Article

Determining the Principal Azimuths of Valles Marineris: A Comparative Analysis of Bezier Spline and GIS Techniques

Dan James [*]

Article

Determining the Principal Azimuths of Valles Marineris: A Comparative Analysis of Bezier Spline and GIS Techniques

Dan James

Independent researcher; treleasemill@hotmail.com

Abstract: Valles Marineris (VM), the most prominent example of a Martian Valley Network, has been extensively studied for over 50 years, yet no detailed examination of the principal azimuths of the system exists. To address this, two methods are presented to precisely calculate the principal azimuths, a Bezier Spline analysis, and a GIS technique. The medial axis of the main canyon of VM was determined analytically from cubic polynomial splines fitted to 93 coordinate points along both north and south edges of the canyon. These splines were optimized, and medial axis points were calculated through numerical techniques that ensured orthogonality between the tangents of each spline and their connecting normal lines. 1,000 medial axis points were extracted, and various regression models constructed, including fitting to sinusoidal and cubic polynomial curves, achieving accuracies with R^2 values of 0.98 and 0.99, respectively. Principal azimuths were obtained using the sinusoidal equation with the slope of the tangent at any point x simply determined by the derivative of the curve's equation. This analytic approach was cross-validated by a GIS method (using QGIS software), where a vector medial axis was obtained which produced principal azimuths that agreed with values from the analytic study with a correlation coefficient of 1.00, and a p value of 6.43e-65. The findings demonstrate that an azimuthal framework can be rigorously constructed as a potential standard reference in VM geoscience, replacing less precise and ambiguous compass bearings with the accurate azimuths necessary for high-resolution spatial analysis for future investigations.

Keywords: Valles Marineris; cubic polynomial splines; multidimensional optimization; medial axis; valley network; mars; principal azimuths; GIS

Bezier Splines

1.1. Introduction

There is conjecture regarding the influence of water in shaping Valles Marineris (VM), the most prominent example of a Martian Valley Network, and interpreted as good evidence of prolonged surface water on Mars (Klokočník et al. 2023, Hynek et al. 2010, Hansjoerg et al. 2018)**, although the dominant hypothesis attributes the formation of VM to tectonics, megashears, subsidence, trough collapse, or a combination** (Brustel et al. **2017,** Fueten et al. **2017,** Andrews-Hanna **2012,** Gurgurewicz et al., 2022)**.**

Previous orientation research has focused on determining azimuths of localised individual features within the VM system, for example dike placement in eastern Coprates Chasma **(**Brustel et al. **2017),** however, despite extensive investigations over 50 years, no survey has to date established the principal azimuths of the VM system, which are essential for a precise understanding the directional forces that have shaped the region.

This study aims to fill this gap by using a Medial Axis Bezier Spline analysis, a method commonly employed in geomorphological modeling for delineating central trajectories (Wang et al., 2023), and cross-validating this approach with a more conventional, algorithmic geographic information system (GIS) technique to determine the principal azimuths of the main canyon. These two analysis methods are rigorously compared to provide statistical evidence confirming the accuracy and reliability of the derived medial axis and azimuthal orientations.

1.2. Main Canyon Definition & CRS

The VM canyon system has distinct morphological complexity with multiple subsidiary canyons, which are ancillary to the main canyon running parallel or subparallel to it, contributing significantly to the overall topology of the system. In this study, the term 'main canyon' is defined as the most pronounced geomorphological feature, trending west to east from Noctis Labyrinth in the west from c15,900 km longitude to Eos Chaos in the east at c18,700 km (see Figure 1, below).

While the main canyon's medial axis, as discussed in this paper, is not the midline of the collective features for the entire VM system, establishing the axial symmetry of the main canyon necessarily preserves congruent symmetry in any parallel or subparallel subsidiary canyon.

Reference images were obtained from the THEMIS daytime IR mosaic (100 m/pixel) and JMARS, with an analytical GIS primarily via ArcMap™. Equidistant Cylindrical Projection was used to preserve accurate distances along meridians via a CRS base: ESRI:103885 (Mars Equidistant 2000 Sphere). Accurate ellipsoid parameters were added with custom code: semimajor axis: 3396.2, semiminor axis: 3376.2.
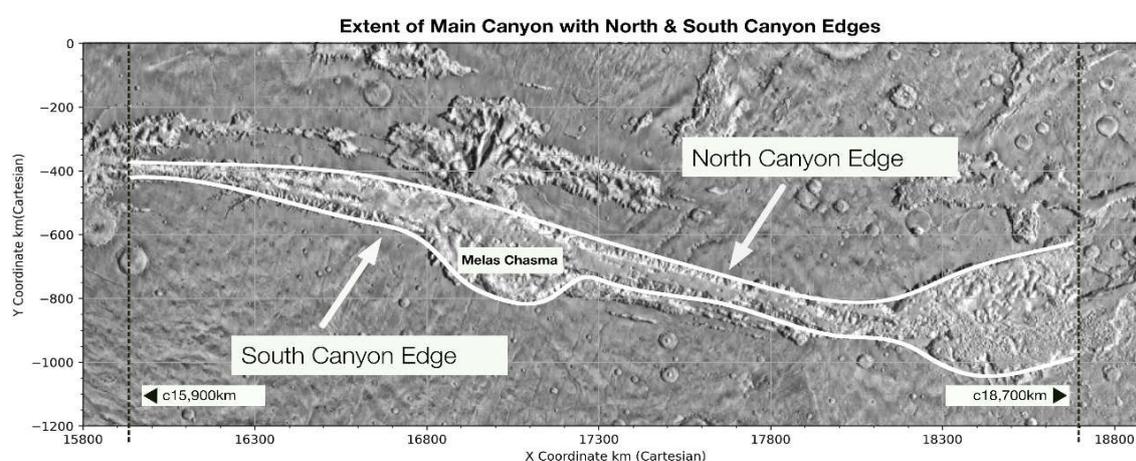

Figure 1. Shaded area represents the main canyon of the VM. The North and South Canyon edges are labeled and marked by white lines. Melas Chasma is at the midpoint of the main canyon. The approximate starting point of the main canyon is at 15,900 km in the west, terminating at approximately 18,700 km in the east.

1.3. Defining Edge Splines

Various methods exist to determine the medial axis of an object, such as Voronoi skeletons, distance transform methods and thinning algorithms such as wave propagation via grassfire algorithms. In this study, a Bezier curve approximation technique was used where univariate spline curves (Dierckx 1981, 1993**) were fitted to the North and South Canyon edges of VMs via the** UnivariateSpline **function from Python's SciPy library (10).**

A total of 93 coordinate points were identified on both North and South Canyon edges, with all points evenly spaced and with longitudinal distances preserved within the Cartesian kilometer-based CRS (ESRI:103885).

The edge splines of the main canyon were constructed with parametric cubic spline continuous functions, ensuring first and second derivative continuity and avoiding sharp changes in the curve's slope or curvature, which are common in lower-degree splines.

Denoting the splines for the North and South edges as *N(t)* and *S(t)*, where *t* is the parameter along the spline or arc length, the spline functions for each Canyon edge in a 2D plane can be written as:

$$N(t) = (x_1(t), y_1(t)) = \sum_{i=1}^{n} B_{i,3}(t) \cdot P_{1i}, \quad t \in [a, b] \qquad (1)$$

$$S(t) = (x_2(t), y_2(t)) = \sum_{j=1}^{m} B_{j,3}(t) \cdot P_{2j}, \quad t \in [a, b] \qquad (2)$$

where:

- $B_{i,3}(t)$ and $B_{j,3}(t)$ are the cubic B-spline basis functions and are defined over a knot vector that is locally supported.
- n and m are the numbers of control points for the North and South edge splines, respectively, determined through the spline fitting process.
- The interval [a, b] represents the domain of the parameter t.
- P_{1i} and P_{2j} are the control points for the North and South edge splines and define the initial shape and direction of the spline, which is then adjusted via the optimization process (see below). While control points do not generally lie on the curve, they act as weights in the combination of basis functions that define the curve.

