An Interdisciplinary Approach to Understanding Volcanoes and Their Processes

Katherine Cosburn

Follow this and additional works at: https://digitalrepository.unm.edu/phyc_etds

Part of the Geophysics and Seismology Commons, and the Physics Commons

Recommended Citation

This Dissertation is brought to you for free and open access by the Electronic Theses and Dissertations at UNM Digital Repository. It has been accepted for inclusion in Physics & Astronomy ETDs by an authorized administrator of UNM Digital Repository. For more information, please contact disc@unm.edu.
Katherine Cosburn  
Candidate  

Physics and Astronomy  
Department  

This dissertation is approved, and it is acceptable in quality and form for publication:

Approved by the Dissertation Committee:

Mousumi Roy, Chairperson  

Richard Rand  

Keith Lidke  

Brandon Schmandt
An Interdisciplinary Approach to Understanding Volcanoes and Their Processes

by

Katherine Cosburn

B.Sc., Physics, University of Toronto, 2016
M.Sc., Physics, University of New Mexico, 2019

Dissertation

Submitted in Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy
Physics

The University of New Mexico

Albuquerque, New Mexico

May, 2022
“Perhaps we attend to a volcano for its elevation, like ballet. How high the molten rocks soar, how far above the mushrooming cloud. The thrill is that the mountain blows itself up, even if it must then like the dancer return to earth; even if it does not simply descend—it falls, falls on us. But first it goes up, it flies.”

Susan Sontag, The Volcano Lover
I would like to thank my adviser, Professor Mousumi Roy, for her unwavering support over the years, for believing in my abilities, and for pushing me towards independent thinking and research. I would also like to thank my family and friends who have made this journey possible through their companionship and encouragement. Thanks go out to my dissertation committee for their constructive feedback, as well as to all of the people I have been fortunate to collaborate with over the years. My research and my writing would not be what it is today without the help of all those mentioned.
An Interdisciplinary Approach to Understanding Volcanoes and Their Processes

by

Katherine Cosburn

B.Sc., Physics, University of Toronto, 2016
M.Sc., Physics, University of New Mexico, 2019
Ph.D., Physics, University of New Mexico, 2022

Abstract

To better understand volcanoes and their processes is important from both a fundamental science perspective and for hazard monitoring purposes. The complexity and limitations we face in pursuing such a science are numerous and this dissertation explores how an interdisciplinary approach combining physics, computer science, and volcanology can address this complexity in a straightforward and meaningful way. This is achieved through various modelling techniques across three studies: (1) a first-order analytic modelling of stratovolcano topographic shape, (2) the use of a Bayesian joint inversion on gravity and novel cosmic-ray muon measurements for imaging flat-lying subsurface density anomalies, and (3) the use of a machine learning model for subsurface density prediction at a volcano. All three studies presented provide important insights into understanding what goes on underneath volcanoes, with the intention that they might guide future hazard monitoring practices.
Contents

List of Figures xi

List of Tables xvii

1 Introduction 1

1.1 Introduction to Volcanology .............................. 2

1.2 Volcano Hazard Monitoring ............................... 5

1.3 Physics of Volcanic Processes ............................ 7

1.3.1 Lava Flow ............................................ 7

1.3.2 Deformation from Magma Emplacement ............... 9

1.4 Gravimetry and Muography ............................... 13

1.4.1 Gravimetry .......................................... 13

1.4.2 Muography ........................................... 15

1.5 Inverse Theory ........................................... 17

1.6 Machine Learning .......................................... 19

1.6.1 Theory .................................................. 19
1.6.2 Relevant Example .................................................. 24
1.7 Overview of Following Chapters ................................. 28

2 Analysing the Topographic Form of Stratovolcanoes 30

2.1 Abstract ................................................................. 30
2.2 Introduction ............................................................ 31
2.3 Ideal Stratovolcano Shape ......................................... 34
  2.3.1 Extracting Topographic Profiles .............................. 34
  2.3.2 Basal Ellipticity and Volcano Axisymmetry ............... 35
  2.3.3 Profile Fitting to Analytic Form ............................. 37
  2.3.4 Physical Explanation for the Analytic Form ............... 39
2.4 Ideal Stratovolcano Shape: Fitting Results .................... 42
2.5 Surface Effects from Subsurface Dyke Intrusions .............. 45
  2.5.1 Dyke Distribution Model ...................................... 46
  2.5.2 Relative Contribution to Surface Displacement ........... 48
2.6 Results from Addition of Dyke Distribution .................... 50
2.7 Discussion ............................................................. 50
  2.7.1 Deviations from Axisymmetry in Ideal Form ............... 52
  2.7.2 Axisymmetric Protuberant Deviations from the Ideal Form .. 53
  2.7.3 Directional Protuberant Deviations from Ideal Form ....... 55
2.8 Conclusions ........................................................... 56
3 Joint Inversion of Cosmic-Ray Muon & Gravity Data 65

3.1 Abstract .............................................. 65
3.2 Introduction .......................................... 66
3.3 Target Density Anomaly ................................. 68
3.4 Data Acquisition and Analysis ......................... 70
3.5 Joint Inversion ........................................ 76
3.6 Results ................................................ 77
   3.6.1 Case 1: Uniform density, isotropic smoothing  ......... 78
   3.6.2 Case 2: Uniform density, anisotropic smoothing .......... 79
   3.6.3 Case 3: Layered density, anisotropic smoothing .......... 79
3.7 Discussion and Conclusions ............................. 80

4 Machine Learning with Cosmic-Ray Muon & Gravity Data 87

4.1 Abstract .............................................. 87
4.2 Introduction .......................................... 88
4.3 Machine Learning Model for Showa-Shinzan ............... 92
   4.3.1 Data and Geological Setting ....................... 92
   4.3.2 Domain Discretization ............................ 93
   4.3.3 Forward Calculation and Traditional Inversion ........... 95
   4.3.4 Machine Learning Joint Inversion ................... 95
4.4 Results ............................................... 101
Contents

4.5 Discussion .................................................. 102

4.6 Conclusion .................................................. 105

5 Conclusion ..................................................... 115

5.0.1 Summary .................................................. 115

5.0.2 Future Work ............................................. 119

Appendices ....................................................... 124

A Chapter 2: Stochastic Model of Edifice Construction ...... 125

B Chapter 2: Published Mendeley Database ................. 130

C Chapter 3: Synthetic & Weighting Tests .................. 139

C.1 Synthetic Tests .............................................. 139

C.2 Weighting Tests ............................................. 141

D Chapter 4: Different Model Parameters & Shuffling Data 150

D.1 Effects of Different Model Parameters .................. 150

D.1.1 Larger and Smaller Domain Discretization ........... 150

D.1.2 Larger Background Density ............................. 151

D.1.3 Smaller Training/Testing Dataset Size ............... 153

D.2 Effects of Shuffling Input Data ........................... 154

E Select Code ..................................................... 156
Contents

E.1 Okada Solutions for Tensile Dislocation .................................. 156
   E.1.1 okada92_ke.m ............................................. 156
   E.1.2 f_dfxyz.m ............................................... 162

E.2 Bayesian Joint Inversion for Density Prediction with Muon & Gravity Data .................................................. 169

E.3 Machine Learning Algorithm for Density Prediction with Muon & Gravity Data ................................................. 178
   E.3.1 Data Generation Algorithm .................................. 179
   E.3.2 Machine Learning Algorithm ................................. 185

References 188
List of Figures

1.1 Examples of monogenetic and polygenetic (both strato and shield) volcanoes from around the world. .......................... 3

1.2 Cartoon of a volcano and the possible products and events that it produces during eruption (modified from USGS). ................. 3

1.3 Cartoon of a stratovolcano underlain by a heterogeneous host rock and system of dykes, sills, conduit, and magma chamber. (modified from Gudmundsson (2020)). ................................. 5

1.4 Cartoon showing the relationship between stress, $\tau$, and strain rate, $\dot{\gamma}$, for a Newtonian versus Bingham fluid. The slope of the line for the Bingham fluid is given by the plastic viscosity, $\mu_p$. .......... 8

1.5 Cartoon showing the relationship between the weight and the shear stress of lava. .................................................. 10

1.6 Cartoon of showing different stress-strain behaviour for loading and unloading of a linearly elastic, nonlinearly elastic, and inelastic material. ......................................................... 11

1.7 Cartoon of showing probabilistic correlation between data and model parameters of the inverse problem and Gaussian distribution of the error (adapted from Tarantola (2005)). ......................... 17
<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.8</td>
<td>(a) Data from Grosse et al. (2014) used in example problem and (b) plot of absolute value of the residuals over two iterations of the gradient boosting algorithm.</td>
</tr>
<tr>
<td>1.9</td>
<td>Decision trees for regression for $m = 1$ and $m = 2$ iterations of the gradient boosting algorithm.</td>
</tr>
<tr>
<td>2.1</td>
<td>Cartoon outlining the important parameters of the analytic model (further explained in text)</td>
</tr>
<tr>
<td>2.2</td>
<td>Map showing $N = 190$ isolated stratovolcanoes assessed in the study on the regular form of stratovolcanoes.</td>
</tr>
<tr>
<td>2.3</td>
<td>DEM of Mayon volcano with shaded regions corresponding to the quadrants bisected by either a semi-major or semi-minor axis over which an average profile is extracted.</td>
</tr>
<tr>
<td>2.4</td>
<td>Three volcanoes representative of each an axisymmetric and nonaxisymmetric good fit to the analytic model.</td>
</tr>
<tr>
<td>2.5</td>
<td>Four volcanoes representative of where there are protuberant deviations to the analytic model.</td>
</tr>
<tr>
<td>2.6</td>
<td>Illustration of dyke parameters and placement on disc-shaped magma reservoir.</td>
</tr>
<tr>
<td>2.7</td>
<td>Views of the uncorrelated and cone models with a sample of dykes distributed on the disc-shaped magma reservoir.</td>
</tr>
<tr>
<td>2.8</td>
<td>Plot at each chamber depth of total intruded volume vs. maximum displacement, as well as example profiles of normalized uplift at the surface.</td>
</tr>
</tbody>
</table>
LIST OF FIGURES

2.9 Cartoon of the scaling problem used to determine the relative contributions of uplift due to dyke distributions that affect a topographic edifice. ................................................................. 62

2.10 Results of the combination model for two stratovolcanoes with protuberant deviations in all quadrants and results of the combination model for two stratovolcanoes with protuberant deviations in one or two quadrants. ................................................................. 63

2.11 Profile of Merapi fit by the ideal analytic form shows a steeper slope in the quadrants bisected by the $a_2$, $b_1$ axes, corresponding to S-SW and the dip angle of the planar trend. ................................................................. 64

3.1 View of hillslope from the TA-41 tunnel, regional stratigraphy of the Bandelier Tuff, and digitized borehole density log from a well on the Pajarito Plateau. ................................................................. 69

3.2 Measured gravity anomalies and 3D histograms of muon range vs zenith angle, azimuthal angle for both outer and inner detector positions. ................................................................. 72

3.3 DEM of study area and optimised discretisation both showing gravity station and muon detector locations. ................................................................. 73

3.4 Model representation of muon trajectories for a selection of $(\theta, \phi)$ bins and muon range (density-length) as a function of attenuation. ................................................................. 74

3.5 Raw muon attenuation data, coloured with respect to $-\ln(N_{j}^{in}/N_{j}^{out})$ using 3-degree angle bins for the detector at inner and outer detector positions. ................................................................. 75

3.6 Results of joint inversion with uniform prior density and isotropic smoothing using the best-fit parameters. ................................................................. 83
<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.7</td>
<td>Results of joint inversion with uniform prior density and anisotropic smoothing using the best-fit parameters</td>
<td>84</td>
</tr>
<tr>
<td>3.8</td>
<td>Results of joint inversion with layered, geologic density prior and anisotropic smoothing using the best-fit parameters</td>
<td>85</td>
</tr>
<tr>
<td>3.9</td>
<td>Two-dimensional slice plot of results from Figures 3.6-3.8</td>
<td>86</td>
</tr>
<tr>
<td>4.1</td>
<td>Map and DEM of Mount Usu and the Showa-Shinzan lava dome</td>
<td>90</td>
</tr>
<tr>
<td>4.2</td>
<td>Figures of gravity and muon data taken at the Showa-Shinzan lava dome</td>
<td>92</td>
</tr>
<tr>
<td>4.3</td>
<td>Discretized domain of the Showa-Shinzan lava dome</td>
<td>94</td>
</tr>
<tr>
<td>4.4</td>
<td>Cartoon illustrating how the machine learning algorithm works</td>
<td>106</td>
</tr>
<tr>
<td>4.5</td>
<td>Workflow for the data generation and machine learning training and testing process</td>
<td>107</td>
</tr>
<tr>
<td>4.6</td>
<td>Four examples of possible synthetic perturbation patterns within the target region</td>
<td>108</td>
</tr>
<tr>
<td>4.7</td>
<td>Results of hyperparameter search to find optimal learning rate, maximum number of leaf nodes, and maximum number of iterations</td>
<td>109</td>
</tr>
<tr>
<td>4.8</td>
<td>Results of machine learning inversion (trained on synthetic data with no added noise) using data from Nishiyama et al. (2017b) for the Showa-Shinzan lava dome</td>
<td>110</td>
</tr>
<tr>
<td>4.9</td>
<td>Results of machine learning inversion (trained on synthetic data with added noise) using data from Nishiyama et al. (2017b) for the Showa-Shinzan lava dome</td>
<td>111</td>
</tr>
<tr>
<td>Figure</td>
<td>Description</td>
<td>Page</td>
</tr>
<tr>
<td>--------</td>
<td>-----------------------------------------------------------------------------</td>
<td>------</td>
</tr>
<tr>
<td>4.10</td>
<td>2D slice plots through the vertical at 40 m increase in elevation comparing results with and without noise to results from Nishiyama et al. (2017b).</td>
<td>112</td>
</tr>
<tr>
<td>4.11</td>
<td>3D structure within the target region of the absolute difference between the density predictions with and without added noise.</td>
<td>113</td>
</tr>
<tr>
<td>4.12</td>
<td>Various plots showing uncertainty index and sensitivity kernel analysis.</td>
<td>114</td>
</tr>
<tr>
<td>A.1</td>
<td>3D model edifice built from ( n = 8000 ) flows, average profiles obtained from the 3D edifices, and contour plots of the absolute difference in ( \varphi_c ) and the ratio ( h_m/R_b ) from our model with the average from our data.</td>
<td>128</td>
</tr>
<tr>
<td>A.2</td>
<td>Average profiles for 1000 to 4000 flows with various parameter bounds from the model.</td>
<td>129</td>
</tr>
<tr>
<td>B.1</td>
<td>Figures for all volcanoes in category 1 with low ellipticity index.</td>
<td>131</td>
</tr>
<tr>
<td>B.2</td>
<td>Figures for all volcanoes in category 1 with intermediate/high ellipticity index.</td>
<td>132</td>
</tr>
<tr>
<td>B.3</td>
<td>Figures for all volcanoes in category 2 with low ellipticity index.</td>
<td>133</td>
</tr>
<tr>
<td>B.4</td>
<td>Figures for all volcanoes in category 2 with intermediate/high ellipticity index.</td>
<td>134</td>
</tr>
<tr>
<td>B.5</td>
<td>Parameters of best fit to the ideal form for volcanoes in all categories (page 1).</td>
<td>135</td>
</tr>
<tr>
<td>B.6</td>
<td>Parameters of best fit to the ideal form for volcanoes in all categories (page 2).</td>
<td>136</td>
</tr>
</tbody>
</table>
**LIST OF FIGURES**

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>B.7</td>
<td>Parameters of best fit to the ideal form for volcanoes in all categories (page 3)</td>
</tr>
<tr>
<td>B.8</td>
<td>Parameters of best fit to the ideal form for volcanoes in all categories (page 4)</td>
</tr>
<tr>
<td>C.1</td>
<td>True density structure for anomaly thickness $d_0 = 2183, h_0 = 30$ m with inversion results</td>
</tr>
<tr>
<td>C.2</td>
<td>True density structure for anomaly thickness $d_0 = 2183, h_0 = 50$ m with inversion results</td>
</tr>
<tr>
<td>C.3</td>
<td>True density structure for anomaly thickness $d_0 = 2171, h_0 = 37$ m with inversion results</td>
</tr>
<tr>
<td>C.4</td>
<td>Inversion results using a uniform prior and anisotropic smoothing with equal and unweighted weighting schemes</td>
</tr>
<tr>
<td>C.5</td>
<td>Inversion results (without tunnel gravity) using a uniform prior and anisotropic smoothing with equal and unweighted weighting schemes</td>
</tr>
<tr>
<td>C.6</td>
<td>Inversion results using a layered prior and anisotropic smoothing with equal and unweighted weighting schemes</td>
</tr>
<tr>
<td>D.1</td>
<td>Results using larger $(100 \times 100 \times 20$ m) and smaller $(50 \times 50 \times 20$ m) discretizations</td>
</tr>
<tr>
<td>D.2</td>
<td>Results using a higher background density of 2200 kg/m$^3$.</td>
</tr>
<tr>
<td>D.3</td>
<td>Results using a higher background density of 1900 kg/m$^3$.</td>
</tr>
<tr>
<td>D.4</td>
<td>Results using a smaller training/testing dataset size, on the order of $10^5$.</td>
</tr>
<tr>
<td>D.5</td>
<td>Results using a shuffled input dataset</td>
</tr>
</tbody>
</table>
List of Tables

1.1 Results for the simple example of the gradient boosting algorithm for predicting volcano volume given height and basal width. 24

2.1 Table of \( N = 102 \) stratovolcanoes with low ellipticity index \((\varepsilon \leq 1.5)\) used in the study on the regular shape of stratovolcanoes. 58

2.2 Table of \( N = 88 \) stratovolcanoes with intermediate \((1.5 < \varepsilon \leq 2)\) or high \((2 < \varepsilon \leq 3)\) ellipticity index used in the study on the regular shape of stratovolcanoes. 59

2.3 Values for dyke parameters (Figure 2.6a) used in both the uncorrelated and cone distribution models (Figure 2.7). 60

4.1 Hyperparameter values used in this study. The search for optimal learning rate, maximum number of leaf nodes, and maximum number of trees (or iterations) can be found in Figure 4.7. 100

A.1 Values for maximum and minimum bounds on the distribution of flow parameters \( s_y, k, \) volume, and \( \sigma_r \). 127
Chapter 1

Introduction

“I looked back; a dense dark mist seemed to be following us, spreading itself over the country like a flood....We had scarcely sat down when night came upon us, not such as when the sky is cloudy, or when there is no moon, but that of a room when it is shut up and all the lights put out.” — Pliny the Younger, c. 104 A.D.

The letters from Pliny the Younger to his friend, the historian Cornelius Tacitus, describing the eruption of Mount Vesuvius on 24 August 79 provide a harrowing account of one of history’s most famous volcanic eruptions. This eruption, however, is only unique in that it was well-documented. For as long as there has been human civilization, there has been human interaction with volcanoes—with our place in that interaction taking the form of a bowing to nature’s will. The the need to observe, predict, and asses the hazards associated with volcanic activity has always been present, and as we have moved into the twenty-first century our abilities to do as such have improved significantly. We live in a time of great technological advance and our arsenal against volcanic threat is furnished with powerful ammunition. We are equipped with sophisticated monitoring devices, improved analogue and numeric modelling, “big” data and new techniques for analysing it, as well as better and faster computers that allow us to achieve these advancements. It has been exciting
for this author to have found herself participating in this field at the interdisciplinary crossroads between physics, computer science, and geology. It is the intention with this dissertation and the work herein to demonstrate a series of studies that exist at this crossroad, and which contribute meaningfully to such an exciting, important, and eclectic field of science.

1.1 Introduction to Volcanology

To begin a discussion on volcanoes and their processes, it is necessary to first outline some key aspects of what a volcano is and the various components that make them into the landforms we recognize. First, we can distinguish between monogenetic and polygenetic volcanoes (Figure 1.1), the latter being what we are most familiar with, as they are what we are taught as children what a volcano looks like. This is rightfully so, since they are largest of the two and typically pose the greatest threat to civilization. The former, monogenetic volcanoes are created from a single eruptive episode (which can last a long time and have periods of high and low mass flux) and usually take the form of spatter or cinder cones. Polygenetic volcanoes, on the other hand, are created over multiple eruptive episodes, as eruptive material like lava, debris, and ash pile up and creates an edifice (Figure 1.2). These volcanoes occur in two main types: shield and stratovolcanoes. Shield volcanoes (or basaltic edifices) like those found in Hawai’i or Iceland, have much different profiles than those of stratovolcanoes, such as Fuji in Japan or Mayon in the Philippines, which have considerably steeper slopes than their shield cousins (see Chapter 2 and Cosburn & Roy (2020b) on the regular form of stratovolcanoes).

This is primarily due to their differing lava compositions—shield volcanoes tend to have more basaltic lava with lower viscosity and yield strength, and thus flow more freely. This causes them to exhibit more effusive eruptions, whilst stratovolcanoes are more likely to exhibit more explosive eruptions—sometimes erupting lava, but
Figure 1.1: Examples of monogenetic and polygenetic (both strato and shield) volcanoes from around the world.

Figure 1.2: Cartoon of a volcano and the possible products and events that it produces during eruption (modified from USGS).
other times erupting pyroclastic flows or tephra, which all influence the mechanical structure of their edifice (Gudmundsson, 2020). For these various reasons, the mechanical strength of stratovolcanoes is much higher than shield volcanoes, causing them to have much steeper slopes and exhibiting the “ideal”, concave form that we often associate with volcanoes (Chapter 2; Cosburn & Roy (2020b)). The higher mechanical strength and more variable mechanical properties of stratovolcanoes also means that fractures are more easily arrested within the edifice, since the stress fields are too heterogeneous to allow for large fractures. The heterogeneity of these stress fields adds to their mechanical strength, which allows them to grow larger without suffering as many deconstructive effects from within the edifice (Gudmundsson, 2020).

Despite all of this, shield volcanoes and stratovolcanoes also have a lot in common. As mentioned before, both are polygenetic volcanoes that are built up into large land masses over time (Figure 1.1). Both are also underlain by complex systems of magma chambers and reservoirs, which are interconnected by magma-filled cracks called dykes or inclined sheets and sills (Figure 1.3). It must be noted that Figure 1.3 shows a magmatic system that is somewhat outdated, since more recent views show magma stored/pressurized within “mush zones” that encompass the whole crust (Cashman et al., 2017). The overpressures within this “plumbing” system are what drive volcanic eruptions, as well as what dictates the lifespan of a volcano. Most edifices are built up due to shallower (less than 5 km deep) magma chambers, which feed eruptions at the surface and are in turn fed by deeper magma reservoirs. Once a reservoir stops providing magma to a shallower chamber—either because it becomes empty or the connection to the chamber, such as a dyke or conduit, disappears—edifice construction tends to stop and a volcano becomes dormant or extinct. It is found that stratovolcanoes are more likely to deviate from their ideal form once this happens and as they advance further in their evolution from birth to extinction (Chapter 2; Cosburn & Roy (2020b); Grosse et al. (2009)).
Figure 1.3: Cartoon of a stratovolcano underlain by a heterogeneous host rock and system of dykes, sills, conduit, and magma chamber. (modified from Gudmundsson (2020)).

1.2 Volcano Hazard Monitoring

Accurate forecasts, measurements, and interpretations of volcanic processes are imperative to the safety and well-being of civilization around the globe. The frequency of volcanic eruptions is high and relatively constant, with around 100 erupting in any given year and more than ten erupting at any given time somewhere on Earth (Loughlin et al., 2015). Volcano science must focus on reducing the impact of volcanic events on humanity, which requires more accurate forecasts of eruptions, made possible by the development and pursuit of new observations, methods of observations, and models of volcano processes. Various geochemical and geophysical techniques can be used to study these processes from very small scales (e.g. magma crystal content) to very large (e.g. lava flow path for a given eruption). Accurate modelling of volcanic processes is imperative for interpreting geochemical and geophysical observations and ultimately gaining more insight into specific hazards and risks. Models can help guide data collection, while data collection provides us with essential information about a volcano’s eruptive behaviour, life cycle, and structure, which, in turn, helps us to build better models (Manga et al., 2017). This feedback loop between
modelling and observation is essential, and strong collaboration between modellers, experimentalists, and field geologists are a must if we are to advance our knowledge and preparedness against the worst volcanic impacts.

Volcanoes can be monitored from both the ground and from space. Ground-based observations are the most direct and monitoring falls into a few different broad categories: seismic, geodetic (deformation), gas, thermal, hydrologic, potential field, and tomography techniques. Other methods include using drones or lightning detection arrays to collect data that might otherwise be difficult for humans to collect directly. In Cosburn et al. (2019) (Chapter 3) and Chapter 4 we focus primarily on the use of traditional gravity field measurements paired with a novel technique of cosmic-ray muon radiography/tomography as a method for monitoring and better understanding subsurface processes. In these cases, we are interesting in imaging shallow subsurface density structures, which help us understand the distribution of magma or hydrothermal areas of interest underneath a volcano. Imaging magma pathways, for example, can help to to better pinpoint where certain hazard events might occur, such as eruptions, flank collapse, or earthquakes (Figure 1.2). More widely used, however, are seismic and magnetotelluric measurements for subsurface imaging and future work to the study done in Chapter 4 would use them with gravity measurements in a machine learning approach. In general, the combinations of disparate datasets is essential to better forecasting and hazard monitoring, as one technique alone rarely provides all of the necessary information (Manga et al., 2017).

In contrast to ground-based observations, in space-based observations, satellites collect data on quantities such as heat flux, gas and ash emissions, and deformation over the course of years or decades and in locations all around the globe. A couple of benefits of satellite observations over ground-based observation are that satellite observations can be done when collecting ground data is too hazardous or if the volcano being monitored is too remote for observing consistently. In Cosburn & Roy (2020b) (Chapter 2), we use digital elevation models from data collected by the Shut-
tle Radar Topography Mission—a radar system that flew on board the Space Shuttle Endeavour for 11 days during February 2000—to develop a database of average topographic profiles (Appendix B) of a large number of stratovolcanoes from around the world. By quantifying an ideal stratovolcano shape, we can fit these profiles to that shape and try to understand deviations as an indirect method of hazard assessment. We focus on protuberant deviations as being due to magma emplacement within a statistical ensemble of subsurface dykes—which can provide information on a volcano’s plumbing structure and help us better understand underlying processes, especially for volcanoes which are difficult to monitor. Destructive processes, such as erosion, landslides, or flank collapse, however, can also be better understood by studying volcano shape (e.g., Karátson et al., 2010, 2012; Favalli et al., 2014).

1.3 Physics of Volcanic Processes

1.3.1 Lava Flow

Lava flow is a complex process that depends on the crystal and bubble content of the magma feeding a particular eruption (e.g., Mueller et al., 2010) and lava flows exhibit large variations in volume, shape and features along their surface (Hulme, 1974). Properties such as temperature, rate of effusion, gravitational field strength, and topography all influence the lava and dictate the final form that a flow takes. To first approximation, modelling lava can be treated as a Bingham fluid (Chapter 2; Cosburn & Roy (2020b); Hulme (1974); Annen et al. (2001)), like toothpaste, which does not flow until internal stresses exceed the yield stress (unlike a Newtonian fluid which flows indefinitely unless encountering a barrier). Mathematically, a Bingham fluid exhibits a linear relationship between shear stress, $\tau$, and shear strain rate, $\dot{\gamma}$, like a Newtonian fluid, only it is offset by the yield stress, $\tau_y$:

$$\tau = \tau_y + \mu_p \dot{\gamma}$$
where $\mu_p$ is the plastic viscosity (Figure 1.4.) This yield stress, along with the viscosity, all determine how long a lava flow might be, whilst the erupted volume determines the thickness and height.

During effusive eruptions, lava flows freely until it has cooled sufficiently such that its viscosity is sufficiently high that it ceases to be able to flow. As an edifice is formed, a first-order model of Bingham fluid flow must also incorporate the effects due to motion on an inclined plane. From Hulme (1974), the width and depth of a flow depend on the supply rate and rheology of the lava, as well as the slope of the plane. In the limit where the lava has ceased to flow, we want to find a set of equations that encapsulates these relationships and which allow us to “pile” flow heights to form an edifice. To start, we can think about the forces that act on a surface area bounding an element of volume. In order for a flow to stop, the yield stress of the fluid must be balanced by the weight of the element of volume (Figure 1.5). For lava flowing on an inclined plane of angle, $\alpha$, this occurs when

$$\tau_y = \rho gh \sin \alpha,$$

where $\rho$ is the density, $g$ is the acceleration due to gravity, and $h$ is
the height of the flow. This can be re-arranged for the height as follows:

\[ h = \frac{\tau_y}{g\rho \sin \alpha} \]  

(1.1)

On small inclines where \( \sin \alpha \to 0 \) and \( h \to \infty \) we postulate a maximum flow height given by:

\[ h(\alpha \leq k) = h_{\text{max}} = \frac{\tau_y}{\rho g k} \]  

(1.2)

where \( k = h/w \) is the ratio of flow thickness to width (where \( h < h_{\text{max}} \)). Thus, as \( \alpha \to 0 \), \( h \to h_{\text{max}} \), which grows as \( 1/k \) instead of \( 1/\sin \alpha \). The aspect ratio \( k \) goes to zero as the flow widens (i.e., \( w \to \infty \)), so we must choose \( k \) such that \( h_{\text{max}} \) depends on the volume and yield stress. In practice, the value \( k \) is empirically chosen, along with volume and yield stress (Appendix A; Annen et al. (2001)), which ensures that the behaviour on small slopes is governed primarily by lateral spreading (Eq. 1.2) instead of the equilibrium condition in Eq. 1.1.

### 1.3.2 Deformation from Magma Emplacement

Once a volcanic edifice is formed via the piling of eruptive products, it can then start to deform in many ways. There is destructive deformation, such as that due to erosion or flank collapse, but there is also constructive deformation, which is due primarily to the emplacement of magma underneath the volcano. This deformation could be due to magma pressurizing an underlying magma chamber, but it could also be due to the injection of magma into the host rock via a dyke/inclined sheet or sill. The Mogi (1958) model provides a set of analytic solutions to address how uplift at the surface may be due to the inflation of a magma chamber and is used in many inversion studies of geodetic data (such as GPS or InSAR measurements) to back out magma chamber parameters from real-time data. As an alternative to focusing on the effects of magma chamber inflation, in Cosburn & Roy (2020b) (Chapter 2), we use the Okada (1985, 1992) analytic solutions for a tensile dislocation to model the
long-term, constructive surface displacement due to a statistical ensemble of dyke intrusions. Both the Mogi (1958) and Okada (1985, 1992) analytic solutions make some important key assumptions about the mechanical properties of the domain. Both assume that the host rock into which magma is being injected is perfectly linearly elastic, homogeneous, and isotropic. Here homogeneous and isotropic means that the material properties are the same at each point and identical in all directions (Turcotte & Schubert, 2014).

That the rock is linearly elastic (Figure 1.6) means that the relationship between stress and strain in the rock is given by the Young's modulus, $E$, of the rock, which is an empirical measure of the stiffness of the rock, and the Poisson ratio, $\nu$, which is a measurement of the expansion/contraction of the rock perpendicular to the direction of loading (other properties such as shear modulus and bulk modulus could be used equivalently). Mathematically, we start with the constitutive equation of Hooke's
Figure 1.6: Cartoon of showing different stress-strain behaviour for loading and unloading of a linearly elastic, nonlinearly elastic, and inelastic material.

Law:

\[ \sigma_{ij} = C_{ijkl} \varepsilon_{kl} \]

where \(\sigma_{ij} = \sigma_{ji}\) is the Cauchy stress tensor, \(\varepsilon_{kl} = \varepsilon_{lk}\) is the infinitesimal strain tensor, and \(C_{ijkl}\) is the fourth-order stiffness tensor. Because of the symmetries in the stress and strain tensors, the stiffness tensor also exhibits key symmetries, namely \(C_{ijkl} = C_{jikl} = C_{ijlk} = C_{klij}\). Isotropic, homogenous material allows for a further reduction and the constitutive Hooke’s Law equation has the form:

\[
\begin{bmatrix}
\varepsilon_{11} \\
\varepsilon_{22} \\
\varepsilon_{33} \\
2\varepsilon_{23} \\
2\varepsilon_{13} \\
2\varepsilon_{12}
\end{bmatrix} = \frac{1}{E} \begin{bmatrix}
1 & -\nu & -\nu & 0 & 0 & 0 \\
-\nu & 1 & -\nu & 0 & 0 & 0 \\
-\nu & -\nu & 1 & 0 & 0 & 0 \\
0 & 0 & 0 & 2 + 2\nu & 0 & 0 \\
0 & 0 & 0 & 0 & 2 + 2\nu & 0 \\
0 & 0 & 0 & 0 & 0 & 2 + 2\nu
\end{bmatrix} \begin{bmatrix}
\sigma_{11} \\
\sigma_{22} \\
\sigma_{33} \\
\sigma_{23} \\
\sigma_{13} \\
\sigma_{12}
\end{bmatrix}
\]

For a finite rectangular source in an elastic half-space, Okada (1985, 1992) derives a closed-form solution for the displacement at the free surface (i.e., where the rock meets the fluid atmosphere) and in the case of Cosburn & Roy (2020b) (Chapter 2) this source is a tensile opening. This is achieved by taking the Green’s function
solutions to the elastic half space problem for a point source and integrating over the rectangle’s length and width dimensions. From Okada (1992), all the displacement formulae are comprised of an infinite medium term (A), a surface deformation term (B), and a depth multiplied term (C). In general form:

\[ u(x, y, 0) = u_B(x, y, 0) \]
\[ \frac{\partial u}{\partial x}(x, y, 0) = \frac{\partial u_B}{\partial x}(x, y, 0) \]
\[ \frac{\partial u}{\partial y}(x, y, 0) = \frac{\partial u_B}{\partial y}(x, y, 0) \]
\[ \frac{\partial u}{\partial z}(x, y, 0) = 2 \frac{\partial u_A}{\partial z}(x, y, 0) + \frac{\partial u_B}{\partial z}(x, y, 0) + u_C(x, y, 0) \]

and are subject to various mathematical singularity conditions that must be accounted for when developing a coded solutions to these equations. Based on the tables of solutions from Okada (1992), these can be found in Appendix E.1. Some key limitations of most analytic closed-form solutions is that real rock is not strictly elastic and is also far from being homogenous and isotropic. Further to this, at least when it come to dyke propagation, the Okada (1985, 1992) solutions do not take into account the effects of topography stresses on dyke propagation and that dykes tend to bend and change their trajectories or become arrested due to these stresses or as they encounter inevitable anisotropies in the host rock (Rivalta et al., 2015).

In Cosburn & Roy (2020b) (Chapter 2) we address this by not allowing our stochastic ensemble of dykes to reach the surface or to propagate through the volcanic edifice. Also, due to the stochastic nature of our algorithm and the fact that their length dimensions are modelled using a power-law distribution, many dykes are relatively short, mimicking the fact that there exists a dyke transition zone, above which only a few percent of dykes can propagate through the rock (Gudmundsson, 2020). Furthermore, due to magma chambers acting to concentrate or raise stresses in the central location of a volcanic edifice (Gudmundsson, 2006, 2009), we are able to model our volcanic system in Cosburn & Roy (2020b) (Chapter 2) as a central
volcano directly underlain by a shallow magma chamber from which dykes/inclined sheets propagate and produce uplift at the surface. This is sufficient for modelling first-order effects on volcano shape, however if one were interested in modelling individual dyke paths in order to better forecast eruptions (e.g. for hazard monitoring), more sophisticated numerical techniques which take into account anisotropies and topographic effects must be used.

1.4 Gravimetry and Muography

1.4.1 Gravimetry

Gravimetry is the measure of the strength of a gravitational field and is one of the oldest and most established techniques in the geosciences for imaging subsurface density. This measurement can be done on the ground using an instrument called a gravimeter, or from space via satellite. The data acquisition for Chapters 3 and 4 used a ground-based relative gravimeter. Within such an instrument is an especially sensitive spring balance carrying a fixed mass. When the gravitational field changes with location of the gravimeter, the weight of the mass changes and so, too, does the length of the spring. The extension of the spring is recorded inside the machine with high precision, resulting in very minute measurements of the gravitational field (on the order of \(10^{-8}\) m/s\(^2\)), relative to a base location (station) within a survey’s domain. The relative gravimeter must be calibrated at regular intervals at a location where the absolute value of gravity is known (which can be found on an international database). After a gravity survey is complete, various corrections to the gravity data (due to solid Earth tides, gravimeter drift, regional trends, and latitude effects) must be implemented. Often, what is used are called “free-air corrected data,” which adjusts gravity measurements to what would have been measured on the geoid (at mean sea level) and which are recast as anomalies relative to a chosen base station.
These kinds of free-air corrected measurements are used Chapters 3 and 4 for the inversions therein.

