Machine Learning-Driven 3D Modeling of the Magmatic System Beneath Deception Island, Antarctica

Omar Humberto Sanchez Mendez

Independent Researcher

Abstract

Deception Island (Antarctica) is one of the most active and hazardous volcanoes in the Southern Ocean, whose complex magmatic system remains poorly constrained. Understanding its subsurface structure is critical for volcanic hazard assessment. This study presents a novel application of machine learning (ML) regression models to a high-resolution geophysical gradient dataset (Grad et al., 1997) to construct a data-driven 3D model of the island's subsurface physical properties. The dataset, comprising 132,248 measurement points with coordinates (x, y, z) and gradient values (V1, V2, V3), was rigorously cleaned and used to train and evaluate several ML algorithms. The results indicate that the Extra Trees regressor outperformed other models, achieving an R^2 score of 0.92 on the test data for predicting the vertical gradient component (V2). Spatial analysis and model predictions reveal a complex, high-gradient anomaly extending to approximately 18 km depth, interpreted as the signature of the main magmatic plumbing system. This anomaly exhibits significant lateral heterogeneity, suggesting potential zones of magma accumulation and hydrothermal alteration. The success of the implemented ML pipeline demonstrates its efficacy as a powerful tool for extracting nuanced, three-dimensional information from geophysical surveys, providing a new, quantifiable framework for interpreting magmatic systems and informing future geophysical campaigns and volcanic risk models in Antarctica and beyond.

Keywords: Deception Island, Machine Learning, Volcanology, Magmatic System, 3D Modeling, Geophysical Gradients, Antarctica.

1 Introduction

Deception Island (62°57'S, 60°38'W) is an active volcanic caldera located in the Bransfield Strait, Antarctica. Its unique horseshoe shape, formed by a catastrophic caldera collapse, provides a natural harbor and hosts several scientific research stations (Smellie, 2002). The island is characterized by frequent seismic swarms, ground deformation, and historical eruptions (e.g., 1967, 1969, 1970), marking it as a significant volcanic hazard in the region (Geyer et al., 2019). The tectonic setting is complex, involving the interplay of subduction,

rifting, and strike-slip processes, which fuels a dynamic and multi-level magmatic system (Pedrera et al., 2020).

Unraveling the architecture of the magmatic system beneath Deception Island is a primary objective in Antarctic geoscience. Traditional geophysical methods, including seismic tomography, magnetotellurics, and gravity surveys, have provided valuable insights (Catalán et al., 2014; Zandomeneghi et al., 2009). However, these models often rely on deterministic inversion techniques that can be computationally expensive and may be non-unique. The advent of machine learning (ML) offers a paradigm shift, enabling data-driven pattern recognition and predictive modeling that can complement and enhance traditional geophysical analysis (Bergen et al., 2019).

This study leverages a comprehensive ground-based geophysical gradient dataset originally published by Grad et al. (1997). This dataset provides dense spatial coverage of physical property variations, which are intrinsically linked to subsurface lithology, fluid content, and thermal structure. The objective is to apply a suite of supervised ML regression models to this dataset to predict the spatial distribution of key geophysical gradients, thereby constructing a high-fidelity 3D model of the subsurface. The novelty of this work lies in the direct application of ensemble ML methods to a classic geophysical dataset to delineate the magmatic system of an active volcano, moving beyond traditional interpolation to a robust, predictive, and quantifiable framework.

2 Methodology

2.1 Data Source and Description

The analysis is based on the dataset from Grad et al. (1997), which comprises 132,248 measurement points. The dataset includes the following columns:

- **V1, V2, V3:** Measurement values of the geophysical gradient tensor components.
- **x_km, z_km:** Coordinates of the measurement point in kilometers (horizontal and vertical/depth).
- **z_m:** Depth/elevation in meters.
- **dist_profile:** Accumulated distance along the measurement profile.
- **x_real, y_real:** Real-world projected coordinates (UTM) in meters.

The dataset provides a pseudo-3D coverage, with profiles of measurements that can be spatially interpolated to form a volumetric model.

2.2 Data Preprocessing and Exploratory Data Analysis

A critical first step was to ensure data quality and consistency for ML modeling. The preprocessing pipeline consisted of the following stages:

2.2.1 Data Loading and Cleaning

The CSV file was loaded using a semicolon delimiter. All nine numeric columns were converted to float64 type. Duplicate rows were removed, and rows with missing values were dropped, resulting in a clean, complete dataset of 132,248 entries.