1.4. Spline Optimization

Univariate spline interpolation involves a piecewise-defined cubic polynomial function that fits a set of data points in a single variable, with knots segmenting the dataset and individual polynomial functions fitted within each segment. The placement of knots and the overall smoothness of the curve in the UnivariateSpline algorithm are regulated by a key parameter: the smoothing factor (*s*).

When *s* is unspecified, UnivariateSpline defaults to place knots at every data point, effectively interpolating the data. However, specifying a value for *s* balances the spline's smoothness with its adherence to the data by a least-squares optimization process, where the spline coefficients are calculated to minimize the sum of squared residuals between the fitted spline and the data points, subject to the smoothness constraint imposed by *s*.

Given a set of data points (x_i, y_i), where $i = 1,2,\ldots,N$, and a spline function $S(x)$, the goal is to find the spline that minimizes the following objective function:

$$J = \sum_{i=1}^{N} [y_i - S(x_i)]^2 + \lambda \int [S''(x)]^2 \, dx \qquad (3)$$

where:

- The sum of the squared residuals, representing the fit of the spline $S(x)$ to the data points (x_i, y_i), (lower values indicate a better fit), is added to the integral of the square of the second derivative of the spline, which acts as a measure of the spline's smoothness. A smoother curve will have a lower value of this integral.
- λ represents the smoothing factor *s* in UnivariateSpline and is a parameter that balances the two objectives: smoothness of the spline and proximity to the data points.

This optimization problem is solved in UnivariateSpline via numerical methods, as curve fitting problems typically lack closed-form solutions. With the user-defined *s* value constraint, the algorithm iteratively adjusts the knots' positions and the coefficients of the spline to minimize J, capturing the essential trends of the coordinate data while mitigating overfitting.

1.5. Canyon Edge Spline Analysis

The sum of the squared residuals (SSR) shown in the legend of Figures 2 and 3 below, although conceptually distinct, has an important relationship with the smoothing factor *s* in the UnivariateSpline algorithm, which explains the high SSR value shown.
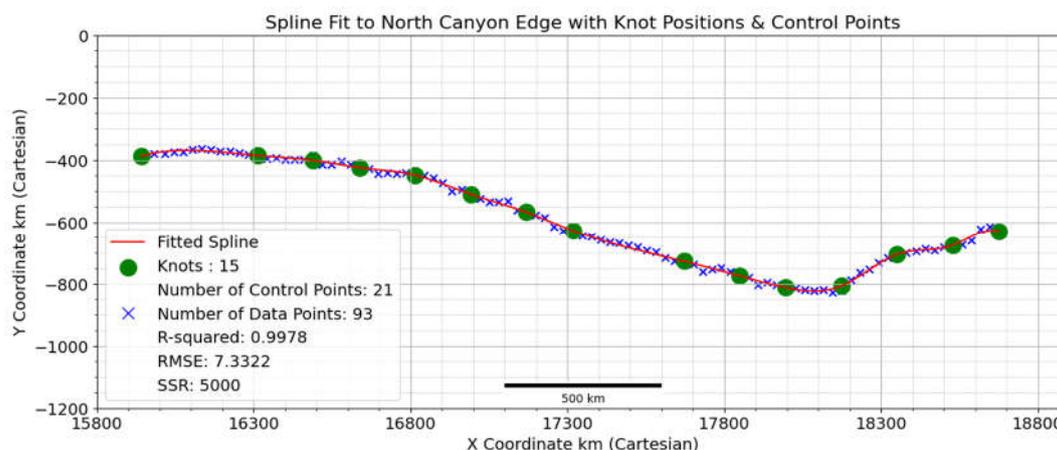
Figure 2. Spline fitted to the North Canyon edge showing 93 coordinate positions and Knot positions. The smoothing factor used was 5,000, resulting in an SSR of 5,000.
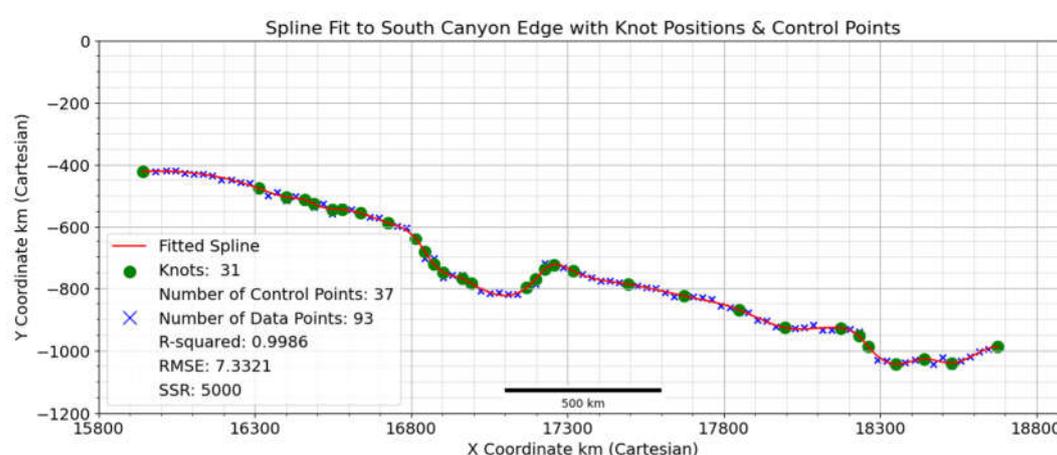

Figure 3. Spline fitted to the South Canyon edge showing 93 coordinate positions and Knot positions. The smoothing factor used was 5,000, resulting in an SSR of 5,000. The South Canyon edge with its higher variability has more generated knots and control points than the North Canyon edge (Figure 2).

The smoothing factor serves as an upper bound for the SSR, guiding the algorithm to adjust the spline's smoothness such that the SSR approaches but does not exceed s. Consequently, the final SSR of a spline often equals the value set for s. Canyon edges, which span >3,000 km, exhibit significant variability due to landslide extensions and subsidiary canyon intrusion, leading to potentially large residuals, cumulatively resulting in a higher SSR.

A higher smoothing factor is therefore required for the Canyon edge data, enabling the spline to adequately curve to the coordinate data's underlying trends and allowing for this higher SSR without being excessively constrained by variability.

SSR is a measure of the total deviation of the response values from the fit to the response values, and while a lower SSR may indicate a better fit, it is not a guide to how good the fit is relative to the spread and variability of the data, whereas the R^2 value, the coefficient of determination, indicates how well the independent variable explains the variability of the dependent variable by calculating the ratio of SSR to SST (total sum of squares) – $R^2 = 1 - SSR/SST$[1].

Both edge splines, which are based on a smoothing factor of 5,000 (SSR 5,000), returned R^2 values close to 1, implying that a significant proportion of the total variation in the data is captured and explained by the spline models with minimal deviation from the actual data points. The RMSE (root mean square error) for the charts shown was 7.33, which, in the context of the Canyon edge data

measured in kilometers, represents 7.33 km—a mean error equivalent of <0.24%, which is low when contrasted against the vast overall length of the Canyon edges (>3,000 km). To assess how different values of the smoothing factor affect the fit, numbers of knots and control points, several iterations were carried out for each Canyon edge, as shown in Tables 1 and 2 below:

Table 1. North Canyon Edge.