One can use Newton’s Law of Gravitation to relate these measurements to subsurface density. In cartesian coordinates, the force between a particle of mass \(m_0\) at point \(P(x, y, z)\) and a particle of mass \(m\) at point \(Q(x', y', z')\) is given by:

\[
F = G \frac{mm_0}{r^2}
\]

where \(G\) is Newton’s gravitational constant and \(r = [(x-x')^2 + (y-y')^2 + (z-z')^2]^{\frac{1}{2}}\). If \(m_0\) is a test particle with unit mass, then dividing this by \(m_0\) yields the gravitational attraction produced by \(m\) at point \(P(x, y, z)\):

\[
g(P) = -G \frac{m(z - z')}{r^3}
\]

If we now consider that our mass is infinitesimally small, \(dm = \rho(x, y, z)dV\) where \(\rho(x, y, z)\) is the density, then we can compute the integral:

\[
g_z(P) = -G \int_V \frac{\rho(x, y, z)(z - z')}{r^3} dV = -G \int_V \frac{\rho(x, y, z)(z - z')dxdydz}{[(x-x')^2 + (y-y')^2 + (z-z')^2]^{\frac{3}{2}}}
\]

For a domain discretized into \(j\) elements, we can cast this as the following forward calculation, where the vertical component of the gravity anomaly at the \(i^{th}\) gravity station can be written:

\[
g_i = \sum_j G_{ij} \rho_j
\]

where the gravity contribution of the \(j^{th}\) element to the \(i^{th}\) gravity station is given by:

\[
G_{ij} = -G \int_{x_j} \int_{y_j} \int_{z_j} \frac{(z - z_i)}{[(x-x_i)^2 + (y-y_i)^2 + (z-z_i)^2]^{\frac{3}{2}}} dxdydz
\]
This forward calculation (along with a similar forward calculation for muon measurements, see next section) is what is used in the Bayesian joint inversion of Chapter 3 and the machine learning joint inversion of Chapter 4.

1.4.2 Muography

The term “muography” (or muon radiography/tomography) refers to an imaging technique that uses cosmic-ray muons to produce a projected image of the density within a target volume. A thorough review paper by Bonechi et al. (2020) explains this technique with a breadth and depth of detail that won’t be explored in this chapter, however, since this methodology used in both Chapters 3 and 4, it is worth a brief overview here. The earliest published field work to use muons as a tool for imaging was by Alvarez et al. (1970) who searched for hidden chambers in the pyramids of Giza. Since then, the methodology has been used within the field of volcanology to image the inside of volcanoes (including the famous Mount Vesuvius in Italy), among many other applications.

Muons are subatomic particles that are produced in the upper atmosphere due to the decay of pions and kaons. They are leptons and experience the weak and electromagnetic forces, as well as gravity. They are also unstable and decay into electrons and neutrinos relatively quickly (with a lifetime of \( \tau = 2.2 \text{s at rest} \)), however, relativistic effects dilate their lifetimes by the Lorentz factor \( \gamma = \sqrt{1 - (p/(m\mu c))^2} \). This effect, and because atmospheric muons are mostly distributed at large momenta, means that many are able to reach sea level (and even penetrate deep underground or through large topographic structures) before decaying. For example, the peak value of the muon momentum distribution is about 4 GeV (Tanabashi et al., 2018), which corresponds to a decay length of 24 km in air. Given that the atmosphere of the Earth is contained within 16 km means that they constitute the vast majority of the charged particles arriving at sea level, arriving at a rate of around 100 Hz/m²
(Tanabashi et al., 2018). As they pass through matter, muons with momenta up to 500 GeV mainly lose energy due to ionization and atomic excitation. The mean energy loss rate can be expressed by the Bethe function:

$$-\frac{dE}{dx} \propto \rho \frac{Z}{A} \beta^2 \ln \left( \frac{K \beta^2}{1 - \beta^2} \right) \text{ + corrections}$$

where $\rho$ is the density of the mass, $Z$ and $A$ are the atomic and mass number, respectively, $K$ is related to the mean excitation potential, and $\beta = \frac{v}{c}$ is the relativistic speed. For the majority of elements that comprise the mass of the Earth’s crust, the ratio $Z/A \approx 0.5$. It is this fact, and that $K$ appears logarithmically in the above equation, that we can consider the energy loss rate and $\rho$ to be quasi-independent of other material properties (Bonechi et al., 2020).

The muon range is then given by integrating the Bethe equation from the starting energy ($\beta \approx 1$) to the stopping energy (where $\beta = 0$). The distance a muon can travel before stopping is given by:

$$x = \int_{E_0}^{0} \left( \frac{dE}{dx} \right)^{-1} dE$$

In practice, we measure the counts of muons that hit a detector positioned either in the subsurface (as in Chapter 3) or directionally pointed towards an area of interest (such as the peak of Mount Showa-Shinzan in Chapter 4). Detectors come in many forms, but a popular type for imaging volcanoes in the field is an emulsion detector, since it doesn’t require power, and it is what is used in Nishiyama et al. (2017b) and subsequently the data of Chapter 4. We then can take the ratio of “on-sky” versus “off-sky” measurements and relate these to the minimum energy needed to traverse matter and reach the detector, where, ideally, the energy spectrum is adequately modelled, taking into account variations that might occur as the muons traverse the matter. This can be converted to range (or density-length or opacity) via a look-up table, such as Groom et al. (2001). The range can be expressed as $R = \int \rho(x)dx$, measured in kg/m$^2$. If we consider a discretized domain, as we did with gravity, then
we can cast this as the following forward calculation:

$$R_i = \sum_j L_{ij} \rho_j$$

where $R_i$ is the range for the $i^{th}$ binned muon trajectory, $L_{ij}$ is the path length of the $i^{th}$ trajectory through the $j^{th}$ element, and $\rho_j$ is the density of the $j^{th}$ element. Given a precise knowledge of the topography, this allows us to turn the range into a measurement of the average density along a line-of-sight trajectory. This forward calculation is used in both Chapters 3 and 4 for determining subsurface density.

## 1.5 Inverse Theory

Inverse theory deals with the quantitative rules used to compare theoretical predictions with observations (e.g., Tarantola, 2005). More specifically, given that a theoretical model exists, we want to infer the best model that predicts the observations subject to measurement errors, given that the model itself also contains errors.
We call the problem of predicting the result of a measurement given a theory as the *forward problem*, whilst the problem of inferring the parameters that characterize a system as the *inverse problem*. The forward problem, \( d = g(m) \), produces exact data measurements, \( d \), given a certain model to produce them, \( m \), however observations are subject to experimental errors and theoretical models are imperfect representations of physical processes. This means that determining the inverse problem is tricky and the treatment of uncertainties must be done carefully. One way to do this is to replace the error-free forward problem with a probabilistic correlation between parameters for the model \( m \) and observations \( d \), \( P(d|m) \) (Figure 1.7). Further to this problem of uncertainties, however, is the problem of non-uniqueness inherent in inverse problems. In this case, sometimes even an infinite number of solutions can be possible, such as when trying to determine the mass distribution inside the Earth given a gravity measurement at the surface. For this reason, one must typically utilize *a priori* information, or additional constraints, such as by combining disparate datasets (Chapters 3 and 4).

In order exploit *a priori* information to solve an inverse problem, we now turn to Bayes’ Theorem, which states:

\[
P(m|d) = \frac{P(m)P(d|m)}{P(d)}
\]

where \( P(m|d) \) is the posterior distribution of the model given the data, \( P(d|m) \) is the data likelihood given the model, and \( P(m) \) is our prior information (since it is independent of \( m \), \( P(d) \) is simply there for scaling/normalization). Now, the most likely model, \( m \), is the model that satisfies the maximum likelihood criterion, and if we assume that the errors on the data and the model parameters are both Gaussian, this then reduces to solving a least-squares problem:

\[
P(m) \propto \exp \left( - \frac{1}{2} (m - m_{\text{prior}})^T C_M^{-1} (m - m_{\text{prior}}) \right)
\]

\[
P(d|m) \propto \exp \left( - \frac{1}{2} (g(m) - d_{\text{obs}})^T C_D^{-1} (g(m) - d_{\text{obs}}) \right)
\]
then \( P(m|d) \propto \exp(-S(m)) \) where twice the misfit function is given by:

\[
2S(m) = (m - m_{\text{prior}})^T C_M^{-1}(m - m_{\text{prior}}) + (g(m) - d_{\text{obs}})^T C_D^{-1}(g(m) - d_{\text{obs}})
\]

\[
= \|g(m) - d_{\text{obs}}\|^2_D + \|m - m_{\text{prior}}\|^2_M
\]

If we further assume that \( g(m) = Gm \) is linear, then

\[
2S(m) = \|((Gm) - d_{\text{obs}}\|^2_D + \|m - m_{\text{prior}}\|^2_M
\]

which is quadratic in \( m \), thus the posterior probability density \( P(m|d) \) is also a Gaussian probability density, with some central point \( \tilde{m} \) and covariance matrix \( \tilde{C}_M \) such that:

\[
P(m|d) \propto \exp\left(-\frac{1}{2}(m - \tilde{m})^T \tilde{C}_M^{-1}(m - \tilde{m})\right)
\]

The solution to this for \( \tilde{m} \) is also the point that minimizes the least-squares function, which means that it is the point closest to \( m_{\text{prior}} \) and the predicted data \( G\tilde{m} \) are closest to the observed data \( d_{\text{obs}} \). This solution is given by Tarantola (2005) as:

\[
\tilde{m} = (G^T C_D^{-1} G + C_M^{-1})^{-1} (G^T C_D^{-1} d_{\text{obs}} + C_M^{-1} m_{\text{prior}})
\]

\[
= m_{\text{prior}} + (G^T C_D^{-1} G + C_M^{-1})^{-1} G^T C_D^{-1} (d_{\text{obs}} - Gm_{\text{prior}})
\]

which is solved for iteratively using a quasi-Newton method when performing the joint inversion in Cosburn et al. (2019) (Chapter 3, Appendix E.2).

## 1.6 Machine Learning

### 1.6.1 Theory

Similar to the Bayesian inversion of the previous section, a supervised machine learning (ML) algorithm aims to minimize a least squares problem, however the parameter being predicted is slightly different. In an inversion problem, given the objective
function to be minimized:

\[ C = \| y - F \hat{x} \|^2 \]

we aim to predict some model estimate \( \hat{x} \) given known observations \( y \) and a known forward operator \( F \). In a supervised ML problem, given an objective function to be minimized:

\[ C = \| x_n - F^\dagger_{\Theta} y_n \|^2 \]

we aim to predict the parameter set \( \Theta \) of some pseudo-inverse operator \( F^\dagger_{\Theta} \) that maps a set of input data \( x_n \) to output data \( y_n \). Here \( \Theta \) depends on the choice of ML algorithm and there are many supervised ML methods that one can choose to solve the problem of determining these parameters. In artificial neural networks (ANNs), this parameter set comes in the form of finding the optimal weights and biases of a network. In other algorithms they are whatever parameters of the model that aren’t tuneable (i.e., that aren’t user-defined hyperparameters) and that the machine must learn on its own. They are unique to the choice of hyperparameters and the training data space \( \{ x_n, y_n \} \).

Popular ML methods in the geosciences include ANNs, naïve Bayes, support vector machines, and random forests. For the content in Chapter 4, we first explored the use of ANNs (including deep neural networks and convolutional neural networks), but eventually settled on a relatively novel approach by implementing a powerful gradient boosting algorithm with decision trees to perform an inversion on gravity and cosmic-ray muon data. A major drawback to ANNs is the computational “artistry” needed to construct one that works well. The key benefit of using a gradient boosting algorithm, instead, is its comparative ease of use. We utilize a well-established function from a popular ML python library, SciKit-Learn, thus reducing the amount of programmatic knowledge needed to create and use an algorithm that makes accurate predictions.
Chapter 1. Introduction

The algorithm for gradient boosting is given below in its theoretical form, but the following subsection aims to illustrate the algorithm with a concrete example from volcano science. To begin, gradient boosting works on both discrete and continuous output (termed a classification or regression problem, respectively), depending on the choice of differentiable loss function. For our inversion, since the output is continuous (regression problem), we choose a half least-squares loss function:

\[ L(y, F(x)) = \frac{1}{2} \sum_{i} (y_i - F(x_i))^2 \]

where \( n \) is the size of the training dataset, \( y_i \) is the observed value, and \( F(x_i) \) is the predicted value. To solve the problem of mapping inputs to outputs (i.e., to find the function \( \hat{F} := F_{\Theta}^{\dagger} \)), the gradient boosting algorithm iterates to combine what are called “weak” learners into a single “strong” ensemble model by minimizing the above loss function at each iteration (where an iteration is defined as the addition of a weak learner). Formally, we are looking to find:

\[ \hat{F}(x) = \arg \min_{F} \mathbb{E}_{x,y} \left[ L(y, F(x)) \right] = \sum_{m=1}^{M} \gamma_m h_m(x) + \text{const}. \]

where \( M \) the maximum number of iterations (a user-specified hyperparameter), \( h_m \in \mathcal{H} \) are the weak learners at each iteration \( m \), and \( \gamma_m \) are some weighting factors. This aggregation of smaller models to make one larger, ensemble model is the “boosting” part of gradient boosting. In this problem, we utilize a decision tree for regression (or “regression tree”) as our weak learners, \( h_m \), so the value of \( M \) is the maximum number of trees allowed for the final ensemble model. A regression tree works by building up and splitting along leaf nodes that discern various features in the data (e.g., \( x_1 < 1.3 \) or \( x_2 > 10.0 \); see Figure 1.9, Figure 4.4 for examples), which it determines using the sum of squares of the residuals between predicted and actual values. Given the user-specified hyperparameters such as maximum number of leaf nodes, minimum number of samples per leaf, and maximum depth, the decision tree algorithm decides how to build a tree that best represents the data.
Chapter 1. Introduction

Given a set of data, which are the input-output pairs \( \{(x_i, y_i)\}_{i=1}^n \), and the loss function above, we first initialize our ensemble model with a constant function:

\[
F_0(x) = \arg\min_{\gamma} \sum_{i=1}^n L(y_i, \gamma)
\]

where \( \gamma \) here ends up simply being the average of the observed values, \( y \), due to the nature of our chosen loss function (i.e., since taking the derivative of \( L \) with respect to \( \gamma \), setting this equal to zero, then solving for \( \gamma \) yields the average). This is equivalent to a predicted value of \( y = \bar{y} \), no matter what the input value is of \( x_i \).

If we are considering that our model learners are regression trees, this corresponds to just a single-leaf tree with a value equal to \( \bar{y} \). One then iterates from \( m = 1 \) to \( M \), carrying out the following steps to fit a regression tree at each iteration:

1. Calculate the pseudo-residual based on the derivative of the loss function with respect to \( F(x_i) \), evaluated for \( F(x) = F_{m-1}(x) \), for each \( i = 1, \ldots, n \) datum in the dataset. In other words, we want to compute:

\[
r_{i,m} = -\left[ \frac{\partial L(y_i, F(x_i))}{\partial F(x_i)} \right]_{F(x)=F_{m-1}(x)} \quad \text{for } i = 1, \ldots, n
\]

This is the “gradient” part of the gradient boosting algorithm.

2. Fit a regression tree (the weak learner), \( h_m \), to the \( r_{i,m} \) values. In other words, we train \( h_m \) on the dataset \( \{(x_i, r_{i,m})\}_{i=1}^n \). Here the weak learner can have variable number of leaf nodes (the maximum number is set as a user-specified hyperparameter), different from the number of leaf nodes for all other weak learners in the ensemble model. We also need to create terminal leaf regions \( R_{j,m} \) for \( j = 1, \ldots, J_m \) where \( J_m \) is the last leaf in the tree \( h_m \). This is necessary since, in practice, multiple outputs may occupy one terminal leaf node and we need a way of combining these into one value.

3. For \( j = 1, \ldots, J_m \), find:

\[
\gamma_{j,m} = \arg\min_{\gamma} \sum_{x_i \in R_{j,m}} L(y_i, F_{m-1}(x_i) + \gamma h_m(x_i))
\]
Figure 1.8: (a) Toy model data taken from Grosse et al. (2014), with volcano height as $x_1$ and basal radius as $x_2$. Colour scale represents the volcano volume. (b) Plot of absolute value of residuals, $r_{i,1}$, $r_{i,2}$, and $r_{i,3}$ from Table 1.1. One can see how these get smaller after each additional regression tree is added to the ensemble model.

where, again, we end up with a $\gamma$ that is simply an average over $x_i \in R_{j,m}$, due to the nature of our chosen loss function.

4. Update the model according to:

$$F_m(x) = F_{m-1}(x) + \eta \sum_{j=1}^{J_m} \gamma_{j,m} h_m(x \in R_{j,m})$$

where $\eta$ is a user-specified learning rate hyperparameter. Here the $\gamma_{j,m}$ are part of the parameter set $\Theta$ that the ML algorithm must learn based on the input-output pairs (the data) and choice of hyperparameters.

Finally, one outputs $F_M(x)$ as the strong, ensemble model for the data. One can then input data into $F_M(x)$ that the algorithm has never seen before and obtain a prediction. This algorithm is at the heart of the ML algorithm of Chapter 4, Appendix E.3.
Table 1.1: Results for the simple example of the gradient boosting algorithm for predicting volcano volume given height and basal width. Here the results for $M = 2$ are presented, along with their residuals. Plots of the residuals after each iteration are found in Figure 1.8b.

### 1.6.2 Relevant Example

To illustrate how the gradient boosting algorithm works, an example relevant to volcano science is presented, where we are looking to predict the volume of a volcano, given its height and basal width. For this, ten volcanoes from the Grosse et al. (2014) database were chosen and, from there, all of the relevant parameters were extracted (Figure 1.8a; Table 1.1). For this example, the same half least-squares loss function from above will be used:

$$L(y, F(x_1, x_2)) = \frac{1}{2} \sum_{i=1}^{n} (y_i - F(x_{1,i}, x_{2,i}))^2$$

For simplicity of the demonstration, only $M = 2$ iterations (trees) and $n = 10$ data points will be used, and only $J_{ns} = 4$ terminal leaf nodes will be allowed (this means only three feature splits will be allowed, see Figure 1.9). The aim of this exercise is to show the workings of the algorithm, not to produce a model that makes accurate predictions. The learning rate is set to be $\eta = 0.1$ and will not be optimized, whereas in practice it would be.

The first step in the algorithm is to determine $F_0(x_1, x_2)$. This ends up being simply the average over all $y$ values (the volcano volume), which gives $F_0(x_1, x_2) = \bar{y}$. The results for each iteration are shown in Table 1.1.
Chapter 1. Introduction

Figure 1.9: See Table 1.1 for list of residual values. (a) A regression tree with \( J_1 = 4 \) terminal leaf nodes \((R_{1,1}, R_{2,1}, R_{3,1}, R_{4,1})\) is fit to the \((x_1, x_2)\) data set and the residuals \( r_{i,1} \). (b) A regression tree with \( J_2 = 4 \) terminal leaf nodes \((R_{1,2}, R_{2,2}, R_{3,2}, R_{4,2})\) is fit to the \((x_1, x_2)\) data set and the residuals \( r_{i,2} \). Note that while the constraints are the same for both trees, they are not identical, and need not be.

\[
\begin{align*}
17.67 \text{ km}^3. & \quad \text{From there, we start the loop from} \quad m = 1 \quad \text{to} \quad M = 2 \quad \text{with steps as follows.} \\
& \quad \text{For} \quad m = 1:
\end{align*}
\]

1. Calculate the pseudo-residuals:
\[
r_{i,1} = - \left[ \frac{\partial L(y_i, F(x_{1,i}, x_{2,i}))}{\partial F(x_{1,i}, x_{2,i})} \right]_{F(x_1, x_2) = F_0(x_1, x_2)} \quad \text{for} \quad i = 1, \ldots, 10 \\
= \{y_i - F_0(x_{1,i}, x_{2,i})\}_{i=1}^{10} = \{y_i - 17.67\}_{i=1}^{10}
\]

The results of this for all ten data points are listed in Table 1.1.

2. A regression tree with \( J_1 = 4 \) terminal leaf nodes \((R_{1,1}, R_{2,1}, R_{3,1}, R_{4,1})\) is fit to the \((x_1, x_2)\) data set and the residuals, \( r_{i,1} \), from the previous step (Figure 1.9a).
3. We are now looking to construct

\[ F_1(x_1, x_2) = F_0(x_1, x_2) + \eta \sum_{j=1}^{4} \gamma_{j,1} \mathbb{1}_{R_{j,1}}(x_1, x_2) \]

where, when using regression trees, \( h_m \) is just an identity operation. To find the \( \gamma_{j,1} \) we compute:

\[ \gamma_{j,1} = \arg \min_{\gamma} \sum_{x_{1,i}, x_{2,i} \in R_{j,1}} L(y_i, F_0(x_{1,i}, x_{2,i}) + \gamma) \]

which can be solved by taking the derivative of \( L \) with respect to \( \gamma \), setting this equal to zero, and solving for \( \gamma \). Each \( \gamma_{1,1}, \gamma_{2,1}, \gamma_{3,1}, \gamma_{4,1} \) then ends up being the average over the difference between \( y_i \) and \( F_0(x_{1,i}, x_{2,i}) \) (i.e., the values of \( r_{i,1} \)) within each terminal leaf node (Figure 1.9a).

4. Now we calculate \( F_1(x_1, x_2) \). For example, for \( i = 1 \), since this data point ends up in terminal leaf node \( R_{2,1} \) this is given by \( F_1(x_{1,1}, x_{2,1}) = F_0(x_{1,1}, x_{2,1}) + \eta \gamma_{2,1} = 17.67 + 0.1 \times (-7.77) = 16.89 \text{ km}^3 \). Note that the actual volume is 16.2 km\(^3\), so this is an improvement over the initial guess of 17.67 km\(^3\). The other nine values of \( F_1(x_1, x_2) \) are similarly calculated and shown along with this value in Table 1.1.

Now we move on to the next iteration. For \( m = 2 \):

1. Calculate the pseudo-residuals:

\[ r_{i,2} = - \left[ \frac{\partial L(y_i, F(x_{1,i}, x_{2,i}))}{\partial F(x_{1,i}, x_{2,i})} \right]_{F(x_{1,i}, x_{2,i})=F_1(x_1, x_2)} \text{ for } i = 1, \ldots, 10 \]

\[ = \{y_i - F_1(x_{1,i}, x_{2,i})\}_{i=1}^{10} \]

The results of this for all ten data points are listed in Table 1.1.

2. A regression tree with \( J_2 = 4 \) terminal leaf nodes \( (R_{1,2}, R_{2,2}, R_{3,2}, R_{4,2}) \) is fit to the \( (x_1, x_2) \) data set and the residuals, \( r_{i,2} \), from the previous step (Figure 1.9b). Note that, although the tree has the same constraints as the one used for \( m = 1 \), it is not identical and does not need to be.
3. Compute the coefficients $\gamma_{j,2}$:

$$\gamma_{j,2} = \arg\min_{\gamma} \sum_{x_{1,i},x_{2,i} \in R_{j,2}} L(y_i, F_1(x_{1,i}, x_{2,i}) + \gamma)$$

Just as before, this ends up being the average over the residuals, $r_{i,2}$, that end up in each terminal leaf node (Figure 1.9b).

4. Determine:

$$F_2(x_1, x_2) = F_1(x_1, x_2) + \eta \sum_{j=1}^{4} \gamma_{j,2} \mathbb{1}_{R_{j,2}}(x_1, x_2)$$

Again, using $i = 1$ as an example, we have that its terminal leaf node is $R_{2,1}$, which corresponds to $\gamma_{2,1} = -8.596$ so $F(1.15, 7.5) = 16.89 + 0.1 \times (-8.596) = 16.03$ km$^3$. Even though we are now “undershooting” the correct value as opposed to “overshooting” it, as before, we still see an improvement in the overall absolute difference between our model prediction and that data. Values of $F_2(x_1, x_2)$ are similarly calculated and all predictions are listed in Table 1.1.

Since $M = 2$, we stop the algorithm here and choose our final, ensemble model to be $F_2(x_1, x_2)$. The residuals for this model are also calculated and put in Table 1.1. A comparison of the absolute value of the residuals for each $F_0(x_1, x_2)$, $F_1(x_1, x_2)$, and $F_2(x_1, x_2)$ models is shown in Figure 1.8b, where one can see that the addition of subsequent models helps the overall difference between prediction and data to decrease (albeit slowly). One can now plug in data that algorithm has never seen before and get a prediction. For example, for Mount Fuji, we have a height of 2.47 km and a basal radius of 11.8. These values correspond to terminal leaf nodes $R_{4,1}$ for $F_1$ and $R_{4,2}$ for $F_2$. The values of $\gamma_{4,1}$ and $\gamma_{4,2}$ are found in Figure 1.9. The prediction would then follow from calculating:

$$F_2(2.47, 11.8) = 17.67 + 0.1 \times (28.88) + 0.1 \times (42.44) = 24.8$$

which is not very accurate, given that the volume from the data is 60 km$^3$. We would expect, however, for this to get better if (a) there were more training data points
than just ten, (b) there were allowed to be more trees (iterations) than just two, and (c) each tree was allowed to have more than just four terminal leaf nodes.

1.7 Overview of Following Chapters

We have covered a few important aspects of volcano science, including a brief introduction to volcanoes and their processes. We have also outlined the importance of hazard assessment of volcanoes, including techniques developed and explored within the following chapters. The relevant physics of volcanic processes and mathematical theory to accompany these techniques have been provided to aid in understanding how we have formulated our approach to addressing these important volcano problems. This dissertation consists of the following chapters:

- Chapter 2: an analysis of stratovolcanoes and their prototypical, or “ideal” shape. We develop a mathematical formula based on the piling of lava flows that can be fit to the average topographic profile of a large number of stratovolcanoes from around the world. We aim to provide a large-scale study of what may cause deviations from such an ideal form, focusing our attention primarily on protuberant deviations that might occur due to a statistical ensemble of dykes which emplace magma beneath an edifice. Our hope is to provide a collection of data that can guide volcano research, particularly for volcanoes that are too remote or difficult to study.

- Chapter 3: the joint inversion of gravity and cosmic-ray muon measurements. We show the viability of combining these two datasets for imagining a regionally extensive, flat-lying structure beneath the town of Los Alamos, New Mexico (not possible if just using gravity alone). We use a novel technique of using both surface and subsurface measurements and cite several papers which use such a joint inversion to image volcanoes and lava domes and suggest that our
approach could be used in a similar setting if an appropriate field or borehole muon detector can be developed. Finally, we suggest that future work involve imaging time-dependent structures, which is further addressed by our group in Chapter 4.

• Chapter 4: the use of machine learning with gravity and cosmic-ray muon measurements for static, subsurface density imaging at a volcano (Mount Showa-Shinzan, Japan). We provide this study as proof-of-concept that using machine learning with two distinctly separate datasets is a viable alternative to a joint inversion and to set the stage for further work using this technique. We compare our results with a previously published joint inversion at the same volcano and successfully image the relatively high-density, cylindrical anomaly that we believe exists beneath the dome. Once trained, the machine learning algorithm takes on the order of seconds to perform a prediction on field data. We anticipate that this would facilitate the continuous monitoring of a volcanic edifice.

• Chapter 5: a brief summary of the previous three chapters along with suggestions for future work on these studies.
Chapter 2

Analysing the Topographic Form of Stratovolcanoes

The contents of this chapter were originally published as part of Cosburn, K. & Roy, M., 2020b. Analysing the topographic form of stratovolcanoes, Journal of Volcanology and Geothermal Research, 407, 107123. Minor edits within this dissertation have been made to clarify the text.

2.1 Abstract

In this study, we aim to explore “ideal” analytic functional forms that best fit the topographic shape of \( N = 190 \) isolated stratovolcanoes from around the world. Using a stochastic model for the piling of lava flows to demonstrate one set of physical processes that give rise to a stratovolcano’s topography, we find that although the ideal form fits well for many stratovolcanoes, there exist deviations from this shape—particularly where the edifice is more protuberant than expected. We explore subsurface factors—such as the emplacement of magma beneath a volcano edifice—that may be responsible for these deviations and estimate the relative contribution of
Chapter 2. Analysing the Topographic Form of Stratovolcanoes

dyke intrusions on the height and shape of a volcanic edifice versus the constructional piling of lava. From this comparison, we are able to gain more insight into the role of important subsurface physical factors, such as chamber depth and total intruded volume.

We find that low basal ellipticity is a characteristic of nearly all volcanoes that fit the ideal stratovolcano form. By fitting along each quadrant bisected by either a semi-major or semi-minor ellipse axis, we find that the volcanoes chosen in this study fall into four groups: (1) volcanoes that are axisymmetric and fit the ideal form well, (2) volcanoes that are not axisymmetric but may be fit to the ideal form using different fitting parameters in each quadrant, (3) volcanoes that have axisymmetric protuberant deviations in all quadrants, and (4) volcanoes that have non-axisymmetric protuberant deviations that are only present in a couple of quadrants.

We show that, in some cases, the non-axisymmetry of volcanoes in category (2) may be understood as the result of the piling of eruptive products along a regional planar trend. For edifices in categories (3) and (4), we model protuberant deviations as the result of a statistical distribution of elastic tensile dislocations representing dykes, which emanate from a magma reservoir and cause uplift at the surface. We support this interpretation of the “excess” topography in a few case studies where independent geologic and geophysical data corroborate our model results.

2.2 Introduction

Both small-scale and large-scale landforms on Earth—such as drumlins, kettles, sand dunes, and volcanoes—exhibit “ideal” regular shapes that can be studied by fitting parametric surfaces that best approximate their real topography (e.g., Davidson et al., 2000; Goudie, 2004; Pike et al., 2009). By pursuing the development of such mathematically-derived surfaces, one can look quantitatively at the extent to which a landform adheres to its ideal form and the causes of any deviations thereof. In
the case of stratovolcanoes, such deviations could be the result of both surface or subsurface processes that can be either constructive (e.g., magma emplacement beneath an edifice) or destructive (e.g., erosion of an edifice, flank collapse). Insight into these processes has important implications for volcano hazard assessment, since they could be indicative of whether or not a volcano might soon erupt or experience catastrophic sector collapse. The practise of analysing a volcano’s topographic form, then, provides a relative “first-order” approximate way of monitoring its current state in cases where real-time data and monitoring may be hard or impossible to obtain.

To start our study on the topographic form of stratovolcanoes, we note the strikingly similar shape exhibited by many of the world’s stratovolcanoes, best epitomized by such regularly-shaped volcanoes as Mayon in the Philippines, Kronotsky in Russia, or the iconic Mount Fuji in Japan. The name “stratocone” often given to these volcanoes suggests that they are best represented by a perfectly conical shape—such as seen with simple cinder cones—however stratovolcanoes actually tend to have more concave profiles. Our study builds upon the idea that their ideal geometric shape is actually best represented by logarithmic (e.g., Milne, 1878; Francis, 1993), exponential (e.g., Davidson et al., 2000; Borgia et al., 2000; Favalli et al., 2014), or polynomial (e.g., Karátson et al., 2012; Castruccio et al., 2017) functions subject to various fitting parameters. In our study, we consider the “splitting” approach of Karátson et al. (2010) and Castruccio et al. (2017) where the two-dimensional average profile of a stratovolcano can be separated into two parts—an inner linear section and an outer nonlinear one (Figure 2.1a). Our results suggest that there is nothing unique about the mathematical form of this nonlinear section, as it is fit equally well by both an exponential and a polynomial function given the same number of fitting parameters. We provide a simple stochastic explanation for where this shape comes from, following the model of Annen et al. (2001) for the spreading of lava flows.

From the morphometric database of Grosse et al. (2014) we select only isolated stratovolcanoes to ensure that nearby topographic features don’t factor in to our
Chapter 2. Analysing the Topographic Form of Stratovolcanoes

Figure 2.1: (a) Cartoon outlining the important parameters of the analytic model (further explained in text): maximum edifice height $h_m$, basal radius $R_b$, critical radius $r_c$, and maximum/critical slope $\varphi_c$. (b) For an axisymmetric form, the profile in (a) is spun around an axis of revolution to create a model edifice.

analysis. For each stratovolcano we extract average radial profiles and analyse the shape of the volcano by fitting to an ideal analytic form. Deviations from the ideal profile are analysed using a simple stochastic dyke model to offer one explanation for the occurrence of protuberant deviations from this shape. Our goal is to estimate the relative contribution these dyke intrusions have on the height and shape of a volcanic edifice compared to the contribution due to the piling up of eruptive products. We assume to first-order that the ideal stratovolcano profile is a result of how eruptive products pile up over time, with the shape they produce eventually reaching some steady-state form. Explanations for cases where deviations from the ideal form are more contracted—rather than protuberant, as in this study—are explored in such studies as Karátson et al. (2012), which looks at the effects of erosion as a source of deviation from the ideal stratovolcano form. Several case studies are presented (Merapi in Java, Karymsky in Kamchatka, and Concepción in Nicaragua) for which our modelling results are corroborated by geologic and geophysical observations.
Chapter 2. Analysing the Topographic Form of Stratovolcanoes

Figure 2.2: (top) Map showing $N = 190$ isolated stratovolcanoes currently being assessed in this study: $N = 102$ with a low ellipticity index ($\varepsilon \leq 1.5$) in yellow and $N = 88$ with an intermediate ($1.5 < \varepsilon \leq 2$) or high ($2 < \varepsilon \leq 3$) in red. (bottom) Histograms of Grosse et al. (2014) ellipticity index for those volcanoes.

2.3 Ideal Stratovolcano Shape

2.3.1 Extracting Topographic Profiles

To examine a representative population of stratovolcanoes from around the world we utilize the morphometric analysis compiled by Grosse et al. (2014). In their study, volcanoes from the Smithsonian Institution Global Volcanism Programme database covered by Shuttle Radar Topography Mission (SRTM) digital elevation
model (DEM) were chosen and their topographies and morphometric properties analysed according to Grosse et al. (2012) and Euillades et al. (2013). Of the 512 stratovolcanoes in the Grosse et al. (2014) database, we consider $N = 190$ isolated volcanoes (Figure 2.2). We extract information about the latitude and longitude of the volcano centre, its basal width, maximum height and slope, and its basal ellipticity and orientation (azimuth of the semi-major elliptical axis). For each chosen stratovolcano, given a central latitude and longitude from Grosse et al. (2014), we download 1 arc-second global SRTMs freely available via the United States Geological Survey EarthExplorer database (Figure 2.3a). The 1 arc-second SRTM data has a horizontal resolution of 30 m and an average vertical uncertainty of 6.5 m (calculated from information across 191 DEM tiles). In terms of deviations to an ideal stratovolcano shape, we do not consider those with a horizontal span less than the SRTM horizontal resolution nor a vertical span less than the vertical uncertainty, as any deviations less than these resolution bounds could be attributed to inaccuracy in the topographic models.

From the DEMs, we use a moving-window average to calculate the average profile along radial lines with length given by the basal radius (half the basal width in the Grosse et al. (2014) database) and centred at the given latitude and longitude in angular step-sizes of 0.5 degrees (Figure 2.3b). We also subtract any planar trend from the data by fitting a plane to the points representing the ends of our radial profile lines and catalogue information—such as planar dip angle and azimuth—in order to see if these factors alone can explain some of the results we obtain (Figure 2.3c).

2.3.2 Basal Ellipticity and Volcano Axisymmetry

Basal ellipticity is well-known and documented from a morphological perspective (e.g. Grosse et al., 2009, 2014; Favalli et al., 2014) with cylindrically symmetric vol-
canoes (i.e., those exhibiting more circular basal geometry) being treated as most representative of the ideal shape (Karátson et al., 2010, 2012). In terms of the effect of evolutionary trends, basal elongation has been explained as a result of a complex combination of processes, including direction of subsurface magma emplacement (Tibaldi, 1995), basement spreading due to a weak underlying substrate (Borgia et al., 2000), edifice strength and structure, as well as nearby tectonic activity (Grosse et al., 2009).

In order to explore the ways in which axisymmetric constructive processes may play a role in the manifestation of the ideal stratovolcanic shape, we categorize our chosen volcanoes by their basal ellipticity, \( \varepsilon \), as defined by Grosse et al. (2012). Volcanoes with “low” basal ellipticity (i.e., the more “circular” ones) are chosen as those with \( \varepsilon \leq 1.5 \) based on analysis done in Grosse et al. (2009) on polygenetic volcano edifices in the Central American Volcanic Front and the southern Central Andes Volcanic Zone (Table 2.1, Figure 2.2). Volcanoes with “intermediate” ellipticity are chosen to be ones where \( 1.5 < \varepsilon \leq 2 \), and those with “high” ellipticity \( 2 < \varepsilon \leq 3 \) (Table 2.2, Figure 2.2). We do not consider volcanoes where \( \varepsilon > 3 \) as they are most likely to be affected by complex combinations of geologic processes contributing to deviations from the ideal shape. We do not seek to reconstruct volcanoes to their ideal form in order to quantify surface processes, e.g. the degree of denudation (Karátson et al., 2012) or other long-term evolutionary effects (Favalli et al., 2014).