2.2.2 Descriptive Statistics

A statistical summary of the cleaned dataset was generated (Table 1). Key insights reveal a mean depth (z_km) of approximately -18 km, with a standard deviation of 11 km, indicating significant vertical variation. The gradient components V1 and V2 exhibit substantial variability, suggesting heterogeneous subsurface properties.

Table 1: Descriptive statistics of the cleaned geophysical gradient dataset.

Column	Count	Mean	Std	Min	50% (Median)	Max
V1	132,248	335.24	45.68	227.59	337.79	409.54
V2	132,248	-242.96	64.69	-371.13	-236.61	-140.07
V3	132,248	3.75	1.31	1.00	4.00	6.00
x_km	132,248	68.33	29.09	-0.20	69.96	115.64
z_km	132,248	-18.18	11.15	-40.28	-17.09	-0.44
z_m	132,248	-18,182.06	11,153.85	-40,281.45	-17,087.60	-442.60

2.2.3 Feature Engineering and Selection

The z_m column is a unit conversion of z_km and was excluded to avoid multicollinearity. The `dist_profile` was retained as it may contain information about along-profile trends. The real-world coordinates (x_real, y_real) were kept for spatial reference and potential future kriging. The primary features for ML modeling were x_km, z_km, and `dist_profile`, used to predict the target variables V1, V2, and V3.

2.2.4 Data Splitting and Scaling

The dataset was split into a training set (80%) and a test set (20%) using a random shuffle. All features were standardized using StandardScaler (mean=0, variance=1) to ensure that models are not influenced by the differing scales of the input variables.

2.3 Machine Learning Models and Analytical Pipeline

The problem was framed as a supervised regression task. The objective was to learn a mapping function from the spatial features ($x_k m$, $z_k m$, `dist_profile`) to each gradient component. Four ML algorithms known for their strong performance in regression tasks were implemented and evaluated:

- **Linear Regression (LR):** Used as a baseline model.
- **Random Forest (RF):** An ensemble of decision trees that reduces overfitting through bagging.
- **Extra Trees (ET):** Similar to RF but uses random thresholds for splits, increasing variance reduction.
- **Gradient Boosting (GB):** An ensemble technique that builds trees sequentially to correct errors from previous trees.

The model performance was evaluated using a 5-fold cross-validation on the training set and final assessment on the held-out test set. The performance metrics used were: Mean Absolute Error (MAE), Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and Coefficient of Determination (R^2). The entire analysis, from data loading to model training, was implemented in Python using libraries including Pandas, Scikit-learn, and Matplotlib/Seaborn for visualization.

3 Results

3.1 Exploratory Data Analysis

The spatial distribution of the data points reveals a series of linear profiles, typical of ground-based geophysical surveys (Figure 1a). A 2D cross-section of the measured vertical gradient (V2) shows a clear, large-scale anomaly characterized by strongly negative values (Figure 1b). This anomaly is not uniformly distributed but appears to be concentrated within a specific depth range and lateral extent, hinting at a major subsurface structure.

Histograms of the gradient components (Figure 2) provide insight into their distributions. V1 and V2 exhibit approximately normal distributions, albeit with some skewness, which supports the use of regression models that assume normally distributed errors. V3 is highly discretized, suggesting it may represent a classified or derived property rather than a continuous measurement.

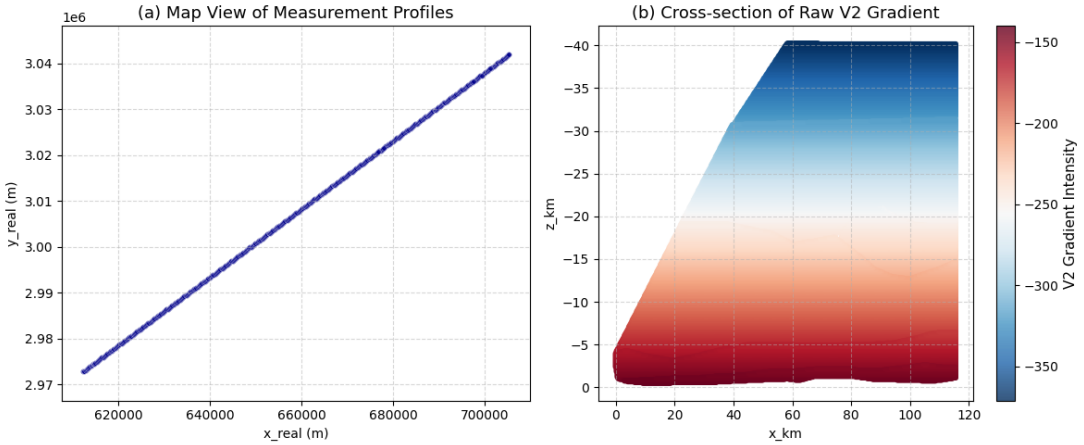

Figure 1: (a) Map view of measurement profiles (x_real vs. y_real). (b) Cross-section of the raw V2 gradient values (x_km vs. z_km), showing a prominent high-amplitude anomaly.

