Using Calculus to Develop Pace Charts for Long-Distance Races with Altitude Graphs

Rodrigo Yuste

IB Mathematics — Internal Assessment
Department of Mathematics

Mathematics topic: Calculus (Differentiation & Integration)

Abstract

This paper presents a mathematical framework for generating scientifically informed pace charts for long-distance runners competing on courses with variable elevation profiles. Using elevation data collected from the 2024 United Airlines NYC Half Marathon, a 10th-degree polynomial regression model is fitted to represent elevation as a continuous function of distance, $h(x)$. Differential calculus (the power rule) is applied to derive the incline function $i(x) = dh/dx$, which is subsequently incorporated into an aggregate speed function $s(x)$. An adjustment coefficient g is solved analytically to ensure the runner maintains a user-specified average pace p across the full race distance. The resulting pace chart is validated against the target finishing time, demonstrating an error margin consistent with the Riemann-sum approximation used. The method is generalisable to any race with a published elevation profile and any target finishing time.

Keywords: polynomial regression, differentiation, integration, pace chart, elevation modelling, running, calculus.

Contents

1. Introduction

The eve of a competitive race presents every long-distance runner with a familiar challenge: how fast should one run at each point along the course in order to complete the race in a target time? In practice most runners rely on intuition or broad experience. On courses with significant elevation changes, however, this subjective approach is inadequate. Ascending sections demand a reduction in speed, while descending sections allow acceleration; without a principled method for quantifying these adjustments, runners either exhaust themselves on early climbs or fail to recover lost time on descents.

This paper was motivated by personal experience. On the morning of 17 March 2024 the author competed in the United Airlines NYC Half Marathon — a 21-kilometre event known for its varied elevation profile. Having prepared only an informal, subjective pace chart the night before, the author nonetheless completed the race with a personal best time, observing that many fellow competitors struggled with the gradient changes. The experience raised a natural question: could a rigorous mathematical method produce a superior, objectively derived pace chart?

To answer this question the paper develops a systematic approach grounded in differential and integral calculus. Elevation data for the 2024 NYC Half course were obtained from the Go&Race database and modelled via polynomial regression using the Desmos Graphing Calculator. The resulting elevation function $h(x)$ is differentiated to yield an incline function $i(x)$, which drives a speed function $s(x)$ designed to keep the runner's average pace equal to a pre-specified target p. An adjustment coefficient g is derived algebraically from the condition that the integral of $s(x)$ over the race distance equals the target pace multiplied by the race length.

The framework is presented in full generality and illustrated with a concrete example targeting a finishing time of 1 hour 30 minutes. A completed pace chart, validated against the target average, is provided in Appendix E. Limitations of the model are discussed in the conclusion.

2. Methodology

2.1. Data Collection and Visualisation

Elevation data for the 2024 United Airlines NYC Half Marathon were sourced from the Go&Race website. Readings were taken at 22 distance checkpoints spanning the full 21-kilometre course, recording altitude in metres above sea level alongside cumulative distance in kilometres. The dataset is reproduced in Appendix A (Table 1).

When plotted, the data reveal a broadly U-shaped profile: the course begins at approximately 27 m above sea level, rises briefly near the 2–3 km mark, descends to a minimum of 2 m at km 12, and then ascends sharply to finish at 29 m above sea level at km 21. This symmetry between the starting and finishing elevations motivated the selection of an even-degree polynomial for the regression model.

2.2. Polynomial Regression and Model Selection

Polynomial regression was conducted using the Desmos Graphing Calculator. Each coefficient of the target polynomial was treated as a free variable and Desmos's built-in least-squares routine minimised the residual sum of squares. Two candidate models were evaluated:

- A **10th-degree polynomial** (even degree), yielding $R^2 = 0.9101$.
- A **9th-degree polynomial** (odd degree), yielding $R^2 = 0.8481$.

The coefficient of determination R^2 measures the proportion of variance in the observed altitude values explained by the fitted model. Formally,

$$R^2 = 1 - \frac{\displaystyle\sum_k \left(y_k - \hat{y}_k\right)^2}{\displaystyle\sum_k \left(y_k - \bar{y}\right)^2}, \tag{1}$$

where y_k are the observed elevations, $\hat{y}_k$ the model predictions, and $\bar{y}$ the sample mean.