SSR/smoothing factor	Knots	Control Points	R²	RMSE
North Canyon Edge – Coordinate Points = 93				
10	84	89	1.0000	0.328
100	71	78	1.0000	1.036
1000	40	49	0.9996	3.277
5000	15	21	0.9978	7.332
10,000	8	14	0.9956	10.369

Table 2. South Canyon Edge.

SSR/smoothing factor	Knots	Control Points	R²	RMSE
South Canyon Edge – Coordinate Points = 93				
10	85	91	1.0000	0.3279
100	72	79	1.0000	1.037
1000	47	55	0.9997	3.277
5000	31	37	0.9986	7.332
10,000	15	21	0.9972	10.368

[1] **In UnivariateSpline documentation, the SSR is only required to be ≤ *s*, so for some splines, it can theoretically be less. For all splines detailed in this paper it was equal.**

1.6. Medial Axis Extraction

The medial axis spline is a path equidistant from two edge splines at any point along its length. If the medial axis spline is denoted by $M(t)$, a point $M(t_i)=(x_m(t), y_m(t))$ on this spline should satisfy the following condition where the distance from any point on the medial axis spline to the nearest point on each edge spline is equal:

$$\| M(t) - N(t) \| \quad = \quad \| M(t) - S(t) \| \tag{4}$$

Finding midpoints between two splines by simply averaging the distance on a vertical bisector so that $Mid_y =(North_y + South_y)/2$, in the context of VM, is not an optimal method for finding the medial points. The VM system trends ≈20° from northwest to southeast, curving in multiple ways; therefore, line segments joining the Canyon edge splines requires orthogonality to fully capture the curvature and direction of the splines.

The goal is to find an equal number of points along each of the Canyon edge splines where lines joining pairs of points are perpendicular (normal) to *both* splines at their respective points of tangency and then find the midpoints $M(t_i)$ of these lines. This multidimensional optimization problem has a 2D parameter space that is required to optimize two variables (t_{north}) and (t_{south}), which are parameters along North and South splines at intersection points with a line segment between both splines (11). The optimization sequence can be written as:

1. ***Smoothing***: Apply *s* = 5000 to both North and South splines.

2. ***Point Pair Identification***: Generate a sequence of point pairs {$N(t_i, S(t_i)$} for *i=1,2,…,n* along the entire length of the North and South splines.

3. *Slope threshold:* Calculate the slope at each point on the southern spline as the first derivative of the spline at each point, denoted as Slope_$S(t_i)$. Define a threshold value for the slope as slope_threshold

4. *Conditional processing:* For each point pair $\{N(t_i, S(t_i)\}$, check the slope at $S(t_i)$:

    If: Slope_$S(t_i)$< slope_threshold, proceed with subsequent steps (5--10).
    Else: Skip to the next point pair $\{N(t_i, S(t_i)\}$ and proceed with subsequent steps (5 to 10)

5. *Vector:* Define L as the vector connecting a point on the North spline at (t_{north}) to a corresponding point on the South spline at (t_{south}) expressed as L= $S(t_{south})-N(t_{north})$.

6. ***Tangent Vectors:*** Calculate the tangent vectors at t_{north} and t_{south} as TN (t_{north}) and TS (t_{south}), respectively.

7. **The perpendicularity function *f* (*t_{north}*, *t_{south}*) is:**

$$f(t_{north}, t_{south}) = [L \cdot TN(t_{north})]^2 + [L \cdot TS(t_{south})]^2$$

8. **This function quantifies the squared sum of the dot products of the vector L with the** vectors at TN (t_{north}) and TS (t_{south}).

9. ***Multidimensional Optimization:*** adjust t_{north} and t_{south} to minimize ***f* (*t_{north}*, *t_{south}*)** to zero and therefore the perpendicularity of L with both TN (t_{north}) and TS (t_{south}).

10. ***Midpoint Calculation:*** Use the perpendicular bisector to calculate the midpoint $M(t_i)$ of L, which represents the optimized medial axis point for each t_i

11. ***Iteration:*** Repeat the process for 1,000 points to compile the set $\{M(ti)\}$.

*(This sequence forms the basis of the logic to produce Python code, which, in this study, was written to determine the Medial Axis with the SciPy-optimize **function from Python's SciPy library (11)).***

Normal line equivalence at points (t_{north}) *and* (t_{south}) requires the tangents of the curve at each point to be parallel. For the original dataset of the Canyon edge coordinate points, this condition is potentially difficult to meet, as there may be significant direction changes in the Canyon edges.

However, the smoothing factor used in the first step of the above procedure helps to regularize the data by reducing noise and minor fluctuations and facilitates the multidimensional optimization process by making the behavior of the spline (especially its tangents) more consistent and less prone to abrupt changes. Too much variability in the coordinate data might otherwise mislead an optimization algorithm to either prevent convergence or alternatively produce a local solution that is not globally optimal. For consistency in this study, a smoothing factor of 5,000 was used for both the North and South splines.

A specific conditionality function is also integrated into the above sequence (steps 3 & 4) to address the anomaly at the midpoint of the South Spline, where the region of Melas Chasma imposes pronounced southward semicircular extrusions along the canyon edge and spline representation (see Figure 3, above). The steeper tangent slopes in this region potentially lead to complexities in the algorithm's ability to resolve accurately. To avoid this, the algorithm employs a threshold-based decision process where if the calculated slope at any given point exceeds a predetermined angle – set for this study at ±25 degrees – the algorithm omits processing for that particular point pair, advancing instead to the subsequent pair. Given the localized nature of the Melas Chasma steep tangent slopes, this adjustment impacts only a negligible portion of the overall spline representation.

The determination of each coordinate point $M(t_i)$ in the saved array $\{M(ti)\}$ can be described more formally by optimizing the objective function J_1, which enforces equidistance and orthogonality constraints for points on the medial axis spline:

$$J_1 = \sum_{i=1}^{N} \left[(d_1(t_i) - d_2(t_i))^2 + \lambda_1 O_1(t_i) + \lambda_2 O_2(t_i) \right]$$

(6)

where:

- $d_1(t_i)$ and $d_2(t_i)$ are the distances from $M(t_i)$ to $N(t_i)$ and $S(t_i)$, respectively.
- $O_1(t_i)$ and $O_2(t_i)$ are measures of the deviation from orthogonality at $M(t_i)$ to $N(t_i)$ and $S(t_i)$, respectively. This is computed as the sum of the squares of the dot product of the vector L with the tangent vectors T_N and T_S.

- λ_1 and λ_2 are weighting parameters for the importance of equidistance and orthogonality in the objective function.
- t_i represents the parameter values at which the spline and constraints are evaluated.
- N is the number of points along the spline where calculations are made.

The array of medial axis coordinate points $\{M(t_i)\}$ was used to construct a medial axis spline **in a 2D plane** via the same method as the Canyon edge splines were created via the UnivariateSpline algorithm, formally expressed as:

$$M(t) = (x_m(t), y_m(t)) = \sum_{k=1}^{p} B_{k,3}(t) \cdot P_{mk}, \quad t \in [c, d]$$

(7)

where:

- $M(t)$ represents the medial axis spline.
- $B_{k,3}(t)$ are the cubic B-spline basis functions for the medial axis spline.
- P_{mk} are the control points of the medial axis spline.
- p is the number of control points for the medial axis spline, which is determined through the spline fitting process.
- The interval $[c, d]$ represents the domain of the parameter t for the medial axis spline.