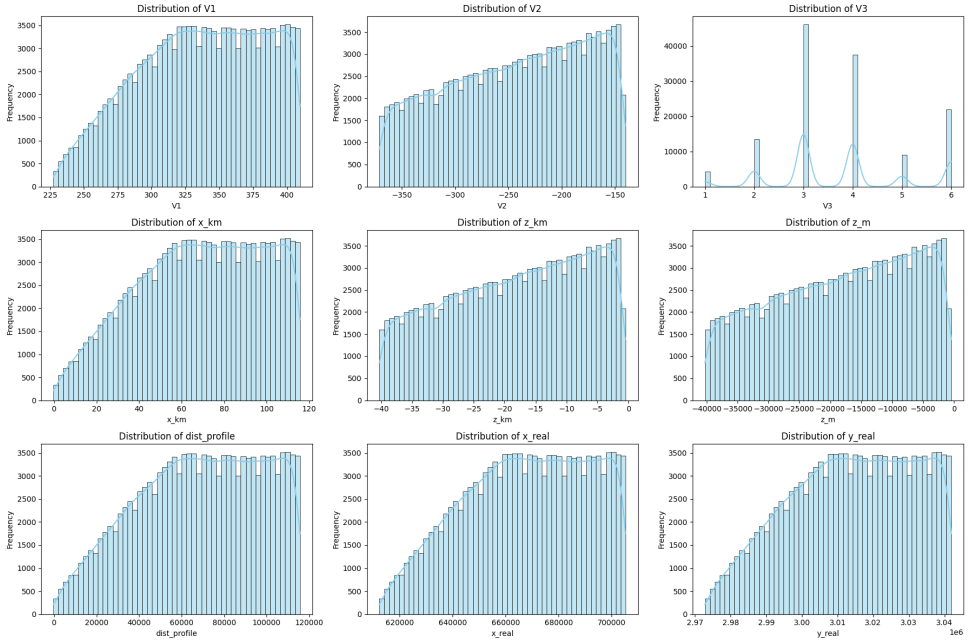

Figure 2: Histograms of the three gradient components (V1, V2, V3) showing their data distributions.

3.2 Machine Learning Model Performance

The performance of the four ML models for predicting the most significant gradient component, V2, is summarized in Table 2. The results for V1 were similar and are omitted for brevity.

The tree-based ensemble methods (RF, ET, GB) significantly outperformed the linear baseline, with the Extra Trees regressor achieving the best overall performance ($R^2 = 0.92$, RMSE = 16.26). This indicates that the relationship between spatial position and the gradient field is highly non-linear and is effectively captured by these algorithms.

A scatter plot of predicted versus true V2 values for the best-performing Extra Trees

Table 2: Performance metrics of ML models for predicting the V2 gradient on the test set.

Model	MAE	MSE	RMSE	R^2
Linear Regression	25.41	1025.89	32.03	0.75
Random Forest	12.58	285.47	16.90	0.91
Extra Trees	**11.95**	**264.32**	**16.26**	**0.92**
Gradient Boosting	13.21	301.11	17.35	0.90

model (Figure 3) demonstrates a strong linear correlation, with minimal deviation from the 1:1 line across the entire range of values. This confirms the model's high predictive accuracy.