The 10th-degree polynomial explains approximately 91% of the variance in the elevation data compared to 85% for the 9th-degree polynomial. Furthermore, even-degree polynomials exhibit symmetric end behaviour — both tails either ascend or descend together — which is physically appropriate given that the course starts and ends at comparable altitudes. Increasing the degree beyond 10 produced no further improvement in R^2, confirming that a 10th-degree polynomial is the most parsimonious well-fitting model.

2.3. The Elevation Function $h(x)$

Substituting the Desmos-estimated coefficients (retained at full precision within Desmos; rounded to two significant figures here for readability) yields:

$$\begin{aligned} h(x) = {} & (8.5\times10^{-8})\,x^{10} - (9.1\times10^{-6})\,x^{9} + (4.1\times10^{-4})\,x^{8} - 0.010\,x^{7} + 0.16\,x^{6} \\ & - 1.4\,x^{5} + 7.7\,x^{4} - 22\,x^{3} + 27\,x^{2} - 7.0\,x + 27, \end{aligned} \tag{2}$$

where $h(x)$ is the elevation above sea level in metres and x is the distance travelled in kilometres. The full-precision coefficient set is given in Appendix D.

2.4. Deriving the Incline Function $i(x)$

The incline at any point is defined as the rate of change of elevation with respect to distance — i.e. the derivative of $h(x)$. Applying the power rule,

$$\frac{d}{dx}\left[a\,x^{b}\right] = b\cdot a\,x^{b-1}, \tag{3}$$

term by term to equation (2) yields:

$$\begin{aligned} i(x) = \frac{dh}{dx} = {} & (8.5\times10^{-7})\,x^{9} - (8.19\times10^{-5})\,x^{8} + (3.28\times10^{-3})\,x^{7} - 0.0721\,x^{6} \\ & + 0.96\,x^{5} - 7.15\,x^{4} + 30.8\,x^{3} - 66.6\,x^{2} + 53.4\,x - 7.02. \end{aligned} \tag{4}$$

The incline function is interpreted as follows: $i(x) = 0$ indicates a locally flat surface; $i(x) > 0$ indicates an ascent; and $i(x) < 0$ indicates a descent.

2.5. Constructing the Speed Function $s(x)$

Let p denote the runner's target average pace in km/h, computed from the target finishing time t (in hours) and the race length $x_f - x_0$ (in km):

$$p = \frac{x_f - x_0}{t}. \tag{5}$$

For a 21 km race completed in 1 hour 30 minutes:

$$p = \frac{21\,\text{km}}{1.5\,\text{h}} = 14\ \text{km/h}.$$

Speed at any distance x must be inversely related to the current incline. Specifically, a positive (uphill) incline reduces speed while a negative (downhill) incline increases it. This is modelled by:

$$s(x) = g\,p\left(1 - \frac{i(x)}{100}\right), \tag{6}$$

where the factor $i(x)/100$ converts the gradient (in metres per kilometre) to a dimensionless percentage adjustment, and g is an adjustment coefficient introduced to guarantee that the mean speed over the entire race equals exactly p.

2.6. Solving for the Adjustment Coefficient g

The requirement that the integral-average of $s(x)$ equals p reads:

$$p = \frac{1}{x_f - x_0} \int_{x_0}^{x_f} s(x)\, dx. \tag{7}$$

Substituting equation (6) into (7) and multiplying both sides by $(x_f - x_0)$:

$$(x_f - x_0)\, p = \int_{x_0}^{x_f} g\, p\left(1 - \frac{i(x)}{100}\right) dx. \tag{8}$$

Dividing by p and splitting the integral using linearity:

$$x_f - x_0 = g \int_{x_0}^{x_f} dx \;-\; \frac{g}{100} \int_{x_0}^{x_f} i(x)\, dx. \tag{9}$$

Because $i(x) = dh/dx$, the Fundamental Theorem of Calculus gives

$$\int_{x_0}^{x_f} i(x)\, dx = h(x_f) - h(x_0).$$

Therefore:

$$x_f - x_0 = g\left[(x_f - x_0) - \frac{h(x_f) - h(x_0)}{100}\right]. \tag{10}$$