We obtain average radial profiles over each quadrant bisected by either a semi-major or semi-minor axis of the basal ellipse (Figure 2.3a) and prescribe the average to that axis (Figure 2.3b). We treat the fitting of each quadrant separately, but note that we would expect more circular and cylindrically symmetric volcanoes, such as Mayon, to be best fit by comparable parameters. Likewise, volcanoes with higher basal ellipticity could still adhere to the ideal functional form, but we would expect that the parameters of best fit to be notably different from one another in the various quadrants.
Figure 2.3: (a) DEM of Mayon ($\varepsilon = 1.11$) with purple and grey shaded regions corresponding to the quadrants bisected by either a semi-major ($a_1, a_2$) or semi-minor ($b_1,b_2$) axis over which an average profile is extracted. (b) The average profiles obtained from the quadrants in (a) prescribed along their corresponding semi-major (purple) or semi-minor (grey) axis. (c) A plane (yellow) is fit to the black dots representing the ends of profile lines drawn from the centre to the edge of the ellipse shown in (a). A horizontal reference plane (red) is depicted to more clearly show the planar trend (yellow) that is then subtracted from the elevation of these profile lines to obtain the final configuration used in our modelling (blue).

### 2.3.3 Profile Fitting to Analytic Form

We consider that a volcano’s shape can be made up of two functions—a linear one for regions less than some critical radius $r_c$, and a nonlinear one for regions beyond (Figure 2.1). We consider two different formulations for the nonlinear regions where $r \geq r_c$, the first being a polynomial of degree $\lambda$ and the second an exponential function with decay factor $\gamma$, both of which we treat as fitting parameters. Castruccio et al. (2017) used a model similar to our polynomial model derived below (Eq. 2.1)
with $\lambda = 1$ and yielded good results, however we find its applicability to be limited. In order to compare it equally with other models, such as one with an exponential formulation (e.g., Davidson et al., 2000; Borgia et al., 2000; Favalli et al., 2014), we treat $\lambda$ as a free parameter to be optimized.

In both models the maximum height and basal radius are defined as $h_m$ and $R_b$, respectively. In regions where $r \geq r_c$, for the polynomial model, we have that $h(r) = Ar^{-\lambda} - Br^\lambda$ subject to the boundary condition that $h(R_b) = 0$ and $h(r_c) = \beta h_m$ where $\beta h_m$ is some fraction of the edifice’s maximum height. With the first boundary condition we get that $B = \frac{A}{R_b^{2\lambda}}$ and using this with the second boundary condition, we find:

$$A = \frac{\beta h_m r_c^{\lambda} R_b^{2\lambda}}{R_b^{2\lambda} - r_c^{2\lambda}}$$

so altogether:

$$h(r \geq r_c) = \left( \frac{\beta h_m r_c^{\lambda}}{R_b^{2\lambda} - r_c^{2\lambda}} \right) \frac{R_b^{2\lambda} - r_c^{2\lambda}}{r^\lambda}$$

(2.1)

For regions where $r < r_c$ we have the linear relation:

$$h(r < r_c) = h_m - \left( \frac{h_m - \beta h_m}{r_c} \right) r = h_m \left( 1 - \frac{r}{r_c} (1 - \beta) \right)$$

(2.2)

where the requirement that $h(0) = h_m$ and $h(r_c) = \beta h_m$ are both satisfied. Now finding the derivatives of each equation at $r = r_c$ gives us an expression for the critical (or max) slope $\varphi_c$ in terms of its tangent:

$$\frac{d}{dr} h(r \geq r_c) \bigg|_{r=r_c} = -\beta h_m \left( \frac{r_c^{\lambda} R_b^{2\lambda}}{R_b^{2\lambda} - r_c^{2\lambda}} \right) \left( \frac{1}{r_c^{\lambda+1}} + \frac{\lambda r_c^{\lambda-1}}{R_b^{2\lambda}} \right) = \tan \varphi_c$$

$$\frac{d}{dr} h(r < r_c) \bigg|_{r=r_c} = -\frac{h_m}{r_c} (1 - \beta) = \tan \varphi_c$$

To ensure continuity at $r = r_c$ we match these derivatives at that point and get the transcendental equation:

$$\frac{\beta}{1 - \beta} \left( \frac{R_b^{2\lambda} + \lambda r_c^{2\lambda}}{R_b^{2\lambda} - r_c^{2\lambda}} \right) - 1 = 0$$
which, given an \( r_c \) at some \( h_c = \beta h_m \), can be solved for \( \lambda \) to find the analytic form that gives us a profile of best fit.

For the exponential model for the outer nonlinear section of an edifice, we consider a treatment of an exponential of the form \( h(r) = Ae^{\gamma r} + B \). Studies like Borgia et al. (2000) and Davidson et al. (2000) fit an exponential over the entire form of the profile, but for our purposes of fitting over some outer region, we require that \( h(R_b) = 0 \) and \( h(r_c) = \beta h_m \) both be satisfied. For this we obtain:

\[
h(r \geq r_c) = \frac{\beta h_m}{1 - \exp\left(\frac{r_c - R_b}{\gamma}\right)} \left( \exp\left(\frac{r_c - r}{\gamma}\right) - \exp\left(\frac{r_c - R_b}{\gamma}\right) \right)
\]

(2.3)

which, when we match slopes to the linear section as before, gives us a transcendental equation for the decay constant \( \gamma \) that we can solve at some \( r_c, h_c \) to find the analytic form corresponding to the best fit to a profile:

\[
\gamma \left(1 - \exp\left(\frac{r_c - R_b}{\gamma}\right)\right) - \frac{\beta r_c}{1 - \beta} = 0
\]

We find that, in general, either Eqs. 2.1 or 2.3, together with Eq. 2.2, provide similar fits to all volcanoes and we conclude that the functional form in the area \( r > r_c \) is not as important as the “splitting” between linear and nonlinear regions. With this splitting we can then find for each volcano—along each ellipse axis where we have prescribed an average profile—a critical radius and maximum slope. This allows us to quantitatively compare how well a volcano adheres to its ideal shape, regardless of basal ellipticity.

### 2.3.4 Physical Explanation for the Analytic Form

Finding a mathematical shape that fits the topography of a landform can provide clues to the processes by which such features arise in the natural world (e.g., Goudie, 2004; Pike et al., 2009; Karátson et al., 2010; Favalli et al., 2014; Karátson et al., 2012). The work done by Castruccio et al. (2017) attempts to do this for volcanoes,
arguing that a polynomial with $\lambda = 1$ for $r \geq r_c$ and a linear function for $r < r_c$ arises due to the piling of lava flows. We find, however, other equally-fitting mathematical formulations, such as our exponential model or polynomial model of variable degree $\lambda$.

Some studies have attempted to explain the shape of stratovolcanoes by treating the growth of an edifice as a porous flow problem, wherein an equilibrium surface drives percolative magma flow, which builds up an edifice as magma reaches the surface in a flank eruption (Lacey et al., 1981; Angevine et al., 1984; Bonafede & Boschi, 1992). In real life, however, a stratovolcano’s main source of eruptive material comes from a central vent, from which lava and pyroclastic flows emanate and pile up. This is addressed more satisfactorily in Karátson et al. (2010), who link information about the chemical composition of the magma of the 19 volcanoes from their study to suggest that morphometric patterns are a result of how silica or water-rich these magmas are. In particular, they focus on the formation of the upper slope portion of the volcano (where $r < r_c$ in our study), which they attribute to long-term, steady-state eruptive activity and suggest that linear behaviour in this region can best be explained by more water-rich magmas, which increase the relative frequency of explosive versus effusive eruptions. They find that a higher water content is more influential on this behaviour compared to how mafic versus silicate the magmas are, which explains the morphometric similarities seen in geographically scattered volcanoes subject to varying mantle compositions and crustal thicknesses.

Several volcanoes in Karátson et al. (2010) that are studied as having a linear near-vent and upper slope morphometric pattern (and subsequently more water-rich magmas) are also included in our study: Inierie, Kliuchevskoi, Koryaksy, Licancabur, Parinacota, Semeru, and Yotei. Other volcanoes from their study (Agua, Klabat, Kronotsky, Mayon, Pavlof, Sumbing, and Tidore) were best-fit by Karátson et al. (2010) with a parabolic functional form fitted to regions for $r < r_c$. We find, however, that when considering a first-order approximation of stratovolcano shape, a
linear form fits these volcanoes just as well. Since the basis of our study is to understand convex, protuberant deviations, however, we are not concerned with this minor discrepancy between fitting with a linear versus a (concave) parabolic function.

To explore a first-order, process-oriented explanation for how the ideal stratovolcano shape arises, we utilize a stochastic model based on the concept of the viscous piling of eruptive material. Annen et al. (2001) provides a basic approach to this problem, by treating a statistical distribution of lava flows modelled as a Bingham fluid undergoing channelized flow on an inclined plane, according to the equations of Hulme (1974). We find that this model does a decent job of explaining how our functional form arises in terms of a statistical ensemble of real physical processes (Appendix A). Although a Bingham rheology is an oversimplification and many works have since incorporated more complex lava flow behaviour such as cooling and crust formation, changes in rheology, influence of existing topographic features, etc. (e.g., Dragoni, 1989; Crisp & Baloga, 1990; Castruccio et al., 2010; Takagi & Huppert, 2010; Tallarico et al., 2011; Chevrel et al., 2018), it is sufficiently robust and computationally inexpensive for a first-order approximation, where we are not interested in lava-flow morphology as a function of time for individual eruptions.

In Annen et al. (2001), the stochastic properties of the lava flows are chosen to model shield volcanoes, thus to consider parameters more representative of stratovolcano flows, we base our bounds on flow volume, aspect ratio, and yield strength (see Appendix A) upon various studies, including those by Mc Birney & Murase (1984); Naranjo et al. (1992); Lyman et al. (2004); Castruccio et al. (2013, 2014); Arnold et al. (2019). We note that to find a set of parameters consistent with all stratovolcanoes in our study would be difficult, but for a first-order approximation of volcano shape, the parameters based on these studies produce physically plausible results.

We find that a linear-nonlinear splitting naturally arises from a stochastic model of lava flows treated as a Bingham fluid and is consistent with the polynomial and exponential forms in Eqs. 2.1 and 2.3.
Chapter 2. Analysing the Topographic Form of Stratovolcanoes

2.4 Ideal Stratovolcano Shape: Fitting Results

We minimize the sum-of-squares error between model and average topographic profile

\[
SSE = \sum_j \left( T_{P_j} - M_{A_j} \right)^2
\]  

where \( T_{P_j} \) is the topographic profile at the \( j^{th} \) data point and \( M_{A_j} \) is the profile at the same point for the analytic model using the observed \( h_m \) and \( R_b \) for a given volcano (Figure 2.1). We find for the \( N = 190 \) volcanoes in this study that the fitting results fall into four main categories (Table 2.1, 2.2):
Chapter 2. Analysing the Topographic Form of Stratovolcanoes

Figure 2.5: Four volcanoes representative of where there are protuberant deviations to the analytic model: (a) two in all four quadrants (Karymsky, Nantai) and (b) two in only two quadrants (Concepción, Momotombo).

1. Functional form fits well; fitting parameters comparable on both semi-major and semi-minor axes (Figure 2.4a): We find that $N = 38$ have topographic profiles that are fit well by either functional form ($\text{SSE average} = 0.2297 \text{ km}^2$), with $r_c$ and $\varphi_c$ values along each each semi-major and semi-minor axis comparable to within 1.5 km and 4 degrees of each other. The volcanoes in this category with a low ellipticity index comprise 65%, those with intermediate 35%, and there are none with a high index. The average ellipticity index for volcanoes in this category is 1.48.
2. Functional form fits well; fitting parameters different on the semi-major vs semi-minor axes (Figure 2.4b): We find that $N = 57$ have topographic profiles that are fit well by either functional form (SSE average = 0.2271 km$^2$) with difference in $r_c$ and $\varphi_c$ values along the semi-major and semi-minor axes greater than 1.5 km and 4 degrees. The volcanoes in this category with a low ellipticity index comprise 44%, those with intermediate 56%, with none having a high index. The average ellipticity index is 1.52.

3. Functional form does not fit well; protuberant deviations occur in all four quadrants (Figure 2.5a): We find that $N = 44$ have topographic profiles that are best described by this category, which suggests an underlying axisymmetric process causes these deviations. The volcanoes in this category with a low ellipticity index comprise 70%, those with intermediate 19%, and 11% with a high index. The average ellipticity index for volcanoes in this category is 1.69.

4. Functional form does not fit well; protuberant deviations occur in 1 or 2 quadrants (Figure 2.5b): We find that $N = 51$ have topographic profiles that are best described by this category, which suggests an underlying directional process causes these deviations. The volcanoes in this category with a low ellipticity index comprise 40%, those with intermediate 56%, and 4% with a high index. The average ellipticity index for volcanoes in this category is 1.6.

In the above categories, we first consider volcanoes with a “good fit” to the functional form (categories 1 and 2) as those with SSE values less than the mean SSE from across the entire dataset. We then visually inspect these volcanoes and discard those that do not qualitatively show a good fit to the functional form. (These volcanoes discarded from categories 1 and 2 are then explored below in terms of categories 3 and 4.) Visual inspection is also done on those volcanoes with SSE values greater than the overall mean SSE, however we do not find that there are any qualitative outliers to the two categories of good fit. We consider the deviations in categories 3
and 4 as arising from the uplift due to an ensemble of subsurface dyke intrusions, as described in the following section. We exclude from these categories volcanoes where there are known craters as documented by Grosse et al. (2014) (see published Data Supplement Cosburn & Roy (2020a); Appendix B).

2.5 Surface Effects from Subsurface Dyke Intrusions

The contribution of subsurface magma intrusion to the growth and evolution of a volcano has been substantiated by numerous field studies (e.g., Kervyn et al., 2008) and remote geodetic techniques such as Interferometric Synthetic Aperture Radar (e.g., Biggs et al., 2013; Biggs & Pritchard, 2017) or seismic reflection data (e.g., Magee et al., 2013). Furthermore, intrusive complexes of thousands of dykes beneath eroded volcano edifices show that it is not uncommon for a volcano to be underlain by such systems (e.g., Burchardt et al., 2013). Although the exact way in which these complexes affect the overlying topography is not well understood, it has been observed that, while the short-term stresses of a series of many such dyke events obey a linear elastic constitutive relation to strains, the long-term deviatoric stresses relax on relatively short time-scales (Hofton & Foulger, 1996). We use a model based on Annen et al. (2001) to try to quantify the relative contribution of internal versus external processes that build a stratovolcano edifice by explaining protuberant deviations from the ideal form in terms of surface displacement due to a statistical ensemble of subsurface dyke events. We note that such deviations could be indicative of other processes not considered here, such as magma chamber inflation.
2.5.1 Dyke Distribution Model

To model the displacement at the surface of a single dyke intrusion, we use the equations of Okada (1985, 1992) for a tensile dislocation in an elastic half-space. Each model dyke has three orientation parameters $\phi$, $\delta$, $(x_0, y_0, z_0)$ corresponding to the strike angle, dip angle, and coordinate of the bottom left-most corner, as well as three length parameters $L$, $W$, $U$ defining its geometry (Figure 2.6a). Each dyke emanates from the top of a disc-shaped magma reservoir of radius $R$ at some depth $z_0$ with its random placement $(x_0, y_0)$ determined by a uniform distribution in $\theta$ and $r^2$ (Figure 2.6b). Here the reservoir is centred on the centre of the volcano at $x = y = 0$. For the length parameters $L$, $W$, and $U$ we draw values from a power-law function proposed by Gutierrez & Youn (2015) from their study of the effects of fracture length scales and distributions on rock masses. The bounds on our distribution are given in Table 2.3 along with the two power law exponents we consider based on a compilation of power law exponents for fracture length distributions by Bonnet et al. (2001). We note that while Annen et al. (2001) considers a variety of different distributions chosen based on the goodness-of-fit to data from the volcanoes in their study, their overall conclusion is that the shape of the distribution does not have a significant quantitative effect on their results compared to the effects of changing

Figure 2.6: (a) Illustration of the main dyke parameters: location $(x_0, y_0, z_0)$, length $L$, width $W$, strike angle $\phi$, and dip angle $\delta$ (dyke opening $U$ not shown). (b) Cartoon depicting how a dyke is placed on the top of a disc-shaped magma reservoir of radius $R$ at some random coordinate $(r, \theta)$. Here $W$ is closest to the $z$-axis and $x = y = 0$ indicates the centre of the volcano.
the bounds. Together with the fact that a power law distribution for fracture length parameters has been widely studied in engineering and geophysics, we use this for our model and do not consider other, perhaps equally-applicable distributions such as half-normal or log-normal. The only special constraint on the length parameter bounds is with the dyke width parameter $W$. Unlike in Annen et al. (2001) we do not consider dykes that reach the surface, since it has been shown in analogue and numeric studies that the effects of volcano loading on the propagation of dykes near the surface are not negligible (e.g., Muller et al., 2001; Maccaferri et al., 2011; Roman & Jaupart, 2014). This behaviour is too complex to incorporate into this stochastic model of first-order effects, so to avoid the problem we fix our choice of upper bound on $W$ so that $W \cos \delta < z_0$ for all values of $W$.

For the angular orientation parameters $\phi$ and $\delta$ we use a simple uniform distribution, since comparison with a wrapped normal distribution did not show a significant quantitative difference and is not as straightforward to implement. We consider two separate models: one in which there is no correlation between strike or dip angle (2.7a, Table 2.3), and another based on a cone-sheet configuration (e.g., Burchardt et al., 2013) wherein the strike angle is set to be tangential to a circle of radius $r = \sqrt{x_0^2 + y_0^2}$ (where $x_0$ and $y_0$ are still chosen randomly on the reservoir disc of radius $R$) and the dip angle is set so that all dykes in this ensemble dip inwards towards $r = 0$ (Figure 2.7b, Table 2.3). For cases where there are deviations in all quadrants, dykes are placed within $\Delta \theta = \theta_{\text{max}} - \theta_{\text{min}} = 360^\circ$ inside the reservoir disc (Figure 2.6). For cases where deviations occur in fewer quadrants, we consider a directional variation of both uncorrelated and cone models by confining the placement of dykes on the disc to one quadrant only with $\Delta \theta = 90^\circ$. The resulting directional model can then be applied to individual quadrants where there are deviations. Both the $360^\circ$ and $90^\circ$ distributions are illustrated for the uncorrelated model in Figure 2.7a.
Figure 2.7: (a) Planar view of the uncorrelated model with a sample of dykes distributed on the disc-shaped magma reservoir (Figure 2.6b) within some (top) $\Delta \theta = 360^\circ$ for cases where deviations occur in all quadrants, and (bottom) a directional distribution with $\Delta \theta = 90^\circ$ for cases where there are deviations in fewer than all quadrants. Black lines show the projection of dykes on to the xy-plane and colourmap represents vertical displacement. (b) 3D view and planar view (inset) of a sample of dykes for the cone model.

2.5.2 Relative Contribution to Surface Displacement

Given a linear elastic medium and assuming that the decay time for short-term stress accumulation in the rock is less than the time between each dyke placement (e.g., Annen et al., 2001), we can simulate the total uplift at the surface due to an ensemble of dykes by summing the effects of each. The result is a linear relationship between intruded volume and maximum uplift, with the proportionality constant related to chamber depth $z_0$ (Figure 2.8a). Using 1000, 2000, 3000, 5000, and 8000 dyke intrusions (Figure 2.8b) we iterate through both distribution models (uncorrelated and cone sheet) and all $z_0$ values, calculating this linear relationship at each iteration to find the volume required to create the maximum uplift determined by some vertical scale factor times $h_m$ (Figure 2.8c). We also scale the range of this uplift by some
horizontal factor times $R_b$. Scaling in the horizontal direction linearly scales both the wavelength of the surface displacement and the amount of intruded volume (Figure 2.8d). These two scalings allow us to control both the horizontal and vertical nature of our surface displacements, which we use to quantify how much contribution these subsurface processes have on a stratovolcano edifice.

By first fitting volcano profiles with a functional form, we choose those volcanoes that deviate from this form where the topography is “plumper” than the ideal shape (Figure 2.5, Table 2.1, 2.2). For stratovolcanoes that deviate in all quadrants (category 3), we find the relative contribution of intrusive processes to this deviation by solving for optimal scaling factors $\chi$ (in range) and $\eta$ (in height) such that the topographic profile of the volcano is best-fit by minimizing:

$$\text{SSE} = \sum_j \left[ (T_{P_j}^x - (M_{A_j}^x + \chi M_{D_j}^x))^2 + (T_{P_j}^y - (\eta M_{A_j}^y + (1 - \eta)M_{D_j}^y))^2 \right]$$

where $T_{P_j}$ is the topographic profile at the $j^{th}$ data point as in Eq. 2.4 and $M_{A_j}$, $M_{D_j}$ are the profiles at the same point for the analytic and dyke distribution models, respectively (Figure 2.9). Here we have used the Euclidean norm of the residuals for the error: $||T - M||_2$ where $T = (T_x, T_y)$ and $M = (M_x, M_y)$ are vector representations containing the range ($x$) and height ($y$) components of the profiles. For stratovolcanoes that deviate in only one or two quadrants (category 4), we use an unscaled analytic model and add to it a dyke distribution model scaled by optimal scaling factors $\chi$ (in range) and $\xi$ (in height). For this we minimize a similar equation as above:

$$\text{SSE} = \sum_j \left[ (T_{P_j}^x - (M_{A_j}^x + \chi M_{D_j}^x))^2 + (T_{P_j}^y - (\eta M_{A_j}^y + (1 - \eta)M_{D_j}^y))^2 \right]$$

In addition to iterating through the two distribution models and each $z_0$ value, we also iterate through a suite of values for $\chi, \eta$, and $\xi$ to find the model of best fit, using a simple F-test to account for the extra degrees of freedom in our scaling model. If a best fit does not pass this test, then we determine that there is no significant improvement to adding a dyke distribution model to the fit from the analytic form.
Chapter 2. Analysing the Topographic Form of Stratovolcanoes

2.6 Results from Addition of Dyke Distribution

For the $N = 45$ volcanoes in category 3 (protuberant deviations in all quadrants), we found that 39 passed our significance test. Of those, we consider the ones with matching fitting parameters ($\eta$, $z_0$ and the dyke distribution model type) along the semi-major and semi-minor axes indicative of the strongest case for an axisymmetric intrusion process. We found 22 of these and show the results for the two volcanoes from Figure 2.5a (Karymsky and Nantai) in Figure 2.10a. For Karymsky, the model of best fit was the uncorrelated model with $\eta = 0.52, 0.5$ (semi-major, semi-minor axes) and $z_0 = 4$ km, with an average SSE of 0.004 km$^2$. For Nantai, the model of best fit was the cone model with $\eta = 0.6, 0.5$ (semi-major, semi-minor axes) and $z_0 = 4$ km, with an average SSE of 0.04 km$^2$.

For the $N = 51$ volcanoes in category 4 (protuberant deviations in one or two quadrants), we found that 35 passed our significance test. In this case, since the intrusion model used is non-axisymmetric by definition, we relax our criterion for the strongest cases as those where the best fitting distribution model (uncorrelated or cone sheet) matches in those quadrants with deviations. Of those, 20 were deemed to be strong fits and results for the two volcanoes from Figure 2.5b are shown in Figure 2.10b. For Concepción, the model of best fit was the cone model with $\xi = 0.5$ on both the $a_2, b_2$ axes and $z_0 = 4$ km, with an average SSE of 0.03 km$^2$. For Momotombo, the model of best fit was the uncorrelated model with $\xi = 0.04$ on both the $a_1, b_2$ axes and $z_0 = 0.05$ km, with an average SSE of 0.012 km$^2$.

2.7 Discussion

In this study we construct a quantitative way of determining the relative contribution of external processes like the piling of eruptive products, versus internal processes like the uplift due to a subsurface dyke distributions, on the construction of a vol-
Chapter 2. Analysing the Topographic Form of Stratovolcanoes

We find that an axisymmetric ideal analytic form—approximated by the steady-state result of the piling of eruptive products with Bingham rheology (according to Appendix A)—does a good job of fitting certain stratovolcano shapes (Figure 2.4a), while a non-axisymmetric ideal analytic form fits provides an equally good fit to others (Figure 2.4b). We note, however, that not all volcanoes that have a morphometric good fit to this form may be geologically good candidates. An important limitation of this work is that we do not consider erosional effects—for example, Puntiagudo in Chile is a volcano that falls into category 2 according to our analysis (Figure 2.4b) despite being an extinct Pleistocene volcano that has undergone significant erosion due to glaciation. Karátson et al. (2010) find that, since the slope profiles and ages of the volcanoes in their study do not correlate, erosion during the relatively short time after extinction does not have a significant overall effect on a profile shape. For Puntiagudo, however, we know that only a volcanic neck in its upper section remains due to extreme erosion, so while these effects can possibly be ignored in many cases, it cannot be done universally and the inclusion of this volcano in category 2—and subsequent explanation for its deviation from axisymmetry—should be considered carefully.

Where there is not a good fit due to protuberant deviations (Figure 2.5) we find that one possible explanation is via a statistical ensemble of dyke intrusions beneath an edifice. We explore this by applying a dyke distribution model to such volcanoes either axisymmetrically in all quadrants bisected by a semi-major or semi-minor ellipse axis, or directionally in one or two of those quadrants (Figure 2.10). For some volcanoes in categories 3 and 4, we recognize that protuberant deviations may be a result of non-axisymmetric surface and eruptive processes, such as what might arise from the effects of cooling, crystallization, or more complex rheologic behaviour (not encapsulated by our first-order Bingham rheology assumption) on the piling of lavas away from the central vent. For the case studies presented below, however, we corroborate our interpretation of the importance of magma intrusion...
with independent observations.

We also consider in our analysis the role of basal ellipticity in the distribution of the volcanoes into the four categories mentioned above. There appears to be a correlation between volcanoes with a lower ellipticity index (around 1.5 on average) and being fit well by either the axisymmetric or non-axisymmetric form. This suggests that deviations from the form, such as the intrusion of magma, may affect the basal ellipticity as well as uplift of the edifice. In our selection of isolated volcanoes and those without significant craters, we found the majority had a low ellipticity index, which corroborates the findings in Grosse et al. (2009) that the less complex stratovolcanoes considered in their study tended to have lower ellipticity indices. While it is hard to tell concretely, since information on the duration of volcanic activity may not be readily available, our results suggest that volcanoes with deviations from our ideal functional form are ones that are further along in their evolution. This can be seen in certain cases in this study, where volcanoes with an intermediate ellipticity index (e.g. Calbuco, Copiapo, Cerro Azul, Nevado de Longavi) are either complex structures made up of lava domes, multiple vents, collapse events, etc. or are extremely eroded with few surviving volcanic primary forms. Any explanation, then, for the protuberant deviations that they exhibit must take these factors into account.

2.7.1 Deviations from Axisymmetry in Ideal Form

For cases where there is a non-axisymmetric good fit to the ideal functional form, since the average ellipticity index ($\varepsilon = 1.48$) only differs marginally from the cases where there is an axisymmetric good fit ($\varepsilon = 1.52$) we don’t consider that these deviations from axisymmetry are due to something inherent in a volcano’s basal ellipticity. Instead we consider the effects of the piling of eruptive products on a slope. Since we fit a planar trend to the ends of the radial lines used to determine our
average profiles, we can obtain information about the dip angle of these trends. For significant regional trends, the gravitational effects may cause a break in axisymmetry in an edifice, without causing overall deviations from the ideal form.

We look at Merapi in Java (Figure 2.4b) as an example and find that there is correlation between the sides of steepest slope and the dip angle of its planar trend (Figure 2.11). The sides of steepest slope occur between the $a_2$ and $b_1$ axes, which correspond to the compass direction of S-SW. The azimuthal angle (from North) of the planar trend dip is calculated to be $209^\circ$, which also corresponds to S-SW. We note that a collapse scar on the western flank of the edifice is controlling many recent eruptive products preferentially in its direction, however due to the primarily southward nature of our anomaly, we consider this effect to be less influential than the effects due to a planar trend. From our findings, we determine that planar trends, in some instances, could be a possible explanation for deviations from axisymmetry in an otherwise ideal volcanic edifice.

### 2.7.2 Axisymmetric Protuberant Deviations from the Ideal Form

The SRTM data span a time period between February 11 - 22, 2000. For this time period, we consider studies on seismicity or deformation patterns for the volcanoes in category 3 that would support our findings regarding the relative contribution of internal magmatic processes to external eruptive processes on the building of these edifices. We find one such case study with Karymsky in Kamchatka (Figure 2.10a). Here we find that the model of best fit predicts a magma chamber at $z_0 = 4$ km and around $.40 - .80$ km$^3$ of intruded volume. From other studies, we know that the most recent cycle of eruptive activity for Karymsky was from January 1996 to October 2016 and that it has been in a state of continuous eruptive activity since 1908 (Ji et al., 2018). In 1996, there was significant seismic activity in the region
preceding the January eruptive events (Zobin et al., 2003; Magus’kin et al., 2008; Gordeev et al., 1998). Izbekov et al. (2004) concluded that the steady character of crustal deformations before 1996 showed that this seismic activity was not a regular process of tectonic stress release, but rather was triggered by magmatic factors such as dyke intrusions.

It has also been concluded that magma intrusions at Karymsky are fed by a shallow magma chamber at 4 km beneath the volcano (Gordeev et al., 1998; Izbekov et al., 2004; Ji et al., 2018), which is consistent with our model results. This shallow chamber is thought to be fed by a larger, deeper magma chamber that connects to the rest of the region’s volcanic centre. During the period of 2000 - 2010 InSAR studies showed that starting in 2000, Karymsky was undergoing a period of deflation due to the inability of this deeper, interconnected magma chamber to continuously supply its shallower chamber (Ji et al., 2018). Following this, was a period of inflation and another period of deflation, which corresponded to around a .008 km$^3$ volumetric change during the 10-year study period. Furthermore, from Ozerov et al. (2003) it was found that the volume of eruptive materials deposited between 1996 - 2000 were estimated to be between 0.02 and 0.03 km$^3$. Both these studies then suggest that deviations from our ideal form are not due to the constructional external processes taking place at the time the SRTM data was taken, nor due to the regular inflation-deflation cycle of the underlying chamber, since the volume changes during these events are much less than what would be required to see the deviations in our profiles. We recognize that lava rheology and composition could be a factor attributing to this anomaly, as Karymsky can be seen to erupt many short, thick lavas that do not reach the base of its edifice and thus might result in the long-term build-up of lava preferentially located closer to the vent. While we do not rule this out as a possibility, we believe—given the evidence presented above—that the axisymmetric protuberant deviations seen with Karymsky are more likely to due to long-term linear effects of magma intrusion from dykes propagating from a known 4 km deep magma chamber.
underneath the volcano edifice.

2.7.3 Directional Protuberant Deviations from Ideal Form

For our case study on volcanoes in category 4, where there are protuberant deviations in one or two quadrants, we look at Concepción in Nicaragua (Figure 2.10b). Here we find the model of best fit on the $a_2$ and $b_2$ axes—corresponding to roughly the entire northern section of the edifice—is the cone-sheet model with a magma chamber at $z_0 = 0.5$ km. The northern positioning of this model of best fit is consistent with the findings of Van Wyk de Vries (1993) that Concepción has a major central vent surrounded by numerous radial vents around its base, a few north-aligned lineaments, and no major non-volcanic faults nearby. It has also been frequently studied as an example of a volcano spreading due to the gravitational loading of its edifice (Borgia, 1994; Borgia et al., 2000; Merle & Borgia, 1996; Borgia & van Wyk de Vries, 2003). Subsequent gravimetric studies by Saballos et al. (2013), however, show that the density profile of Concepción is inconsistent with a theory of gravitational spreading. They find that the density required for the characteristic spreading time determined in the previous studies is much too low and instead believe that the deformation of Concepción is due to complex underlying magmatic processes. They suggest that the gravity map they obtain can be explained by the analogue models of Galland (2012) on the effects of more complex intrusion processes—such as cone-sheet distributions of dykes—on the surface deformation of an edifice. Along with the known north-aligned lineaments, this is also consistent with our results. Finally, we note that the gravimetric study done in Saballos et al. (2013) and the seismic study done in Saballos et al. (2014) predict a shallow magmatic source at 2 km depth. The model they use for their prediction is one of a simple cylindrical column of magma rising from a shallow reservoir along an open-topped conduit, which is enough of a simplification from the complex structures studied in Galland (2012) that our prediction of a source that is 1.5 km shallower should not be discounted completely. We
conclude that the underlying plumbing structure of Concepción consists of a shallow magma chamber around 0.5 - 2 km deep that feeds a complex dyke system—such as a cone-sheet distribution—that is preferentially oriented towards the Northern section of the volcano.

2.8 Conclusions

The processes that make up the iconic shape of most stratovolcanoes that we see today are numerous and complex. Insight into how this shape is formed, however, can give us a better understanding of the relative contribution of internal and external processes that go in to stratovolcano construction, which could have implications for bettering hazard assessment. In this study we fit a series of stratovolcano edifice profiles to an axisymmetric ideal mathematical form based on a stochastic piling of eruptive products. We find that this form fits many stratovolcanoes well, but that there are notable deviations—either from axisymmetry or from the form itself, or both (see data supplement: Appendix B; Cosburn & Roy (2020a)). For deviations from the form itself, we consider those that are more protuberant as possibly arising from the internal emplacement of magma due to distributions of dykes below an edifice. We find that our results are consistent with previous studies done for the volcanoes Karymsky in Kamchatka and Concepción in Nicaragua. For cases where there is a good fit to the ideal form, but a deviation from axisymmetry, we conclude that this could arise from the effects of a regional trend on how eruptive products such as lava pile up. For this we find that the profile of Merapi in Java has potentially arisen due to such an effect. Finally, we look at basal ellipticity and how it affects how a volcanic profile adheres to the ideal functional form. We find that there is a better fit, in general, by volcanoes with a lower ellipticity index ($\bar{e} = 1.5$) and conclude that this is an important factor in a volcano’s overall morphometric shape and evolution. In summary, the models and categorization presented here are a way
to formulate hypotheses about the constructive processes that build the topographic form of a stratovolcano. Misfits and deviations from our models, therefore, provide an important way to explore other processes not considered here, such as destructive ones like erosion.
### Table 2.1: Table of $N = 102$ stratovolcanoes with low ellipticity index ($\varepsilon \leq 1.5$) used in this study, along with their region, Grosse et al. (2014) ellipticity index, and functional form fitting category (1 = axisymmetric good fit, 2 = nonaxisymmetric good fit, 3 = protuberant deviations all quadrants, 4 = protuberant deviations 1-2 quadrants). See also: the Data Supplement (Appendix B; Cosburn & Roy (2020a)).