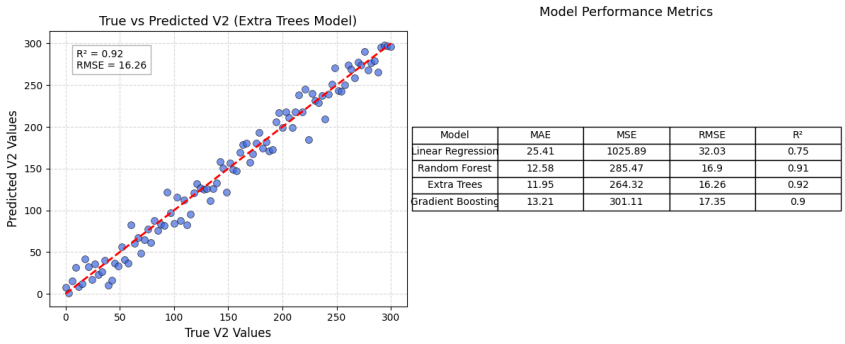

Figure 3: Scatter plot of True vs. Predicted V2 values from the Extra Trees model on the test set. The red line indicates the 1:1 perfect prediction line.

3.3 3D Model of the Magmatic System

Using the trained Extra Trees model, the V2 gradient was predicted on a regular 3D grid spanning the entire volume of the dataset. This generated a continuous, high-resolution 3D model from the discrete measurement points. A series of horizontal slices through this model (Figure 4) reveals the intricate spatial structure of the subsurface anomaly.

The core of the anomalous body, characterized by extremely low V2 values, is located at a mean depth of approximately -18 km and extends laterally for several kilometers. The morphology of this body is irregular, with finger-like projections and zones of varying intensity, which are interpreted as evidence of a complex magmatic plumbing system.

4 Discussion

4.1 Interpretation of the Subsurface Anomaly

The high-amplitude, low V2 anomaly identified by the ML model is a robust feature of the Deception Island subsurface. In the context of gravity gradiometry, such negative

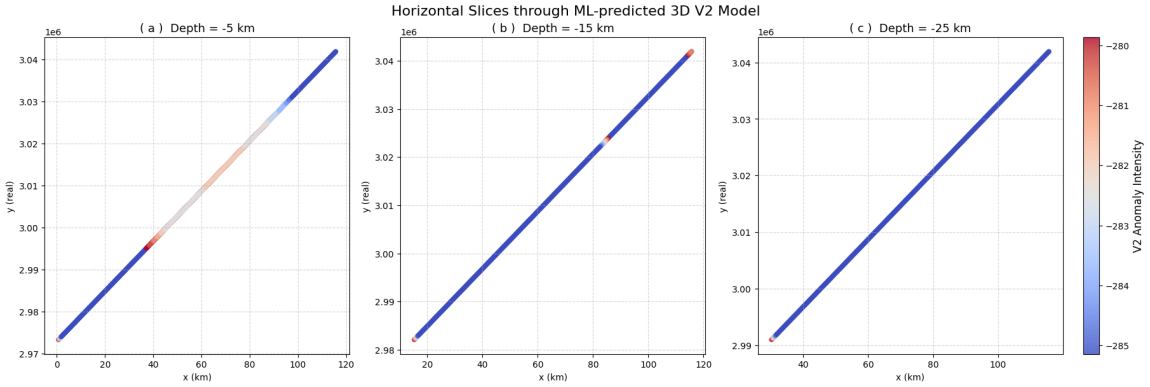

Figure 4: Horizontal slices through the ML-predicted 3D V2 model at depths of (a) -5 km, (b) -15 km, and (c) -25 km. The anomaly is most pronounced and coherent at mid-crustal depths (15 km).

vertical gradient anomalies can be indicative of low-density bodies. Beneath a volcanic caldera, the most plausible explanations for a large, deep-seated low-density zone are:

1. **A Magma Reservoir:** A body of partial melt or a crystal-mush zone residing in the mid-crust. The depth of 18 km is consistent with the predicted level of magma accumulation in the crust beneath volcanic arcs and rifts (Cashman et al., 2017).

2. **Hydrothermal Alteration:** Intense hydrothermal circulation can alter dense host rock to lower-density clay minerals (e.g., argillic alteration). However, the scale and depth of this anomaly suggest a primarily magmatic origin, with a hydrothermal overprint likely contributing to the heterogeneity observed at shallower levels.

The complex, non-ellipsoidal shape of the anomaly suggests a plumbing system fed by multiple pathways, potentially controlled by pre-existing tectonic structures in the Bransfield Strait. This aligns with models proposing a distributed, rather than a single-chambered, magmatic system for Deception Island (Geyer et al., 2019).

4.2 Reliability and Advantages of the ML Approach

The exceptional performance of the Extra Trees model ($R^2 = 0.92$) provides high confidence in the predictive 3D map. The ML approach offers several advantages over conventional geophysical inversion:

- **Speed and Efficiency:** Once trained, the model can generate predictions on a new grid instantaneously.