For the medial axis spline optimization the objective function J_M balances the dual goals of fitting the medial axis spline closely to the determined medial axis points while also maintaining a smooth curve controlled by the choice of the smoothing parameter:

$$J_M = \sum_{i=1}^{q} [y_i - M(x_i)]^2 + \lambda \int [M''(x)]^2 dx$$

(8)

where:

- (x_i, y_i) are the coordinate points that the medial axis spline is intended to fit.
- q is the number of data points.
- The sum of the squared residuals, representing the fit of the medial axis spline to the data points (x_i, y_i), is added to $\int [M''(x)]^2 dx$, the integral of the square of the second derivative of the spline, and acts as a measure of the spline's smoothness.
- λ is the smoothing parameter that balances the trade-off between the fit to the data points and the smoothness of the spline (represented by the smoothing factor s in UnivariateSpline).
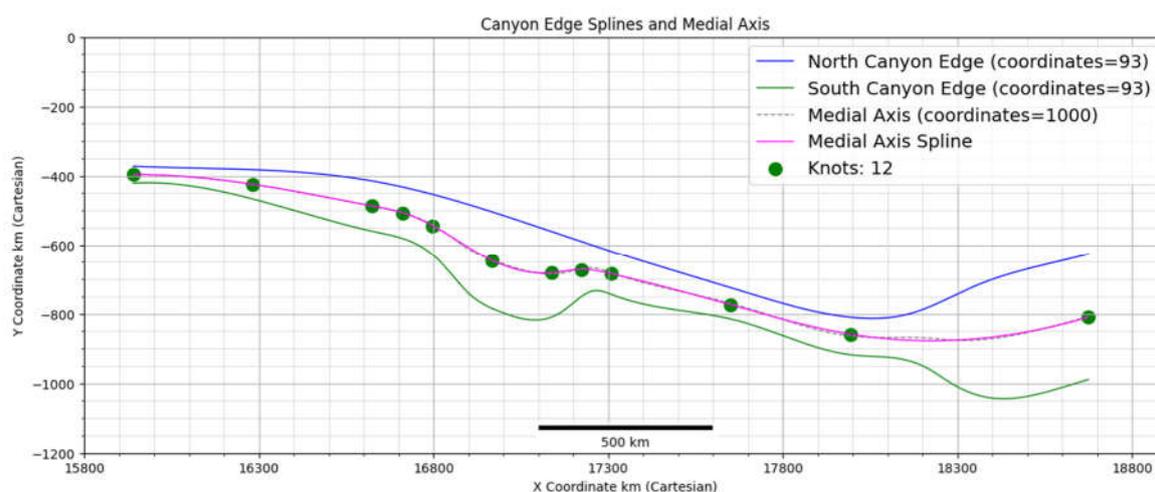
1.7. Medial Axis Spline Charts

Figure 4. Medial axis spline fitted to 1,000 median axis points. The smoothing factor of 5,000, resulted in 12 knots and an SSR of 5,000.
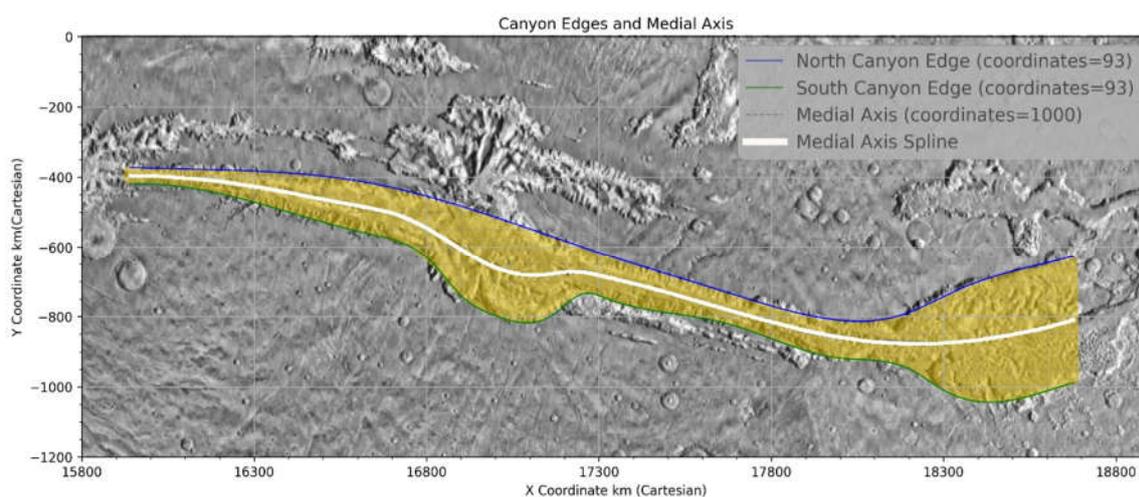

Figure 5. Positions relative to a THEMIS daytime IR mosaic composite image of Valles Marineris of the fitted splines (North, South & Medial). The shaded area represents the extent of the VM Main Canyon.

Regression Analysis

2.1. Model: Sinusoidal

Various curve fitting models were considered for fitting to the median axis coordinate data of 1000 points. A visual inspection of the medial axis plot (see Figure 6} suggested a possible sinusoidal fit to the data, and an initial guess for this fit, its determination and comparison to final fit values, is shown in Table 3 below:
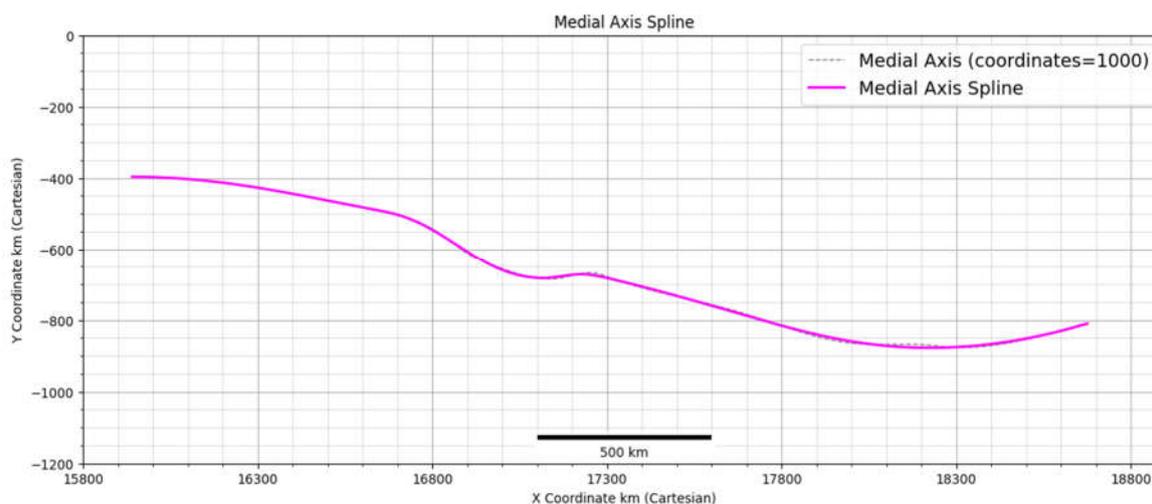
Figure 6. Chart showing the medial axis spline based on 1,000 medial axis coordinates, without showing canyon edge splines, for evaluation of potential medial axis symmetries.