<table>
<thead>
<tr>
<th>Name</th>
<th>Region</th>
<th>Ellip. Index</th>
<th>Fit. Cat.</th>
<th>Name</th>
<th>Region</th>
<th>Ellip. Index</th>
<th>Fit. Cat.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Agua</td>
<td>Guatemala</td>
<td>1.04</td>
<td>1</td>
<td>Latukan</td>
<td>Mindanao-Philippines</td>
<td>1.5</td>
<td>2</td>
</tr>
<tr>
<td>Alaid</td>
<td>Kuril-Is</td>
<td>1.24</td>
<td>1</td>
<td>Lava</td>
<td>Java</td>
<td>1.46</td>
<td>2</td>
</tr>
<tr>
<td>Ale Bagu</td>
<td>Ethiopia</td>
<td>1.26</td>
<td>3</td>
<td>Lewotolo</td>
<td>Lesser-Sunda-Is</td>
<td>1.27</td>
<td>2</td>
</tr>
<tr>
<td>Anam</td>
<td>Kamchatka</td>
<td>1.36</td>
<td>4</td>
<td>Licancabur</td>
<td>Chile-N</td>
<td>1.21</td>
<td>4</td>
</tr>
<tr>
<td>Antuco</td>
<td>Chile-C</td>
<td>1.25</td>
<td>3</td>
<td>Llullaillaco</td>
<td>Chile-N</td>
<td>1.49</td>
<td>2</td>
</tr>
<tr>
<td>Arena1</td>
<td>Costa-Rica</td>
<td>1.25</td>
<td>4</td>
<td>Lolo</td>
<td>New-Britain-SW-Pac</td>
<td>1.48</td>
<td>4</td>
</tr>
<tr>
<td>Astoumpiri</td>
<td>Kuril-Is</td>
<td>1.19</td>
<td>1</td>
<td>Maca</td>
<td>Chile-S</td>
<td>1.43</td>
<td>3</td>
</tr>
<tr>
<td>Ayelu</td>
<td>Ethiopia</td>
<td>1.36</td>
<td>3</td>
<td>Maderas</td>
<td>Nicaragua</td>
<td>1.25</td>
<td>2</td>
</tr>
<tr>
<td>Bachelor</td>
<td>US-Oregon</td>
<td>1.32</td>
<td>3</td>
<td>Matutum</td>
<td>Mindanao-Philippines</td>
<td>1.41</td>
<td>2</td>
</tr>
<tr>
<td>Bam</td>
<td>New-Guinea-NE</td>
<td>1.41</td>
<td>3</td>
<td>Mayon</td>
<td>Luzon-Philippines</td>
<td>1.11</td>
<td>1</td>
</tr>
<tr>
<td>Bamus</td>
<td>New-Britain-SW-Pac</td>
<td>1.44</td>
<td>3</td>
<td>Merapi</td>
<td>Java</td>
<td>1.31</td>
<td>2</td>
</tr>
<tr>
<td>Belenkaya</td>
<td>Kamchatka</td>
<td>1.36</td>
<td>3</td>
<td>Momotombo</td>
<td>Nicaragua</td>
<td>1.18</td>
<td>4</td>
</tr>
<tr>
<td>Bobrof</td>
<td>Aleutian-Is</td>
<td>1.45</td>
<td>3</td>
<td>Nantai</td>
<td>Honshu-Japan</td>
<td>1.36</td>
<td>3</td>
</tr>
<tr>
<td>Bola</td>
<td>New-Britain-SW-Pac</td>
<td>1.27</td>
<td>4</td>
<td>Arenal</td>
<td>Anadaman-Is-Indian-O</td>
<td>1.46</td>
<td>3</td>
</tr>
<tr>
<td>Borisov</td>
<td>Ethiopia</td>
<td>1.34</td>
<td>4</td>
<td>Nevado del Tolima</td>
<td>Colombia</td>
<td>1.45</td>
<td>3</td>
</tr>
<tr>
<td>Carlisle</td>
<td>Aleutian-Is</td>
<td>1.19</td>
<td>1</td>
<td>O Denyo</td>
<td>Indonesia</td>
<td>1.27</td>
<td>4</td>
</tr>
<tr>
<td>Cerro</td>
<td>Java</td>
<td>1.42</td>
<td>2</td>
<td>Osorno</td>
<td>Chile-S</td>
<td>1.27</td>
<td>1</td>
</tr>
<tr>
<td>Cerro Tugule</td>
<td>Argentina</td>
<td>1.31</td>
<td>3</td>
<td>Ostry</td>
<td>Kamchatka</td>
<td>1.33</td>
<td>4</td>
</tr>
<tr>
<td>Chaqlak</td>
<td>Aleutian-Is</td>
<td>1.41</td>
<td>3</td>
<td>Pakomo</td>
<td>Chile-C</td>
<td>1.35</td>
<td>4</td>
</tr>
<tr>
<td>Chilen</td>
<td>Chile-N</td>
<td>1.19</td>
<td>2</td>
<td>Parinacota</td>
<td>Chile-N</td>
<td>1.19</td>
<td>4</td>
</tr>
<tr>
<td>Cleveland</td>
<td>Aleutian-Is</td>
<td>1.14</td>
<td>1</td>
<td>Parol Sister</td>
<td>Alaskan-Peninsula</td>
<td>1.15</td>
<td>4</td>
</tr>
<tr>
<td>Cobachi</td>
<td>Chile-N</td>
<td>1.43</td>
<td>3</td>
<td>Peinado</td>
<td>Argentina</td>
<td>1.25</td>
<td>3</td>
</tr>
<tr>
<td>Concepcion</td>
<td>Nicaragua</td>
<td>1.15</td>
<td>4</td>
<td>Penanggungan</td>
<td>Java</td>
<td>1.38</td>
<td>2</td>
</tr>
<tr>
<td>Damaund</td>
<td>Iran</td>
<td>1.35</td>
<td>1</td>
<td>Pico</td>
<td>Azores</td>
<td>1.25</td>
<td>3</td>
</tr>
<tr>
<td>Ebulobo</td>
<td>Lesser-Sunda-Is</td>
<td>1.2</td>
<td>1</td>
<td>Puliseri</td>
<td>Java</td>
<td>1.39</td>
<td>1</td>
</tr>
<tr>
<td>El Misti</td>
<td>Peru</td>
<td>1.34</td>
<td>2</td>
<td>Puntinggal</td>
<td>Chile-C</td>
<td>1.39</td>
<td>2</td>
</tr>
<tr>
<td>Fiji</td>
<td>Honshu-Japan</td>
<td>1.27</td>
<td>1</td>
<td>Saka</td>
<td>W-Indies</td>
<td>1.34</td>
<td>3</td>
</tr>
<tr>
<td>Fuso Peak</td>
<td>Kuril-Is</td>
<td>1.29</td>
<td>3</td>
<td>Sakar</td>
<td>New-Guinea-NE</td>
<td>1.33</td>
<td>1</td>
</tr>
<tr>
<td>Gamkonora</td>
<td>Halmahera-Indonesia</td>
<td>1.34</td>
<td>1</td>
<td>Santa Maria</td>
<td>Guatemala</td>
<td>1.27</td>
<td>2</td>
</tr>
<tr>
<td>Hary</td>
<td>New-Britain-SW-Pac</td>
<td>1.24</td>
<td>1</td>
<td>Senaru</td>
<td>Java</td>
<td>1.27</td>
<td>4</td>
</tr>
<tr>
<td>Hood</td>
<td>US-Oregon</td>
<td>1.41</td>
<td>2</td>
<td>Shiraldin</td>
<td>Aleutian-Is</td>
<td>1.38</td>
<td>1</td>
</tr>
<tr>
<td>Hou</td>
<td>Halmahera-Indonesia</td>
<td>1.41</td>
<td>2</td>
<td>Sinabung</td>
<td>Sumatra</td>
<td>1.25</td>
<td>3</td>
</tr>
<tr>
<td>Hiloobeng</td>
<td>Lesser-Sunda-Is</td>
<td>1.35</td>
<td>2</td>
<td>Soputan</td>
<td>Sulawesi-Indonesia</td>
<td>1.39</td>
<td>3</td>
</tr>
<tr>
<td>Ilyndey</td>
<td>Kamchatka</td>
<td>1.32</td>
<td>4</td>
<td>Sumbung</td>
<td>Java</td>
<td>1.15</td>
<td>1</td>
</tr>
<tr>
<td>Inerie</td>
<td>Lesser-Sunda-Is</td>
<td>1.15</td>
<td>1</td>
<td>Tacora</td>
<td>Chile-N</td>
<td>1.38</td>
<td>2</td>
</tr>
<tr>
<td>Isla El Tigre</td>
<td>Honduras</td>
<td>1.27</td>
<td>3</td>
<td>Tafahi</td>
<td>Tonga-SW-Pac</td>
<td>1.48</td>
<td>2</td>
</tr>
<tr>
<td>Izalgo</td>
<td>El-Salvador</td>
<td>1.15</td>
<td>3</td>
<td>Tamponas</td>
<td>Java</td>
<td>1.37</td>
<td>2</td>
</tr>
<tr>
<td>Kanajja</td>
<td>Aleutian-Is</td>
<td>1.11</td>
<td>4</td>
<td>Tarso Tonsuid</td>
<td>Africa-N</td>
<td>1.4</td>
<td>1</td>
</tr>
<tr>
<td>Kao</td>
<td>Tonga-SW-Pac</td>
<td>1.48</td>
<td>3</td>
<td>Tata Sabaya</td>
<td>Bolivia</td>
<td>1.31</td>
<td>2</td>
</tr>
<tr>
<td>Karangantang</td>
<td>Sangihe-Is-Indonesia</td>
<td>1.5</td>
<td>4</td>
<td>Tikore</td>
<td>Halmahera-Indonesia</td>
<td>1.16</td>
<td>1</td>
</tr>
<tr>
<td>Karymsky</td>
<td>Kamchatka</td>
<td>1.18</td>
<td>3</td>
<td>Timakula</td>
<td>Santa-Cruz-Is-SW-Pac</td>
<td>1.33</td>
<td>4</td>
</tr>
<tr>
<td>Kermi</td>
<td>Sumatra</td>
<td>1.3</td>
<td>2</td>
<td>To-shima</td>
<td>Indonesia</td>
<td>1.31</td>
<td>3</td>
</tr>
<tr>
<td>Kinyo</td>
<td>Africa-E</td>
<td>1.44</td>
<td>3</td>
<td>Ulawun</td>
<td>New-Britain-SW-Pac</td>
<td>1.34</td>
<td>4</td>
</tr>
<tr>
<td>Kizimen</td>
<td>Kamchatka</td>
<td>1.2</td>
<td>4</td>
<td>Uliaga</td>
<td>Aleutian-Is</td>
<td>1.49</td>
<td>1</td>
</tr>
<tr>
<td>Klubat</td>
<td>Sulawesi-Indonesia</td>
<td>1.33</td>
<td>2</td>
<td>Villarrica</td>
<td>Chile-C</td>
<td>1.3</td>
<td>3</td>
</tr>
<tr>
<td>Klinchevski</td>
<td>Kamchatka</td>
<td>1.09</td>
<td>2</td>
<td>Viluchik</td>
<td>Kamchatka</td>
<td>1.31</td>
<td>1</td>
</tr>
<tr>
<td>Koryalsky</td>
<td>Kamchatka</td>
<td>1.41</td>
<td>1</td>
<td>Visoke</td>
<td>Africa-C</td>
<td>1.25</td>
<td>3</td>
</tr>
<tr>
<td>Krenitzyn Peak</td>
<td>Kuril-Is</td>
<td>1.22</td>
<td>3</td>
<td>Vsevidof</td>
<td>Aleutian-Is</td>
<td>1.13</td>
<td>2</td>
</tr>
<tr>
<td>Kronotsky</td>
<td>Kamchatka</td>
<td>1.33</td>
<td>1</td>
<td>Yotei</td>
<td>Hokkaido-Japan</td>
<td>1.15</td>
<td>1</td>
</tr>
<tr>
<td>Lamn</td>
<td>Chile-C</td>
<td>1.24</td>
<td>2</td>
<td>Zelitovsky</td>
<td>Kamchatka</td>
<td>1.21</td>
<td>1</td>
</tr>
</tbody>
</table>
## Table 2.2: Table of $N = 88$ stratovolcanoes with intermediate ($1.5 < \varepsilon \leq 2$) or high ($2 < \varepsilon \leq 3$) ellipticity index used in this study, along with their region, Grosse et al. (2014) ellipticity index, and functional form fitting category (1 = axisymmetric good fit, 2 = nonaxisymmetric good fit, 3 = protuberant deviations all quadrants, 4 = protuberant deviations 1-2 quadrants). See also: the Data Supplement (Appendix B; Cosburn & Roy (2020a)).

<table>
<thead>
<tr>
<th>Name</th>
<th>Region</th>
<th>Ellip. Index</th>
<th>Fit Cat.</th>
<th>Name</th>
<th>Region</th>
<th>Ellip. Index</th>
<th>Fit Cat.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Adams</td>
<td>US-Washington</td>
<td>1.75</td>
<td>4</td>
<td>Kagamil</td>
<td>Aleutian-Is</td>
<td>1.92</td>
<td>2</td>
</tr>
<tr>
<td>Agung</td>
<td>Lesser-Sunda-Is</td>
<td>1.78</td>
<td>1</td>
<td>Kamchatka</td>
<td>Kamchatka</td>
<td>1.76</td>
<td>1</td>
</tr>
<tr>
<td>Antisana</td>
<td>Ecuador</td>
<td>1.72</td>
<td>2</td>
<td>Karang</td>
<td>Java</td>
<td>1.61</td>
<td>3</td>
</tr>
<tr>
<td>Aragots</td>
<td>Armenia</td>
<td>2.35</td>
<td>3</td>
<td>Kharimkotan</td>
<td>Kurl-Is</td>
<td>1.72</td>
<td>4</td>
</tr>
<tr>
<td>Ararat</td>
<td>Turkey</td>
<td>1.65</td>
<td>2</td>
<td>Komarov</td>
<td>Kamchatka</td>
<td>1.86</td>
<td>4</td>
</tr>
<tr>
<td>Ararat</td>
<td>Luzon-Philippines</td>
<td>1.77</td>
<td>2</td>
<td>Krunyt</td>
<td>Sumatra</td>
<td>1.92</td>
<td>4</td>
</tr>
<tr>
<td>Bakering</td>
<td>Kamchatka</td>
<td>1.52</td>
<td>2</td>
<td>Kurilema</td>
<td>Honshu-Japan</td>
<td>2.44</td>
<td>4</td>
</tr>
<tr>
<td>Bandai</td>
<td>Honshu-Japan</td>
<td>1.95</td>
<td>2</td>
<td>Late</td>
<td>Tonga-SW-Pac</td>
<td>1.59</td>
<td>3</td>
</tr>
<tr>
<td>Baranbey</td>
<td>Kurl-Is</td>
<td>1.7</td>
<td>2</td>
<td>Laima</td>
<td>Chile-C</td>
<td>1.78</td>
<td>4</td>
</tr>
<tr>
<td>Baru</td>
<td>Panama</td>
<td>1.58</td>
<td>2</td>
<td>Lopevi</td>
<td>Vanatu-SW-Pac</td>
<td>1.59</td>
<td>4</td>
</tr>
<tr>
<td>Baturi</td>
<td>Ateian-Is</td>
<td>1.76</td>
<td>2</td>
<td>Maeroves</td>
<td>Luzon-Philippines</td>
<td>1.69</td>
<td>2</td>
</tr>
<tr>
<td>Bur Ni Telong</td>
<td>Sumatra</td>
<td>2.12</td>
<td>4</td>
<td>Masaraga</td>
<td>Luzon-Philippines</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>Calbrazo</td>
<td>Chile-S</td>
<td>1.62</td>
<td>4</td>
<td>Mentolat</td>
<td>Chile-S</td>
<td>1.69</td>
<td>4</td>
</tr>
<tr>
<td>Calaguei</td>
<td>Chile-C</td>
<td>1.84</td>
<td>4</td>
<td>Merbabu</td>
<td>Java</td>
<td>1.78</td>
<td>4</td>
</tr>
<tr>
<td>Cerro del Aalto</td>
<td>Chile-C</td>
<td>1.7</td>
<td>2</td>
<td>Minhinniavida</td>
<td>Chile-S</td>
<td>1.74</td>
<td>3</td>
</tr>
<tr>
<td>Cerro del Condor</td>
<td>Argentina</td>
<td>1.93</td>
<td>2</td>
<td>Mozua Alli</td>
<td>Ethiopia</td>
<td>2.37</td>
<td>4</td>
</tr>
<tr>
<td>Chinginanggik</td>
<td>Aletian-Peninsula</td>
<td>1.75</td>
<td>4</td>
<td>Nevado de Ruiz</td>
<td>Colombia</td>
<td>1.54</td>
<td>3</td>
</tr>
<tr>
<td>Chimborazo</td>
<td>Ecuador</td>
<td>1.69</td>
<td>4</td>
<td>Nevus Peak</td>
<td>W-Indies</td>
<td>1.68</td>
<td>2</td>
</tr>
<tr>
<td>Chirinkoten</td>
<td>Kurl-Is</td>
<td>1.6</td>
<td>2</td>
<td>Nila</td>
<td>Banda-Sea</td>
<td>1.54</td>
<td>4</td>
</tr>
<tr>
<td>Copiapo</td>
<td>Chile-N</td>
<td>1.59</td>
<td>2</td>
<td>Paluwech</td>
<td>Lesser-Sunda-Is</td>
<td>1.72</td>
<td>3</td>
</tr>
<tr>
<td>Cumbal</td>
<td>Colombia</td>
<td>1.95</td>
<td>2</td>
<td>Pavlof</td>
<td>Aletian-Peninsula</td>
<td>1.75</td>
<td>1</td>
</tr>
<tr>
<td>Dabulhnu</td>
<td>Ethiopia</td>
<td>1.59</td>
<td>3</td>
<td>Pedee</td>
<td>W-Indies</td>
<td>1.96</td>
<td>2</td>
</tr>
<tr>
<td>Dutton</td>
<td>Aletian-Peninsula</td>
<td>1.8</td>
<td>2</td>
<td>Quetupilinan</td>
<td>Chile-C</td>
<td>1.61</td>
<td>4</td>
</tr>
<tr>
<td>El Tice</td>
<td>El-Salvador</td>
<td>1.85</td>
<td>2</td>
<td>Rainier</td>
<td>US-Washington</td>
<td>1.73</td>
<td>1</td>
</tr>
<tr>
<td>Ercues Dagi</td>
<td>Turkey</td>
<td>2.08</td>
<td>3</td>
<td>Nishiri</td>
<td>Hokkaido-Japan</td>
<td>1.85</td>
<td>1</td>
</tr>
<tr>
<td>Fourpeaked</td>
<td>Aletian-Peninsula</td>
<td>1.82</td>
<td>3</td>
<td>Rota</td>
<td>Nicaragua</td>
<td>1.65</td>
<td>2</td>
</tr>
<tr>
<td>Gabillemna</td>
<td>Ethiopia</td>
<td>1.77</td>
<td>4</td>
<td>Ruapehu</td>
<td>New-Zealand</td>
<td>1.79</td>
<td>4</td>
</tr>
<tr>
<td>Garelo</td>
<td>Aletian-Is</td>
<td>1.57</td>
<td>4</td>
<td>Rudakoff</td>
<td>Kurl-Is</td>
<td>1.62</td>
<td>3</td>
</tr>
<tr>
<td>Glacier Peak</td>
<td>US-Washington</td>
<td>1.69</td>
<td>1</td>
<td>Saint Helena</td>
<td>US-Washington</td>
<td>1.57</td>
<td>1</td>
</tr>
<tr>
<td>Golaya</td>
<td>Kamchatka</td>
<td>1.91</td>
<td>2</td>
<td>San Vicente</td>
<td>El-Salvador</td>
<td>1.86</td>
<td>3</td>
</tr>
<tr>
<td>Guadalupe</td>
<td>Chile-N</td>
<td>1.52</td>
<td>4</td>
<td>Sangangnuey</td>
<td>Mexico</td>
<td>1.83</td>
<td>1</td>
</tr>
<tr>
<td>Heard</td>
<td>Indian-O-S</td>
<td>1.75</td>
<td>4</td>
<td>Sarychev Peak</td>
<td>Kurl-Is</td>
<td>1.51</td>
<td>4</td>
</tr>
<tr>
<td>Hecchi</td>
<td>Honshu-Japan</td>
<td>1.72</td>
<td>4</td>
<td>Seniauwah Agam</td>
<td>Sumatra</td>
<td>1.7</td>
<td>2</td>
</tr>
<tr>
<td>Hornopiren</td>
<td>Chile-S</td>
<td>1.8</td>
<td>4</td>
<td>Slamet</td>
<td>Java</td>
<td>1.58</td>
<td>4</td>
</tr>
<tr>
<td>Ichinsky</td>
<td>Kamchatka</td>
<td>1.61</td>
<td>4</td>
<td>Sumaco</td>
<td>Ecuador</td>
<td>1.52</td>
<td>2</td>
</tr>
<tr>
<td>Illalabekan</td>
<td>Lesser-Sunda-Is</td>
<td>1.55</td>
<td>2</td>
<td>Sumbing</td>
<td>Sumatra</td>
<td>1.53</td>
<td>4</td>
</tr>
<tr>
<td>Illinza</td>
<td>Ecuador</td>
<td>1.85</td>
<td>4</td>
<td>Tamao</td>
<td>Mexico</td>
<td>1.71</td>
<td>2</td>
</tr>
<tr>
<td>Iriga</td>
<td>Luzon-Philippines</td>
<td>1.72</td>
<td>1</td>
<td>Tajumulco</td>
<td>Guatemala</td>
<td>1.64</td>
<td>2</td>
</tr>
<tr>
<td>Isanotski</td>
<td>Aletian-Is</td>
<td>1.65</td>
<td>4</td>
<td>Teunerfe</td>
<td>Canary-Is</td>
<td>1.65</td>
<td>2</td>
</tr>
<tr>
<td>Iwakoll</td>
<td>Honshu-Japan</td>
<td>1.53</td>
<td>1</td>
<td>Tungurahuna</td>
<td>Ecuador</td>
<td>1.52</td>
<td>2</td>
</tr>
<tr>
<td>Jabilo</td>
<td>Holmaharta-Indonesia</td>
<td>1.52</td>
<td>1</td>
<td>Usulutan</td>
<td>El-Salvador</td>
<td>1.84</td>
<td>2</td>
</tr>
<tr>
<td>Jefferson</td>
<td>US-Oregon</td>
<td>1.78</td>
<td>2</td>
<td>Uturuncu</td>
<td>Bolivia</td>
<td>1.84</td>
<td>2</td>
</tr>
<tr>
<td>Jocotitlan</td>
<td>Mexico</td>
<td>1.74</td>
<td>3</td>
<td>Visoiky</td>
<td>Kamchatka</td>
<td>1.57</td>
<td>3</td>
</tr>
</tbody>
</table>
Chapter 2. Analysing the Topographic Form of Stratovolcanoes

<table>
<thead>
<tr>
<th>Uncorrelated Model</th>
<th>Cone Model</th>
</tr>
</thead>
<tbody>
<tr>
<td>L (length)</td>
<td>50 - 5000 (m)</td>
</tr>
<tr>
<td>W (width)</td>
<td>50 - z₀ (m)</td>
</tr>
<tr>
<td>U (opening)</td>
<td>0.25 - 1.0 (m)</td>
</tr>
<tr>
<td>φ (strike angle)</td>
<td>0° - 360°</td>
</tr>
<tr>
<td>δ (dip angle)</td>
<td>50° - 130°</td>
</tr>
</tbody>
</table>

Table 2.3: Values for dyke parameters (Figure 2.6a) used in both the uncorrelated and cone distribution models (Figure 2.7).
Figure 2.8: (a) Plot at each chamber depth $z_0$ of total intruded volume vs. maximum displacement (cone model) for 3 stochastic instances each of 1000, 2000, 3000, 5000, and 8000 dyke intrusions, as illustrated in (b). (c) Profiles of uplift at the surface for $z_0 = 0.5, 2, 5$ km for both the uncorrelated and cone models and 3 stochastic instances each of 1000, 2000, 3000, 5000, and 8000 dyke intrusions, as also illustrated in (b). The range is normalized in order to scale horizontally to account for variable chamber radii. Shallower chambers tend to produce uplift with more complex shapes and deeper chambers tend to produce those that are more broad and flat. (d) Example profiles of normalized uplift (uncorrelated model) showing the effects of larger vs. smaller chamber radii. Blue lines show the effects of doubling a chamber radius (dashed lines to solid lines) for a shallow $z_0 = 0.5$ km chamber and red lines show the same effect but for a deeper $z_0 = 2$ km chamber. In both cases, the overall shape is preserved, but the wavelength and amount of intruded volume doubles. This means that for volcanoes with larger basal radii, larger reservoirs are need to feed any dyke system that would produce significant protuberant deviations to the functional form.
Figure 2.9: Cartoon of the scaling problem we consider to determine the relative contributions of uplift due to dyke distributions that affect the topographic edifice. The blue line, $M_A$, shows a profile of the analytic ideal form, and the red line, $M_D$, shows one possible surface deformation that might arise from a subsurface ensemble of dykes. The sum-of-squares error between the topographic profile and the $M_A, M_D$ combination model is minimized for horizontal scaling factor $\chi$ and either vertical scaling factor $\eta$ (for volcanoes in category 3) or vertical scaling factor $\xi$ (for volcanoes in category 4).
Figure 2.10: (a) Results for the combination model (Figure 2.9) for the two stratovolcanoes in category 3 (protuberant deviations in all quadrants) from Figure 2.5a. For Karymsky, the model of best fit was the uncorrelated model with $\eta = 0.52, 0.5$ (semi-major, semi-minor axes) and $z_0 = 4$ km, with an average SSE of 0.004 km$^2$. For Nantai, the model of best fit was the cone model with $\eta_3 = 0.6, 0.5$ (semi-major, semi-minor axes) and $z_0 = 4$ km, with an average SSE of 0.04 km$^2$. (b) Results for the combination model (Figure 2.9) for the two stratovolcanoes in category 4 (protuberant deviations in one or two quadrants) from Figure 2.5b. For Concepción, the model of best fit was the cone model with $\xi = 0.5$ on both the $a_2, b_2$ axes and $z_0 = 4$ km, with an average SSE of 0.03 km$^2$. For Momotombo, the model of best fit was the uncorrelated model with $\xi = 0.04$ on both the $a_1, b_2$ axes and $z_0 = 0.05$ km, with an average SSE of 0.012 km$^2$. (all insets) Thick red lines indicate the semi-major or semi-minor ellipse axis or axes where a profile deviates from the functional form.
Figure 2.11: (left) Profile of Merapi fit by the ideal analytic form shows a steeper slope in the quadrants bisected by the $a_2$, $b_1$ axes, corresponding to S-SW and the dip angle of the planar trend (right).
Chapter 3

Joint Inversion of Cosmic-Ray Muon & Gravity Data

The contents of this chapter were originally published as part of Cosburn, K., Roy, M., Guardincerri, E., & Rowe, C., 2019. Joint inversion of gravity with cosmic ray muon data at a well-characterized site for shallow subsurface density prediction, Geophysical Journal International, 217(3), 1988–2002. Minor edits within this dissertation have been made to clarify the text.

3.1 Abstract

It is well known in the geophysics community that the estimation of subsurface density is important for imaging various geologic structures, such as volcanic edifices, reservoirs, and aquifers. Muon tomography has recently been used to complement traditional gravity measurements as a powerful method for probing shallow subsurface density structure beneath volcanoes and lava domes. As a result of its markedly different spatial sensitivity, gravity has proven to be a powerful combination to use with muon data, especially when imaging structures on spatial scales larger than
the area encompassed by crossing muon trajectories. In Cosburn et al. (2019), we explore and test a joint inversion of gravity and muon data in a study area containing an independently-characterized target anomaly: a regionally extensive, high-density layer of ash-flow tuff beneath Los Alamos, New Mexico, USA. We resolve the nearly flat-lying structure using a unique experimental set-up wherein surface and subsurface gravity and muon measurements are obtained above and below the target volume. Our results show that with minimal geologic (prior) constraints, the joint inversion correctly recovers salient features of the expected density structure. The results of our study illustrate the potential of combining surface and subsurface (e.g., borehole) gravity and muon measurements to invert for shallow geologic structures.

### 3.2 Introduction

The development of accurate static and time-dependent models for resolving shallow subsurface density is important to a wide range of earth science problems, such as volcano hazard monitoring (e.g., Bagnardi et al., 2014; Kazama et al., 2015; Tanaka, 2015), reservoir imaging (e.g., Lumley, 2001; Ringrose & Bentley, 2015), and discerning time-dependent fluid flow in hydrology (e.g., Kennedy et al., 2016), injection wells (e.g., Ellsworth, 2013; Keranen et al., 2014), and carbon sequestration (e.g., Zhang et al., 2012). Gravity measurements are traditional in geophysical imaging of subsurface density, typically in conjunction with constraints from seismic (e.g., Daggett et al., 1986; Roecker et al., 2004; Roy et al., 2005) or electromagnetic data (e.g., O’Neill & England, 1994; Hautot et al., 2007). Integration and interpretation of these geophysical datasets is complicated by the fact that, whereas gravity measurements depend linearly on density, observables such as seismic wavespeeds and electrical conductivity are complex functions of density, temperature, composition, porosity, and saturation. In contrast, the combination of gravity data with muon attenuation is ideal because, like gravity, this attenuation depends linearly on material
density (e.g., Leo, 1987; Groom et al., 2001), which can thus be interpreted directly from the integrated dataset.

Cosmic-ray muons are produced by the decay of pions (a product of interactions between cosmic-ray protons and atmospheric nuclei) and kaons in the upper atmosphere, and arrive at the surface of the Earth with a well-known energy spectrum (e.g., Shukla & Sankrith, 2016). The processes that lead to muon stopping, or attenuation, in matter are well-understood (e.g., Leo, 1987) and provide information about the integrated density-length encountered by the muons along their path (e.g., application to the imaging of caverns in the pyramids of Giza by Alvarez et al. (1970)). Attenuation of muons through matter depends primarily on the density of the material, with additional effects arising from the chemical composition and water content of the target volume. Interpretation of a formal joint inversion of gravity and muon attenuation combines their inherently distinct spatial sensitivities and allows for a tomographic approach in places where there are crossing muon trajectories, thus mitigating the non-uniqueness problem intrinsic to gravity measurements (Jourde et al., 2015; Nishiyama et al., 2014). Recent studies have demonstrated the power of using muon tomography in joint inversions with gravity for predicting volcanic subsurface density structures (e.g., Jourde et al., 2015; Nishiyama et al., 2014; Tanaka, 2015; Rosas-Carbajal et al., 2017), as well as experiments exploring borehole muon tomography as an application for wider subsurface exploration (e.g., Bonneville et al., 2017).

At the time of publishing Cosburn et al. (2019), gravity-muon joint inversion methodology has not been tested using a known and independently characterised target. Additionally, the combination of gravity and muon data from both surface and subsurface measurements has not been previously explored, and here we present joint inversions of gravity and muon data that exploit both types of measurements, allowing us to image a laterally uniform density anomaly, which could not be detected using surface measurements alone. Our study has two unique features designed to
best evaluate and test the joint inversion: (1) availability of quantitative constraints on the density anomaly to be imaged from independent geologic observations, and (2) an advantageous experimental set-up wherein our target volume is sandwiched above and below by muon and gravity measurements. The goal of our study is to determine whether the joint inversion of surface and subsurface gravity and muon measurements may be used to image vertically stratified structures with minimal prior constraints. Using available geologic constraints on the expected structure, we test our joint-inversion methodology and explore the sensitivity of our model to various facets of the two datasets. Our approach is generally applicable to imaging subsurface targets where gravity data can provide more information about the regional context of a density anomaly outside the (typically limited) region of highest muon sensitivity (as determined by the acceptance angles of the detector).

3.3 Target Density Anomaly

Our study area is located on the eastern flank of the Jemez volcano and west of the Neogene-Quaternary Rio Grande Rift near the town of Los Alamos, New Mexico, USA. Our target volume comprises the rocks within the mesa beneath Los Alamos and above a 100 m long tunnel within Technical Area 41 (TA-41), part of the Los Alamos National Laboratory (Figure 3.1a). A well-characterized regional stratigraphy of cooling units within the Tshirege member of the Bandelier Tuff (Gardner & Goff, 1984) provides the structural setting for our study (Figure 3.1b). These cooling units have been shown to be regionally extensive and nearly flat-lying (Lewis et al., 2009). Those relevant to our target volume are, stratigraphically from highest to lowest: Qbt3, Qbt2, and Qbt1 (Figure 3.1b). The densities of each unit have been measured in a borehole 4-5 km southeast of our study area (Broxton & Vaniman, 2005), with density averages of $\rho_3 = 1800 \text{ kg/m}^3$, $\rho_2 = 2100 \text{ kg/m}^3$, and $\rho_1 = 1400 \text{ kg/m}^3$ for Qbt3, Qbt2, and Qbt1, respectively (Figure 3.1c).
Our target anomaly is the higher-density Qbt2 layer, located between the lower-density Qbt3 and Qbt1 layers (Figures 3.1b; 3.1c). The upper interface between Qbt2 and Qbt3 is exposed along the cliff face within the Los Alamos canyon, at an elevation of 2183 m (Figure 3.1a). The lower interface between Qbt1 and Qbt2 is hidden, but inferred, to be at or above the elevation of the TA-41 tunnel, based on prior geologic mapping (Figure 3.1b; Lewis et al. (2009)), borehole measurements (Figure 3.1c; Broxton & Vaniman (2005)), and gravity modelling (Roy et al., 2017). The Qbt2 layer is thus a nearly horizontal, positive anomaly within our target rock volume.
As is common in many basin/reservoir settings, or on the flanks of volcanic edifices such as our study area, stratigraphic units may be separated by regionally-extensive, sub-planar interfaces. In this setting, the laterally-uniform, higher-density Qbt2 layer described above would not be detectable if we only made measurements at the surface. Instead, our experimental geometry exploits the location of the TA-41 tunnel and combines measurements of gravity and muon intensity (countrate of muons per angular bin) both inside the tunnel and on the mesa surface above it. The combination of surface and subsurface gravity measurements are crucial for resolving the depth-variability of density (Roy et al., 2017). Density can also be found through muon attenuation within our target volume, obtained by taking the ratio of muon intensity from within the tunnel to the intensity above (Guardincerri et al., 2017). This experimental set-up is ideal for testing and validation of a gravity-muon joint inversion and, while subsurface measurements may not be attainable in all geologic settings, such as at volcanic edifices, our findings have applicability for future borehole muon-gravity studies (e.g., Bonneville et al., 2017) and for generally understanding the trade-offs between gravity and muons for resolving subsurface targets.

### 3.4 Data Acquisition and Analysis

The gravity data acquisition, processing, and corrections are detailed in Roy et al. (2017), while the muon detector specifications, muon data acquisition, and first muon tomographic image are described by Guardincerri et al. (2017). Here we summarize each.

Gravity measurements are made using a Lacoste and Romberg model D meter at 27 stations on top of the mesa, 6 stations on the hillslope within Los Alamos Canyon and 27 stations within the TA-41 tunnel (Figure 3.2a). Precise positioning for the survey is provided by a Topcon RTK GPS system for stations with access.
to satellite signals, and by using differential levelling for stations within the TA-41 tunnel. Gravity measurements are first corrected for solid Earth tides, instrument drift, and latitude, after which the free-air correction is applied. The corrected data are then recast as anomalies relative to a chosen base-station, which we take to be station HV-1: the station at the lowest elevation located outside and slightly below the tunnel entrance.

After applying the free-air correction, the remaining variability is attributed to anomalous masses that are either above our reference station and below the mesa surface—but still within our target volume—or below both the reference station and our target volume. To account for the effect of deep-seated anomalous masses below our target volume, we remove a long-wavelength regional trend obtained by fitting a planar surface to a regional database of Bouguer gravity. Upon removal of this planar trend, and after accounting for the uncertainty arising from detrending, the remaining variability (Figure 3.2a) is what we study, as it may be primarily attributed to the gravitational effect of the mass of the rocks of the mesa above the tunnel (Roy et al., 2017). We discretize our model domain using a digital elevation model (DEM) generated from LIDAR data (Figure 3.3a) and optimize the discretization following Roy et al. (2017) so that it is finer near the gravity station locations and coarser in the surrounding area (Figure 3.3b). Each voxel or cell is chosen to be small enough so that any further reduction in size leads to a < 1% change in the predicted gravity at any station (Roy et al., 2017). This produces 7575 cells consisting of a more finely-discretised inner region (4247 voxels, 16 × 16 × 16 m³ each) surrounded by a coarser outer region (3328 cells, 200 × 200 × 16 m³ each) (Figure 3.3b). Muon intensity measurements are obtained using the Los Alamos National Laboratory mini muon tracker (MMT), constructed of orthogonal drift tubes that track muons with an angular resolution of 2.5 milliradians across a surface of 1.4 m². Measurements are made at four locations within the TA-41 tunnel and one on the surface. In this paper, for reasons described below, we confine our attention to the two innermost
Figure 3.2: (a) Measured gravity anomalies relative to the base station (red), plotted at each station location. (b,c) Three-dimensional histograms of muon range (as defined in the text) vs zenith angle \( \theta \) and azimuthal angle \( \phi \) (\( \phi = 0 \) corresponds to due North) for the (b) outer detector position (40.2 m from tunnel entrance), and (c) inner detector position (77.82 m from tunnel entrance).

MMT locations in the tunnel (corresponding to positions 0 and 1 in Guardincerri et al. (2017)). We track the number of muons arriving per unit time for a given \((\theta, \phi)\) direction, where \(\theta\) is the polar angle spanning the detector acceptance range (0° to 45°) and \(\phi\) is the azimuthal angle measured from North (−180° to 180°). In the tunnel stations, the countrate of incoming muons along a given \((\theta, \phi)\) direction, \(N^{\text{in}}(\theta, \phi)\), is inversely proportional to the integrated density along that path (Leo, 1987).