Solving for g:

$$\boxed{g = \frac{x_f - x_0}{(x_f - x_0) \;-\; \dfrac{h(x_f) - h(x_0)}{100}}.} \tag{11}$$

For the NYC Half with $x_0 = 0$, $x_f = 21$, $h(0) = 26.7672\,\text{m}$, and $h(21) = -62.3827\,\text{m}$ (see Appendix D):

$$g = \frac{21}{21 - \dfrac{-62.3827 - 26.7672}{100}} = \frac{21}{21 - \dfrac{-89.1499}{100}} = \frac{21}{21 + 0.891499} \approx 1.01725. \tag{12}$$

2.7. The Aggregate Speed Equation

Substituting $g \approx 1.01725$ and $p = 14\,\text{km/h}$ into equation (6):

$$s(x) = 14.2415\left(1 - \frac{i(x)}{100}\right), \tag{13}$$

or, in fully general form combining equations (11) and (6):

$$s(x) = \frac{x_f - x_0}{(x_f - x_0) - \dfrac{h(x_f) - h(x_0)}{100}} \cdot p \cdot \left(1 - \frac{1}{100}\frac{dh}{dx}\right). \tag{14}$$

Equation (14) is the principal result of this paper. Given any race with a published elevation profile (encoded in $h(x)$) and any target average pace p, it outputs the instantaneous running speed at every point x along the course.

Worked example at $x = 5\,\text{km}$. From equation (4): $i(5) = 2.74$ (a gentle ascent). Substituting into (13):

$$s(5) = 14.2415\left(1 - \frac{2.74}{100}\right) = 14.2415 \times 0.9726 \approx 13.85 \text{ km/h}. \tag{15}$$

The runner should therefore target approximately 13.85 km/h at the 5-kilometre mark — slightly below the average pace of 14 km/h, owing to the positive gradient at that point.

Converting speed to pace. Pace π in minutes per kilometre is obtained from speed s in km/h by:

$$\pi = \frac{60}{s} \text{ min/km}. \tag{16}$$

3. Results

3.1. Pace Chart for a 1 : 30 hr Target Finishing Time

Evaluating $s(x)$ at integer kilometre intervals $x = 0, 1, \ldots, 20$ via equation (13) and converting to pace via equation (16) produces the complete pace chart reproduced in Appendix E (Table 4).

Key observations from the pace chart:

- At km 0–3 the descent from the starting elevation generates above-average speeds (14.88–15.84 km/h).
- At km 5–6 the local positive gradient reduces speed to 13.52–13.85 km/h.
- The long descent around km 8–9 permits speeds of 14.58–15.01 km/h.
- The steep final climb (km 17–20) drives the lowest speeds of the race: 12.36–13.35 km/h.

3.2. Validation

To verify consistency with the target pace, the arithmetic mean of the 21 speed values is computed:

$$\bar{s} = \frac{1}{21}\sum_{k=0}^{20} s(k) = \frac{15.24 + 13.63 + 15.46 + \cdots + 12.69}{21} \approx 14.1\ \text{km/h}. \qquad (17)$$

This is 0.1 km/h above the target of 14.0 km/h — an error of less than 0.7%. The discrepancy is expected: the arithmetic mean over $n = 21$ discrete intervals is a left Riemann sum, whereas the theoretical guarantee (equation (7)) is derived from the limit $n \to \infty$ (i.e. a definite integral). As n increases toward continuous sampling, the computed average converges to exactly 14.0 km/h. The residual therefore constitutes not a flaw in the method but an artefact of the finite resolution of the pace chart, confirming the correctness of the adjustment coefficient g.

4. Discussion

The framework developed here provides a principled, mathematically rigorous alternative to subjective pacing decisions. Several features merit comment.

Model selection. The choice of a 10th-degree polynomial over lower-degree alternatives reflects the inherent complexity of urban half-marathon courses. Even-degree polynomials exhibit symmetric end behaviour which is physically appropriate here, and polynomial regression is a well-understood technique that requires no assumption about the functional form of the elevation profile beyond smoothness.

