

Personalised Hiking Time Estimation

SÁNDOR APÁTHY M.

Department of Mathematics
Corvinus University of Budapest,
H-1828 Budapest Hungary

(Received: August 27, 2016)

Abstract

There are numerous attempts to estimate hiking time since the age of the ancient Roman Empire, the new digital era calls for more precise and exact solutions to be implemented in mobile applications. The importance of the topic lies in the fact that route planning algorithms and shortest path problems apply time estimations as cost functions. Our intention is to design a hiking time estimation method that accounts for terrain circumstances as well as personal factors, while the level of accuracy and the simplicity of the algorithm should enable the solution to be utilised in the practice. We refine Tobler's earlier results to estimate a relation between terrain steepness and hiker's velocity. Later we use fitted curve to design our novel, personalised hiking time estimation method.

Mathematics Subject Classifications (2015). 91B84, 62L12

Keywords. Time estimation, Operations research, Model Construction and Estimation, Forecasting and Prediction Methods, Tourism

1 Introduction

In today's technoid era people strive to live their life more planned and organised, including their free time, that supposed to be spent effectively. By fast-spreading of smartphone technology these scheduling mobile applications, that intend to facilitate our various activities, became more and more an integrated part of our everyday life. This study aims to refine the accuracy of route planning applications. First we briefly introduce the existing hiking time estimation methods, then a novel method will be presented to estimate the travel time of a hiker considering many circumstances. We utilise more than 2400 tracklogs to perform our estimations that is - as per our knowledge - outrageous in the literature. Based on NASA data on Earth's surface we developed a DEM model to substitute GPS elevation data that is rather unreliable. Longitude-latitude pairs of the raw GPS tracklogs have been smoothened by Kalman-filter to eliminate GPS noises. We follow Tobler's earlier results estimate a function describing the relationship between terrain steepness and hiker's velocity, then the fitted curve will be personalised by two novel hiking time estimation method. The test results has shown 11.5% and 12.9% Mean Absolute Relative Error, that clearly outperforms existing solutions.

2 Hiking time estimation - Related work

Since the ancient Roman Empire there has been several attempts to estimate the speed of hikers [8]. The most commonly used rule of thumb dates back to 1892, when a Scottish mountaineer, William Naismith described his estimation [6]: a man with average physical conditions can walk 3 miles (4827.9 m) in 1 hour on flat terrain, while additional 1 hour is required for every 2000ft (632m) elevation. So practically 1 unit of elevation equals to 7.92 units of horizontal walk. There has been many attempts to refine Naismith's estimation (Langmuir [5], Aitken [1]), including Waldo Tobler [11], who conjectured exponential relationship between steepness of the terrain and the velocity of the hiker. We compared the hiking time estimations on Figure 1.

By fast spreading of GPS devices several routing applications have been designed (e.g. Komoot, Strava or Locus), though there is no exact description of the estimation methods used by these applications. Unfortunately the results are rather unreliable due to the high dilution of precision (DOP) of GPS signals under certain disposition of satellites and weather circumstances [10]. In the next section we briefly describe how we strived to overcome the challenge of poor GPS information. One of the most recent estimation methods has been developed by Pitman et al. [7], that applies multivariate regression model that accounts for several circumstance of the trail (such as total hiking distance, cumulated ascent, gradient angle, elapsed distance, etc.). As per our experience the explanatory power of these models are still very low (resulted in less than 10% adjusted R^2), despite of the high number of variables involved, hence a novel method has been designed, that is to be presented in the following section.