Due to the relatively coarse nature of our model discretization and the steep cliff face, muon data from the outer two tunnel detector positions (located 4.95 m and...
31.57 m from the tunnel entrance) do not contain any crossing paths within our model geometry. Since this is the condition necessary for 3D tomography, we only consider data from the two innermost detector positions (located 40.2 m and 77.82 m from the tunnel entrance) in our inversion (green stars; Figure 3.4a). The muon countrates at each detector position are binned using $3^\circ$ angular steps for both $\theta$ and $\phi$, giving $16 \times 120$ bins per detector and $16 \times 120 \times 2 = 3840$ total angular bins for the muon measurements (Figure 3.2b; 3.2c). The central axis through each angular bin may be then thought of as a single muon trajectory. Within the $j^{th}$ angular bin, the ratio
Figure 3.4: (a) Model representation of muon trajectories for a selection of \((\theta, \phi)\) bins; colours correspond to values of \(\theta\) in (b). Inset shows orientation of \((\theta, \phi)\) with respect to model geometry. Yellow dots mark voxels with crossing muon trajectories. Green stars mark the two tunnel detector positions used in this paper, while red stars mark the two positions used in the experiment, but omitted from this paper for reasons described in the text. (b) Muon range (density-length) as a function of attenuation (from Eq. 3.1) for four values of \(\theta\). Each dot represents a binned \(\phi\) value for each \(\theta\) at both detector positions and coincides with a different attenuation. The range is shown for each trajectory, separated by detector position, in Figure 3.2b; 3.2c. Here the error in the range is only shown for \(\theta = 45^\circ\) for clarity.

of the countrate inside the tunnel to outside, \(N_{in}^j(\theta, \phi)/N_{out}^j(\theta, \phi)\), gives a measure of muon attenuation for that trajectory (Figure 3.5). Because we take a ratio here, we can, as a first-order approximation, neglect the effects of the acceptance pattern of the detector. Following Guardincerri et al. (2017), we calculate the minimum kinetic energy, \(E_{min}^j\), that a muon must possess to reach a tunnel detector location without being absorbed using the relation

\[
\frac{N_{in}^j(\theta, \phi)}{N_{out}^j(\theta, \phi)} = \frac{\int_{E_{min}^j}^{\infty} f(\theta_j, E)dE}{\int_{0}^{\infty} f(\theta_j, E)dE} \tag{3.1}
\]

where \(f(\theta_j, E)\) is the muon probability distribution at the Earth’s surface for kinetic energy \(E\) and polar angle \(\theta_j\). We generate \(f(\theta_j, E)\) at the 2100 m elevation of Los Alamos using the Cosmic Ray Shower (CRY) Monte Carlo Software (Hagmann
Chapter 3. Joint Inversion of Cosmic-Ray Muon & Gravity Data

Figure 3.5: Raw muon attenuation data, coloured with respect to $-\ln(N_{j}^{in}/N_{j}^{out})$ using 3-degree angle bins for the detector at (a) the inner position with $4.926 \times 10^6$ total number of counts, observed for 564 hours, and at (b) the outer position with $4.585 \times 10^6$ total number of counts, observed for 392 hours. Muon count rate measurements made on the mesa surface were obtained over 348 hours for a total of $1.405 \times 10^8$ counts. The dotted line in (a) shows the direction of due North, which corresponds to the direction containing the most rock and, therefore, the highest muon attenuation.

For each $E_{j}^{min}$, the path-integrated density (the density-length) or range, $R_j$, of a particle incident on a tunnel detector position is interpolated from a set of empirical calculations for standard rock, where $\rho = 2650$ kg/m$^3$ and $Z/A = 0.50$ (Groom et al., 2001). For each angular bin (trajectory), we then obtain a range (Figure 3.2b; 3.2c) corresponding to a measure of muon attenuation, $N_{j}^{in}/N_{j}^{out}$ (Figure 3.4b). Angular bins with lower values of $N_{j}^{in}/N_{j}^{out}$ (i.e., higher values of $E_{j}^{min}$) correspond to greater attenuation and therefore a higher range.
3.5 Joint Inversion

We use a Bayesian joint inversion technique with the combined gravity-muon dataset, similar to the methods in Nishiyama et al. (2014) and Rosas-Carbajal et al. (2017). The observed gravity at the \( j \)th station, \( g_j \), is a function of the \( i \)th matter volume density, \( \rho_i \), through \( g_j = \sum_i G_{ij} \rho_i \), where \( G_{ij} \) is a matrix representing the gravity effect of the \( i \)th voxel at the \( j \)th gravity station. This matrix is generated based on the Nagy et al. (2000) theoretical model for the gravity effect of a rectangular prism of uniform density. For the muons, there is the similar relationship: \( R_j = \sum_i L_{ij} \rho_i \), where \( L_{ij} \) represents the path length that a muon travels in the \( i \)th voxel along its central trajectory in the \( j \)th angular bin (Nishiyama et al., 2014; Guardincerri et al., 2017). The combined gravity and muon observations are given by \( \mathbf{d}_{\text{obs}} = [\mathbf{g} \ \mathbf{R}]^T \), while the predicted joint observations \( \mathbf{d} \) are a linear function of the joint inversion operator \( \mathbf{J} \) and density \( \rho \):

\[
\mathbf{d} = \sum_i \begin{bmatrix} G_{ij} \\ L_{ij} \end{bmatrix} \rho_i = \sum_i J_{ij} \rho_i = \mathbf{J} \rho
\]  

(3.2)

This system is inverted for density \( \rho \) by evaluating:

\[
\rho = \rho_0 + (\mathbf{J}^T \mathbf{C}_d^{-1} \mathbf{J} + \mathbf{C}_m^{-1})^{-1} \mathbf{J}^T \mathbf{C}_m^{-1} (\mathbf{d}_{\text{obs}} - \mathbf{d}_{\text{pred}})
\]  

(3.3)

where \( \rho_0 \) is the starting (prior) density and \( \mathbf{C}_d, \mathbf{C}_m \) are the data and model covariances, respectively. We solve this iteratively using regularization in a quasi-Newton method that minimizes the \( L^2 \)-norm of \( \rho_0 - \rho \), subject to the model covariance \( \mathbf{C}_m \) (Tarantola, 2005).

Following Roy et al. (2017), this variability in \( \rho \) is constrained by defining \( \mathbf{C}_m \) with two types of exponential smoothing: an isotropic function parameterised by a single correlation length \( \lambda \) (Eq. 3.4) and an anisotropic function with two correlation lengths, \( \lambda_{xy} \) and \( \lambda_z \), that allow for independent depth and lateral variations (Eq. 3.5). For a pair of voxels \((i, j)\), the possible smoothing functions are:

\[
C_{m,ij} = \sigma^2 \rho \exp \left( -\frac{r_{ij}}{\lambda} \right)
\]  

(3.4)
and

\[ C_{m,ij} = \sigma^2 \exp \left( -\frac{r_{ij}^2}{\lambda_{xy}} - \frac{z_{ij}}{\lambda_z} \right) \]  

(3.5)

where \( r_{ij} \) is the distance between voxels and \( \sigma^2 \) is the variance of the model. Assuming that all muon and gravity observational data are uncorrelated, the data covariance \( C_d \) is assumed to be an \((n_g + n_\mu) \times (n_g + n_\mu)\) diagonal matrix, where \( n_g = 60, n_\mu = 3840 \) are the number of gravity and muon measurements. The error in \( R_j \) (Figure 3.4b) is calculated with standard error propagation, assuming a Poisson distribution for muon count rates \( N^{\text{in}}_j \) and \( N^{\text{out}}_j \). The correlated systematic errors are not taken into consideration in this study, since they have yet to be evaluated. The gravity error is taken to be the individual station errors, which average about 0.200 mgals. These errors vary from \( \sim 0.020 \) to 0.800 mgals, as some of our stations (such as those on the mesa) suffer from more cultural noise than others. Finally, in order to account for the orders of magnitude differences between the numerical values of muon and gravity observations, we normalize both \( d_{\text{obs}} \) and \( C_d \) by the errors of these two datasets. This normalization methodology may be improved upon in the future by taking into account the full error matrix for the muon data.

### 3.6 Results

We present a suite of inversions that progress from least restrictive (3.6.1), where we ignore the known geology by using a uniform starting density and an isotropic smoothing function (Eqn. 3.4), to models that successively incorporate prior knowledge to a greater degree. In our most restrictive model (3.6.3), we use a stratified prior density structure based on geology (Figure 3.1c) and anisotropic smoothing (Eqn 3.5) to favour a flat-lying, regionally-extensive anomalous structure. As an intermediate case, we consider models with an (unrealistic) uniform starting density but favour layered structure by using anisotropic smoothing (3.6.2).
each case, to determine goodness-of-fit, we find the parameter values that keep $\chi_2^2/\text{DOF} = ((d_{\text{obs}} - d_{\text{pred}})^T C_a^{-1} (d_{\text{obs}} - d_{\text{pred}}))/(n_g + n_\mu) \approx 1$, while minimizing $\sum_{i=1}^{N_{\text{vox}}} |\rho_i - \rho_0|^2$.

In conjunction with the joint inversions of the experimental observations, we explore a set of inversions using synthetic gravity and muon datasets. These are generated from a forward-calculation using a series of vertically stratified, synthetic density structures, with the aim of determining the ability of the inversion, along with the optimal parameters found using our experimental data, to adequately resolve each one. These synthetic tests are described in Appendix C.1 and we note that the optimal parameters for inversion results presented below are substantiated by the results of the synthetic tests.

### 3.6.1 Case 1: Uniform density, isotropic smoothing

Here we start with a uniform density $\rho_0$ and the isotropic smoothing kernel of Eq. 3.4. We vary $\rho_0$ from 1400-2100 kg/m$^3$ based on the expected density range for ash-flow tuffs and consistent with those obtained from the borehole observations (Figure 3.1c). For this density range, we optimise $\sigma_\rho$ and $\lambda$ from Eq. 3.4 based on the goodness-of-fit criteria (Figure 3.6c). The optimal parameters found are: $\rho_0 = 2000$ kg/m$^3$, $\sigma_\rho = 125$ kg/m$^3$, and $\lambda = 826$ m (with $\chi_2^2/\text{DOF} = 0.981$ and $\sum_{i=1}^{N_{\text{vox}}} |\rho_i - \rho_0| = 7.87 \times 10^3$ kg/m$^3$). The model generated using these optimal parameters (Figure 3.6a; Figure 3.9a) shows a thinner, higher-density body ($\rho_{\text{avg}} = 2185$ kg/m$^3$) within a lower-density region, but does not produce a three-layered structure. The low-density region ($\rho_{\text{avg}} = 2040$ kg/m$^3$) is higher in density than that expected for Qbt3 and Qbt1 ($\rho_3 = 1800$ kg/m$^3$ and $\rho_1 = 1400$ kg/m$^3$).
Chapter 3. Joint Inversion of Cosmic-Ray Muon & Gravity Data

3.6.2 Case 2: Uniform density, anisotropic smoothing

As in Case 1, $\rho_0$ is taken to be uniform, however we now utilise an anisotropic smoothing kernel (Eq. 3.5). This allows us to include knowledge that the study area comprises rocks that are nearly flat-lying and regionally extensive (Figure 3.1b) by permitting $\lambda_{xy} > \lambda_z$. Note that this does not impose any conditions on the number of layers or their density variations. The optimal parameter values of $\rho_0$, $\sigma_\rho$, and $\lambda$ from Case 1 are used in our parameter search on $\lambda_{xy}$, $\lambda_z$. We take the correlation length $\lambda = 826$ m and optimise for $c_{xy}$, $c_z$ such that $\lambda_{xy} = \lambda c_{xy}$ and $\lambda_z = \lambda c_z$. The optimal model for Case 1 corresponds to $c_{xy} = c_z = 1$ and our parameter search ($c_{xy} \in [1, 10^3]$, $c_z \in [10^{-3}, 1]$) departs from this base model while keeping $c_{xy} > c_z$. The chosen ranges allow for a maximum difference of $\lambda_{xy} = 10^6 \lambda_z$.

The optimal parameters from this search are found to be $c_{xy} = 2.8$ and $c_z = 0.15$ (with $\chi_d^2/DOF = 1.00$ and $\sum_{i=1}^{N_{\text{vox}}} |\rho_i - \rho_0| = 1.45 \times 10^4 \text{ kg/m}^3$) so that $\lambda_{xy} = 2.29 \times 10^3$ m and $\lambda_z = 1.29 \times 10^2$ m (Figure 3.7, inset). The inverted structure shows that, by introducing anisotropy between $\lambda_{xy}$ and $\lambda_z$, we readily recover a 3-layered model wherein a high-density layer is sandwiched between two lower-density layers (Figure 3.7; Figure 3.9b). The high-density anomaly is resolved as nearly flat-lying with $\rho_{\text{avg}} = 2240 \text{ kg/m}^3$, which is higher than expected for the Qbt2 layer, but in keeping with the results found in Roy et al. (2017). The lowest layer has $\rho_{\text{avg}} = 1850 \text{ kg/m}^3$ and the highest layer $\rho_{\text{avg}} = 2050 \text{ kg/m}^3$, which are both higher than expected for Qbt1 and Qbt3 respectively.

3.6.3 Case 3: Layered density, anisotropic smoothing

We start with a layered prior using the average densities from the borehole observations (Figure 3.1c) and choose anisotropic smoothing. This is the most restrictive model, as it relies on the highest amount of information about the known geology
(average density and attitude of layers). We use the optimal values from Cases 1 and 2, however instead of a constant value of $\rho_0$, we formulate a layered prior using the values of $\rho_3$, $\rho_2$, $\rho_1$ inferred for Qbt3, Qbt2, and Qbt1 (Figure 3.8, inset). This inversion (Figure 3.8; Figure 3.9c) preserves the initial layered density structure, however, $\rho_{\text{avg}}$ for each layer is $\sim$200 kg/m$^3$ higher than those given by the geologic prior. The goodness-of-fit results yield $\chi^2_{\text{df}}$/DOF = 1.074, which is poorer than that of the best fit model, Case 2. We note that this value may not, to some, be considered “significantly poorer,” thus, while we still consider Case 2 to be our best-fit model, we cannot discard the results of Case 3 entirely.

### 3.7 Discussion and Conclusions

Our results show that by combining gravity and muon measurements, the differing spatial sensitivities in the datasets (Jourde et al., 2015) may be exploited, allowing us to recover a realistic density structure using minimal geologic (prior) constraints. Our best-fit model (Case 2; Figure 3.7; Figure 3.9b) recovers a three-layered structure with reasonable density variations starting from a nonphysical constant density prior but subject to anisotropic smoothing. By comparison, if all geologic knowledge is included (a three-layered prior with anisotropic smoothing), the recovered structure is similar (Case 3; Figure 3.8; Figure 3.9c) but with a worse fit to the data. While it is not unexpected that a less restrictive model would fit the data better in a Bayesian inversion, it is instructive to note that recovery of a three-layer structure with fewer constraints in Case 2 (Figure 3.7; Figure 3.9b) depends upon our unique surface-subsurface experimental geometry and the methodological robustness of a gravity-muon joint inversion.

We note that the average predicted densities for all inversions are higher than those from the borehole logs (Figure 3.1c), however the variations in this data are calculated to be 210 kg/m$^3$ for Qbt3 and 270 kg/m$^3$ for Qbt2 and Qbt1, thus our
results fall within expected variability. Discrepancies may also be attributed to lateral density variations within the strata, since the borehole data were collected 4-5 km southeast of our study area (in well CdV-16-2(i), from Broxton & Vaniman (2005)). We also note that our analysis of the muon data does not account for range straggling, which diverts them from the assumed linear trajectories. This would have the effect of increasing their apparent attenuation along an assumed linear trajectory, leading to an overestimation of the density, as observed. In future work, we plan to improve our simulation of muon attenuation to include a fuller treatment of such effects.

Since our target anomaly is expected to be a regionally extensive, nearly flat-lying layer—and that the region of greatest muon sensitivity is a narrow, nearly vertical swath cutting across it—imaging this anomaly should depend on information from gravity data to adequately resolve its lateral continuity. The sensitivity of the inversion to gravity is both determined by the gravity measurement errors, and by the respective sizes of the two datasets (Gelman et al., 2013). In our inversions above, the variability and accuracy of our gravity measurement errors are what drive this sensitivity, however in Appendix C.2, we explore how a relative weighting based on the vastly different sizes of the two datasets can also be a powerful tool for providing this sensitivity, particularly if the gravity measurements do not have the requisite precision. Recovery of the expected three-layer structure in the inversions above, as well as in our inversions in Appendix C.2, shows that the gravity measurements are crucial for sensing the lateral extent of our target anomaly.

We recognize that in our study we used both surface and subsurface gravity measurements, whereas in a typical setting (e.g., at a volcano), subsurface measurements are not likely to be available. In such areas, surface gravity measurements complement information from muons, especially outside of the region of crossing muon paths. To determine now much of our successful imaging of a flat-lying anomaly is controlled by surface vs. subsurface gravity, we conducted a number of hypothetical tests where
we ignore any information from the tunnel gravity. Our results (Appendix C.2) show that information from the muons and the surface gravity measurements alone can recover a three-layered structure, but only if the gravity and muon data are given a relative equal weighting in the inversion. The amplitude of the density anomaly, however, is better predicted if we include both surface and subsurface gravity measurements, thus we conclude that the gravity data within the tunnel is needed for resolving density in the lowest Qbt1 layer. When the gravity data are used alone (i.e., without muons), the inversions only produce the requisite structure when a layered prior is used in conjunction with anisotropic smoothing (as demonstrated in Roy et al. (2017), which uses a similar model to our most restrictive Case 3). It is thus clear that the combination of muons and gravity is a significantly more powerful tool than either dataset by itself, provided the inversions are adequately sensitive to information from the gravity data.

In summary, our unique experimental geometry (a target volume sandwiched above and below by muon and gravity measurements), together with the inversion methodology outlined above, successfully resolves an independently characterized density structure using minimal prior geologic information. We demonstrate that the combination of gravity and muon data from surface and subsurface measurements—paired with the introduction of anisotropic smoothing in the model covariance—could be a powerful method for imaging strata in reservoirs or aquifers, provided a more rigorous treatment of the muon data (e.g., Oláh et al., 2018) accounts for the water content of the rock within the study area. Although this study targets a static density structure, our approach may be extended to time-dependent density variations, such as those associated with fluid flow.
Figure 3.6: (a) Case 1: Slice plot of inversion results (view looking NE; DEM of view shown in (b) for clarification, with gravity station locations marked in red) using uniform prior density and isotropic smoothing for the best-fit parameters: $\rho_0 = 2000$ kg/m$^3$, $\sigma_\rho = 125$ kg/m$^3$, and $\lambda = 826$ m (white star in (c)). Inset in (a) shows histogram of normalised data residuals. Parameter search results for $\chi^2_{\text{dof}}/\text{DOF}$ are shown by contours of (c), where black dots mark the parameter values of the search.
Figure 3.7: Case 2: Inversion results for uniform prior density, anisotropic smoothing yielding the best-fit model. Insets show histogram of normalised data residuals and contour plot of parameter search results for $\chi^2_{\text{red}}/\text{DOF}$, with $\lambda_{xy} = 2.29 \times 10^3$, $\lambda_z = 1.29 \times 10^2$ (white star).
Figure 3.8: Case 3: Inversion results for anisotropic smoothing with $\lambda_{xy} = 2.29 \times 10^3$, $\lambda_z = 1.29 \times 10^2$ (as in Case 2) and a layered, geologic prior. Insets show histogram of normalised data residuals and the colour-coded densities of the geologic prior used in the inversion.
Figure 3.9: Two-dimensional representation (slice at Northing = $5.4109 \times 10^5$ m, as in Figures 3.6-3.8) of the inversion results for (a) Case 1, (b) Case 2, and (c) Case 3.
Chapter 4

Machine Learning with Cosmic-Ray Muon & Gravity Data

The contents of this chapter have been submitted to the *Geophysical Journal International* and are currently under review. First author is K. Cosburn, with co-authors M. Roy, and R. Nishiyama.

4.1 Abstract

The ability to accurately and reliably obtain images of shallow subsurface anomalies within the Earth is important for hazard monitoring and a fundamental understanding of many geologic structures, such as volcanic edifices. In recent years, the use of machine learning (ML) as a novel approach for addressing complex problems in the geosciences has gained increasing attention. Here we present an ML-based inversion method to integrate cosmic-ray muon and gravity datasets for shallow subsurface density imaging at a volcano. Starting with an ensemble of random density anomalies, we use physics-based forward calculations to find the corresponding set of expected gravity and muon attenuation observations. Given a large enough en-
semblence of synthetic density patterns and observations, the ML algorithm is trained to recognize the expected spatial relations within the synthetic input-output pairs, learning the inherent physical relationships between them. Once trained, the ML algorithm can then interpolate the best-fit anomalous pattern given data that it has never seen before, such as those obtained from field measurements. We test the validity of our ML algorithm using field data from the Showa-Shinzan lava dome (Mount Usu, Japan) and show that our model produces results consistent with those obtained using a more traditional Bayesian joint inversion. Our results are similar to the previously published inversion, and suggest that the Showa-Shinzan lava dome consists of a relatively high-density \( (2200 \text{ – } 2400 \text{ kg/m}^3) \) cylindrical anomaly, about 300 m in diameter. Adding noise to synthetic training and testing datasets shows that, as expected, the ML algorithm is most robust in areas of high sensitivity, as determined by the forward kernels. A key benefit to using an ML approach to an inversion is its relative ease of use and, once trained, the speed at which a prediction on field data can be obtained. We anticipate that this latter factor has the potential to facilitate continuous monitoring of a volcano, provided the sampling of data is also fast (which may not be the case for muography, but could be the case for gravimetry and other methods).

### 4.2 Introduction

The merging of disparate datasets in geophysical imaging is imperative for developing an accurate image of subsurface structure and the physical processes that affect it. In particular, there is appreciable need to better understand volcanoes and their processes, from both hazard assessment and fundamental science perspectives. Time-dependent imaging is the ideal in volcano science, however, our ability to understand even static structures allows us to image magma beneath an edifice, which can help predict eruptions or other hazards. In terms of this study, we are interested in
imaging subsurface density, which is an important physical property and indicator of magma or hydrothermal fluids within a volcanic structure.

Muon radiography, or “muography,” is a novel method that can illuminate such subsurface density structure, especially when used alongside traditional gravity measurements. It is a technique that has been used increasingly in recent years to study the interior of volcanoes (e.g., Okubo & Tanaka, 2012; Nishiyama et al., 2014, 2017b; Rosas-Carbajal et al., 2017; Oláh et al., 2018; Tioukov et al., 2019) and pyramids (e.g., Alvarez et al., 1970; Morishima et al., 2017), as well as to investigate regions of interest within reservoirs (e.g., Pieczonka et al., 2020), mines (e.g., Schouten, 2018), tunnels (e.g., Guardincerri et al., 2017; Thompson et al., 2020), and glaciers (Nishiyama et al., 2017a, 2019). For a thorough overview of muography, we point the reader to the review paper of Bonechi et al. (2020), however, we summarize the methodology here. Cosmic-ray muons are produced in the upper atmosphere due to the decay of pions and kaons and arrive at the surface of the Earth with a well-documented and broad energy spectrum (e.g., Shukla & Sankrith, 2016). Depending on their angle of incidence, incoming flux, and energy, they are able to penetrate several hundreds of metres through rock. The muography method relies on measurements of the absorption of cosmic-ray muons that pass through matter and are attenuated according to the density (and other material properties) of the structure they pass through (Lechmann et al., 2021). Our study works primarily with the density-length, also known as the range, opacity, or column density, which unlike seismic wavespeeds, for example, is linearly proportional to density and the path length of the structure it passes through. From the muon flux that arrives at some detector pointed towards a volcano of interest, such as the peak of the Showa-Shinzan lava dome (Figure 4.1), the range through the material along a line-of-sight trajectory can be derived (Groom et al., 2001). For this study, we are interested in a joint inversion of muon range with gravity data, which also depends linearly on subsurface density. One can use Newton’s Law of Gravitation to relate a gravity
measurement at the surface to some density anomaly in the subsurface. This is a non-unique problem, however, and combining gravity with muon measurements can mitigate the issue in regions of the volcano traversed by muon trajectories that fall within the acceptance angles of the detector (Jourde et al., 2015; Barnoud et al., 2019).

The advent of machine learning (ML) in the geosciences has brought forth a new way to approach many geophysical problems, including subsurface imaging. The most commonly used ML algorithms tend to be convolutional neural networks, due to their ability to work with image data. ML has been used in such studies as determining subsurface resistivity from the inversion of electromagnetic data (Puzyrev, 2019; Puzyrev & Swidinsky, 2021), in lieu of a standard cross-gradient objective technique to map seismic velocities to magnetotelluric resistivities (Guo et al., 2020), for forecasting eruptions using muographs (Nomura et al., 2020), for image reconstruction in muon imaging problems (Yang et al., 2018), and for determining volcano deformation...
from unwrapped InSAR images (Anantrasirichai et al., 2018). Other geophysical inversion studies have used deep neural networks for seismic reflectivity inversion (Kim & Nakata, 2018) and invertible neural networks for seismic one-dimensional surface wave dispersion inversion and two-dimensional travel time tomography (Zhang & Curtis, 2021). Our study departs from these other machine learning studies in three main ways: (1) the application of a broadly accessible machine learning model based on a well-established machine learning algorithm from Sci-kit Learn; (2) a joint inversion on more than one type of dataset using machine learning alone; and (3) the prediction of a full 3D density structure of the Showa-Shinzan lava dome, as opposed to 1D or 2D subsurface predictions.

Point number (2) is particularly important, as we hope to provide a proof-of-concept study that machine learning can be used effectively with disparate datasets for predicting subsurface structure. By utilising the data collected in Nishiyama et al. (2014, 2017b), we compare our machine learning approach with results from a traditional Bayesian joint inversion (Nishiyama et al., 2017b). When the observations are used in our machine learning algorithm—trained on an ensemble of possible density structures and their corresponding synthetic gravity and muon measurements—we find comparable results to a traditional inversion. This study demonstrates that ML may be used as a powerful tool to jointly invert disparate datasets. Unlike a traditional inversion, however, the relationships between model parameters (e.g., density) and the datasets to be assimilated (e.g., muon range and gravity), do not need to be analytically known. Instead, ML may be used to recognize patterns and relate a set of model parameters to observations for which there may be no known forward kernel, such as in the case of heatflow or carbon dioxide emissions. In these cases, the methodology of the current study would not be appropriate, however, and a different ML approach should be considered. This study sets the stage for future applications to time-dependent variations, for example, relating temporally varying subsurface density to changes in gravity and muon attenuation for better
Figure 4.2: (a) Gravity data from Nishiyama et al. (2014) used in this study overlain on a contour plot of Mount Showa-Shinzan. Each of the 35 stations is represented by a circle and colour-coded by its free-air anomaly value. The muon detector is indicated as a grey square and the base station is outlined in red. The acceptance angle of the muon detector is shown by red lines and is what defines the target region. (b) Image of muography data from Nishiyama et al. (2017b) and its error for each trajectory allocated to 30 discrete bins.

understanding volcanoes and their processes.

4.3 Machine Learning Model for Showa-Shinzan

4.3.1 Data and Geological Setting

The present work aims at applying a machine learning approach to published muography (Nishiyama et al., 2017b) and gravity (Nishiyama et al., 2014) data at Mount Showa-Shinzan. Mount Showa-Shinzan is a parasitic lava dome of the Mount Usu volcano, formed between 1943 and 1945 and located at the southern rim of the Toya Caldera, Hokkaido, Japan. Its present topography consists of the dome itself—approximately 300 m in diameter with a maximum elevation of 398 m—surrounded
by a plateau. A digital elevation model (DEM) was extracted for the dome (Figure 4.1b), published by the Geospatial Information Authority of Japan, with spatial resolution of 4.6 m in the east-west direction and 6.2 m in the north-south direction.

Gravity measurements are described by Nishiyama et al. (2014) and we summarize the data collection process here. The gravity survey used in this study was performed in 2011 using a LaCoste and Romberg relative gravimeter G-875. The gravity survey covered approximately 600×600 m of area via 35 gravity stations. The free-air gravity anomaly is used for this study, relative to the local free-air anomaly of the base station (BS, Figure 4.2a). Muography measurements are described by Nishiyama et al. (2017b) and we summarize the process here. Muography measurements were taken so that the field of view of the detector encapsulated both the plateau and the dome of Mount Showa-Shinzan. The muon detector used in this study had a multilayer architecture of OPERA-type (Nakamura et al., 2006) emulsion films (20 films in total) with 1-mm-thick lead plates inserted in between. This muon detector was installed at 189 m altitude and 500 m west of the dome summit (Figure 4.1b; MD, Figure 4.2a) and exposed for 168 days between November 2011 and May 2012. After exposure, the films were developed and the recorded tracks were read using a special microscope. The reconstructed tracks were then assigned to 30 rectangular bins and converted to the average density along the radial direction from the detector (Figure 4.2b).

### 4.3.2 Domain Discretization

In this study, we settled on a discretization of the volume of Mt. Showa-Shinzan of 50×50×40 m elements (Figure 4.3). We tried smaller element sizes (50×50×20 m), however, we found that the machine learning algorithm did not do as well at training on datasets with a larger number of output (Appendix D.1). We also tried larger elements sizes (100×100×20 m) but found that this did not adequately
discretize the topography in the xy-directions. An element size of $50 \times 50 \times 40$ m, therefore, was found to be the ideal discretization between coarse enough so that the machine learning algorithm still performs well and fine enough to adequately image the shape of subsurface anomalies. Within this discretized domain—and since the muon range is not sensitive outside the acceptance angles of the detector—we choose a sub-domain encapsulating these acceptance angles as our target region (Figure 4.2a; inset of Figure 4.3). This target region contains 217 elements, as opposed to the entire 1839, thus reducing the number of elements for which our machine learning algorithm needs to predict a density and improving its performance.
4.3.3 Forward Calculation and Traditional Inversion

With the domain of interest discretized into $n$ elements with density $\rho_j$ for $j = 1, 2, ..., n$ we can write the gravity and muon data as a linear function of the density values. We are interested in the vertical component of the gravity anomaly at each $i^{th}$ gravity station, which can be denoted as:

$$\Delta g_i = \Delta g(x_i, y_i, z_i) = \sum_{j=1}^{n} G_{ij} \rho_j$$

(4.1)

where $G_{ij}$ is a matrix representing the effects of the vertical gravity from the $j^{th}$ voxel at the $i^{th}$ gravity station, generated based on the Nagy et al. (2000) theoretical model for the gravity effect from a rectangular prism of uniform density at a point. The density-length, or range, which has been derived from muon flux measurements, can be expressed in a similar relationship:

$$R_i = \sum_{j=1}^{n} L_{ij} \rho_j$$

(4.2)

where $L_{ij}$ is the path length of the $i^{th}$ muon trajectory through $j^{th}$ element in the discretized domain. In the form of a linear equation, eqs. 4.1 and 4.2 can be combined as:

$$d = \begin{bmatrix} g_i \\ R_i \end{bmatrix} = \sum_j \begin{bmatrix} G_{ij} \\ L_{ij} \end{bmatrix} \rho_j = \sum_j J_{ij} \rho_j = J \rho$$

(4.3)

In a traditional Bayesian joint inversion (e.g., Nishiyama et al., 2014, 2017b), this system in Eq. 4.3 can be inverted using a prior and subject to smoothness constraints (e.g., Nishiyama et al., 2014, 2017b; Guardincerri et al., 2017; Roy et al., 2017; Cosburn et al., 2019).

4.3.4 Machine Learning Joint Inversion

After trying several machine learning algorithms (including deep neural networks) we settled on the Sci-kit Learn Histogram-based Gradient Boosting Regression Tree
algorithm (https://scikit-learn.org/) for its ease-of-use and robustness. Part of our motivation in this study was to provide a workflow that could be easy-to-use by anyone interested in applying a machine learning approach to an inversion at their volcano of interest. One drawback of a deep neural network, for example, is the computational knowledge needed to build a model that performs well. By using a well-established routine that has been rigorously developed by computer scientists, we are able to harness its power and simplicity to focus more on the science and less on the programmatic aspect of our study.

The Histogram-based Gradient Boosting Regression Tree algorithm is a faster alternative to a regular gradient boosting algorithm, thus making it ideal for large (greater than $10^5$) sized datasets. Since our dataset is on the order of $10^7$, we have chosen the histogram-based version of this algorithm. Boosting in this context refers to a general ensemble technique that improves a model by the sequential addition of smaller models (in this case, decision trees with a given depth and breadth) to the ensemble, where the addition of subsequent models corrects the performance of previous models (Figure 4.4). Gradient boosting is an extension of boosting algorithms that allows for the optimization of arbitrary loss functions, which means that it can be used for either regression or classification problems. For reasons explained in more detail below, since our outputs are a continuous value, we treat this as a regression problem. Since the output for the Scikit-Learn HistGradientBoostingRegressor object is one-dimensional, we utilize a wrapper function, MultiOutputRegressor (also a part of the Scikit-Learn library), to cast the number of our outputs from 1 to 217 (the number of elements in the target region; Section 4.3.2). This function fits a separate instance of the HistGradientBoostingRegressor object to each output and is responsible for the reason we are able to “shuffle” the input data and still obtain the same results as the case where the data is not shuffled (Appendix D.2). The combination of the HistGradientBoostingRegressor and MultiOutputRegressor functions is what is referred to as the “ML algorithm” throughout this paper (Figure 4.5).
Synthetic Data Generation

The first step of the data generation process is to produce a large ensemble of synthetic input-output pairs \((\rho, d)\) to train on our machine learning model. We generate on the order of \(10^7\) possible perturbation patterns, and for the \(n^{th}\) such pattern, \(\rho_{n,j}\), we generate a stacked array \(d_n = [g_{n,i} R_{n,i}]^T\) of gravity and range measurements via the forward calculation of Eq. 4.3 (Figure 4.5). For these \(\rho_{n,j}\) perturbation patterns, within the target region (217 elements), we consider perturbations that may be spatially isolated (single element) or “clumped” (two or more connected elements). In this part of the workflow, a user defines some background density \(\rho_{bg}\) and a range of densities that are representative of possible density anomalies for their region of interest \(\rho_{rng}\). They also define the total number of independent perturbation patterns, \(N_p\), the maximum number of clumps per perturbation pattern, \(N_c\), and the maximum number of elements per clump, \(N_e\). We then design it so that the elements in each clump (a random number \(\leq N_e\)) are all connected together by their faces and each clump (a random number \(\leq N_c\)) does not overlap with other, existing clumps. Each clump, then, gets assigned a random density value picked from \(\rho_{rng}\), while the remaining elements that have not been perturbed get set to \(\rho_{bg}\). In this way, we can build up a set of \(N_p\) random, independent perturbation patterns and perform the forward calculation from Eq. 4.3 to obtain their corresponding synthetic gravity and range measurements.

In the case of our study, we set \(\rho_{bg} = 1700 \text{ kg/m}^3\), and choose a range of density anomaly values between 1200–2600 kg/m\(^3\). The results below are somewhat sensitive to the choice of background density and we explore other values in Appendix D.1. Since we wish to emulate the sensitivity checkerboard tests of Nishiyama et al. (2014, 2017b), which have density anomalies alternating in 200 × 200 m squares in the xy-directions, we choose for our training and testing datasets (for 50 × 50 × 40 m elements) values of \(N_c = 20\) and \(N_e = 5\). This means that, at minimum, there is one perturbed element within the entire discretized domain (217 elements), and
at maximum roughly 50% of the domain could be perturbed. The outputs $\rho_{n,j}$ are, in real space, an array describing a three-dimensional density structure, however, our outputs to the machine learning algorithm are a stretched-out, 1-D array of this 3-D structure (Figure 4.5). In this manner, each entry in the 1-D array contains the density value at the corresponding place in three dimensions. These values can be equal to, higher than, or lower than the background density, depending on the random selection generated for each perturbation pattern (Figure 4.6).

**Training and Testing**

In the case of our study, since each of the $\rho_{n,j}$ outputs are a continuous value, we treat this as a regression problem and choose a half least squares loss function when training and testing. Out of the three loss functions available from Sklearn, we find that this choice performs the best at generalizing to data its never seen before. The half least squares loss function is expressed mathematically as:

$$L(y, \hat{y}) = \frac{1}{2} \sum_j (y_j - \hat{y}_j)^2$$

(4.4)

where $y$ is the “true” output value and $\hat{y}$ is the predicted. The entire data set consists of $N_p = 3 \times 10^7$ independent perturbation pattern outputs and their corresponding gravity-muon synthetic data inputs. We also look at the results of a smaller training/testing dataset with $N_p = 3 \times 10^5$, and find that the ML algorithm produces a comparable structure to that obtained with a larger dataset size (Appendix D.1). We reserve 20% of this $N_p = 3 \times 10^7$ for testing the generalizability of our machine learning model. The goodness-of-fit score is decided by the coefficient of determination, $R^2$, which is given by one minus the ratio of the total sum of squares to the residual sum of squares, or:

$$R^2 = 1 - \frac{\sum(y - \hat{y})^2}{\sum(y - \bar{y})^2}$$

(4.5)

where, as before, $y$ is the “true” output value (with mean $\bar{y}$) and $\hat{y}$ is the predicted output value ($R^2 = 1$ indicates a perfect fit to that data).
Using this criterion for goodness-of-fit on the reserved test dataset we then perform a hyperparameter search to determine the optimal learning rate, maximum number of iterations, and maximum number of leaf nodes (Figure 4.7; Table 4.1). Throughout each run, we choose an L2-regularization, which helps prevent overfitting. Also to prevent over-fitting (and to improve computational time) we choose values of the maximum number of leaf nodes and maximum number of iterations (trees) such that we do not see an improvement in $R^2$ above these values. For maximum number of leaf nodes, we choose a value of 256 since it performs better than the smaller values of 64 and 128 and because values of 512 and 1024 do not show an improvement in generalizability (Figure 4.7a). Similarly, for maximum number of iterations (trees), we choose 500, since a value of 1000 does not show an improvement in $R^2$ and smaller values do not perform as well (Figure 4.7b). We do not treat the maximum number of histogram bins of the input as a hyperparameter and keep it at the default value of 255 (the maximum value allowed by Scikit-Learn).