- **Transparent Performance Metrics:** The use of standard regression metrics (R^2, RMSE) provides a clear, quantitative measure of model accuracy.

- **Data-Driven Non-linearity:** The model captures complex, non-linear relationships in the data without requiring a pre-defined physical law, potentially revealing subtleties missed by linearized inversions.

A key limitation is that the model is purely data-driven and interpolative; it does not incorporate physical laws like Poisson's equation for gravity fields. Future work could integrate this ML-predicted model as a starting or reference model for full physical inversion, potentially leading to more geologically realistic and computationally efficient results.

4.3 Comparison with Previous Studies

The presented model corroborates the findings of previous geophysical studies that inferred a significant magmatic body beneath Deception Island. For instance, seismic tomography studies have identified low-velocity zones in the mid-crust (Zandomeneghi et al., 2009), and magnetotelluric surveys have revealed conductive bodies interpreted as hydrothermal and magmatic fluids (Catalán et al., 2014). This ML-driven gradient model provides an independent and highly detailed spatial delineation of this magmatic system, refining its geometry and highlighting its internal heterogeneity. The depth and extent of the identified anomaly are in broad agreement with these independent geophysical constraints, strengthening the overall geological interpretation.

5 Conclusion

This study successfully demonstrates the power of machine learning as a primary tool for 3D geological modeling. By applying ensemble regression algorithms to a geophysical gradient dataset from Deception Island, the following was achieved:

1. The development of a high-performance predictive model ($R^2 = 0.92$) for subsurface geophysical gradients.
2. The construction of a detailed 3D model that reveals a large, complex, low-density anomaly at a mean depth of 18 km, interpreted as the island's central magmatic plumbing system.
3. The provision of a new, quantifiable framework for interpreting volcanic structures that complements and enhances traditional geophysical methods.

The methodology presented is transferable to other volcanic systems worldwide where dense geophysical data are available. Future work will focus on integrating multiple geophysical datasets (seismic, magnetic, MT) within a multi-input ML framework and combining the ML predictions with physical inversion to create a unified, physically consistent

model of the Deception Island magmatic system. This approach marks a significant step forward in quantitative volcanic hazard assessment.

Data Availability Statement

The geophysical gradient dataset used in this study is from the published work of Grad1997 and was obtained from their supplementary materials. The machine learning code, processed datasets, and analysis scripts developed for this research are publicly available in the GitHub repository: https://github.com/humbersanme/MachineL_Magma. Additional results and 3D model data are available from the corresponding author upon reasonable request.

References

Bergen, K. J., Johnson, P. A., de Hoop, M. V., & Beroza, G. C. (2019). Machine learning for data-driven discovery in solid earth geoscience. *Science*, *363*(6433), eaau0323.

Cashman, K. V., Sparks, R. S. J., & Blundy, J. D. (2017). Vertically extensive and unstable magmatic systems: A unified view of igneous processes. *Science*, *355*(6331), eaag3055.

Catalán, M., Martos, Y. M., Galindo-Zaldívar, J., et al. (2014). Initial stages of oceanic spreading in the bransfield strait from geophysical evidence. *Marine Geophysical Research*, *35*(3), 283–297.

Geyer, A., Álvarez-Valero, A. M., Gisbert, G., Aulinas, M., Hernández-Barreña, D., Lobo, A., & Marti, J. (2019). Deciphering the evolution of deception island's magmatic system. *Scientific Reports*, *9*(1), 373.

Grad, M., Guterch, A., & Środa, P. (1997). Upper crustal structure of deception island area, bransfield strait, west antarctica. *Antarctic Science*, *9*(2), 201–208.

Pedrera, A., Galindo-Zaldívar, J., Ruiz-Constán, A., et al. (2020). Strain partitioning in the bransfield strait, west antarctica. *Journal of Geophysical Research: Solid Earth*, *125*(7), e2019JB019017.

Smellie, J. L. (2002). The 1969 subglacial eruption on deception island (antarctica): Events and processes during an eruption beneath a thin glacier and implications for volcanic hazards. *Geological Society, London, Memoirs*, *22*(1), 187–212.

Zandomeneghi, D., Barclay, A., Almendros, J., Ibañez, J. M., Wilcock, W. S. D., & Ben-Zvi, T. (2009). Crustal structure of deception island volcano from p-wave seismic tomography: Tectonic and volcanic implications. *Journal of Geophysical Research: Solid Earth*, *114*(B6).