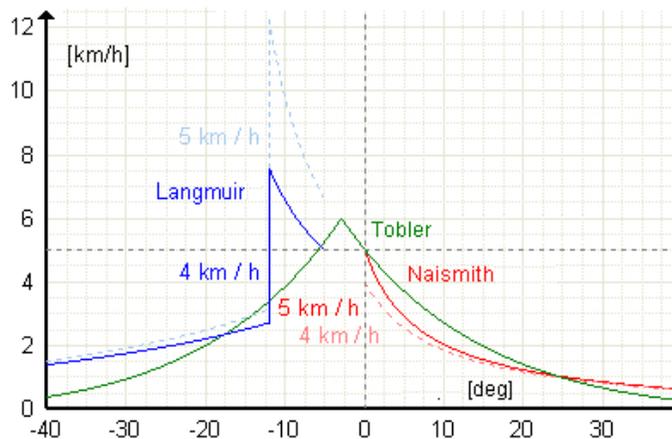


Figure 1: Comparison of hiking time estimations¹

3 Personalised hiking time estimation method

By following Tobler’s concept we performed a novel function that refines Tobler’s results and enables us to obtain more accurate estimation of hiking times.

Our database consists of 2400 tracklogs (after eliminating outliers). In case of each tracklog we removed the raw GPS elevation data and substituted with the elevation values calculated by our Digital Elevation Model (DEM). DEM is based on NASA Earth’s surface data (that assigns a single elevation value to every $30\text{m} \times 30\text{m}$ square of the surface). Our DEM model uses bilinear interpolation to calculate estimated elevation and as a result it provides approximation of the Earth’s surface. The summary of recently used Digital Elevation Models is presented in Hirt et al. [4], and the calculation methods are compared in Skidmore [9]. According to our comparison between DEM and Google Elevation API the average difference between the two estimations is (13 cm). It is an excellent result (by assuming Google as a good benchmark) considering that elevation error of mobile devices can be higher than 150 m. The longitude and latitude values of raw GPS tracklogs have been smoothened by Kalman-filtering method, (Goh et al. [2] provides a comprehensive study on KF method). Thus we obtained adjusted 3D tracklogs, and standardised to 20 second long logs. Based on these logs’ steepness (m) and velocity (v) values we calculated the average speed of hikers for every $1/8$ degree of steepness (with radius $1/8$ degree) and fitted a curve in R software with linear model package (that uses QR matrix decomposition method, [3]):

$$v(m) = \left(\begin{array}{l|l} \exp(2.3203m + 0.4462) & m \in (-\text{inf}; -0.15) \\ p(m) & m \in [-0.15; 0.15] \\ \exp(-2.4672m + 0.3769) & m \in (0.15; \text{inf}) \end{array} \right)$$

The fitted $v(m)$ curve indicated plausible results: the maximum velocity of 1.4 m/s is reached at slight slope -2° and decreases fast by changing steepness of the terrain (see Fig. 2). Test statistics of $v(m)$ are summarised in the Appendix.

3.1 The steepness based method

Based on the fitted $v(m)$ function we developed a novel hiking time estimation method calculated as follows: Let m_i denote the steepness of the terrain for the i -th log of a trail. Let us estimate the velocity of the first log by $v(m_1)$ then after measuring back the actual results (v_1) on the first log, we calculate the ratio of actual and estimated values, that is $b_1 = v_1/v(m_1)$. We consider this as the estimated personal fitness factor of a particular hiker, that enables us to adjust our estimations dynamically through the trail. Thus we estimate the velocity of the second log by $b_1v(m_2)$. After observing the fitness factor of the second log, we can estimate the 3rd log’s velocity by $(b_1 + b_2)v(m_3)/2$, i.e. we calculate the overall fitness factor as the arithmetic mean of the observed fitness

¹source: https://en.wikipedia.org/wiki/Naismith%27s_rule

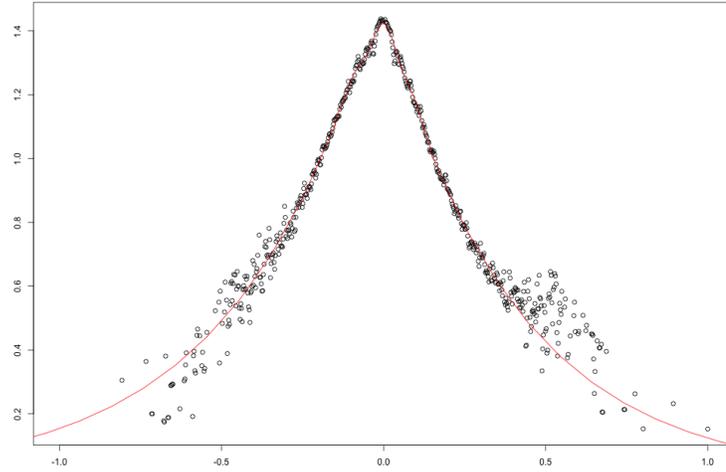


Figure 2: Fitted steepness-velocity function (radian - m/s)