After determining the optimal hyperparameters (using a smaller dataset on the order of $N_p = 10^5$) we run our training on 80% of the $N_p = 3 \times 10^7$ sized dataset. Training and testing is done in parallel on 32 CPU cores and takes around 30 hours to run. Once we determine that our algorithm performs well on the 20% test dataset (by obtaining an $R^2$ score greater than 0.5), we plug in the field data collected at the Showa-Shinzan lava dome and find a resulting density formation that may represent the true structure. The benefit of an ML approach is that, once the data generation and training/testing has been completed, the time taken for calculation on a single field dataset is on the order of seconds.

**Adding Noise**

We examine the effects of how errors in the data might affect the ways in which our machine learning algorithm learns and the resulting density structure. The errors in
Chapter 4. Machine Learning with Cosmic-Ray Muon & Gravity Data

<table>
<thead>
<tr>
<th>Hyperparameter</th>
<th>Optimal Value</th>
<th>Values Tried</th>
</tr>
</thead>
<tbody>
<tr>
<td>Loss Function</td>
<td>Half Least Squares</td>
<td>Half Least Squares, Half Poisson</td>
</tr>
<tr>
<td>Learning Rate</td>
<td>0.12</td>
<td>Between 0.1 and 0.3</td>
</tr>
<tr>
<td>Max. No. Leaf Nodes</td>
<td>256</td>
<td>[64, 128, 256, 512, 1024]</td>
</tr>
<tr>
<td>Max. No. Trees</td>
<td>500</td>
<td>[50, 100, 200, 500, 1000]</td>
</tr>
<tr>
<td>L2-Regularization</td>
<td>20</td>
<td>[0, 10, 20, 30]</td>
</tr>
<tr>
<td>Dataset Size</td>
<td>$3 \times 10^7$</td>
<td>$3 \times 10^5, 3 \times 10^7$</td>
</tr>
</tbody>
</table>

Table 4.1: Hyperparameter values used in this study. The search for optimal learning rate, maximum number of leaf nodes, and maximum number of trees (or iterations) can be found in Figure 4.7.

The gravity measurements are primarily due to inaccuracies in the DEM and have a total value of around $\sigma_g = 300\mu$gals per station (Nishiyama et al., 2014, 2017b). The muon range data error, $\sigma_{\mu,i}$, varies for each trajectory and has an absolute maximum of around $1.1 \times 10^5$ kg/m$^2$, minimum of $3.5 \times 10^3$ kg/m$^2$, and average of $4.3 \times 10^4$ kg/m$^2$, according to the muon study performed in Nishiyama et al. (2017b). For the uncertainty estimation, we choose an amount of error to randomly add or subtract from the synthetic training and testing datasets before training our ML algorithm on the noisy data (note that, unlike a least-squares inversion, we do not need to assume Gaussian noise). Here the gravity error is given by a flat value $\Delta g = \pm 0.5\sigma_g$ for each measurement and the muon error is a trajectory-wise value $\Delta R_i = \pm 0.5\sigma_{\mu,i}$. By adding or subtracting a constant value of error, we have designed a simple noise model that can explore maximum versus minimum (error-free) variations in the predicted density structure. In this way, we would expect that the “true” density structure lies somewhere between the two predictions, with and without noise.
4.4 Results

We first consider the results without measurement error and train our algorithm with noise-free synthetic data. Without added noise, the machine learning algorithm obtained an $R^2 = 0.70$, based on the reserve 20% testing dataset. If the algorithm is given the observations from Nishiyama et al. (2017b), it predicts a density structure that is in agreement with that obtained using a Bayesian joint inversion. Both the ML and traditional inversion methods suggest that the Showa-Shinzan lava dome is underlain by a cylindrical density anomaly, around 300 m in diameter and consisting of lava around $2200 - 2400 \text{ kg/m}^3$ in average density (Figure 4.8, 4.10). Next we add error to the training and testing synthetic datasets according to the noise model described above (Section 4.3.4). Here we find a lower $R^2 = 0.55$, however the overall predicted structure does not change much, and still indicates to an anomaly around 300 m in diameter and consisting of lava around $2200 - 2400 \text{ kg/m}^3$ in average density (Figure 4.9, 4.10).

To look at the spatial distribution of uncertainty in our density estimates, we define an uncertainty index as the absolute value of the difference between the noise-free and noisy models (Figure 4.11), normalized by the maximum value. We also define a theoretical sensitivity measurement by first normalizing the $G_{ij}$ and $L_{ij}$ kernels by their respective maximum values. We then take the $j^{th}$ column of a stacked array of these two normalized kernels and take the Euclidean norm, $s_j$, of it and then normalize, again, by the maximum of $s_j$. In this way we can get a number between zero and one that is indicative of the sensitivity of the combined $G_{ij}$ and $L_{ij}$ kernels on an element-by-element basis. Comparing these two measurements, we find that there is an inverse relationship between the theoretical sensitivity and our uncertainty estimate (Figure 4.12), indicating that the machine learning model is more robust to the effects of noise in areas of high sensitivity. This implies that the ML algorithm has uncovered the expected relationships between density and gravity.
and muon range measurements.

4.5 Discussion

In this study, we tested the feasibility of using a machine learning approach to jointly interpret gravity and muon data to determine subsurface density structure at a volcano. By comparing our results to those obtained in a more traditional Bayesian joint inversion (Figure 4.10), we have been able to provide a proof-of-concept that disparate datasets can be combined effectively and easily in a machine learning context in order to image subsurface features, such as density anomalies, in a volcanic setting. We focused on gravity and muon data in this study for two reasons: we have previous results from Nishiyama et al. (2017b) and we have known forward kernels that are easy to compute and map each data set linearly to density. Our approach, however, is generalizable and would be well suited for data without known forward kernels. In this study, we opted for a well-established Histogram Gradient Boosting algorithm (Figure 4.4), developed by Sci-kit Learn (a popular python machine learning library), for its ability to work robustly, with less programmatic knowledge needed to set up the problem than with a deep neural network. In this way, we offer a user-accessible approach for geophysical inverse problems. The benefit of our approach, also, is that we need not consider the order of the input data. In fact, since the machine learning algorithm simply learns a mapping from inputs to outputs, we can “shuffle” the input data and still get out a comparable density structure (Appendix D.2) to that obtained with un-shuffled data. The machine learning algorithm in this case just learns a different mapping between data and density structure, whilst preserving the overall empirical relationships between these two quantities.

In this study we consider two main cases: (1) the case where the data is perfect and contains no noise; and (2) the case where the data is maximally noisy. In this sense we have both a best-case scenario (Figure 4.8) and a worst-case scenario (Figure
4.9). In case (1) we find that the Showa-Shinzan lava dome appears to be underlain by a cylindrical density anomaly around 300 m in diameter with average density around $2200 - 2400 \text{ kg/m}^3$. This is consistent with the results from Nishiyama et al. (2017b), as well as Nemoto et al. (1957), Nishida & Miyajima (1984), and Goto & Johmori (2014). In the case of Nemoto et al. (1957), the anomaly is predicted to increase in diameter with depth, however, we notice that the anomaly stays approximately the same diameter down to 240 m a.s.l. Because of the muon detector’s acceptance angle and the coarse discretization of our model domain (Section 4.3.2), we do not image further down from this level, so we are not sure whether the anomaly increases in size below this elevation, based on this study alone. The main discrepancy between our study and Nishiyama et al. (2017b) is the existence of a low-density structure to the East near the plateau part of the region at an elevation of 240 m. This is a region of low sensitivity (Figure 4.12c) and we anticipate that this would be better resolved if more gravity measurements or another muon detector could be placed in this area. Unlike Nakamura & Mori (1949), Nemoto et al. (1957), and Nishida & Miyajima (1984), we do not notice a high-density anomaly protruding from the East. This was considered one reason that the plateau could have been formed, however, our results are more consistent with Komazawa et al. (2010) and Goto & Johmori (2014), which also do not predict such an anomaly, and instead suggest that the plateau was formed via uplift of the ground and lateral migration from the dome region of rock and sediment due to the intrusion of dacite magma. It should be noted that recent work (Takeo et al., 2022) using a passive seismic campaign of Mount Showa-Shinzan inferred densities from S-wave velocities that were significantly lower than that our predicted density for the dome. They attribute this to potential fracturing within the dome, which could reduce shear strength without a large decrease in density.

Finally for case (2), where we look at the effect of error in the data, we note that our noisy model also resolves an approximately 300 m in diameter, higher-than-average density anomaly. The predicted density is around $2200 - 2400 \text{ kg/m}^3$, 
which is consistent with the results for case (1). The anomaly is not identical to the case without noise, however, but this is not surprising, as the areas where we see a large differences tend to be areas of low model sensitivity. This inverse relationship is seen most clearly in regions of the topography towards the West (closest to the muon detector) and at the peak of the dome (Figure 4.12). By comparing our results with and without noise to a mapping of the sensitivity kernels we discover that the machine learning algorithm is able to predict the kernels themselves, simply by mapping inputs to outputs, where the inputs were generated based on a suite of random, yet representative, density outputs (Section 4.3.4). This is a rather remarkable finding and proves the power of the simplicity of the machine learning algorithm to be able to recover the physics behind the forward calculation, despite the fact that $N_p = 3 \times 10^7$ is just a subset of the entire space of possible input-output pairs for this domain.

It would be worthwhile to see in future work if similar results can be obtained for nonlinear forward calculations or for cases where the forward calculation is not known. For the latter case, one is limited by the amount observational data available for a particular volcano. For this study, we are able to generate on the order of $10^7$ input-output pairs using the forward calculations, however, for cases where we don’t know the forward calculation, this amount of data is likely unfeasible to obtain and researchers will need to be creative about data augmentation or creating models that generalize across multiple volcanoes with similar characteristics. This being said, since the actual calculation on field data takes just seconds (after the algorithm is trained), continuous monitoring of a volcano may be facilitated if the experimental set-up is kept constant at all snapshots in time, allowing for imaging time-dependent variations. Of course, the time it takes to collect data is a major limiting factor, as well as how topography might change as magma or other fluids move within an edifice (which would require a new discretization of the topography and re-training of the ML algorithm). We note that in determining the optimal
discretization and background density, we have been guided by the results of the previous joint inversion. For cases where such previous results are not available, we suggest a thorough exploration of model parameters and the judicious use of any *a priori* information to determine the best model. Sensitivity to model parameters (Appendix D.1) is a key limitation to using an ML approach alone as a black box method, as well as the computational intensiveness needed to run the workflow from start to finish. Nevertheless, we believe that machine learning can be considered a powerful tool for studying shallow subsurface density at volcanoes.

### 4.6 Conclusion

A three-dimensional prediction of the density structure of the Showa-Shinzan lava dome (Mount Usu, Japan) was obtained using a machine learning approach to the inversion of gravity and cosmic-ray muon datasets. The results of this study show that the volcano is underlain by a cylindrical high-density anomaly, roughly 300 m in diameter and with an average density between $2200 - 2400 \text{ kg/m}^3$, which is consistent with previous results from a Bayesian joint inversion. By using a well-established gradient-boosting algorithm as our machine learning model, we are able to reduce the amount of expertise needed to approach this problem from a machine learning perspective, and show the power and simplicity inherent in it. We suggest that machine learning can be a relatively straightforward and effective tool to use alongside traditional inversions for subsurface density prediction. This study lays the ground for future work with other kinds of disparate datasets than gravity and muon range, and other subsurface properties than density.
Chapter 4. Machine Learning with Cosmic-Ray Muon & Gravity Data

Figure 4.4: (a) Cartoon showing how a decision tree for regression splits along leaf nodes to extract hidden features from data. Here the percentages count the red versus white dots in each area delineated by the value of a leaf node split. We note that the dimension of the data space here is just 2, while it is 65 (35 gravity stations plus 30 binned muon trajectories) in this study. The maximum number of leaf nodes is depicted as 5 in this cartoon for simplicity, while it could be much larger for a larger data space (see results of hyperparameter search; Figure 4.7, Table 4.1). (b) Cartoon showing how the Gradient Boosting Regression Tree algorithm, together with the wrapper function MultiOutputRegressor, adds a decision tree at each iteration and fits to the multi-output data better after each iteration. The fit to the data is illustrated in the inset as a fit to a single element perturbation, where the black dots are the actual density values and the blue triangles, green circles, red squares, and yellow stars represent the predicted values as subsequent trees are added to the model. One can see that as the iteration number increases, the error between actual and predicted value decreases with each iteration. The output space in this cartoon is 5, while it is 217 in terms of this study (the size of the number of elements in the target region). What is not shown is that each added tree may have a variable number of leaves (here they are all depicted as being the same).
Chapter 4. Machine Learning with Cosmic-Ray Muon & Gravity Data

Figure 4.5: Workflow for the data generation and machine learning training/testing process. Here the “ML algorithm” represents the combination of the Scikit-Learn HistGradientBoostingRegressor functions together with MultiOutputRegressor function. The inputs are a stack of synthetic gravity and muon range measurements that are generated via the forward calculation (Section 4.3.3) performed on a representative suite of density models with varying values and patterns (Figure 4.6). The output is multidimensional (equal to 217, the number of elements in the target region), motivating the use of the MultiOutputRegressor wrapper function. We reserve 20% of our entire data set for testing the generalizability of the ML algorithm, which is determined via the coefficient of determination, or $R^2$ (Section 4.3.4).
Figure 4.6: Four examples of possible synthetic perturbation patterns within the target region (Figure 4.3) used as the output to the machine learning algorithm. A forward calculation (Section 4.3.3) is performed on these density structures in order to obtain the synthetic inputs (a stack of gravity and muon range). Shown here are 3D structures, but the outputs to the ML algorithm are 1D arrays as illustrated in Figure 4.5.
Figure 4.7: (a) Results of the hyperparameter search to find the optimal learning rate and maximum number of leaf nodes using the values from Table 4.1. The search was performed on a representative, but smaller data set, thus the $R^2$ score does not reflect the results using the larger dataset (it improves with dataset size). Given these results, we set the optimal learning rate to 0.12 and maximum number of leaf nodes to 256. We note that the $R^2$ does not improve above 256 leaf nodes, thus we choose this value to improve computation time and prevent over-fitting. It must also be noted that a previous grid-search (not shown here) found that $R^2$ falls rapidly for learning rates below 0.1. (b) Results of hyperparameter search to find optimal maximum number of iterations (trees). With the parameters from (a) and the values from Table 4.1, we take 500 to be the optimal value for the maximum number of iterations, since the $R^2$ does not improve above it.
Figure 4.8: Results of machine learning inversion (trained on synthetic data with no added noise) using data from Nishiyama et al. (2017b) for the Showa-Shinzan lava dome. (a) The entire 3D structure of the density prediction within the target region. (b) Slice plots of the density structure at 40 m intervals showing the high-density anomaly through each layer.
Figure 4.9: Results of machine learning inversion (trained on synthetic data with added noise; Section 4.3.4) using data from Nishiyama et al. (2017b) for the Showa-Shinzan lava dome. (a) The entire 3D structure of the density prediction within the target region. (b) Slice plots of the density structure at 40 m intervals showing that overall shape of the high-density anomaly does not change much with the addition of noise.
Figure 4.10: 2D slice plots through the vertical at 40 m increase in elevation shown for case with no noise (left) and noise (centre), as compared in same colour scheme with the results from Nishiyama et al. (2017b) (right).
Figure 4.11: (a) The entire 3D structure within the target region of the absolute difference between the density predictions with and without added noise in the synthetic data in the target region (inset: histogram of residuals for this difference). (b) Slice plots showing this difference at 40 m intervals.
Figure 4.12: (a) Normalized difference between the models with and without noise (the “uncertainty index”) compared with the normalized sensitivity kernel (see Section 4.4 for definition) for each element in the target region (see Figure 4.2a and inset of Figure 4.3 for definition of the target region). Shaded regions, in particular, show an inverse relationship between sensitivity and uncertainty. For context, lower element numbers correspond to elements that are more West and at the bottom of the topography (closer to the muon detector), while the higher element numbers correspond to elements in the peak region of the topography. (b) Plot showing the inversely proportional trend between uncertainty index and kernel sensitivity, indicating that the machine learning algorithm is more robust to noise in areas of higher sensitivity. (c) 3D plot of the normalized $G_{ij}$, $L_{ij}$ sensitivity shown within the target region. Darker colours correspond to areas of higher sensitivity, which is seen in elements at the bottom of the topography that are more towards the West, as well as in the peak region.
Chapter 5

Conclusion

5.0.1 Summary

This dissertation has focused on studying volcanoes and their processes from an interdisciplinary, physics-based perspective. In particular, it has aimed to provide meaningful insight into how a volcanic edifice is formed and imaging (either directly or indirectly) what goes on underneath. Three studies that illuminate methods for pursuing such problems have been presented and a summary of each study with their main conclusions are as follows:

Topographic Form of Stratovolcanoes

This research study involves the development of a first-order approximation to stratovolcano shape and an examination of how deviations may be interpreted. Stratovolcanoes are notable for many reasons, but their iconic shape is an almost universal aspect of them that make them so recognizably different from their shield volcano cousins. The internal and external processes that make up a stratovolcano edifice are numerous and complex, and insight into how their shape is formed can give us a better understanding of these processes, which could have implications for bettering hazard assessment. For this study, we fit $N = 190$ stratovolcano edifice profiles to
Chapter 5. Conclusion

an axisymmetric ideal mathematical form based on a stochastic piling of eruptive products. We find that this form fits many stratovolcanoes well, however, there are pronounced deviations from either axisymmetry or from the ideal form entirely (or both). Deviations from the form itself come in two main forms, concave or protuberant, and we focus our study on the latter. We posit that these protuberant deviations could possibly arise from the internal emplacement of magma due to distributions of dykes below an edifice. Several case studies are explored in order to validate this interpretation, with the volcanoes Karymsky in Kamchatka and Concepción in Nicaragua. We find good agreement that the protuberant deviations found in these two volcanoes are due to subsurface dykes.

For cases where there is a deviation from axisymmetry (but overall a good fit to the ideal form), we hypothesize that this might arise from the effects of a regional trend on how eruptive products such as lava pile up. For this we find that the profile of Merapi in Java could potentially be explained by such an effect. Finally, we look at basal ellipticity and any correlations between this measurement and how well a stratovolcano profile adheres to the ideal form. We find that a volcano is more likely to deviate from its ideal form as basal ellipticity increases, suggesting a correlation between the two that may be linked to where the volcano is along its evolutionary timeline. The categorization of volcanoes in this study are presented as a way to hypothesize about the constructive processes that build the topographic form of a stratovolcano. Misfits and deviations, then, provide insight into the various surface (e.g., erosion) and subsurface (e.g., magma emplacement) processes that affect an edifice. In particular, we focus on subsurface processes and present our methodology as a way of systematically studying these effects—especially useful for volcanoes which are too remote or dangerous to obtain data for directly.

Bayesian Inversion of Gravity and Muon Data

Estimation of subsurface density is important for imaging various geologic structures, such as volcanic edifices. In this study we utilize a Bayesian joint inversion to
combine gravity and cosmic-ray muon data to image the shallow subsurface density of the mesa beneath the town of Los Alamos, New Mexico. The density of this area is well-characterized, with constraints given by borehole logs from a previous study. In this manner, we are able to test the validity of our methodology against previous information and get an understanding of the benefits and limitations. Traditionally, gravity data has been used to image subsurface density, however, muography (muon radiography or tomography) has been used in recent years as a complement to gravity measurements, especially for probing the subsurface density structures beneath volcanoes.

The markedly different spatial sensitivities of gravity and muon data make them a powerful combination, especially when imaging static, flat-lying density anomalies, which could not be resolved using gravity alone. Here we are able to image a regionally extensive, high-density layer of ash-flow tuff. We resolve this nearly flat-lying structure using a unique experimental set-up of surface and subsurface gravity and muon measurements and our results show that with minimal prior (geologic) constraints, this methodology is able to recover the main features of the expected density structure. We show how the use muon and gravity measurements together together with the surface-subsurface experimental design, can be a powerful tool for subsurface imaging. While this study does not explicitly look at imaging a volcano, we posit that this methodology could be used in such a setting, provided borehole measurements could be obtained.

Machine Learning Inversion of Gravity and Muon Data

Our machine learning (ML) methodology approaches the same problem as the Bayesian inversion study, but as a pattern recognition tool, instead. To reiterate, faithful imaging of Earth’s subsurface is a critical task in many important geologic endeavours, such as volcano hazard monitoring. In particular, subsurface density can provide insight into where magmatic or hydrothermal areas of interest exist within a volcanic edifice. This can not only aid in hazard monitoring, but is important for
a fundamental understanding of a volcano’s plumbing structure. This study aims to demonstrate the use of an ML approach for imaging the three-dimensional subsurface density at Mount Showa-Shinzan (a parasitic lava dome of the Mount Usu volcano), by integrating gravity measurements and novel comic-ray muon radiography.

The workflow starts by generating an ensemble of random density anomalies and uses physics-based forward calculations to find the corresponding gravity and muon attenuation observations. The ML algorithm is trained to recognize the expected spatial relations within the synthetic input-output pairs, learning the inherent physical relationships between them (i.e., the forward kernels). Once trained, it is expected that the ML algorithm can then interpolate the best-fit anomalous pattern given data that it has never seen before, such as those from field measurements. To test the validity of our ML algorithm, we use field data from the Showa-Shinzan lava dome, collected in a previous study. We find that our results are consistent with those obtained using a more traditional Bayesian joint inversion and that the lava dome consists mainly of a relatively high-density, cylindrical anomaly.

We then add noise to the synthetic training and testing datasets and show that, as expected, the ML algorithm is most robust in areas of high sensitivity (where sensitivity is determined by the forward kernels). A key benefit to using an ML approach to an inversion, as opposed to Bayesian methods, is its relative ease of use and, once trained, the speed at which a prediction on field data can be obtained. Moreover, by using a well-established gradient-boosting algorithm as our ML model (as opposed to artificial neural networks, for example), we are able to reduce the amount of programmatic expertise needed to approach this problem from a machine learning standpoint. We are further able to show the power and simplicity inherent in this methodology and suggest that machine learning can be a straightforward and efficacious tool to use alongside traditional inversions for subsurface density prediction at volcanoes.
5.0.2 Future Work

The work within this dissertation has set the stage for many potential future studies which may further the field of volcano science. For each of the three studies presented, some ideas on how to improve and build upon each are proposed as follows:

**Topographic Form of Stratovolcanoes**

In this study, we make a series of key assumptions that may not provide the most physical interpretation of real-life conditions. When building a volcanic edifice, we assume that (1) it is only constructed through the piling of lavas and not any other eruptive products, and (2) the rheology of the lava is that of a Bingham fluid, when in actuality its rheology is more complicated (due to the existence of bubbles and crystals). We are able to show that, as a first-order approximation, these are good enough assumptions, however, an avenue of future work would be to move beyond a first-order approximation and incorporate more elaborate modelling of edifice-building behaviour. In particular, it would be beneficial to include the effects of pyroclastic flows and ballistic projectiles on how a stratovolcano is formed. These eruptive products are an important part of the heterogeneity that a stratovolcano edifice contains, which in turn is an important part of the size and shape that they exhibit. Furthermore, we assume that all protuberant deviations are due to subsurface magma intrusions, when they could also be due to non-axisymmetric piling of eruptive products. It would be worthwhile to examine the conditions upon which we might observe protuberant deviations that are due to these external, constructive effects. This would offer a compliment to our current study and other studies that examine the external, destructive effects of erosion on stratovolcano shape (e.g., Karátson et al., 2012).

When considering the subsurface effects of dyke intrusions, we also make some key assumptions. The biggest of these is that the domain into which a dyke is injected behaves as a linear, elastic material and that the injection of each dyke
has independent effects from all others. Also what is missing from our model is the effects of topography on dykes close to the surface (e.g., Rivalta et al., 2015). To circumvent this in our study, we ensure that dykes are not allowed to reach the surface or propagate within a volcanic edifice, however, this is not an entirely physical simplification and more sophisticated numerical techniques (such as finite elements) could potentially model this behaviour more satisfactorily. Moreover, the linear, elastic assumption of the host rock could be better modelled by considering inelastic rheologic behaviour, such as viscoelasticity (e.g., Yamasaki et al., 2018). The depths of our model’s magma chambers (from where our model dykes propagate) are allowed to exist as deep as the brittle-ductile transition zone of the Earth’s crust. It would be beneficial to examine ductile behaviour in how dykes propagate and their subsequent effects on surface deformation (e.g., Kjøll et al., 2019). Also worth examining is how the magma chambers themselves (how they inflate, etc.) effect an edifice’s overall shape, which is another factor we do not consider. Altogether, there are many avenues of potential work that could build upon this study meaningfully and rigorously.

Bayesian Inversion of Gravity and Muon Data

The most pressing question that this study raises is whether or not an application of our surface-subsurface methodology to image a volcano is feasible. One key drawback of this study is the detector that is used to make measurements of muon attenuation. The detector is a drift-tube muon detector made of 576 sealed, aluminum drift tubes arranged in planes (Guardincerri et al., 2017) and is not only large, but requires substantial power to operate. The reason we were able to make subsurface measurements with this detector is that we utilized a decommissioned Cold War tunnel on the Los Alamos National Laboratory property, which was both large enough to contain the detector and also able to supply it with power throughout the duration of the experiment (around 2000 hours total). In the volcanic field, one would need (1) a borehole within which a detector could be placed, and (2) a lightweight
detector that could operate within a borehole and be powered easily with, e.g., solar power for the duration of the study (or one could utilize an emulsion detector, such as in Nishiyama et al. (2017b)). Such borehole detectors have been developed (e.g., Bonneville et al., 2017), however, they have yet to be used to image a volcano. This is likely due to a lack of boreholes near volcanic edifices, however, Mounts Meager and Cayley in Canada might be good contenders, as they have several boreholes near them which were drilled in search of hydrothermal areas of interest in the 1970s and later in 2020 (Klaasen et al., 2021).

Further to this, our study only images a static structure, when we are ultimately interested in time-dependent imaging, especially for its implication for volcano hazard monitoring and imaging fluid motion beneath an edifice. To explore this, a more rigorous treatment of the muon data to account for water content of the rock (e.g., Oláh et al., 2018) must first be implemented, as the current assumption is that the muons are passing through a bulk composition of “standard” rock, with no further heterogeneity. If this is addressed, one may take various “snapshots” in time (e.g., Tanaka, 2015) to get a time-dependent picture of how density changes beneath a volcano, which may allow one to pinpoint eruption locations and forecast potential hazards. Other applications may include imaging the time-dependent migration of hydrothermal areas of interest, which can better our ability to harness such systems as a renewable energy resource. Overall, there is great promise that the muon-gravity, surface-subsurface methodology of this study can be applied to time-dependent imaging of volcanoes, provided some overriding limitations are satisfactorily addressed in future work.

Machine Learning Inversion of Gravity and Muon Data

The advent of machine learning (ML) in the geosciences—and volcanology, in particular—has opened up the door for many studies where this technique may be feasibly used. One study has been presented within this dissertation, however there are a myriad of potential ways in which it can be built and improved upon. As
in the other muon-gravity study with a Bayesian joint inversion, the methodology within this study seeks only to image a static density structure. This methodology, however, is very well-suited for time-dependent monitoring of a volcano. While the training of the ML algorithm may take some time, it is true that, once trained, a single prediction takes on the order of seconds to complete. If all factors of the experimental set-up are thus kept constant (muon detector location and positioning, gravity station location, etc.) this has great potential to facilitate the continuous monitoring of a volcano. To use a Bayesian joint inversion, on the other hand, requires a re-running and optimization of the algorithm each time a “snapshot” of the volcano is taken.

More than with single datasets alone, the combination of disparate datasets is important for obtaining accurate predictions of subsurface phenomena, which can aid in better hazard monitoring of volcanoes (Manga et al., 2017). This study focuses primarily on merging the two disparate datasets of muon range and gravity, however, the beauty of the ML algorithm is its use as a pattern recognition tool. We anticipate that the ML methodology from this study can be extended to easily integrate other measurements where the forward calculation is known, such as with seismic and magnetotelluric data. We focus on subsurface density, but believe that this technique can be extended to other anomalous physical properties, such as temperature, resistivity or conductivity, composition, porosity, and saturation. This would be a very challenging inversion problem, however, it would simply mean a re-working of the input-output pairs of a ML model to include these other factors. One could input a stack of many different datasets and output a stack of many different subsurface properties and we would expect that the ML algorithm would learn the corresponding relationships (such as we saw with gravity and muon range with density). The main drawback would be the computational power needed for such a study, however, with a judicious choice of domain discretization, this could very well be feasible.
Additionally, since the ML algorithm simply maps inputs to outputs, it is possible to combine other kinds of datasets where the forward kernels are unknown, such as with thermal anomalies or carbon dioxide emissions. For this kind of study, one would need enough data to reliably train the ML algorithm, which is the main limitation to this approach. Given that, however, it could be a powerful way of combining datasets that have not previously been combined, or that require a certain level of expertise to jointly interpret. The ML algorithm could automate a difficult problem for scientists, or even find patterns in data that are completely hidden in the eyes of more traditional techniques.

Another key undertaking, too, would be to develop field studies that are catered to a machine learning approach to the analysis. For example, in the muon-gravity study, we anticipate that the ML algorithm would do a better job of resolving the low-density anomaly towards the East that is seen with a previous published Bayesian joint inversion (Nishiyama et al., 2017b) if another muon detector or more gravity stations were placed in that area. Knowing that the sensitivity (forward) kernels dictate how robustly the ML algorithm performs, one can design a study beforehand to capitalize on this fact. It would be worth it in future work to perform an entirely theoretical study of this phenomena, in the spirit of studies such as Barnoud et al. (2019) or Jourde et al. (2015) performed with a Bayesian joint inversion. In all, the ML approach for subsurface imaging at volcanoes is rife with possibilities and, as algorithms improve and computational power increases, we expect volcano scientists will find such an approach increasingly relevant.
Appendices

A Chapter 2: Stochastic Model of Edifice Construction 125

B Chapter 2: Published Mendeley Database 130

C Chapter 3: Synthetic & Weighting Tests 139

D Chapter 4: Different Model Parameters & Shuffling Data 150

E Select Code 156
Appendix A

Chapter 2: Stochastic Model of Edifice Construction

Following Annen et al. (2001) we use the equation of Hulme (1974) for the height of an ideal Bingham fluid undergoing channelized flow on an inclined plane:

\[ h(\alpha) = \frac{sy}{\rho g \sin \alpha} \]

where \( h \) is the flow thickness, \( s_y \) is the yield stress, \( \rho \) is the density, and \( \alpha \) is the slope. For small slopes where \( 1/\sin \alpha \to \infty \) the flow thickness is governed by lateral spreading and is thus determined by its aspect ratio \( k = h/w \) (where \( w \) is the flow width). Explicitly, this is given by:

\[ h(\alpha \leq k) = \frac{s_y}{\rho g k} \]

Used together with the previous equation, we consider \( n = 8000 \) flows and “build” an edifice by piling them on a \( 30 \times 30 \) km grid, with grid spacing of 100 m (Figure A.1 a). After this we calculate an average profile for the 3D edifice (Figure A.1 b) and find the exponential and polynomial model parameters of best fit, \( r_c \) and \( \varphi_c \). For each flow, values of \( k \), \( s_y \), and volume are drawn from a truncated normal distribution with bounds based on the studies of McBirney & Murase (1984); Naranjo et al. (1992);
Lyman et al. (2004); Castruccio et al. (2013, 2014); Arnold et al. (2019). We use a half-normal distribution for the flow starting position from grid centre and consider the effects of varying the spread \( \sigma_r \) of this distribution. We find that the radius of the inner linear region of the edifice, \( r_c \), is controlled by \( \sigma_r \) (Figure A.2a).

We also consider the effects of varying the maximum bounds on \( k, s_y \), and volume (Table A.1). Keeping \( s_y, k \), and \( \sigma_r \) fixed we find that \( r_c \) and \( \varphi_c \) do not change as the maximum bound on volume is increased from \( 2.5 \times 10^6 \) m\(^3\) to \( 10 \times 10^6 \) m\(^3\), but that there is a linear increase in the basal radius \( R_b \). Keeping volume and \( \sigma_r \) fixed we look at the interdependence of \( s_y \) and \( k \) on \( \varphi_c \) (Figure A.1 c) and the ratio of maximum edifice height to basal radius, \( h_m/R_b \) (Figure A.1 d). We compare these values to the average values from the \( N = 190 \) volcanoes in our database (avg. \( \varphi_c = 27.53^o \) and avg. \( h_m/R_b = 0.3 \)) and find that there is a greater sensitivity of these values to variability in the yield strength rather than the aspect ratio.

Finally, we keep all maximum bounds on parameters fixed with the values in Table A.1 and vary the number of flows between 1000 to 4000. Here we find linear relationships between the number of flows with \( \varphi_c \) and the number of flows with the ratio \( h_m/R_b \) (Figure A.2b). This tells us that the trends found by varying the other parameters (Figure A.1 c,d) will not change with the number of flows, thus to obtain the results in Figure A.1 we use \( n = 8000 \) flows across all simulations—a large enough number to ensure an adequate sampling of the stochastic distributions whilst remaining computationally efficient.
<table>
<thead>
<tr>
<th>Parameter</th>
<th>Min. Bound</th>
<th>Max. Bound</th>
</tr>
</thead>
<tbody>
<tr>
<td>yield strength, $s_y$ (kPa)</td>
<td>10</td>
<td>100</td>
</tr>
<tr>
<td>aspect ratio, $k$</td>
<td>0.05</td>
<td>0.35</td>
</tr>
<tr>
<td>volume ($10^6$ m$^3$)</td>
<td>0.5</td>
<td>10</td>
</tr>
<tr>
<td>$\sigma_r$ (m)</td>
<td>100</td>
<td>1000</td>
</tr>
</tbody>
</table>

Table A.1: Values for maximum and minimum bounds on the distribution of flow parameters $s_y$, $k$, volume, and $\sigma_r$ based on McBirney & Murase (1984); Naranjo et al. (1992); Lyman et al. (2004); Castruccio et al. (2013, 2014); Arnold et al. (2019). The minimum bound value is fixed for all simulations, while the maximum bound value listed here represents the overall highest maximum bound across all simulations.
Figure A.1: (a) 3D model edifice built from \( n = 8000 \) flows with the minimum and maximum bounds on parameters from Table A.1. (b) The average profiles obtained from the 3D edifices for various maximum bound values of \( s_y \) and \( k \) below or equal to the maximum bound value given in Table A.1. (c) Contour plot of the absolute difference in \( \varphi_c \) from our model with the average from our data of 27.53° for each maximum bound value of \( s_y \) and \( k \). (d) The same as in (c) but for the ratio \( h_m/R_b \) compared with the average data value of 0.3. The dark blue regions indicate values of \( s_y \) and \( k \) that produce values of \( \varphi_c \) and \( h_m/R_b \) most indicative of real-world values.
Figure A.2: (a) Average profiles for $\sigma_r = 100, 200, 500, 1000$ m with $n = 4000$ flows and the parameter bounds on $s_y, k, \text{and volume from Table A.1}$; (inset) the relationship between $\sigma_r$ and $r_c$ for these curves. (b) Average profiles for 1000 to 4000 flows with the parameter bounds on $s_y, k, \text{and volume from Table A.1 and } \sigma_r = 500 \text{ m}$; (insets) the relationship between number of flows with $h_m/R_b$ and number of flows with $\varphi_c$ for these curves.
Appendix B

Chapter 2: Published Mendeley Database

We study the topographic shape of isolated stratovolcanoes around the world by formulating an ideal mathematical form based on the constructional piling of eruptive products that fits the quintessential stratovolcano shape. For cases when the edifice is more protuberant than our ideal form, we consider this to be the effect of uplift at the surface from a statistical distribution of dykes emanating from a magma reservoir, which cause displacement at the surface. Our aim is to estimate the relative contribution of these dyke intrusions on the height and shape of a volcanic edifice compared to the contribution due to the constructional piling up of eruptive products.