Table 3. Methodology for Initial Guess (Sinusoidal).

	Initial Guess	Methodology for Initial Guess	Final Fit Values	% Change
Parameter				
Amplitude (A)	-236	A ≈ (max(y) - min(y))/2	-239.01	-1.27%
Angular Frequency (ω)	-0.0018	ω ≈ 2π * (Cycles/Range of x)	-0.00128	28.90%
Phase Shift (φ)	261.02	Align first peak with model	251.1	-3.80%
Vertical Shift (B)	-631.57	B ≈ mean(y)	-624.05	1.20%

To optimize the sinusoidal model parameters, the iterative least-squares Levenberg–Marquardt method was used, implemented via SciPy's 'optimize' and 'curve_fit' functions. This method minimizes the residual sum of squares, a common objective function (loss function) in regression analysis. The initial parameter estimates, shown in Table 3 above, resulted in reliable model convergence, with final values closely matching the initial guess. For comparative analysis, linear, geodesic path, and cubic polynomial models were included, as shown in both Figure 7 and Table 4 below. The fitted sinusoidal curve was determined as:

$$y \ = 239.01x \sin(-\ 0.0013\ x\ (x + 251.1)) - 624.05 \tag{9}$$

2.2. Model: Cubic Polynomial

Cubic polynomial models are widely used in situations where a balance is needed between model complexity and the ability to accurately represent the underlying trends of the data without the risk of overfitting associated with higher degree models. The loss function for *n* data points can be defined as:

$$L(a, b, c, d) = \sum_{i=1}^{n}(y_i - (ax_i^3 + bx_i^2 + cx_i + d))^2 \tag{10}$$

This loss function was minimized by a least squares algorithm implemented in Python by scipy.optimize.curve, and the fitted cubic polynomial was determined for 1000 coordinate points as:

$$y = 0.0000x^3 + -0.0035x^2 + 59.0218x + -332751.1942 \tag{11}$$

2.3. Model: Linear

VM is typically described as exhibiting linear characteristics or even, more casually, as being a linear feature. To critically evaluate this assertion, a linear model was included as a reference or baseline in this analysis. This approach allows for a direct comparison with other models to demonstrate that a more complex model structure than a purely linear one, better represents the underlying data patterns.

2.4. Model: Geodesic

A geodesic model was chosen to evaluate any significant divergence from the geodesic trajectory by the medial axis, which would indicate influences or additional factors beyond simple distance minimization. Substantial deviations from the geodesic require a causal explanation, whether stochastic, deterministic or a combination (e.g., tectonic split in the crust, water/lava erosion, or other causal factors).

The 'Geod' object from the 'pyproj' library, which is optimized for handling latitude and longitude data for distance and angle calculations on a spherical model (Mars), was used to define the geodesic path. Longitude values were normalized to the 0-360° range in a standard positive east longitudinal system, and paths were optimized to prevent wrapping around the sphere. The start and end coordinates were set at -6.9° latitude, 270° longitude, and -13.8° latitude, 316° longitude, respectively.

Based on Mars's mean radius of 3396.2 kms, the 'Geod' object generated 100 intermediate coordinate points along the calculated great circle route. For quantitative analysis and plotting in linear units, a conversion from latitude and longitude to kilometers was conducted via a 'numpy' function applying the Equidistant Cylindrical projection (ESRI:103885). The latitude and longitude were adjusted relative to a central point (defaulted to 0°) and then converted to radians multiplied by the mean radius of Mars to compute the Cartesian coordinates, where x represents the horizontal distance and y represents the vertical distance from the central point in a planar projection of Mars. The geodesic path, converted from original latitude/longitude coordinates to Cartesian coordinates, was modeled via a quadratic polynomial to accommodate the expected slight curvature. The best-fit polynomial, determined through least squares fitting, is given by the following equation:

$$y = 2.87482 \times 10^{-5}\, x^2 - 1.18077x + 11168.8393 \tag{12}$$
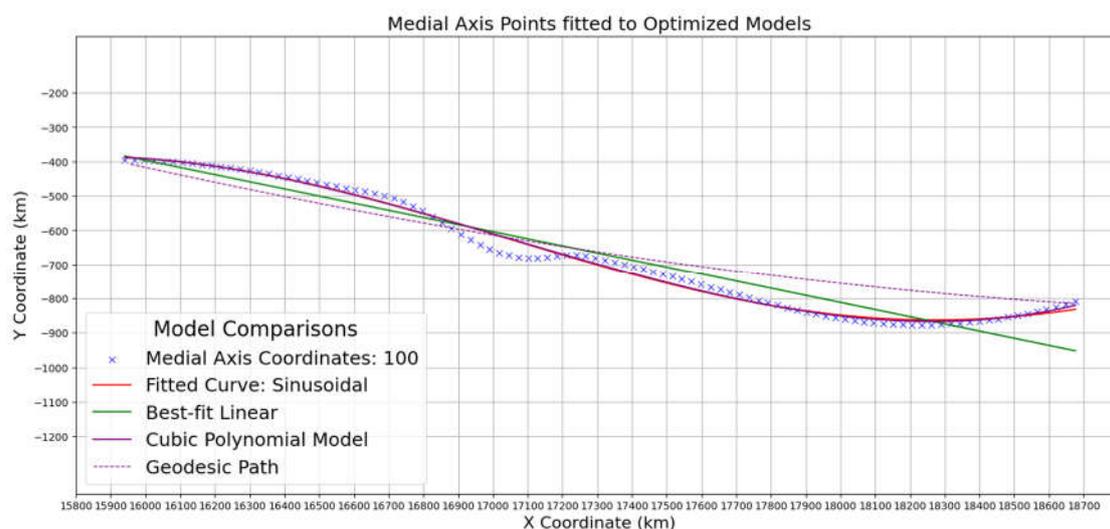
2.5. Model Comparison - 100 Coordinate Points

Figure 7. Chart showing 4 Models fitted to medial axis coordinates (100).

Table 4. Comparison of regression metrics.

	Log-Likelihood	AIC	BIC	SD of Residuals	R²	Adjusted R²	RMSE
Model							
Cubic Polynomial Model	-425.196	858.392	868.812	17.082	0.990	0.990	16.996
Sinusoidal Model	-429.416	866.832	877.253	17.818	0.989	0.989	17.729
Linear Regression Model	-519.145	1042.290	1047.500	43.707	0.935	0.935	43.488
Geodesic Path	-552.370	1108.740	1113.950	35.240	0.876	0.873	60.319

The plot shown in Figure 7 above, is not at a sufficient scale to distinguish separate sinusoidal and cubic polynomial fitted curves - there is considerable overlap. However, a clearer way of discriminating between each model can be seen in Table 4 above, where a dataset of 100 medial axis points was examined according to a variety of standard metrics. The cubic polynomial model's ability to adapt to more complex patterns, although numerically showing the best performance, does not provide a significant advantage, suggesting that the complexity it adds may not be necessary to capture the primary dynamics of the data, thereby tending to validate the simpler sinusoidal model.

Cubic polynomial models have more coefficients for each degree of the polynomial, allowing a fit that captures more fluctuations in the data than a sinusoidal model, which is characterized by a single frequency, amplitude, and phase. The Akaike information criterion (AIC) and the Bayesian information criterion (BIC) are standard techniques for comparing models with different complexities, with the AIC penalizing complexity by adding twice the number of parameters (2k) and the BIC further multiplying the number of parameters by the logarithm of the data points (log(n) *k). However, in both cases, while complexity is penalized, a good fit is always rewarded, and if the polynomial model's complexity leads to a better fit, this can suggest improved performance despite a greater number of parameters.