factors assigned to each log. More generally we estimate the velocity of the n -th log as follows:

$$v_n^* = \frac{\sum_{i=1}^n b_{i-1}}{n-1} v(m_i)$$

By adjusting our initial estimation with the dynamically calculated fitness factor we can amend the estimated hiking and account for other circumstances, as the actual physical state of the hiker or the weather. Though during the first several logs (namely in the first 10%) we observed that the estimation has poorer results due to the high variance of the fitness factors, after obtaining a stable fitness estimation we performed far better results comparing to earlier estimation methods (e.g. [7]).

3.2 Mean velocity based method

The idea of this method is excessively simple: we estimate the first 20% of the trail with the $v(m)$ fitted steepness-velocity function and gather the observed velocity values of each log. After this test period we estimate the velocity of the n th log of the trail with the arithmetic mean velocity of the $n-1$ previous trails:

$$v_n^* = \frac{\sum_{i=1}^n v_{i-1}}{n-1}$$

During the first 20% of the trail this estimation has typically yield poor estimation, hence we substituted the mean velocity based estimations with the $v(m)$ values.)

4 Evaluation of Results

We compared our results to the existing solutions (we considered Tobler’s curve as the most sophisticated and accurate method) and performed a test on the available 2400 tracklogs (splitting it to a 75-25% learning and test dataset). On the learning dataset we fitted the $v(m)$ curve, then applied $v(m)$ on the test dataset to estimate hiking time). 10 iterations have been performed. To measure the goodness of our estimations we partitioned each tracklog separately to 100 equal pieces and for the p -th percentile of the tracklog the estimated hiking time of the remaining trail has been calculated (denoted by r_{ip}^* for the i -th tracklog) comparing with the actual remaining time (r_{ip}). We used the Mean Absolute Relative Error (MARE) measure for evaluation:

$$\text{MARE}(p) = \frac{1}{n} \sum_{i=1}^n \left| \frac{r_{ip} - r_{ip}^*}{r_{ip}} \right|$$

In Table 1. the results of the 10 test iterations have been summarised. The Welch-tests performed (based on 100 iterations) has rejected the null hypothesis that the mean of MARE values of the two methods are equal (see References), hence the proposed first hiking time estimation method yield significantly better results (11.58%) than Tobler’s estimation (17.36%). By eliminating tracklogs resulted in MARE value higher than 20%, we obtain 9.37% mean MARE value for the 10 iterations.

The mean velocity based method yielded slightly poorer results (12.93%), though clearly outperforms Tobler’s estimation.

	MARE_steep	MARE_st.out	MARE_vel	MARE_vel_out	MARE_Tobler
Test_1	11.90%	9.51%	13.00%	9.47%	17.06%
Test_2	11.41%	9.38%	12.87%	9.56%	16.74%
Test_3	11.01%	9.36%	12.89%	9.66%	17.38%
Test_4	11.95%	9.00%	12.74%	9.98%	16.72%
Test_5	11.44%	9.38%	12.55%	10.03%	17.80%
Test_6	11.64%	9.44%	13.26%	9.60%	18.25%
Test_7	11.59%	9.63%	13.56%	9.92%	16.35%
Test_8	11.91%	9.30%	12.73%	10.24%	19.48%
Test_9	11.55%	9.15%	12.49%	10.18%	17.09%
Test_10	11.36%	9.53%	13.33%	9.85%	16.76%
mean	11.58%	9.37%	12.94%	9.85%	17.36%

Table 1: Comparison of MARE values of the 3 methods

On Fig.3 we illustrated the results (red: steepness based model, black: mean velocity based model, blue: Tobler’s estimation). Our methods are performing relatively better on the 15-80% percent of the trail. At the beginning due to the variance of the estimated factors the errors are higher, while at the last part of the trail the smaller errors seem to be relatively high as we compare them to a smaller value of the remaining hiking time.