We classify our volcanoes into four categories: (1) those with an axisymmetric good fit to the functional form, (2) those with a non-axisymmetric good fit to the functional form, (3) those with protuberant deviations in all quadrants, and (4) those with protuberant deviations in one or two quadrants. The data presented here gives the parameters of best fit to the ideal form for volcanoes in all categories, figures of the best fits for those in categories (1) and (2), and the parameters of the stochastic dyke model of best fit for those in categories (3) and (4).
Figure B.1: Figures for all volcanoes in category 1 with low ellipticity index.
Figure B.2: Figures for all volcanoes in category 1 with intermediate/high ellipticity index.
Figure B.3: Figures for all volcanoes in category 2 with low ellipticity index.
Figure B.4: Figures for all volcanoes in category 2 with intermediate/high ellipticity index.
## Figure B.5: Parameters of best fit to the ideal form for volcanoes in all categories (page 1).
Figure B.6: Parameters of best fit to the ideal form for volcanoes in all categories (page 2).
## Parameters of best fit to the ideal form for volcanoes in all categories (page 3)

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
<th>Unit</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>Parameter 1</td>
<td>0.123</td>
<td>Unit1</td>
<td>Reference 1</td>
</tr>
<tr>
<td>Parameter 2</td>
<td>45.67</td>
<td>Unit2</td>
<td>Reference 2</td>
</tr>
<tr>
<td>Parameter 3</td>
<td>89.01</td>
<td>Unit3</td>
<td>Reference 3</td>
</tr>
</tbody>
</table>

Figure B.7: Parameters of best fit to the ideal form for volcanoes in all categories (page 3)
<table>
<thead>
<tr>
<th>Name</th>
<th>Region</th>
<th>Volcano Information</th>
<th>Truncated (1000)</th>
<th>Polynorm (trunc) mean age (My)</th>
<th>Exponential (trunc) mean age (My)</th>
<th>Exponential (cum) mean age (My)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mount Etna</td>
<td>Sicily</td>
<td>Etna</td>
<td>3.76</td>
<td>28,125.35</td>
<td>2,197.52</td>
<td>3,687.35</td>
</tr>
<tr>
<td>Mount Etna</td>
<td>Sicily</td>
<td>Eastern Etna</td>
<td>3.76</td>
<td>28,125.35</td>
<td>2,197.52</td>
<td>3,687.35</td>
</tr>
<tr>
<td>Mount Etna</td>
<td>Sicily</td>
<td>Southern Etna</td>
<td>3.76</td>
<td>28,125.35</td>
<td>2,197.52</td>
<td>3,687.35</td>
</tr>
<tr>
<td>Mount Etna</td>
<td>Sicily</td>
<td>Central Etna</td>
<td>3.76</td>
<td>28,125.35</td>
<td>2,197.52</td>
<td>3,687.35</td>
</tr>
<tr>
<td>Mount Etna</td>
<td>Sicily</td>
<td>Southeastern Etna</td>
<td>3.76</td>
<td>28,125.35</td>
<td>2,197.52</td>
<td>3,687.35</td>
</tr>
<tr>
<td>Mount Etna</td>
<td>Sicily</td>
<td>Western Etna</td>
<td>3.76</td>
<td>28,125.35</td>
<td>2,197.52</td>
<td>3,687.35</td>
</tr>
<tr>
<td>Mount Etna</td>
<td>Sicily</td>
<td>Eastern Paricutin</td>
<td>3.76</td>
<td>28,125.35</td>
<td>2,197.52</td>
<td>3,687.35</td>
</tr>
<tr>
<td>Mount Etna</td>
<td>Sicily</td>
<td>Southern Paricutin</td>
<td>3.76</td>
<td>28,125.35</td>
<td>2,197.52</td>
<td>3,687.35</td>
</tr>
<tr>
<td>Mount Etna</td>
<td>Sicily</td>
<td>Central Paricutin</td>
<td>3.76</td>
<td>28,125.35</td>
<td>2,197.52</td>
<td>3,687.35</td>
</tr>
<tr>
<td>Mount Etna</td>
<td>Sicily</td>
<td>Southeastern Paricutin</td>
<td>3.76</td>
<td>28,125.35</td>
<td>2,197.52</td>
<td>3,687.35</td>
</tr>
<tr>
<td>Mount Etna</td>
<td>Sicily</td>
<td>Western Paricutin</td>
<td>3.76</td>
<td>28,125.35</td>
<td>2,197.52</td>
<td>3,687.35</td>
</tr>
<tr>
<td>Mount Etna</td>
<td>Sicily</td>
<td>Eastern Tavurvur</td>
<td>3.76</td>
<td>28,125.35</td>
<td>2,197.52</td>
<td>3,687.35</td>
</tr>
<tr>
<td>Mount Etna</td>
<td>Sicily</td>
<td>Southern Tavurvur</td>
<td>3.76</td>
<td>28,125.35</td>
<td>2,197.52</td>
<td>3,687.35</td>
</tr>
<tr>
<td>Mount Etna</td>
<td>Sicily</td>
<td>Central Tavurvur</td>
<td>3.76</td>
<td>28,125.35</td>
<td>2,197.52</td>
<td>3,687.35</td>
</tr>
<tr>
<td>Mount Etna</td>
<td>Sicily</td>
<td>Southeastern Tavurvur</td>
<td>3.76</td>
<td>28,125.35</td>
<td>2,197.52</td>
<td>3,687.35</td>
</tr>
<tr>
<td>Mount Etna</td>
<td>Sicily</td>
<td>Western Tavurvur</td>
<td>3.76</td>
<td>28,125.35</td>
<td>2,197.52</td>
<td>3,687.35</td>
</tr>
<tr>
<td>Mount Etna</td>
<td>Sicily</td>
<td>Eastern Ubinay</td>
<td>3.76</td>
<td>28,125.35</td>
<td>2,197.52</td>
<td>3,687.35</td>
</tr>
<tr>
<td>Mount Etna</td>
<td>Sicily</td>
<td>Southern Ubinay</td>
<td>3.76</td>
<td>28,125.35</td>
<td>2,197.52</td>
<td>3,687.35</td>
</tr>
<tr>
<td>Mount Etna</td>
<td>Sicily</td>
<td>Central Ubinay</td>
<td>3.76</td>
<td>28,125.35</td>
<td>2,197.52</td>
<td>3,687.35</td>
</tr>
<tr>
<td>Mount Etna</td>
<td>Sicily</td>
<td>Southeastern Ubinay</td>
<td>3.76</td>
<td>28,125.35</td>
<td>2,197.52</td>
<td>3,687.35</td>
</tr>
<tr>
<td>Mount Etna</td>
<td>Sicily</td>
<td>Western Ubinay</td>
<td>3.76</td>
<td>28,125.35</td>
<td>2,197.52</td>
<td>3,687.35</td>
</tr>
<tr>
<td>Mount Etna</td>
<td>Sicily</td>
<td>Eastern Ubinay</td>
<td>3.76</td>
<td>28,125.35</td>
<td>2,197.52</td>
<td>3,687.35</td>
</tr>
<tr>
<td>Mount Etna</td>
<td>Sicily</td>
<td>Southern Ubinay</td>
<td>3.76</td>
<td>28,125.35</td>
<td>2,197.52</td>
<td>3,687.35</td>
</tr>
<tr>
<td>Mount Etna</td>
<td>Sicily</td>
<td>Central Ubinay</td>
<td>3.76</td>
<td>28,125.35</td>
<td>2,197.52</td>
<td>3,687.35</td>
</tr>
<tr>
<td>Mount Etna</td>
<td>Sicily</td>
<td>Southeastern Ubinay</td>
<td>3.76</td>
<td>28,125.35</td>
<td>2,197.52</td>
<td>3,687.35</td>
</tr>
<tr>
<td>Mount Etna</td>
<td>Sicily</td>
<td>Western Ubinay</td>
<td>3.76</td>
<td>28,125.35</td>
<td>2,197.52</td>
<td>3,687.35</td>
</tr>
<tr>
<td>Mount Etna</td>
<td>Sicily</td>
<td>Eastern Ubinay</td>
<td>3.76</td>
<td>28,125.35</td>
<td>2,197.52</td>
<td>3,687.35</td>
</tr>
<tr>
<td>Mount Etna</td>
<td>Sicily</td>
<td>Southern Ubinay</td>
<td>3.76</td>
<td>28,125.35</td>
<td>2,197.52</td>
<td>3,687.35</td>
</tr>
<tr>
<td>Mount Etna</td>
<td>Sicily</td>
<td>Central Ubinay</td>
<td>3.76</td>
<td>28,125.35</td>
<td>2,197.52</td>
<td>3,687.35</td>
</tr>
<tr>
<td>Mount Etna</td>
<td>Sicily</td>
<td>Southeastern Ubinay</td>
<td>3.76</td>
<td>28,125.35</td>
<td>2,197.52</td>
<td>3,687.35</td>
</tr>
<tr>
<td>Mount Etna</td>
<td>Sicily</td>
<td>Western Ubinay</td>
<td>3.76</td>
<td>28,125.35</td>
<td>2,197.52</td>
<td>3,687.35</td>
</tr>
<tr>
<td>Mount Etna</td>
<td>Sicily</td>
<td>Eastern Ubinay</td>
<td>3.76</td>
<td>28,125.35</td>
<td>2,197.52</td>
<td>3,687.35</td>
</tr>
<tr>
<td>Mount Etna</td>
<td>Sicily</td>
<td>Southern Ubinay</td>
<td>3.76</td>
<td>28,125.35</td>
<td>2,197.52</td>
<td>3,687.35</td>
</tr>
<tr>
<td>Mount Etna</td>
<td>Sicily</td>
<td>Central Ubinay</td>
<td>3.76</td>
<td>28,125.35</td>
<td>2,197.52</td>
<td>3,687.35</td>
</tr>
<tr>
<td>Mount Etna</td>
<td>Sicily</td>
<td>Southeastern Ubinay</td>
<td>3.76</td>
<td>28,125.35</td>
<td>2,197.52</td>
<td>3,687.35</td>
</tr>
<tr>
<td>Mount Etna</td>
<td>Sicily</td>
<td>Western Ubinay</td>
<td>3.76</td>
<td>28,125.35</td>
<td>2,197.52</td>
<td>3,687.35</td>
</tr>
<tr>
<td>Mount Etna</td>
<td>Sicily</td>
<td>Eastern Ubinay</td>
<td>3.76</td>
<td>28,125.35</td>
<td>2,197.52</td>
<td>3,687.35</td>
</tr>
<tr>
<td>Mount Etna</td>
<td>Sicily</td>
<td>Southern Ubinay</td>
<td>3.76</td>
<td>28,125.35</td>
<td>2,197.52</td>
<td>3,687.35</td>
</tr>
<tr>
<td>Mount Etna</td>
<td>Sicily</td>
<td>Central Ubinay</td>
<td>3.76</td>
<td>28,125.35</td>
<td>2,197.52</td>
<td>3,687.35</td>
</tr>
<tr>
<td>Mount Etna</td>
<td>Sicily</td>
<td>Southeastern Ubinay</td>
<td>3.76</td>
<td>28,125.35</td>
<td>2,197.52</td>
<td>3,687.35</td>
</tr>
<tr>
<td>Mount Etna</td>
<td>Sicily</td>
<td>Western Ubinay</td>
<td>3.76</td>
<td>28,125.35</td>
<td>2,197.52</td>
<td>3,687.35</td>
</tr>
</tbody>
</table>
Appendix C

Chapter 3: Synthetic & Weighting Tests

C.1 Synthetic Tests

We explore a set of synthetic structures that are based on the density log data from Broxton & Vaniman (2005) and the observed elevation of 2183 m for the Qbt3-2 interface exposed on the cliff-face above the TA-41 tunnel. Therefore, our synthetic structures consist of a high density layer of density 2100 kg/m$^3$ with an upper interface at $d_0 = 2183$ m and variable thickness $h_0$, sandwiched between an upper layer (Qbt3) of density 1800 kg/m$^3$, and a lower layer (Qbt1) of density 1400 kg/m$^3$.

For each synthetic structure, we estimate the gravity effect of the rocks within our study volume by using the discretisation of the terrain into rectangular prisms as in Roy et al. (2017). For an individual $i^{th}$ prism, we calculate the predicted gravity anomaly at a $j^{th}$ station, $g_j$, by summing the gravity effect according to $g_j = \sum_i G_{ij} \rho_i$, where the $G_{ij}$ matrix is generated based on the Nagy et al. (2000) theoretical model for the gravity effect of a rectangular prism of uniform density.
This provides a higher resolution and more accurate estimate of the terrain effects and allows us to prescribe arbitrary density structures in our models (Roy et al., 2017).

We also calculate, for each detector position, the expected muon density-length (or range), $R_j$, given a muon trajectory in a $j^{th}$ angular bin, considered here in three-degree steps for polar angle $\theta \in [0, 45^\circ]$ and azimuthal angle $\phi \in [-180^\circ, 180^\circ]$. This is done according to $R_j = \sum_i L_{ij} \rho_i$, where $L_{ij}$ represents the path length of the $j^{th}$ muon trajectory, corresponding to the angular bin $(\theta_j, \phi_j)$, through the $i^{th}$ prism in our model discretisation (Guardincerri et al., 2017).

For the two generated data sets, $g_{syn}$ and $R_{syn}$, we add random Gaussian noise, $\epsilon$, to account for observational variability (Nishiyama et al., 2014), with standard deviations determined by the variability in the range and gravity measurements. The error in the gravity data is taken as the mean of our experimental gravity data, set to be uniform across all stations. We calculate the error in the muon data assuming Poisson statistics and using the Groom et al. (2001) energy vs range table for standard rock. We construct the joint data set $d_{obs} = [g_{syn} \hspace{1cm} R_{syn}]^T + [\epsilon_g \hspace{1cm} \epsilon_R]^T$ and for the system $d = \sum_i [G_{ij} \hspace{1cm} L_{ij}]^T \rho_i = \sum_i J_{ij} \rho_i = J \rho$, we invert for density according to Tarantola (2005), as described in detail in the main text. Similar to the inversions presented in the text, we utilise a uniform prior and compare results using isotropic and anisotropic smoothing. In each case, we determine the fit based on the minimization of $\Delta \rho = \sqrt{\sum_i N_{vox}(\rho_{inv,i} - \rho_{syn,i})^2}$, the Euclidean norm between the inverted and synthetic density structures.

Using the same optimal parameters as in the inversions with actual data presented in the text, we invert the above synthetic data to recover the predicted structures. We test our inversions for a significant parameter space of $h_0$, but present here synthetic and recovered structures for two cases: $h_0 = 30$ m (Figure C.1), and $h_0 = 50$ m (Figure C.2). Further tests, wherein we shift the elevation of the upper interface of the high-density layer downward to $d_0 = 2171$ m and vary $h_0$, produce similar
results, and we present one case here for $h_0 = 37$ m (Figure C.3). In all cases, isotropic smoothing does not recover the low-high-low three-layer density structure, whereas anisotropic smoothing is confirmed to be the best-fit.

### C.2 Weighting Tests

Considering that our target anomaly is expected to be a regionally extensive, flat-lying layer—and that the muon data are confined a narrow, almost vertical section of our model domain—we foresee that imaging the lateral continuity of such an anomaly through a gravity-muon joint inversion would depend more heavily on information provided by the gravity data, given its long-range sensitivity (Jourde et al., 2015). Since the importance of the gravity data in such an inversion is both determined by the gravity measurement errors and by the respective sizes of the two datasets, we explore how each option provides the needed gravity measurement sensitivity. In the main text, it is driven completely by the gravity measurement errors alone. Due to the inter-station variability and high precision of our measurements, the inversions in Figures 3.6-3.9 do not take into account the fact that the number of muon measurements are far greater than those of the gravity ($n_\mu = 3840$, $n_g = 60$). Here, on the other hand, in order to understand the dependence of the inversions on the number of each type of measurement, we explore a relative weighting between gravity and muon observations, assuming a realistic, but constant (station-independent) error in the gravity. Jourde et al. (2015) provides a thorough theoretical analysis of the sensitivities and relative importance of each of these datasets in a gravity-muon joint inversion, but here we aim to use our unique experimental set-up, wherein we have a well-characterised anomaly to image, to explore these sensitivities using experimental data.

To begin our proof-of-principles exercise, we set the error in our gravity measurements to be a uniform 0.300 mgals across all stations and apply a weighting matrix,
Appendix C. Chapter 3: Synthetic & Weighting Tests

\( \alpha \), to the data residuals, \((d_{\text{obs}} - d_{\text{pred}})\), which allows us to normalise the relative importance each type of data by the number of measurements (Gelman et al., 2013). With the given station-independent gravity errors, and since \( n_{\mu} \gg n_g \), without such a normalisation the inversion is “naturally” driven by the muon observations. The corresponding \((n_g+n_{\mu}) \times (n_g+n_{\mu})\) weighting matrix has the values \([\alpha_g^{1} \ldots \alpha_g^{n_g}, \alpha_{\mu}^{1} \ldots \alpha_{\mu}^{n_{\mu}}]\) along its diagonal and the weighted data residuals are then set to be:

\[
(d_{\text{obs}} - d_{\text{pred}})_{\text{weighted}} = \left( \frac{\|d_{\text{obs}} - d_{\text{pred}}\|}{\|\alpha (d_{\text{obs}} - d_{\text{pred}})\|} \right) (\alpha (d_{\text{obs}} - d_{\text{pred}})) \tag{C.1}
\]

where \( \| \| \) represents the Euclidean L2-norm given by \( \sqrt{\sum_{k=1}^{N} |v_k|^2} \) for some vector \( v \). We then investigate a variable relative weighting, parameterised by \( \kappa \) such that:

\[
\alpha_{\mu} = \kappa \frac{n_g}{n_{\mu}} = \begin{cases} 
\frac{n_g}{n_{\mu}}, & \kappa = 1 \\
1, & \kappa = \frac{n_{\mu}}{n_g} 
\end{cases} \tag{C.2}
\]

We explore models that fix \( \alpha_g = 1 \) and allow \( \kappa \) to range from the “equally-weighted” \( \kappa = 1 \), which normalises for the number of each type of measurement, to the “unweighted” \( \kappa = \frac{n_{\mu}}{n_g} \), which more highly favours the muons. The inversions in the main text correspond to this latter, “unweighted” scheme, whilst here we test our inversions using both equally-weighted and unweighted schemes, as well as those with other values of \( \kappa \). Starting with a uniform prior, anisotropic smoothing, and the optimal parameters from the main text, the results using a station-independent gravity error with equally-weighted and unweighted inversions are shown in Figure C.4. We find that the optimal weighting occurs with the equal weighting scheme \((\kappa = 1)\), where both the reduced chi-squared fit and the recovery of a three-layered structure are better than with the unweighted scheme \((\kappa = \frac{n_{\mu}}{n_g})\). In particular, the area of our model at and below the tunnel (where the lowest-density layer is expected) is not adequately resolved when the muons are given more importance in the inversion (Figure C.4b). This is not surprising, since we are limited to muon trajectories that are incident on the top of the detector between \( \theta = 0^\circ \) and \( 45^\circ \).
Appendix C. Chapter 3: Synthetic & Weighting Tests

The power of the weighting methodology can be seen in the results in Figure C.5, which is the same inversion as in Figure C.4, but without any information from the tunnel gravity stations. From Roy et al. (2017), we know that surface and subsurface gravity measurements (without muons) are able to resolve a flat-lying anomaly, albeit with more model constraints. Here we see that using the muon measurements (surface and subsurface) together with just the surface gravity measurements allows the inversion to sense a three-layered structure, provided there is equal weighting of the datasets in the inversion (Figure C.5a). Comparing Figure C.4a to Figure C.5a, it is clear that the tunnel measurements are important for predicting more accurate densities in the lower Qbt1 region. When we include the tunnel gravity measurements, the predicted density in the lowest layer is $\rho_{\text{avg}} = 1470$ kg/m$^3$ (Figure C.4a), comparable to the borehole logs for this region (Figure 3.1c). Thus it would appear that the subsurface gravity is best for recovering more physical density values in regions without crossing muon trajectories.

We now aim to determine the extent to which the gravity sensitivity, using both surface and subsurface measurements, is important in recovering our three-layered structure. We do this by testing our weighting schemes using the most restrictive model of a layered, geologically-based prior and anisotropic smoothing. From Figure C.6b, it is clear that even when we provide prior information about the predicted layering of our structure—in particular, information about the lower-density region in the area at and below the tunnel—we are unable to recover a three-layered structure when the muons are given more weight in the inversion. We conclude from this that a higher sensitivity to gravity data, given here by an equal weighting in the datasets, is essential for obtaining accurate predictions in areas of a model where there can be no muon tomography. By using this relative equal weighting of the datasets, we leverage the inherent lateral spatial resolution of the gravity to capture the regionally extensive nature of the density anomaly without the need for stricter model constraints. This is seen clearly where both the less restrictive model in Figure
C.4a and the more restrictive model in Figure C.6a recover similar layered structures.

In addition to the tests done with $\kappa = 1$ and $\kappa = \frac{n\mu}{m_g}$ (Figures C.4-C.6), we further investigate how favouring the muons less, relative to the gravity, helps recover the three-layered structure through the exploration of a suite of models using intermediate values of $\kappa \in [1, \frac{n\mu}{m_g}]$. For a uniform prior with isotropic smoothing, we begin with $\kappa = 1$ and find that, as it is increased, we start to lose this structure at $\kappa = 3$. For the model with a uniform prior and anisotropic smoothing there is a built-in tendency for layering, thus in comparison to the case of isotropic smoothing, the layered structure seen with $\kappa = 1$ (Figure C.4a) remains intact until it starts to disappear at a much higher value, at $\kappa = 15$. From these variable-weighting tests, we conclude that spatial structure at volcanoes and other targets where subsurface data acquisition is not possible may be resolved by exploiting the inherently different spatial sensitivities of gravity and muon data through a relative weighting of the datasets in the joint inversion.
Figure C.1: (a) True density structure for anomaly thickness $d_0 = 2183$ m, $h_0 = 30$ m. Inversion results using a uniform prior with (b) anisotropic ($\Delta \rho = 8.93 \times 10^3$), and (c) isotropic smoothing ($\Delta \rho = 1.88 \times 10^4$). The optimal isotropic parameters are $\rho_0 = 2000$ kg/m$^3$, $\sigma_\rho = 125$ kg/m$^3$, and $\lambda = 826$ m (same as in Figure 3.6) and the optimal anisotropic parameters are $\lambda_{xy} = 2.29 \times 10^3$, $\lambda_z = 1.29 \times 10^2$ (same as in Figure 3.7).
Appendix C. Chapter 3: Synthetic & Weighting Tests

Figure C.2: (a) True density structure for anomaly thickness $d_0 = 2183$, $h_0 = 50$ m. Inversion results using a uniform prior with (b) anisotropic ($\Delta \rho = 8.78 \times 10^3$), and (c) isotropic ($\Delta \rho = 1.74 \times 10^4$) smoothing. The optimal isotropic parameters are $\rho_0 = 2000$ kg/m$^3$, $\sigma_\rho = 125$ kg/m$^3$, and $\lambda = 826$ m (same as in Figure 3.6) and the optimal anisotropic parameters are $\lambda_{xy} = 2.29 \times 10^3$, $\lambda_z = 1.29 \times 10^2$ (same as in Figure 3.7).
Figure C.3: (a) True density structure for anomaly thickness $d_0 = 2171$ m, $h_0 = 37$ m. Inversion results using a uniform prior with (b) anisotropic ($\Delta \rho = 8.41 \times 10^3$), and (c) isotropic ($\Delta \rho = 1.67 \times 10^4$) smoothing.
Figure C.4: Inversion results using a uniform prior and anisotropic smoothing with (a) “equal” ($\kappa = 1$) weighting between datasets, and (b) an “unweighted” ($\kappa = \frac{N_\mu}{N_\gamma}$), muon-favoured weighting scheme.

Figure C.5: Results for inversions without tunnel gravity using a uniform prior and anisotropic smoothing with (a) “equal” ($\kappa = 1$) weighting between datasets, and (b) an “unweighted” ($\kappa = \frac{N_\mu}{N_\gamma}$), muon-favoured weighting scheme.
Figure C.6: Inversion results using a layered, geologically-based prior and anisotropic smoothing with (a) “equal” ($\kappa = 1$) weighting between datasets, and (b) an “unweighted” ($\kappa = \frac{N_b}{N_g}$), muon-favoured weighting scheme.
Appendix D

Chapter 4: Different Model Parameters & Shuffling Data

D.1 Effects of Different Model Parameters

D.1.1 Larger and Smaller Domain Discretization

In this section, we look at the results of both a finer and coarser discretization than the $50 \times 50 \times 40$ m elements used in the results of the main text. What we find is that the coarser discretization ($100 \times 100 \times 20$ m elements) does not capture well the topography of the Showa-Shinzan lava dome in the xy-directions within the target region (Figure D.1a). In particular, the lower density seen towards the West is not well resolved, resulting in an anomaly that seems skewed in that direction. There does, however, appear to be a high-density anomaly that extends through it down to 240 m a.s.l. as in the case of the model with $50 \times 50 \times 40$ m elements. In the case of a finer ($50 \times 50 \times 20$ m elements) discretization, we do get a better resolution of the shape of the topography within the target region (Figure D.1b), however, because the machine learning algorithm must perform a prediction on twice as many elements,
Appendix D. Chapter 4: Different Model Parameters & Shuffling Data

Figure D.1: (a) 3D density results using a larger 100 × 100 × 20 m discretization. (b) 3D density results using a smaller 50 × 50 × 20 discretization.

It does not perform quite as well (the $R^2 = 0.59$, as opposed to $R^2 = 0.70$ as in the case of 50 × 50 × 40 m elements). This decrease in model generalization accuracy also adds to the uncertainty in the model prediction on the Nishiyama et al. (2017b) data. For these reasons, we determine that 50 × 50 × 40 m elements is the “sweet spot” of model discretization in terms of trade-off between true representation of the topography and performance of the machine learning algorithm.

### D.1.2 Larger Background Density

The assumption of a background density of 1700 kg/m$^3$ is low, compared to what we know of the bulk density of the Showa-Shinzan lava dome, which has been found to be around 2200 kg/m$^3$ (Komazawa et al., 2010). For this reason, we look at the effects of making our background density higher to 2200 kg/m$^3$. We find that the resulting structure (using an ML algorithm trained on data without added noise) does change from the case with a background density of 1700 kg/m$^3$ (Figure 4.8), however, there still appears to be an anomalous cylindrical, higher-than-average density structure beneath the dome between 2200 – 2400 kg/m$^3$ (Figure D.2). We note that these
results are not as clear as in the case where the background density is 1700 kg/m$^3$, thus we believe that a better background density for our synthetic training/testing dataset is 1700 kg/m$^3$. We also look at an “intermediate” background density of 1900 kg/m$^3$ and find that the results are much more comparable to the case where the background density is 1700 kg/m$^3$ (Figure D.3), suggesting that even a choice of 1900 kg/m$^3$ would be appropriate for this set-up. Overall, however, the choice of background density is a hyperparameter that must be chosen judiciously (perhaps with guidance from a traditional inversion) and explored before settling on the best results.
Appendix D. Chapter 4: Different Model Parameters & Shuffling Data

D.1.3 Smaller Training/Testing Dataset Size

We look at the effect of a smaller training/testing dataset size on how well the ML algorithm learns and the resulting structure that is predicted using the data from Nishiyama et al. (2017b). In the main text, our choice of \( N_p = 3 \times 10^7 \), however, here we look at an \( N_p = 3 \times 10^5 \). We note that this dataset is what is used to perform the hyperparameter search (Figure 4.7). We see that the smaller dataset size does not perform as well, as its \( R^2 = 0.63 \), compared to \( R^2 = 0.70 \) as what is obtained with the larger dataset size. We do see, however, that the resulting density structure of the Showa-Shinzan lava dome is very similar to that obtained with the larger dataset in the case where we consider error-free data (Figure D.4). For this reason, we conclude that, although the ML algorithm doesn’t perform as well on the testing data, it still interprets field data similarly to the case where the \( R^2 \) is

Figure D.3: (a) 3D density results using a higher background density of 1900 kg/m\(^3\). (b) Slice plot of the same density structure shown at 40 m intervals.
higher (when the dataset used for training/testing is larger) and could be considered a viable sized dataset to use in realistic situations. This is good news for those who do not have the time or resources to generate a full $10^7$ sized dataset. On the other hand, we do not want our results to be compromised by too small a training set, which is what has guided us in choosing the dataset size from the main part of the paper.

### D.2 Effects of Shuffling Input Data

To study the effects of how the machine learning algorithm learns the physics of the forward calculation just by mapping inputs to outputs on a representative set of possible perturbation patterns, we look at the effect of shuffling the input data...
(while keeping the output data in the same order). The results of this show an almost identical density structure, when the data for Mount Showa-Shinzan is input into this newly trained machine learning model (Figure D.5). Any deviations thereof, we posit, are most likely due to statistical fluctuations and not due to the actual act of shuffling the data. This shows how the machine learning algorithm can effectively “discover” the physics behind the forward calculation, no matter the order of the gravity and muon range data. This also justifies our use of a machine learning algorithm where there is no dependence on spatial relationships within the flattened 1-D input data (unlike convolutional neural networks, for example).
Appendix E

Select Code

All code in this chapter can be found on Github at https://github.com/katcosburn.

E.1 Okada Solutions for Tensile Dislocation

Coded solutions to the Okada (1992) analytical model for internal deformation due to a tensile crack. Solves for the internal deformation due to a rectangular tensile crack opening in an elastic medium at some point below the free surface (can also solve at the free surface as per Okada (1985)). It outputs both displacements and their derivatives and is used to simulate the subsurface dyke intrusions of Chapter 2. There are two main parts to the code: okada92_kc.m and f_dfxyz.m. It is written for MATLAB.

E.1.1 okada92_kc.m

function [u, v, w, dudx, dvdy, dwdz, dudx, dwdx, dvdz, dwdy, dudy, dvdx] = okada92_kc(x0, y0, z0, x, y, z, L, W, U, phi, delta, mu, nu)
\begin{verbatim}
lambda = 2*mu*nu/(1-2*nu);
alpha = (lambda + mu)/(lambda + 2*mu);

% From the reference system of Fig. 3 (Okada, 1992):
% -> centre at (x0,y0) and rotate by strike angle phi

rotmat = [sin(phi), cos(phi); -cos(phi), sin(phi)];
xyrot = rotmat*[x(:) - x0, y(:) - y0]';
xc = xyrot(1,:);
yc = xyrot(2,:);

%%% Collect fA, fB, fC functions from Table 6

d = [z0 - z, z0 + z];

for i = 1:length(d)
    p = yc*cos(delta) + d(i)*sin(delta);
    q = yc*sin(delta) - d(i)*cos(delta);

    if i == 1
        [fAa, dfAa_dx, dfAa_dy, dfAa_dz] = f_dfxyz('A', xc, p, q, z, delta, alpha);
        [fAb, dfAb_dx, dfAb_dy, dfAb_dz] = f_dfxyz('A', xc, p-W, q, z, delta, alpha);
        [fAc, dfAc_dx, dfAc_dy, dfAc_dz] = f_dfxyz('A', xc-L, p, q, z, delta, alpha);
        [fAd, dfAd_dx, dfAd_dy, dfAd_dz] = f_dfxyz('A', xc-L, p-W, q, z, delta, alpha);
    end
end
\end{verbatim}
Appendix E. Select Code

\[ [f_{Ba}, df_{Ba\_dx}, df_{Ba\_dy}, df_{Ba\_dz}] = f_{\text{dfxyz}}('B', xc, p, q, z, \text{delta}, \text{alpha}); \]
\[ [f_{Bb}, df_{Bb\_dx}, df_{Bb\_dy}, df_{Bb\_dz}] = f_{\text{dfxyz}}('B', xc, p-W, q, z, \text{delta}, \text{alpha}); \]
\[ [f_{Bc}, df_{Bc\_dx}, df_{Bc\_dy}, df_{Bc\_dz}] = f_{\text{dfxyz}}('B', xc-L, p, q, z, \text{delta}, \text{alpha}); \]
\[ [f_{Bd}, df_{Bd\_dx}, df_{Bd\_dy}, df_{Bd\_dz}] = f_{\text{dfxyz}}('B', xc-L, p-W, q, z, \text{delta}, \text{alpha}); \]

\[ [f_{Ca}, df_{Ca\_dx}, df_{Ca\_dy}, df_{Ca\_dz}] = f_{\text{dfxyz}}('C', xc, p, q, z, \text{delta}, \text{alpha}); \]
\[ [f_{Cb}, df_{Cb\_dx}, df_{Cb\_dy}, df_{Cb\_dz}] = f_{\text{dfxyz}}('C', xc, p-W, q, z, \text{delta}, \text{alpha}); \]
\[ [f_{Cc}, df_{Cc\_dx}, df_{Cc\_dy}, df_{Cc\_dz}] = f_{\text{dfxyz}}('C', xc-L, p, q, z, \text{delta}, \text{alpha}); \]
\[ [f_{Cd}, df_{Cd\_dx}, df_{Cd\_dy}, df_{Cd\_dz}] = f_{\text{dfxyz}}('C', xc-L, p-W, q, z, \text{delta}, \text{alpha}); \]

else
\[ [f_{ATa}, df_{ATa\_dx}, df_{ATa\_dy}, df_{ATa\_dz}] = f_{\text{dfxyz}}('A', xc, p, q, -z, \text{delta}, \text{alpha}); \]
\[ [f_{ATb}, df_{ATb\_dx}, df_{ATb\_dy}, df_{ATb\_dz}] = f_{\text{dfxyz}}('A', xc, p-W, q, -z, \text{delta}, \text{alpha}); \]
\[ [f_{ATc}, df_{ATc\_dx}, df_{ATc\_dy}, df_{ATc\_dz}] = f_{\text{dfxyz}}('A', xc-L, p, q, -z, \text{delta}, \text{alpha}); \]
\[ [f_{ATd}, df_{ATd\_dx}, df_{ATd\_dy}, df_{ATd\_dz}] = f_{\text{dfxyz}}('A', xc-L, p-W, q, -z, \text{delta}, \text{alpha}); \]
%% From Equation 15 (Chinnery's expression)

\begin{align*}
A &= f_{Aa} - f_{Ab} - f_{Ac} + f_{Ad}; \\
AT &= f_{ATa} - f_{ATb} - f_{ATc} + f_{ATd}; \\
B &= f_{Ba} - f_{Bb} - f_{Bc} + f_{Bd}; \\
C &= f_{Ca} - f_{Cb} - f_{Cc} + f_{Cd}; \\

\frac{dA}{dx} &= df_{Aa}/dx - df_{Ab}/dx - df_{Ac}/dx + df_{Ad}/dx; \\
\frac{dAT}{dx} &= df_{ATa}/dx - df_{ATb}/dx - df_{ATc}/dx + df_{ATd}/dx; \\
\frac{dB}{dx} &= df_{Ba}/dx - df_{Bb}/dx - df_{Bc}/dx + df_{Bd}/dx; \\
\frac{dC}{dx} &= df_{Ca}/dx - df_{Cb}/dx - df_{Cc}/dx + df_{Cd}/dx; \\

\frac{dA}{dy} &= df_{Aa}/dy - df_{Ab}/dy - df_{Ac}/dy + df_{Ad}/dy; \\
\frac{dAT}{dy} &= df_{ATa}/dy - df_{ATb}/dy - df_{ATc}/dy + df_{ATd}/dy; \\
\frac{dB}{dy} &= df_{Ba}/dy - df_{Bb}/dy - df_{Bc}/dy + df_{Bd}/dy; \\
\frac{dC}{dy} &= df_{Ca}/dy - df_{Cb}/dy - df_{Cc}/dy + df_{Cd}/dy; \\

\frac{dA}{dz} &= df_{Aa}/dz - df_{Ab}/dz - df_{Ac}/dz + df_{Ad}/dz; \\
\frac{dAT}{dz} &= df_{ATa}/dz - df_{ATb}/dz - df_{ATc}/dz + df_{ATd}/dz; \\
\frac{dB}{dz} &= df_{Ba}/dz - df_{Bb}/dz - df_{Bc}/dz + df_{Bd}/dz; \\
\frac{dC}{dz} &= df_{Ca}/dz - df_{Cb}/dz - df_{Cc}/dz + df_{Cd}/dz; \\
\end{align*}

