 1;
26                else
27                    if cat(i,5) > 0.028318,
28                        cat(i,6) = 2;
29                    else
30                        cat(i,6) = 1;
31                    end
32                end
33            end
34        end
35    end
36 end
37
38
39
40 dlmwrite('fullhighland_prune.dat',cat,'delimiter',' ','precision',16)
```

# Appendix B

## Crater Degradation Classification

### Crater Profile Extraction and Chebyshev Fitting

```
1  %{
2  This routine, when given a DIM and catalog of crater candidates
   in the format {ID, x_c, y_c, r, depth} is designed to
   automatically extract the craters 2-D topo profile and
   perform a Chebyshev polynomial reconstruction of the profile.
   It identifies the line of greatest gradient going through
   the center of the crater to a radial distance of 2R (total
   line length 4R).
3
4  This software uses two external functions: Bresenham's line
   algorithm (by Aaron Wetzler) and an overloaded Polyfit
   function from the Chebfun package (Oxford/Chebfun Developers)
   .
5
6  Written by Ian Pritchard, 2016.
7  %{
8  %% Read in the catalog, DIM, and initialize variables
9
10
11  cat = dlmread('arabia_catalog.txt',',',1,0);
12  fulldtm = dlmread('arabia_dtm.txt',',',6,0);
13  res = 0.463;
14  circs = cell(1,size(cat,1));
15
16  %% Calculate all perimeter points for each crater
17  for row = 1:size(cat,1)
18      r = cat(row,4)/res;
19      x_c = cat(row,3);
20      y_c = cat(row,2);
```

```

21     circ_points = [];
22     for th = 0:pi/50:(2*pi-(pi/50))
23         xunit = 2*r * cos(th) + x_c;
24         yunit = 2*r * sin(th) + y_c;
25         circ_points = [circ_points; xunit yunit];
26         circ_points = round(circ_points);
27         hold on
28         plot(yunit, xunit);
29         set(gca, 'xaxislocation', 'top', 'yaxislocation', 'left', '
           xdir', 'normal', 'ydir', 'reverse')
30     end
31     circs{1,row} = circ_points;
32 end
33
34 %% Calculate highest gradient point pairs
35
36 slopes = cell(1, size(cat,1));
37 maxindices = zeros(1, size(cat,1));
38 linepairs = cell(1, size(cat,1));
39
40 for j = 1:length(circs)
41     pointpairs = circs{j};
42     pointpairs = round(pointpairs);
43     slope = zeros(50,1);
44     for k = 1:(length(pointpairs)-50)
45         elev1 = fulldtm(pointpairs(k,1), pointpairs(k,2));
46         elev2 = fulldtm(pointpairs(k+50,1), pointpairs(k+50,2));
47         slope(k,1) = abs((elev2 - elev1) / (sqrt(((pointpairs(k
           +50,1)-pointpairs(k,1))^2)+((pointpairs(k+50,2)-
           pointpairs(k,2))^2)*res)));
48         slopes{1,j} = slope;
49     end
50     [steepest, ind] = max(slope);
51     maxindices(1,j) = ind;
52     pointlist = circs{1,j};
53     maxpts = pointlist(maxindices(1,j),:);
54     maxpts = [maxpts; pointlist(maxindices(1,j)+50,:)];
55     linepairs{1,j} = maxpts;
56 end
57
58 %% Use maximum indices to build line with Bresenham's algorithm.
59 linepoints = {};
60
61 for n = 1:length(linepairs)
62     maxpts2 = linepairs{1,n};

```