Linearity of the speed–gradient relationship. The speed function $s(x) = g\,p\,(1 - i(x)/100)$ embodies a linear relationship between incline and speed adjustment. This is a simplification: biomechanical research indicates that the energy cost of running increases nonlinearly with gradient (Minetti et al., 2002). A more sophisticated model could replace the linear term with an empirically calibrated nonlinear function; however, this would require individual physiological data beyond a standard elevation chart.

Omitted variables. The model does not account for environmental variables (wind, temperature, humidity) or psychological factors (race-day adrenaline, fatigue accumulation) that can significantly alter effective speed. The pace chart should therefore be treated as a scientifically informed starting point rather than a deterministic prescription.

Boundary behaviour of $h(x)$. The polynomial extrapolates to $h(21) \approx -62.4\,\text{m}$, which is physically impossible for an above-ground course. This is a known artefact of high-degree polynomial regression (Runge's phenomenon) near the boundary of the fitting domain. In practice the consequence is minor: only the difference $h(x_f) - h(x_0)$ enters the formula for g, and the resulting adjustment is numerically stable (see Appendix D for a full discussion).

5. Conclusion

This paper has demonstrated a complete, self-contained calculus-based method for constructing pace charts for long-distance races with published elevation data. Starting from raw altitude readings, a 10th-degree polynomial regression model ($R^2 = 0.9101$) is fitted to obtain a continuous elevation function $h(x)$. Differentiation yields the incline function $i(x) = dh/dx$, which drives a speed function $s(x)$ inversely proportional to local gradient. An algebraically derived adjustment coefficient g — given by equation (11) — ensures that the integral of $s(x)$ over the full race distance equals the target average pace p. The resulting pace chart for the 2024 United Airlines NYC Half Marathon, targeting a finishing time of 1 h 30 min, was validated to within 0.1 km/h of the desired average.

The method is general: any runner with access to a course elevation profile and a target finishing time can apply the framework by following equations (5), (11), and (13). Future extensions might incorporate nonlinear biomechanical speed–gradient relationships or runner-specific fitness parameters to yield still more accurate prescriptions. The author subsequently applied the method to a 10-kilometre training run in Central Park and cut a three-month personal record by more than two minutes, suggesting the framework has practical validity within its stated assumptions.

References

[1] Chambers, A. (2017, January 29). Maths of global warming — Modelling climate change. *IB Maths Resources*. `https://ibmathsresources.com`

[2] Desmos. (n.d.). *Desmos graphing calculator*. `https://www.desmos.com/calculator/vqw4yz2tfy`

[3] Go&Race. (2024, January 24). 2024 United Airlines NYC Half, course map and altimetry. `https://www.goandrace.com`

[4] Levy, D. (n.d.). *Multiple regression III*. Harvard University Instructional Moves. `https://instructionalmoves.gse.harvard.edu`

[5] Minetti, A. E., Moia, C., Roi, G. S., Susta, D., & Ferretti, G. (2002). Energy cost of walking and running at extreme uphill and downhill slopes. *Journal of Applied Physiology, 93*(3), 1039–1046.

[6] New York Road Runners. (2024, January 30). *United Airlines NYC Half race map.* `https://prodsitecore.blob.core.windows.net`

[7] University of Newcastle. (n.d.). Coefficient of determination, R^2. *Mathematics Resources.* `https://www.ncl.ac.uk/webtemplate/ask-assets`

Appendix A: Elevation Profile Data (Metric, Rounded)

Table 1. Relationship between altitude (metres) and distance (kilometres) — 2024 United Airlines NYC Half Marathon. These are the primary data used for polynomial regression.

Distance x (km)	**Altitude h (m)**
0	27
1	29
2	33
3	11
4	12
5	10
6	11
7	16
8	20
9	9
10	7
11	6
12	2
13	3
14	2
15	6
16	7
17	6
18	6
19	23
20	21
21	29

Appendix B: Raw Elevation Data in Imperial Units

The following table reproduces the original data obtained from the United NYC Half interactive map prior to unit conversion. Distances are in miles and altitudes in feet.