2.6. Kernel Density Estimate for Residuals

A kernel density estimation plot of all four models (Figure 8, below),also reveals a close match for both sinusoidal and cubic polynomial models, with a unimodal distribution of sharp high central peaks for both indicating a high concentration of residuals near zero and therefore a good fit with

minimal error. The other two comparative models, linear and geodesic path, display pronounced bimodal distributions, both have two distinct residual groups where the models are underfitting or overfitting and cannot capture the curvature or pattern in the data.
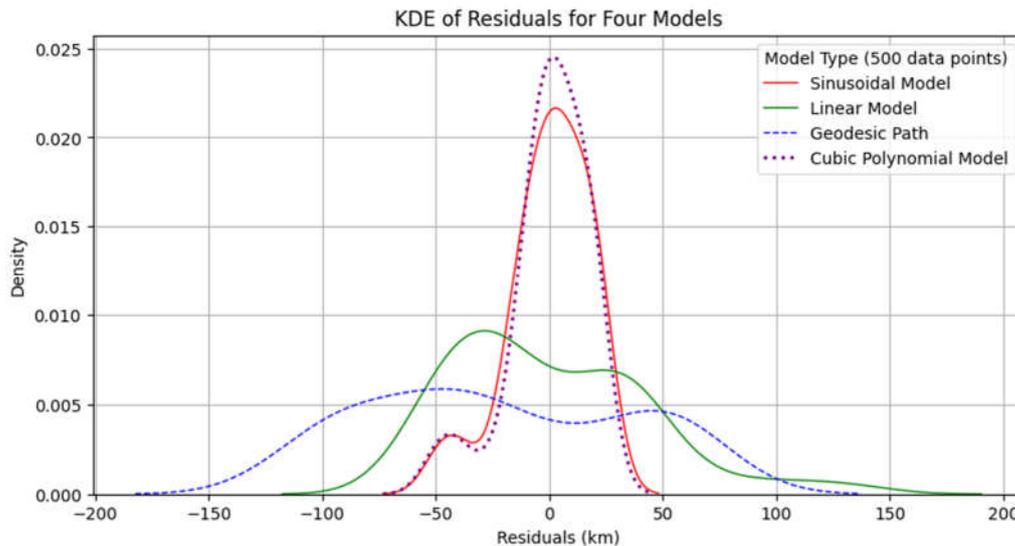

Figure 8. KDE of the residuals for the four models.

2.7. Data Scaling

To evaluate model robustness and efficiency across a spectrum of data volumes, an AIC analysis was used based on a range from 25 coordinate points to 1,000 points, which were extracted from the original medial spline calculation sinusoidal curve. The linear trend on Figure 9 below, shows that as the dataset size increases, the complexity of the models scales up linearly, indicating no sudden increase in complexity, which might otherwise imply overfitting or poor scalability.
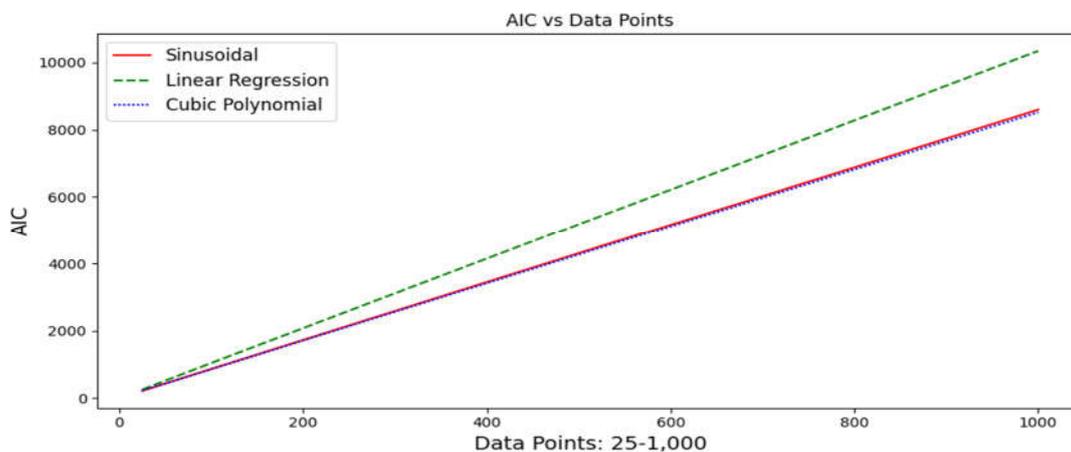

Figure 9. Data scaling for 25–1,000 coordinates compared with the AIC value.

The cubic polynomial model consistently maintains a marginal advantage over the sinusoidal model, yet both overlap substantially at the scale of the chart (Figure 9). Both significantly outperform the linear model used as a reference.

2.8. Azimuths

Determining the principal azimuths of the main canyon is crucial for understanding the geomorphology of Valles Marineris. By fitting a sinusoidal curve to the medial axis coordinate data, the azimuths can be efficiently calculated using the derivative of the curve, with the slope of the tangent at any point x simply determined by the derivative of the curve's equation. Using this

method, potentially unlimited datasets of azimuths can be generated for the curve; however, only 20 azimuths were generated for display clarity on a chart (see Figure 10, below). A total of 1,000 medial axis coordinates were curve fitted to a sinusoidal function to derive the best fit equation:

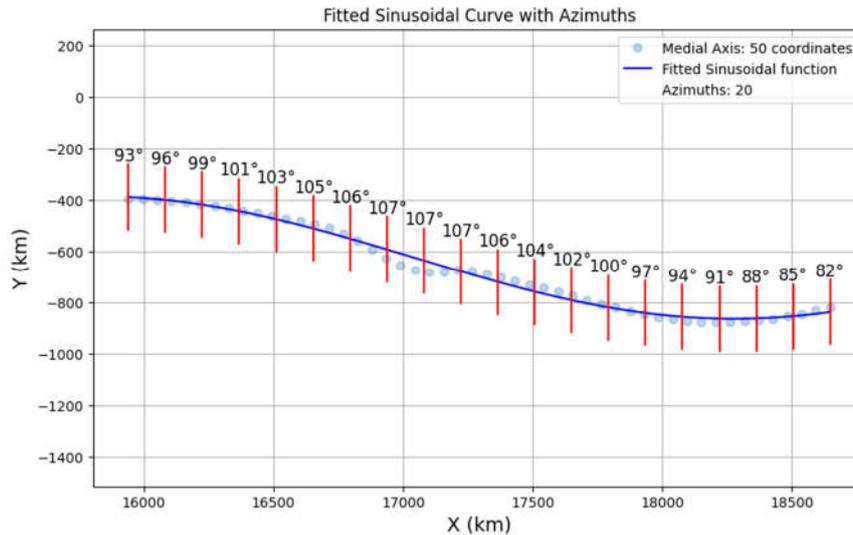$$y = 239.01x \sin(-0.0013\,x\,(x + 251.1)) - 624.05 \quad (13)$$

Figure 10. Extraction of 20 azimuth values in a Cartesian CRS.