```

63     [x,y] = bresenham(maxpts2(1,1),maxpts2(1,2),maxpts2(2,1),
64                     maxpts2(2,2));
65     map = plot(y,x,'or');
66     linepoints{1,n} = [x,y];
67 end
68 saveas(map,'circmap.jpg');
69
70 hold off
71 %% Build 2-D profiles using lines, normalize to between -1,1
72
73 profiles = cell(1,size(cat,1));
74 coeffs = zeros(1,17);
75 coeffs_list = [];
76
77 for crat = 1:length(linepoints)
78     profilepts = linepoints{1,crat};
79     for i = 1:length(profilepts);
80         profilepts(i,3) = (-1+(i-1)*2/(size(profilepts,1)-1));
81         profilepts(i,4) = fulldtm(profilepts(i,1),profilepts(i,2));
82     end
83     profiles{1,crat} = profilepts;
84     figure('visible','off')
85     topline = plot(profilepts(:,3),profilepts(:,4),'LineWidth',2);
86     hold on
87     %% Chebfitting
88     profilepts = profiles{1,crat};
89     f = polyfit(profilepts(:,3),profilepts(:,4),16,domain(-1,1))
90         ;
91     chebfitplot = plot(f,'-r','LineWidth',2);
92     xlabel('Scaled Radial Distance','FontSize',16)
93     ylabel('Elevation (m)','FontSize',18)
94     legend({'Elevation Profile','Chebyshev Approximation'},'FontSize',14,'Location','northeast')
95     saveas(chebfitplot,[num2str(cat(crats,1)) '.jpg']);
96     hold off
97     coeffs = chebcoeffs(f);
98     coeffs = coeffs';
99     coeffs = [cat(crats,1) coeffs];
100     coeffs_list = [coeffs_list;coeffs];
101 end
102
103 dlmwrite('chebyshev_coeffs.csv',coeffs_list,'delimiter',' ','precision',8)

```

**List of Chebyshev Polynomials, M=8**

$$T_0(x) = 1$$

$$T_1(x) = x$$

$$T_2(x) = 2x^2 - 1$$

$$T_3(x) = 4x^3 - 3x$$

$$T_4(x) = 8x^4 - 8x^2 + 1$$

$$T_5(x) = 16x^5 - 20x^3 + 5x$$

$$T_6(x) = 32x^6 - 48x^4 + 18x^2 - 1$$

$$T_7(x) = 64x^7 - 112x^5 + 56x^3 - 7x$$

$$T_8(x) = 128x^8 - 256x^6 + 160x^4 - 32x^2 + 1$$

# Curriculum Vitae

**Name:** Ian Pritchard

**Post-Secondary Education and Degrees:** Queen's University  
Kingston, ON  
2008 - 2012, BScH Physics and Mathematics

University of Western Ontario  
London, ON  
2014 - 2016, MSc Planetary Science and Exploration

**Honours and Awards:** Western Graduate Research Scholarship  
2014-2016

Center for Planetary Science and Exploration (CPSX)  
Interdisciplinary Graduate Scholarship  
2016

**Related Work Experience:** Research & Teaching Assistant  
The University of Western Ontario  
2014 - 2016

## Presentations:

**Pritchard, I.M.**, Wang, J. & Stooke, P. (2016, March) *Automated detection of lunar impact craters in elevation data - insights and related accuracies*, Poster presented at the Lunar and Planetary Science Conference, The Woodlands, TX

Cross, M.D.G., **Pritchard, I.M.**, Francis, R. & Osinski, G.R. (2016, March) *Science activity resource planning for 2015 CANMARS MSR analogue mission*, Poster presented at the Lunar and Planetary Science Conference, The Woodlands, TX

Silber, E.A., Osinski, G.R., Francis, R., Cross, M.D.G. & **Pritchard, I.M.** (2016, March) *2015 CANMARS MSR analogue mission: An overview of the mission control tactical team*, Poster presented at the Lunar and Planetary Science Conference, The

Woodlands, TX

**Pritchard, I.M.**, Wang, J. & Stooke, P. (2016, April) *Lunar impact crater detection using digital terrain models (DTMs) - terrain dependence*, Poster presented at the Conference of the American Association of Geographers, San Francisco, CA

**Pritchard, I.M.** (2015, April) *Image Processing in Planetary Science*, public lecture given at Western Space Day, London, ON