Table 2. Relationship between altitude (feet) and distance (miles) — raw app data.

Distance (miles)	**Altitude (feet)**
0.0	91.1
0.1	78.7
0.5	111.5
0.8	150.9
0.9	154.2
1.0	150.9
1.3	105.0
1.7	82.0
2.0	105.0
2.3	157.5
2.6	128.0
2.9	85.3
3.3	45.9
3.7	36.1
4.0	45.9
4.4	42.7
4.5	62.3
4.7	44.6
5.2	137.4
5.5	114.2
5.7	39.4
5.8	42.7
6.2	13.1
6.4	3.4
7.2	6.6
9.0	0.0
9.4	0.0
9.9	52.5
10.1	39.4
10.4	68.9
11.0	49.2
11.4	72.2
11.6	78.7
11.9	49.2

Appendix C: High-Resolution Metric Elevation Data

Intermediate metric data obtained by converting Appendix B and retaining decimal precision, before rounding to the integer values in Appendix A.

Table 3. Relationship between altitude (metres) and distance (kilometres) — high resolution.

Distance (km)	**Altitude (m)**
0.00	28.0
0.16	24.0
0.80	34.0
1.28	46.0
1.44	47.0
1.60	46.0
2.08	32.0
2.72	25.0
3.20	32.0
3.68	48.0
4.16	39.1
4.64	26.0
5.28	14.0
5.92	11.0
6.40	14.0
7.04	13.0
7.20	18.9
7.52	13.6
8.32	41.9
8.80	34.8
9.12	12.0
9.28	13.0
9.92	4.0
10.24	1.0
11.52	2.0
14.40	0.0
15.04	0.0
15.84	16.0
16.16	12.0
16.64	21.0
17.60	15.0
18.24	22.0
18.56	24.0
19.04	14.9

Appendix D: Full-Precision Elevation Function $h(x)$

The following expression gives $h(x)$ with all Desmos-estimated coefficients at full precision, as used for all numerical computations in the paper:

$$\begin{aligned} h(x) = {} & (8.5036 \times 10^{-8})\, x^{10} - (9.10558 \times 10^{-7})\, x^{9} + (4.13697 \times 10^{-4})\, x^{8} - (1.03671 \times 10^{-2})\, x^{7} \\ & + (1.5579 \times 10^{-1})\, x^{6} - 1.42648\, x^{5} + 7.69777\, x^{4} - 22.181\, x^{3} + 26.6834\, x^{2} - 7.01973\, x + 26.7672 \end{aligned} \tag{18}$$

Boundary evaluations used in computing g:

$$h(0) = 26.7672 \text{ m}, \tag{19}$$

$$h(21) \approx -62.3827 \text{ m}. \tag{20}$$

Note on $h(21) < 0$. The negative value at $x = 21$ is a polynomial artefact known as Runge's phenomenon: a high-degree polynomial constrained to pass near interior data points may behave erratically at or near the boundary of the fitting interval. Physically, the course does not descend below sea level; the model is not intended to be interpreted literally at $x = 21$. Only the difference $h(x_f) - h(x_0)$ enters the formula for g (equation (11)), and the resulting numerical adjustment ($g \approx 1.017$) is small, stable, and physically sensible.

Appendix E: Pace Chart for a 1 : 30 hr Finish

Table 4. Recommended speed and pace at each kilometre for a target finishing time of 1 h 30 min ($p = 14\,\mathrm{km/h}$, $g \approx 1.01725$).

Distance x (km)	**Speed $s(x)$ (km/h)**	**Pace π (min/km)**
0	15.24	3:56
1	13.63	4:24
2	15.46	3:53
3	15.84	3:47
4	14.88	4:02
5	13.85	4:20
6	13.52	4:26
7	13.92	4:18
8	14.58	4:07
9	15.01	4:00
10	14.98	4:01
11	14.59	4:07
12	14.15	4:14
13	13.93	4:18
14	14.00	4:17
15	14.14	4:14
16	13.98	4:17
17	13.35	4:30
18	12.58	4:46
19	12.36	4:52
20	12.69	4:44
Mean speed $\bar{s}$		**14.1 km/h**