To find the slope $y'(x)$ at any point, we differentiate $y(x)$ with respect to x, using standard differentiation techniques (product rule and chain rule), the derivative of the sinusoidal function, which represents the slope at any point x, is found to be:

$$\begin{aligned} y'(x) = {}& 239.01 \sin(-0.0013x(x + 251.1)) \\ & - 239.01x \cdot 0.0013\,(2x + 251.1) \cos(-0.0013x\,(x + 251.1)) \end{aligned} \quad (14)$$

Assuming an initial position (x_0, y_0), the calculated slope of the tangent is converted to an angle via the arctangent function, which yields an angle θ in radians from the positive x-axis:

$$\begin{aligned} \theta_0 = {}& \arctan\,(239.01 \sin(-0.0013x_0\,(x_0 + 251.1)) - 239.01x_0 \cdot 0.0013\,(2x_0 + 251.1) \\ & \cos(-0.0013x_0\,(x_0 + 251.1))) \times 180/\pi \end{aligned} \quad (15)$$

To convert this Cartesian angle into an azimuth (north-based bearing), the coordinate system is rotated by $90° - \theta_0$, and to ensure that all azimuths are positive and appropriately normalized, the azimuth is adjusted via the modulo operation: azimuth = (azimuth + 360) mod 360. For the iteration of azimuths, the change in azimuth over distance $\triangle x$ uses the derivative:

$$\begin{aligned} \triangle\theta = {}& (239.01 \sin(-0.0013x(x + 251.1)) - 239.01x \cdot 0.0013\,(2x + 251.1) \\ & \cos(-0.0013x\,(x + 251.1)))\triangle x \end{aligned} \quad (16)$$

with the next azimuth value predicted by:

$$\theta_{n+1} = (\theta_n + \triangle\theta)\ \bmod\ 360 \quad (17)$$

and to ensure that θ_{n+1} is positive:

$$\theta_{n+1} = (\theta_{n+1} + 360)\ \bmod\ 360 \quad (18)$$

where:

- x: Independent variable representing the horizontal coordinate.
- $y(x)$:   The sinusoidal function describing the vertical coordinate as a function of x.
- $y'(x)$:   The derivative of $y(x)$ represents the slope of the function at any point x.
- θ: The azimuth angle in degrees, measured clockwise from north.
- θ_0:   The initial azimuth angle at the starting position.
- $\triangle\theta$:   The change in the azimuth angle over a distance $\varDelta x$

- Δx: A change in the horizontal coordinate x.

2.9. *Implications of the Azimuth Calculations*

By applying the method described in section 2.8, detailed azimuth data along the entire length of the main canyon can be generated. This high-resolution azimuthal information is crucial for:

- *Analyzing Structural Trends:* Understanding how the orientation of the canyon changes provides insights into tectonic stresses and faulting patterns.
- *Comparing with Geological Features:* Correlating azimuth variations with specific geological formations may reveal relationships between structural orientation and depositional or erosional processes.

Advantages of the Method:

- *Efficiency:* The analytical approach allows for rapid computation of azimuths without the need for extensive GIS processing.
- *Precision:* Calculations are based on a mathematical model fitted to the data, reducing potential errors from manual measurements.
- *Scalability:* The method can be applied to any number of points along the canyon, facilitating studies at different resolutions.

3.1. *GIS Validation*

In addition to the above detailed analytic techniques, a computational GIS approach was also undertaken via QGIS software. The aim was to cross-validate the analytic results, particularly the azimuth calculations, with azimuths derived algorithmically and in an alternative CRS geodetic coordinate system. Several tools and plugins were used to automatically detect edges and vectorize. A brief description of the workflow in the QGIS is as follows:

- **Obtain high-resolution DEM data (JMARS) and georeference with CRS ESRI:107971 (Mars 2000 Sphere)**.
- **Geotrace North and South Canyon edges of the main canyon** with an edge detection algorithm and appropriate threshold values (Canny Edge Detection)
- **Buffer each raster canyon edge, merge raster layers to vectors, extract vertices, and then process with the Voronoi polygon tool.**
- **From the vectorized Voronoi skeleton extract the medial axis line**
- **Segment medial axis vector via a field calculator and derive azimuths for each segment**

3.2. *GIS/Analytic Comparison*

Two sets of azimuths, one from the analytic approach (Cartesian) and one from the GIS methodology (Geodetic), (Figures 10 and 11 above), were compared to assess the accuracy and correlation of both methods. 50 azimuths were extracted for each set (Table 5, below).
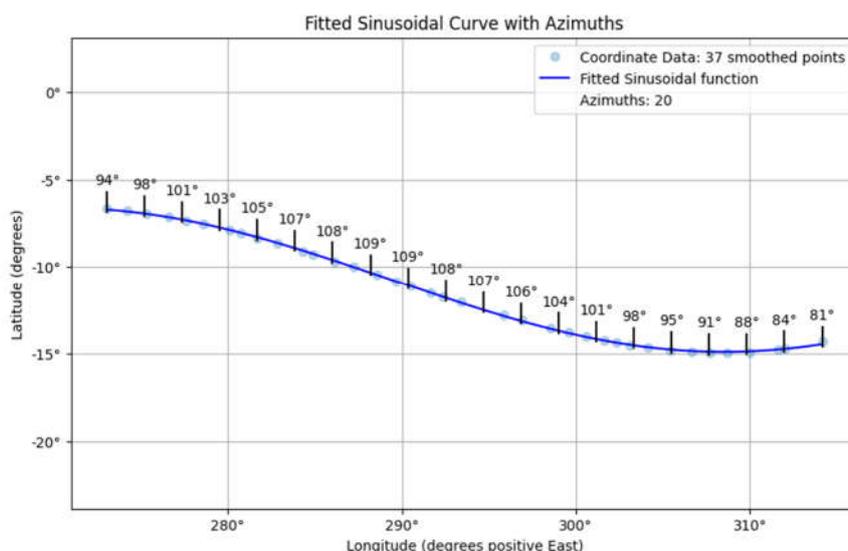
Figure 11. Extraction of 20 azimuth values in a Geodetic CRS. Medial axis spline smoothed with an SSR of 5,000.

Table 5. Mean summary and correlation metrics.

Mean Summary

Metric	Geodetic Azimuths	Cartesian Azimuths	Comparison
Mean (degrees)	100.25°	99.07°	Geodetic Mean slightly higher
Standard Deviation (degrees)	7.95°	7.16°	Similar variability

Correlation Metrics

Metric	Value	Interpretation
Correlation Coefficient (r)	1.00	Perfect Linear relationship
p value (Correlation)	6.45e-65	Highly significant correlation
Paired t test (t-statistic)	9.590	Significant difference in means
Paired t test (p value)	7.94e-13	Highly significant difference

The azimuths derived independently from the geodetic and Cartesian CRS exhibit a high degree of similarity; however, a slight difference in the mean azimuths was observed (Geodetic: 100.25 °, Cartesian: 99.07 °). This variation can be attributed to several factors inherent to the different methods employed. Aside from potential CRS projection transformation discrepancies (which can introduce minor errors due to the approximations used in the conversion processes), the principal way discrepancies are likely to have occurred is with the collection of data for each set.

The **geodetic azimuths** were calculated algorithmically via the GIS program QGIS, whereas the **Cartesian azimuths** were determined analytically via Bezier splines. Slight discrepancies in the edge detection of canyons and spline derivation from selected coordinate points on canyon edges affect the final azimuth calculations. However, the significant correlation between the two sets of azimuths (correlation coefficient = 1.00, p value = 6.43e-65), despite the methodological differences in obtaining them, indicates strong agreement between the two approaches.