%% Calculate displacements in x, y, z directions using convention in Eqs. 16, 17

\[ u_P = (0.5*U/pi)*(A(1,:) - AT(1,:)) + B(1,:) + z*C(1,:)); \]
\[v_P = \left(0.5 \frac{U}{\pi}\right) \left( (A(2,:) - AT(2,:) + B(2,:) + z \cdot C(2,:)) \cdot \cos(\delta) - (A(3,:) - AT(3,:) + B(3,:) + z \cdot C(3,:)) \cdot \sin(\delta) \right)\]

\[w_P = \left(0.5 \frac{U}{\pi}\right) \left( (A(2,:) - AT(2,:) + B(2,:) - z \cdot C(2,:)) \cdot \sin(\delta) + (A(3,:) - AT(3,:) + B(3,:) - z \cdot C(3,:)) \cdot \cos(\delta) \right)\]

\[uvrot = \text{rotmat}\backslash[u_P; v_P];\]

\[u = \text{reshape}(uvrot(1,:), \text{size}(x));\]

\[v = \text{reshape}(uvrot(2,:), \text{size}(y));\]

\[w = \text{reshape}(w_P, \text{size}(x));\]

%%% Calculate x, y, z derivatives in x, y, z directions using
% convention in Eqs. 16, 17

\[
\text{rotmat_dvs} = [\sin(\phi)^2, \cos(\phi)^2, 0, \sin(\phi) \cdot \cos(\phi), \ldots
\sin(\phi) \cdot \cos(\phi), 0, \sin(\phi) \cdot \cos(\phi), 0, 0, 0; \ldots
\cos(\phi)^2, \sin(\phi)^2, 0, -\sin(\phi) \cdot \cos(\phi), 0, \ldots
-\sin(\phi) \cdot \cos(\phi), 0, 0, 0; \ldots
0, 0, 1, 0, 0, 0, 0, 0, 0; \ldots
-\sin(\phi) \cdot \cos(\phi), \sin(\phi) \cdot \cos(\phi), 0, \ldots
\sin(\phi)^2, 0, -\cos(\phi)^2, 0, 0, 0; \ldots
0, 0, 0, 0, \sin(\phi), 0, \cos(\phi), 0, 0; \ldots
-\sin(\phi) \cdot \cos(\phi), \sin(\phi) \cdot \cos(\phi), 0, \ldots
-\cos(\phi)^2, 0, \sin(\phi)^2, 0, 0, 0; \ldots
0, 0, 0, 0, -\cos(\phi), 0, \sin(\phi), 0, 0; \ldots
0, 0, 0, 0, 0, 0, \sin(\phi), \cos(\phi); \ldots
0, 0, 0, 0, 0, 0, -\cos(\phi), \sin(\phi)];\]

\[dudxP = \left(0.5 \frac{U}{\pi}\right) \left( dA_dx(1,:) - dAT_dx(1,:) + dB_dx(1,:) + \ldots\right)\]


\[ z \cdot dCdx(1,:) \] 
\[ dvdxP = (0.5*U/pi)*((dAdx(2,:) - dATdx(2,:) + dBdx(2,:)) + z*dCdx(2,:))*cos(delta) - (dAdx(3,:) - dATdx(3,:) + dBdx(3,:)) + z*dCdx(3,:))*sin(delta); \]
\[ dwdxP = (0.5*U/pi)*((dAdx(2,:) - dATdx(2,:) + dBdx(2,:)) - z*dCdx(2,:))*sin(delta) + (dAdx(3,:) - dATdx(3,:) + dBdx(3,:) - z*dCdx(3,:))*cos(delta); \]
\[ dudyP = (0.5*U/pi)*(dAdy(1,:) - dATdy(1,:) + dBdy(1,:)) + z*dCdy(1,:)); \]
\[ dvdyP = (0.5*U/pi)*((dAdy(2,:) - dATdy(2,:) + dBdy(2,:)) + z*dCdy(2,:))*cos(delta) - (dAdy(3,:) - dATdy(3,:) + dBdy(3,:)) + z*dCdy(3,:)*sin(delta); \]
\[ dwdyP = (0.5*U/pi)*((dAdy(2,:) - dATdy(2,:) + dBdy(2,:)) - z*dCdy(2,:))*sin(delta) + (dAdy(3,:) - dATdy(3,:) + dBdy(3,:) - z*dCdy(3,:))*cos(delta); \]
\[ dudzP = (0.5*U/pi)*(dAdz(1,:) + dATdz(1,:) + dBdz(1,:) + C(1,:)) + z*dCdz(1,:)); \]
\[ dvdzP = (0.5*U/pi)*((dAdz(2,:) + dATdz(2,:) + dBdz(2,:) + C(2,:)) + z*dCdz(2,:))*cos(delta) - (dAdz(3,:) + dATdz(3,:) + dBdz(3,:) + C(3,:)) + z*dCdz(3,:))*sin(delta); \]
\[ dwdzP = (0.5*U/pi)*((dAdz(2,:) + dATdz(2,:) + dBdz(2,:) - C(2,:) - z*dCdz(2,:))*sin(delta) + (dAdz(3,:) + dATdz(3,:) + dBdz(3,:) - C(3,:)) - z*dCdz(3,:))*cos(delta); \]
\[ dvs_rot = rotmat_dvs[dudxP; dvdyP; dwdzP; dudyP; dudzP; dvdxP; dvdzP; dwdyP]; \]
Appendix E. Select Code

\[ \text{dudx} = \text{reshape}(\text{dvs}_\text{rot}(1,:), \text{size}(x)); \]
\[ \text{dvdy} = \text{reshape}(\text{dvs}_\text{rot}(2,:), \text{size}(y)); \]
\[ \text{dwdz} = \text{reshape}(\text{dvs}_\text{rot}(3,:), \text{size}(x)); \]
\[ \text{dudy} = \text{reshape}(\text{dvs}_\text{rot}(4,:), \text{size}(y)); \]
\[ \text{dudz} = \text{reshape}(\text{dvs}_\text{rot}(5,:), \text{size}(y)); \]
\[ \text{dvdx} = \text{reshape}(\text{dvs}_\text{rot}(6,:), \text{size}(y)); \]
\[ \text{dvdz} = \text{reshape}(\text{dvs}_\text{rot}(7,:), \text{size}(x)); \]
\[ \text{dwdx} = \text{reshape}(\text{dvs}_\text{rot}(8,:), \text{size}(y)); \]
\[ \text{dwdy} = \text{reshape}(\text{dvs}_\text{rot}(9,:), \text{size}(x)); \]
end

E.1.2 f_dfxyz.m

function \( [f, dfdx, dfdy, dfdz] = f_dfxyz(\text{letter, xi, eta, q, z, delta, alpha}) \)

\[ \text{epsilon} = 1e-8; \% \text{criterion for determining if a variable is zero} \]
\[ \text{R} = \sqrt{\text{xi}^2 + \text{eta}^2 + \text{q}^2}; \]
\[ \text{X} = \sqrt{\text{xi}^2 + \text{q}^2}; \]
\[ \text{ytilde} = \text{eta} \cdot \cos(\text{delta}) + \text{q} \cdot \sin(\text{delta}); \]
\[ \text{dtilde} = \text{eta} \cdot \sin(\text{delta}) - \text{q} \cdot \cos(\text{delta}); \]
\[ \text{ctilde} = \text{dtilde} + \text{z}; \]
\[ \text{Rdtilde} = \text{R} + \text{dtilde}; \]
\[ \text{alpha2} = 1 - \text{alpha}; \]
\[ \text{D11} = 1/(\text{R} \cdot \text{Rdtilde}); \]
%% Check singularity conditions

% condition (i)
check_i = abs(q) > epsilon;
theta = check_i.*atan(xi.*eta./(q.*R));

% condition (ii)
check_ii = abs(xi) > epsilon;

% condition (iii)
check_iii = abs(R + xi) > epsilon;
infcheck_Rxi_1 = -1*log(R - xi);
infcheck_Rxi_2 = log(R + xi);
infcheck_Rxi_1(isinf(infcheck_Rxi_1)) = 0;
infcheck_Rxi_2(isinf(infcheck_Rxi_2)) = 0;
ln_Rxi = ~check_iii.*infcheck_Rxi_1 + check_iii.*infcheck_Rxi_2;
X11 = check_iii./R./(R + xi);
X32 = check_iii.*(2*R + xi)./(R.^3)./(R + xi).^2;
X53 = check_iii.*(8*R.^2 + 9*R.*xi + 3*xi.^2)./(R.^5)./(R + xi).^3;

% condition (iv)
check_iv = abs(R + eta) > epsilon;
infcheck_Reta_1 = -1*log(R - eta);
infcheck_Reta_2 = log(R + eta);
infcheck_Reta_1(isinf(infcheck_Reta_1)) = 0;
infcheck_Reta_2(isinf(infcheck_Reta_2)) = 0;
ln_Reta = ~check_iv.*infcheck_Reta_1 + check_iv.*infcheck_Reta_2;
Y11 = check_iv./R./exp(infcheck_Reta_2);
Y32 = check_iv.*(2*R + eta)./(R.^3)./exp(infcheck_Reta_2).^2;
Y53 = check_iv.*(8*R.^2 + 9*R.*eta + 3*eta.^2)./(R.^5)
    ./exp(infcheck_Reta_2).^3;

%% Define other variables from equation 14
h = q*cos(delta) - z;
Z32 = sin(delta)./(R.^3) - h.*Y32;
Z53 = 3*sin(delta)./(R.^5) - h.*Y53;
Y0 = Y11 - Y32.*xi.^2;
Z0 = Z32 - Z53.*xi.^2;

%% Define I parameters (for displacement equations, Table 6)
if cos(delta) < epsilon
    I3 = 0.5*(eta./Rdtilde + ytilde.*q./(Rdtilde.^2) - ln_Reta);
    I4 = check_ii*0.5.*xi.*ytilde./(Rdtilde.^2);
else
    I3 = ytilde./Rdtilde/cos(delta) - (ln_Reta -
        sin(delta)*log(Rdtilde))/cos(delta)^2;
    I4 = check_ii.*(xi*tan(delta)./Rdtilde + 2*atan((eta.*(X +
        q*cos(delta)) + X.*(R + X)*sin(delta))...
        ./xi.*(R + X)*cos(delta)))/cos(delta)^2);
end
I1 = -(xi./Rdtilde)*cos(delta) - I4*sin(delta);
I2 = log(Rdtilde) + I3*sin(delta);

%% Define J and K parameters (for x-derivative equations, Table 7)
J2 = xi.*ytilde.*D11./Rdtilde;
J5 = -D11.*(dtilde + (ytilde.^2)./Rdtilde);

if cos(delta) < epsilon
    K1 = xi.*q.*D11./Rdtilde;
    K3 = sin(delta).*(D11.*xi.^2 - 1.0)./Rdtilde;
    J3 = -xi.*(D11.*q.^2 - 0.5)./Rdtilde.^2;
    J6 = -ytilde.*(D11.*xi.^2 - 0.5)./Rdtilde.^2;
else
    K1 = xi.*(D11 - Y11*sin(delta))/cos(delta);
    K3 = (q.*Y11 - ytilde.*D11)/cos(delta);
    J3 = (K1 - J2*sin(delta))/cos(delta);
    J6 = (K3 - J5*sin(delta))/cos(delta);
end

K2 = 1./R + K3*sin(delta);
K4 = xi.*Y11*cos(delta) - K1*sin(delta);
J1 = J5*cos(delta) - J6*sin(delta);
J4 = -xi.*Y11 - J2*cos(delta) + J3*sin(delta);

%% Define E, F, G etc. parameters (for y-derivative equations, Table 8)

E = sin(delta)./R - ytilde.*q./(R.^3);
F = dtilde./(R.^3) + Y32*sin(delta).*xi.^2;
G = 2*X11*sin(delta) - ytilde.*q.*X32;
H = dtilde.*q.*X32 + xi.*q.*Y32*sin(delta);
P = cos(delta)./(R.^3) + q.*Y32*sin(delta);
Q = 3*ctilde.*dtilde./(R.^5) - sin(delta)*(z*Y32 + Z32 + Z0);
%% Define E', F', G' etc. parameters (for z-derivative equations, Table 9)

\[ E_p = \cos(\delta) / R + \tilde{d} \cdot q / (R^3); \]
\[ F_p = \tilde{y} / (R^3) + Y32 \cos(\delta) \cdot \xi^2; \]
\[ G_p = 2 \cdot X11 \cos(\delta) + \tilde{d} \cdot q \cdot X32; \]
\[ H_p = \tilde{y} \cdot q \cdot X32 + \xi \cdot q \cdot Y32 \cos(\delta); \]
\[ P_p = \sin(\delta) / (R^3) - q \cdot Y32 \cos(\delta); \]
\[ Q_p = 3 \cdot \tilde{c} \cdot \tilde{y} / (R^5) + q \cdot Y32 - \cos(\delta) \cdot (z \cdot Y32 + Z32 + Z0); \]

%% Create f, df/dx, df/dy, df/dz arrays depending on letter 'A', 'B', or 'C'

if strcmp(letter, 'A')
    f1 = -0.5*alpha2*ln_Reta - 0.5*alpha*Y11.*q.^2;
    f2 = -0.5*alpha2*ln_Rxi - 0.5*alpha*X11.*q.^2;
    f3 = 0.5*(theta - alpha*q.*(eta.*X11 + xi.*Y11));
    df1x = -0.5*alpha2*xi.*Y11 + 0.5*alpha*xi.*Y32.*q.^2;
    df2x = -0.5*alpha2./R + 0.5*alpha*(q.^2)./(R.^3);
    df3x = -0.5*alpha2*q.*Y11 - 0.5*alpha*Y32.*q.^3;
    df1y = -0.5*alpha2*(cos(\delta)./R + q.*Y11*sin(\delta))
            - 0.5*alpha*q.*F;
    df2y = -0.5*alpha2*ytilde.*X11 - 0.5*alpha*q.*G;
    df3y = 0.5*alpha2*(dtilde.*X11 + xi.*Y11*sin(\delta))
            + 0.5*alpha*q.*H;
Appendix E. Select Code

\[ df1z = 0.5*\alpha_2*(\sin(\delta)/R - q.*Y11*cos(\delta)) \]
\[ - 0.5*\alpha*q.*Fp; \]
\[ df2z = 0.5*\alpha_2*d\tilde{\text{x}}11 - 0.5*\alpha*q.*Gp; \]
\[ df3z = 0.5*\alpha_2*(y\tilde{\text{x}}11 + \xi*Y11*cos(\delta)) \]
\[ + 0.5*\alpha*Fp; \]

\[
\text{elseif strcmp(letter, 'B')}
\]
\[ f1 = Y11.*q.^2 - (\alpha_2*I3*sin(\delta)^2)/\alpha; \]
\[ f2 = X11.*q.^2 + (\alpha_2*\xi*sin(\delta)^2)/\alpha./Rd\tilde{\text{x}}; \]
\[ f3 = q.*(\eta.*X11 + \xi.*Y11) - \theta - \]
\[ (\alpha_2*I4*sin(\delta)^2)/\alpha; \]
\[ df1x = -\xi.*Y32.*q.^2 - \alpha_2*J4*(\sin(\delta)^2)/\alpha; \]
\[ df2x = -(q.^2)/(R.^3) - \alpha_2*J5*(\sin(\delta)^2)/\alpha; \]
\[ df3x = Y32.*q.^3 - \alpha_2*J6*(\sin(\delta)^2)/\alpha; \]
\[ df1y = q.*F - \alpha_2*J1*(\sin(\delta)^2)/\alpha; \]
\[ df2y = q.*G - \alpha_2*J2*(\sin(\delta)^2)/\alpha; \]
\[ df3y = -q.*H - \alpha_2*J3*(\sin(\delta)^2)/\alpha; \]
\[ df1z = q.*Fp + \alpha_2*K3*(\sin(\delta)^2)/\alpha; \]
\[ df2z = q.*Gp + \alpha_2*\xi*D11*(\sin(\delta)^2)/\alpha; \]
\[ df3z = -q.*Hp + \alpha_2*K4*(\sin(\delta)^2)/\alpha; \]

\[
\text{elseif strcmp(letter, 'C')}
\]
\[ f1 = -\alpha_2*(\sin(\delta)/R + q.*Y11*cos(\delta)) \]
\[ - \alpha*(z*Y11 - Z32.*q.^2); \]
\[ f2 = 2*\alpha_2*\xi.*Y11*sin(\delta) + d\tilde{\text{x}}le.*X11 - \]
\[ \alpha*ct\tilde{\text{e}}.*(X11 - X32.*q.^2); \]
f3 = alpha2*(ytilde.*X11 + xi.*Y11*cos(delta)) +
alpha*q.*(ctilde.*eta.*X32 + xi.*Z32);

df1x = alpha2*xi*sin(delta)./(R.^3) + xi.*q.*Y32*cos(delta) +
alpha*xi..*(3*ctilde.*eta./)(R.^5) - 2*Z32 - Z0);
df2x = 2*alpha2*Y0*sin(delta) - dtilde./((R.^3) +
alpha*ctilde.*(1 - 3*(q./R).^2)./((R.^3));
df3x = -alpha2*(ytilde./((R.^3) - Y0*cos(delta)) -
alpha*(3*ctilde.*eta./q./((R.^5) - q.*Z0);

df1y = alpha2*(q./((R.^3) + Y0*sin(delta)*cos(delta)) +
alpha*(z*cos(delta)/((R.^3) + 3*ctilde.*dtilde.*q./((R.^5) -
q.*Z0*sin(delta));
df2y = -2*alpha2*xi.*P*sin(delta) - ytilde.*dtilde.*X32 +
alpha*ctilde.*(X32.*(ytilde + 2*q*sin(delta)) -
X53.*ytilde.*q.^2);
df3y = -alpha2*(xi.*P*cos(delta) - X11 + X32.*ytilde.^2) +
alpha*ctilde.*(X32.*(dtilde - 2*q*cos(delta)) -
X53.*dtilde.*eta.*q) + alpha*xi.*Q;

df1z = -eta./((R.^3) + Y0*cos(delta)^2 -
alpha*(z*sin(delta)/((R.^3) - 3*ctilde.*ytilde.*q./((R.^5) -
Y0*sin(delta)^2 + q.*Z0*cos(delta));
df2z = 2*alpha2*xi.*Pp*sin(delta) - X11 + X32.*dtilde.^2 -
alpha*ctilde.*(X32.*(dtilde - 2*q*cos(delta)) -
X53.*dtilde.*q.^2);
df3z = alpha2*(xi.*Pp*cos(delta) + ytilde.*dtilde.*X32) +
alpha*ctilde.*(X32.*(ytilde - 2*q*sin(delta)) +
X53.*dtilde.*eta.*q) + alpha*xi.*Qp;
else
   error('please enter letter A, B, or C');
end

f = [f1; f2; f3];
dfdx = [df1x; df2x; df3x];
dfdy = [df1y; df2y; df3y];
dfdz = [df1z; df2z; df3z];

E.2 Bayesian Joint Inversion for Density Prediction with Muon & Gravity Data

The code used for the Bayesian joint inversion of gravity-muon data taken for the mesa below the town of Los Alamos, New Mexico is given below. It is for the case of a uniform prior and anisotropic smoothing. It assumes that all data has been pre-processed and the domain has already been discretized.

%% Load in necessary data

vars = load('vars.mat');
vox_top = load('voxel_top_corner.mat');
vox_cen = load('voxel_cen.mat');
xj = load('xj_01.mat','xj');
Lij = load('Lij_01.mat','Lij');
G = load('G.mat','G');

cnt_in0 = load('counts_in_0.mat');
cnt_in1 = load('counts_in_1.mat');
cnt_out = load('cntdt_out.mat');
cnt_out_mask = load('cnt_out_mask.mat');

TU = load('Trend_Upper.dat');
TL = load('Trend_Lower.dat');

%% Pick out the arrays from the structures

voxel_corner = vars.voxel_corner;
voxel_top_corner = vox_top.voxel_top_corner;
voxel_cen = vox_cen.voxel_cen;
voxel_diag = vars.voxel_diag;

voxel_sep_x = vars.voxel_sep_x;
voxel_sep_y = vars.voxel_sep_y;
voxel_sep_z = vars.voxel_sep_z;
voxel_sep_xy = vars.voxel_sep_xy;

layerbot = vars.layerbot;
layertop = vars.layertop;
nlayers = vars.nlayers;
mu_d_pos = vars.mu_d_pos;

XI = vars.XI;
YI = vars.YI;
ElevI = vars.ElevI;
xnodes = vars.xnodes;
ynodes = vars.ynodes;

xj = xj.xj;
Lij = Lij.Lij;
G = G.G;

cnt_in0 = cnt_in0.cntdt;
cnt_in1 = cnt_in1.cntdt;
cnt_out = cnt_out.cntdt_outside;
cnt_out_mask = cnt_out_mask.cnt_out_mask;

%%% Define necessary constants

min_z = min(ElevI(:));
max_z = max(ElevI(:));

num_voxels = sum(nlayers(:));
maxlayers = max(nlayers(:));

dx = 16;      % Use 16m spacing of voxels in inner region
dy = dx;
DX = 200;     % Use 200m spacing of voxels in outer region
DY = DX;
dz = dx;

Xrange = max(voxel_corner(1,:)) - min(voxel_corner(1,:));
Yrange = max(voxel_corner(2,:)) - min(voxel_corner(2,:));
\begin{verbatim}
Zrange = maxlayers*dz;

rho_base      = 2000;
rho_std       = 125.5556;
corr_length   = 826.491;

XYcorr_fac = 2.333;
Zcorr_fac   = .1;

XYcorr_length = corr_length*XYcorr_fac;
Zcorr_length  = corr_length*Zcorr_fac;

niter = 50;
mu    = .1;

%% Load in and work with station locations in the study area
[point_table, measured_points] = build_table();

eval_pts = point_table{measured_points, Constants.xyz_index}';

measured_values = point_table{measured_points, 'Measurements'};
measure_errors  = point_table{measured_points, 'Errors'};

gz_avg_at_stations  = cellfun(@mean, measured_values);
gz_error_at_stations = cellfun(@norm, measure_errors);

% Subtract off the upper and lower trends
\end{verbatim}
below_cutoff_height = eval_pts(3,:) < 2150;

UpObs = gz_avg_at_stations(~below_cutoff_height);
LowObs = gz_avg_at_stations(below_cutoff_height);

UpObs = UpObs -TU(:,3) + mean(TU(:,3));
LowObs = LowObs -TL(:,3) + mean(TL(:,3));

UpLowError = [TU(:,4); TL(:,4)];
Trend_error = [UpLowError(1:23); UpLowError(29:32);
               UpLowError(24:28); UpLowError(33:end)];

UpLow = [UpObs; LowObs];
Trend_gz = [UpLow(1:23); UpLow(29:32);
           UpLow(24:28); UpLow(33:end)];

% Convert gravity data from mgal to m/s^2
measured_gz = Trend_gz*(1e-5);
error_gz = sqrt(Trend_error.^2 + gz_error_at_stations.^2)*(1e-5);

%% Get variance in muon measurements

atn0 = -log(cnt_in0(:,1:16)./cnt_out(:,1:16));
atn1 = -log(cnt_in1(:,1:16)./cnt_out(:,1:16));

xj0 = xj(1:length(atn0(:)));
xj1 = xj(length(atn1(:))+1:end);
% Cut the counts arrays so theta = (0:3:45)

cnt_in0_cut = cnt_in0(:,1:16);
cnt_in1_cut = cnt_in1(:,1:16);
cnt_out_cut = cnt_out(:,1:16);

Nin0 = cnt_in0_cut;
Nin1 = cnt_in1_cut;
Nout = cnt_out_cut;

Nratio0 = Nin0./Nout;
Nratio1 = Nin1./Nout;

delta_N = 1e-4;

Nplus0 = Nratio0 + delta_N;
Nplus1 = Nratio1 + delta_N;

xj0_reshape = reshape(xj0, size(Nratio0));
xj1_reshape = reshape(xj1, size(Nratio1));

% Get variance (sigma_R) for inner and outer det. locations

sigma_R0 = zeros(size(Nratio0));
sigma_R1 = zeros(size(Nratio1));

for th = 1:size(cnt_out_cut, 2)
    logNO = log(Nratio0(:,th));
logN1 = log(Nratio1(:,th));

[p0,S0] = polyfit(logN0, log(xj0_reshape(:,th)), 1);
[p1,S1] = polyfit(logN1, log(xj1_reshape(:,th)), 1);

logRplus0 = polyval(p0, log(Nplus0(:,th)));
logRplus1 = polyval(p1, log(Nplus1(:,th)));

logR0 = polyval(p0, logN0);
logR1 = polyval(p1, logN1);

DeltaR_DeltaN0 = (exp(logRplus0) - exp(logR0))/delta_N;
DeltaR_DeltaN1 = (exp(logRplus1) - exp(logR1))/delta_N;

Nout = cnt_out_cut(end,th);

sigma_R0(:,th) = abs(DeltaR_DeltaN0).*sqrt(Nratio0(:,th).*(1 + Nratio0(:,th)).*(1/Nout));
sigma_R1(:,th) = abs(DeltaR_DeltaN1).*sqrt(Nratio1(:,th).*(1 + Nratio1(:,th)).*(1/Nout));

end

xj_error = [sigma_R0(:); sigma_R1(:)]; % error in the range

%% Concatenate G to Lij matrix
% Scale everything to approx. same order of magnitude so that
inversion is stable
scale_gz = 1./error_gz;
scale_mu = 1./xj_error;

Ng  = length(measured_gz);
Nmu = length(xj);

measured_obs = [measured_gz.*scale_gz; xj.*scale_mu];
GL = [G.*scale_gz; Lij.*scale_mu];

%%% Make data covariance matrix and inverse

covar_data       = diag(ones(length(measured_obs),1));
covar_data_inv  = inv(covar_data);

%%% Make prior model/rho covariance matrix

covar_rho = (rho_std^2)*exp(-voxel_sep_xy/XYcorr_length -
voxel_sep_z/Zcorr_length);
inv_covar_rho  = inv(covar_rho);

%%% Make posterior model covariance

inv_C_post = GL.'*(covar_data\GL) + inv_covar_rho;
operator1 = (inv_C_post\GL.')/covar_data;
operator2 = inv_C_post\inv_covar_rho;

%%% Create weighting matrix for observation residuals

alpha_mu = Ng/Nmu;
alpha_mu = 1;

weight_mu = ones(Nmu,1)*alpha_mu;
weight_g = ones(Ng,1)*alpha_g;

weight_matrix = [weight_g; weight_mu];

%% Iterate with rho_inv as new input model

rho_uniform = repmat(rho_base, num_voxels, 1);
rho_guess = rho_uniform; % Uniform prior

for i=1:niter

  % Forward calculation
  pred_obs = GL * rho_guess;

  % Subtract off the base station from gravity data
  pred_g = pred_obs(1:Ng);
  pred_g = pred_g - pred_g(strcmp(measured_points, 'BS_TN_1'));
  pred_obs(1:Ng) = pred_g;

  residual = pred_obs - measured_obs;
  fac = norm(residual)/norm(weight_matrix.*residual);
  residual = fac*weight_matrix.*residual;
Appendix E. Select Code

% Update rho according to inversion step

term1 = operator1*residual;
term2 = operator2*(rho_guess - rho_uniform);
term = term1 + term2;

rho_inv = rho_guess - mu*term;
rho_old = rho_guess;
rho_guess = rho_inv;

% Save RMS of residuals to check convergence

rmsobs = rms(residual);
rmslist(i) = rmsobs;

end

E.3 Machine Learning Algorithm for Density Prediction with Muon & Gravity Data

Below is the machine learning algorithm used for the subsurface density prediction with gravity and muon data at Mount Showa-Shinzan (Japan) found in Chapter 4. It is written for the noise-free model (with no input shuffling). The first script is written for MATLAB and it generates the data used in the second script, which contains the machine learning algorithm and is written in Python. It assumes that all data has been pre-processed and the domain has been discretized.
E.3.1 Data Generation Algorithm

clear all; close all; rng('shuffle');

var_filename = 'var_filename.mat';
input_filename = 'filename_in.csv';
output_filename = 'filename_out.csv';

dz_cut = 40;
percent_noise = 0;

n_perts = 5e6;
max_cluster_nvox = 5;  % max # of voxels in each cluster
nmax_clusters_per_iter = 20;  % max # of clusters per iteration

rho_background = 1700;

drho_min = 1400;

rhorho_max = 2600;

rhorho_step = 1;

vars = load(var_filename);
Gij = vars.Gij;
Lij = vars.Lij;

global_index = vars.global_index;
muon_det_coords = vars.mu_det;
grav_stn_coords = vars.grav_stns;
voxel_bot_corner = vars.voxel_corner;
theta_arr = vars.theta_arr;
Appendix E. Select Code

\texttt{ind\_cut = (voxel\_bot\_corner(3,:) >= muon\_det\_coords(3)+dz\_cut \&
voxel\_bot\_corner(3,:) <= max(voxel\_bot\_corner(3,:)));}
\texttt{vox\_cut\_inds = find(ind\_cut);} 
\texttt{voxel\_bot\_corner\_cut = voxel\_bot\_corner(:,ind\_cut);} 

\texttt{pert\_cell = cell(1, n\_pers);}

\texttt{for n = 1:n\_pers}
\texttt{    nclusters\_per\_iter = randi(nmax\_clusters\_per\_iter);} 
\texttt{    iter\_pert\_cell = cell(1, nclusters\_per\_iter);} 
\texttt{    iter\_pert\_voxnums = [];} 

\texttt{for i = 1:nclusters\_per\_iter}
\texttt{    cluster\_nvox = randi([1, max\_cluster\_nvox]);}
\texttt{    np = randsample(vox\_cut\_inds, 1);} 
\texttt{    drho = randsample((drho\_min:drho\_step:drho\_max), 1);} 

\texttt{while ismember(np, iter\_pert\_voxnums)} 
\texttt{    np = randsample(vox\_cut\_inds, 1);} 
\texttt{end}

\texttt{if cluster\_nvox == 1}
\texttt{    iter\_pert\_voxnums = [iter\_pert\_voxnums; np];}
\texttt{    iter\_pert\_cell{i} = [np; drho];}
\texttt{    iter\_pert\_voxnums = [iter\_pert\_voxnums; np];}
\texttt{else}
\texttt{    iter\_pert = [];} 
\texttt{    iter\_pert = [iter\_pert; np];}
\texttt{    iter\_pert\_voxnums = [iter\_pert\_voxnums; np];}
adj_cells = get_adjacent_cells(global_index(:, np));

adj_tf = ismember(global_index', adj_cells', 'rows');
adj_ind_inglob = find(adj_tf);

adj_tf_invox = ismember(adj_ind_inglob, vox_cut_inds);
adj_ind_invox = adj_ind_inglob(adj_tf_invox);

adj_tf_norepeat = ismember(adj_ind_invox, iter_pert_voxnums);
adj_ind_invox_norepeat = adj_ind_invox(~adj_tf_norepeat);

if isempty(adj_ind_invox_norepeat)
    cluster_nvox = 1;
    iter_pert_voxnums = [iter_pert_voxnums; np];
    iter_pert_cell{i} = [np; drho];
    iter_pert_voxnums = [iter_pert_voxnums; np];

elseif cluster_nvox == 2
    rand_adj_cell =
    randsample(adj_ind_invox_norepeat, 1);
    iter_pert = [iter_pert; rand_adj_cell];
    iter_pert_voxnums = [iter_pert_voxnums; rand_adj_cell];
    iter_pert_cell{i} = [iter_pert';
    drho*ones(size(iter_pert))''];
else
    cluster_nvox_count = 1;
while cluster_nvox_count <= cluster_nvox

    rand_adj_cell =
    randsample(adj_ind_invox_norepeat, 1);

    if ~ismember(rand_adj_cell, iter_pert_voxnums)
        iter_pert = [iter_pert; rand_adj_cell];
        iter_pert_voxnums = [iter_pert_voxnums;
                         rand_adj_cell];
        cluster_nvox_count = cluster_nvox_count + 1;
    end

    if cluster_nvox_count == cluster_nvox; break;
    end

    adj_adj_cells = get_adjacent_cells(global_index
           (:,rand_adj_cell));

    adj_adj_tf = ismember(global_index',
           adj_adj_cells', 'rows');
    adj_adj_ind_inglob = find(adj_adj_tf);

    adj_adj_tf_invox =
    ismember(adj_adj_ind_inglob, vox_cut_inds);
    adj_adj_ind_invox =
    adj_adj_ind_inglob(adj_adj_tf_invox);

    adj_adj_tf_norepeat =
    ismember(adj_adj_ind_invox,
iter_pert_voxnums);
adj_adj_ind_invox_norepeat =
adj_adj_ind_invox(~adj_adj_tf_norepeat);

if ~isempty(adj_adj_ind_invox_norepeat)
    rand_adj_adj_cell =
randsample(adj_adj_ind_invox_norepeat, 1);
    iter_pert = [iter_pert,
                 rand_adj_adj_cell];
    iter_pert_voxnums = [iter_pert_voxnums,
                         rand_adj_adj_cell];
    cluster_nvox_count = cluster_nvox_count + 1;
else
    cluster_nvox = cluster_nvox_count;
    break
end
if cluster_nvox_count == cluster_nvox; break;
end
end
iter_pert_cell{i} = [iter_pert';
                    drho*ones(size(iter_pert))'];
end
end
end
iter_pert_cell_array = [];
for j = 1:length(iter_pert_cell)
    iter_pert_cell_array = [iter_pert_cell_array,
                            iter_pert_cell{j}];
end
pca_tf_invox = ismember(iter_pert_cell_array(1,:), vox_cut_inds);
pert_cell{n} = iter_pert_cell_array(:,pca_tf_invox);
end

%% Calculate gj and Rj and write to file
for n = 1:n_perts
    nf = pert_cell{n};
    rho_model = rho_background*ones(size(voxel_bot_corner, 2),1);
    rho_model(nf(1,:)) = nf(2,:);
    
    dgz = Gij*rho_model;
    dgz_noise = .01*percent_noise*std(dgz)*randn(size(dgz));
    dgz_model = dgz + dgz_noise;
    dgz_model = dgz_model - dgz_model(1);
    
    drj = Lij*rho_model;
    drj_noise = .01*percent_noise*std(drj)*randn(size(drj));
    drj_model = drj + drj_noise;
    
    writematrix(rho_model(ind_cut)', ['path_to_output' output_filename], 'WriteMode','append')
    writematrix([dgz_model', drj_model'], ['path_to_output' input_filename], 'WriteMode','append')
end

%% Define some functions
function adj_cells = get_adjacent_cells(array)
    adj_cells = [array(1)-1, array(2), array(3); ...
                 array(1)+1, array(2), array(3); ...]
array(1), array(2)-1, array(3); ... 
array(1), array(2)+1, array(3); ... 
array(1), array(2), array(3)-1; ... 
array(1), array(2), array(3)+1]’;

end

E.3.2 Machine Learning Algorithm

def transform_data(data, norm_type):
    transformer = norm_type.fit(data)
    data_transform = transformer.transform(data)
    return data_transform, transformer

if __name__ == '__main__':

    ngrav = 35

    data_path = Path('/path/to/data/directory')
    input_data = pd.read_csv(data_path/'input_filename').to_numpy()
    output_data = pd.read_csv(data_path/'output_filename').to_numpy()

    xtrain, xtest, ytrain, ytest = train_test_split(input_data, output_data, test_size=0.2, random_state=42)
Appendix E. Select Code

```python
grav_train = xtrain[:, :ngrav]
grav_test = xtest[:, :ngrav]
muon_train = xtrain[:, ngrav:]
muon_test = xtest[:, ngrav:]

transformer_in = Normalizer()

grav_train_transform, _ = transform_data(grav_train, transformer_in)
grav_test_transform, _ = transform_data(grav_test, transformer_in)
muon_train_transform, _ = transform_data(muon_train, transformer_in)
muon_test_transform, _ = transform_data(muon_test, transformer_in)

xtrain = np.concatenate((grav_train_transform, muon_train_transform), axis=1)
xtest = np.concatenate((grav_test_transform, muon_test_transform), axis=1)

model = MultiOutputRegressor(HistGradientBoostingRegressor(learning_rate=0.1, max_iter=500, max_leaf_nodes=256, l2_regularization=20.0), n_jobs=-1)

model.fit(xtrain, ytrain)
ypred = model.predict(xtest)
score = model.score(xtest, ytest)
```
print(score)

np.save('ypred.npy', ypred)
np.save('ytest.npy', ytest)

xnish = np.load('xnish_input_50x50x40mvox.npy')
xnish = np.expand_dims(xnish, 1)

grav_nish, __ = transform_data(xnish[:ngrav], Normalizer())
muon_nish, __ = transform_data(xnish[ngrav:], Normalizer())

xnish = np.concatenate((grav_nish, muon_nish), axis=0)
ynish = model.predict(xnish.reshape(1,-1))
np.save('ynishpred.npy', ynish)
References


REFERENCES


and range tables 10 MeV-100 TeV, *Atomic Data and Nuclear Data Tables*, **78**(2), 183–356.


Kjøll, H. J., Galland, O., Labrousse, L., & Andersen, T. B., 2019. Emplacement mechanisms of a dyke swarm across the brittle-ductile transition and the geody-


Lyman, A. W., Koenig, E., & Fink, J. H., 2004. Predicting yield strengths and
REFERENCES


REFERENCES