Although the geometric and Cartesian azimuths differ slightly in terms of their exact values, they still follow the same general sinusoidal trend, which is reflected in the high correlation coefficient and demonstrates that the two datasets have a very strong linear relationship, i.e., effective cross-validation. The small difference in means, while statistically significant (paired t test: t statistic = 9.59, p value = 7.94e-13), indicates that even minor differences can be detected with high precision due to the paired nature of the data and the large sample size.

4. Discussion

This study derived the medial axis for the main canyon of the VM via both an analytic approach with Bezier splines and an algorithmic approach with GIS software (QGIS). The results were compared statistically to assess the correlation between the two methods by comparing the azimuths produced by both methods. The statistical analysis revealed a correlation coefficient for both sets of azimuths of 1.00 and a p value of 6.43e-65. This strong correlation validates the accuracy and reliability of both approaches in mapping the medial axis of the VM.

Determining the principal azimuths for Valles Marineris introduces the possibility of standardized directional references across future VM geoscience studies. Defining orientations with precise azimuthal degrees eliminates the ambiguity and imprecision of compass bearings, aligning with contemporary geospatial good practice, and paving the way for a universally applicable VM azimuth standard. Using azimuths helps minimize errors and misinterpretations caused by the vagueness of compass bearings, and is essential for detailed spatial analysis, with the clarity offered by azimuths ensuring that data is analyzed and understood consistently. Adopting a precise azimuthal framework could significantly improve the reproducibility of high resolution VM research findings and foster more robust cross-disciplinary collaborations.

In the regression analysis, four models were compared: sinusoidal, cubic polynomial, linear, and geodesic. The sinusoidal and cubic polynomial models provided the best fit, demonstrating significant alignment with the derived medial axis at distances greater than 3,500 km. The fit of the sinusoidal and cubic polynomial models over such a large distance is unexpected and suggests that underlying mechanisms influence the formation of the main canyon and its medial axis in a predictable manner.

The challenge lies in understanding which geological processes, such as planetary tectonics, might create patterns that align so accurately with these mathematical models. The data from the azimuthal calculations potentially support the role of **large-scale tectonic forces** in shaping the landscape, paralleling studies of Earth's own tectonic systems (Ghosh et al., 2019) or a possible megashear at a dichotomy boundary (Gurgurewicz et al., 2022). An accurate fitted curve model, particularly one that fits well over a large distance, implies a deterministic causation process, as such models typically represent regular, periodic, and predictable phenomena rather than stochastic, unpredictable phenomena. A potential focus for further research is to examine the principal azimuths of the main canyon for their applicability to all features of the VM system by quantifying the exact degree of congruent symmetry or parallelism in all subsidiary canyons or features.

5. Limitations

The selection of coordinate data on the Canyon edges to construct Bezier splines has inherent subjectivity; however, the choice of evenly spaced x-axis coordinates (longitudes) was precisely aimed at minimizing any subjective bias.

Selection effects are also inevitably involved in the choice of where precisely the main canyon of VM originates and ends (see Figure 2). VM trends west to east, and the first identifiable main canyon edges can be seen in the west, as they emerge from the heavily fractured area of the Noctis Labyrinth at approximately 15,900 km on the x-axis in a Cartesian CRS or 265° east of a positive east reporting system in a Geodetic CRS.

The termination of the main canyon edges is less clear in the east, where large outflow channels from VM have created a poorly defined chaotic terrain at Eos Chasma and Capri Chasma, leading to Aurorae Chaos. This study takes the last well-defined main canyon edge to be at approximately

18,700 km or 320° east, after which the canyon edges become highly irregular. This easternmost point can be seen as the boundary between the main canyon and the beginning of the outflow networks that eventually lead to the northern expanse of Chryse Planitia.

6. Conclusion

This research presents two methodologically robust approaches to calculating the medial axis of the main canyon of Valles Marineris. A Bezier approximation spline technique coupled with multidimensional optimization was compared with an algorithmic GIS approach. Both methods are highly reproducible and yield accurate values for the medial axis coordinates. In addition, both methods can determine the principal azimuths of the main canyon of the VM and statistically show a high degree of correlation, therefore effectively cross validating each approach. and introducing the potential for standardized directional references across future VM geoscience studies. That a sinusoidal model is an accurate fit to the medial axis is an unexpected result and suggests that deterministic processes were primarily involved in the formation of Valles Marineris.

The data that support the findings of this study are available from the corresponding author, D James, upon request: treleasemill@hotmail.com

References

1. Klokočník J, Kletetschka G, Kostelecký J, Bezděk A, 2023. Gravity aspects for Mars, Icarus, Volume 406, 2023, 115729, ISSN 0019-1035, https://doi.org/10.1016/j.icarus..115729.
2. Hynek, B. M., Beach, M., Hoke M. R. 2010. Updated global map of Martian valley networks and implications for climate and hydrologic processes. Journal of Geophysical Research: Planets, 115(E9). https://doi.org/10.1029/2009JE003548
3. Hansjoerg J. Seybold *et al. 2018.* Branching geometry of valley networks on Mars and Earth and its implications for early Martian climate. Sci. Adv.4, eaar6692 DOI:10.1126/sciadv.aar6692
4. Brustel, C. , Flahaut, J., Hauber, E., Fueten, F., Quantin, C., Stesky, R. Davies, G. R., 2017. Valles Marineris tectonic and volcanic history inferred from dikes in eastern Coprates Chasma,   Journal of Geophysical Research: Planets, 10.1002/2016 JE005231, 122, 6, (1353-1371).
5. Fueten, F., Novakovic, N., Stesky, R.,   Flahaut, J., Hauber, E., 2017. The Evolution of Juventae Chasma, Valles Marineris, Mars: Progressive Collapse and Sedimentation. Journal of Geophysical Research. Planets, 122 (11), pp.2223-2249. ⟨10.1002/2017JE005334⟩. ⟨hal-02130274⟩
6. Andrews-Hanna, J. C. 2012. The formation of Valles Marineris: 1. Tectonic architecture and the relative roles of extension and subsidence. Journal of Geophysical Research: Planets, 117(E3). https://doi.org/10.1029/2011JE003953
7. Gurgurewicz, J., Mège, D., Schmidt, F. *et al.* Megashears and hydrothermalism at the Martian crustal dichotomy in Valles Marineris. *Commun Earth Environ* **3**, 282 (2022).
8. https://doi.org/10.1038/s43247-022-00612-
9. Wang, S., Liu, C., Panozzo, D., Zorin, D., & Jacobson, A. (2023). Bézier Spline Simplification Using Locally Integrated Error Metrics. SIGGRAPH Asia.
10. Dierckx, P.,1981. An improved algorithm for curve fitting with spline functions, report tw54, Dept. Computer Science, K.U. Leuven,
11. Dierckx, P., 1993. Curve and surface fitting with splines, Monographs on Numerical Analysis, Oxford University Press, 1993.
12. scipy.interpolate.UnivariateSpline.   SciPy   v1.11.4   Manual https://docs.scipy.org/doc/scipy/reference/generated/scipy.interpolate.UnivariateSpline.
13. Optimization   and   root   finding.   SciPy   v1.11.4   Manual https://docs.scipy.org/doc/scipy/reference/optimize.html
14. Ghosh, A., Holt, W. E., & Bahadori, A. (2019). Role of Large-Scale Tectonic Forces in Intraplate Earthquakes of Central and Eastern North America. *Geochemistry, Geophysics, Geosystems*, 20(4), 2134-2156. https://doi.org/10.1029/2018GC008060

18

disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.