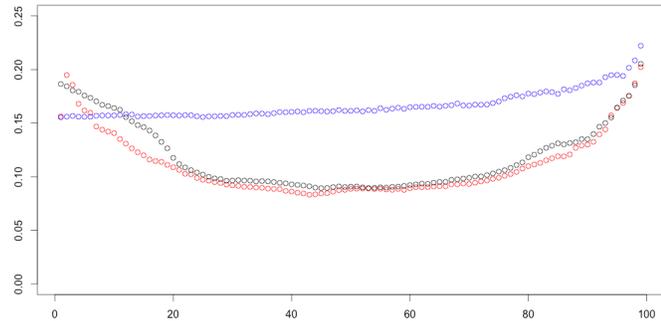


Figure 3: MARE values of the two estimation methods

We also tested whether we can reach better MARE results by applying our steepness based method replacing $v(m)$ by Tobler-curve. According to the test results our method's mean MARE value (11.12%) was significantly better than the Tobler-curve based solution (13.48%), see Figure 4.

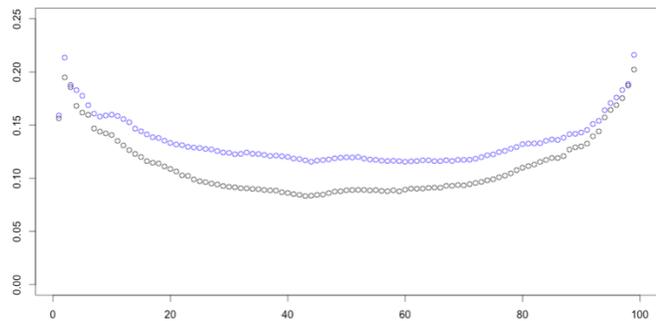


Figure 4: MARE values comparison for $v(m)$ and Tobler-curve based methods

5 Conclusions and future work

By following Tobler's initial idea to conjecture a relationship between steepness of the terrain and velocity of the hiker we refined his earlier results. We developed two novel methods to estimate hiking time using only our fitted steepness-velocity curve and the observed velocity values of previous logs. The tests were performed on an outrageously large dataset (2400 tracklogs) and resulted in 11.58% and 12.94% Mean Absolute Relative Error compared to Tobler's result of 17.36%. In the near future we intend to broaden the list of explanatory variables used in our model to enhance the accuracy of our estimations. In case of acquiring a larger dataset where tracklogs related to the same user can be identified, we could probably design a collaborative estimation method, that is

previous trails of the user and other users who are similar to him may be utilised to adjust our predictions.

Appendix

	coeff.	std. err.	t-value
0	1.422E + 00	4.011E - 03	354.590
1	-1.708E - 01	1.466E - 01	-1.165
2	-6.566E + 01	5.107E + 00	-1.2858
3	-8.371E + 01	7.463E + 01	-1.122
4	1.129E + 04	1.689E + 03	6.686
5	1.156E + 04	1.201E + 04	0.962
6	-1.075E + 06	2.132E + 05	-5.041
7	-6.130E + 05	7.470E + 05	-0.821
8	4.643E + 07	1.131E + 07	4.104
9	1.112E + 07	1.570E + 07	0.708
10	-7.398E + 08	2.126E + 08	-3.480
exp1_coeff	2.3203	0.05709	40.65
exp1_interc	0.4462	0.01617	27.60
exp2_coeff	-2.4672	0.04407	-55.99
exp2_interc	0.3769	0.01253	30.08
	p(m)	exp1	exp2
RSE	0.01648	0.04387	0.03422
adj. R-sq.	0.9774	0.9359	0.9649
df	124	112	113
F-stat	581.4	1652	3135
p-value	< 2.2E - 16	< 2.2E - 16	< 2.2E - 16

Table 2: Test statistics of fitted steepness-velocity function

Steepness method - Tobler's estimation	
t-value	17.43
p-value	< 2.2E - 16
mer_mean	0.1158
Tobler_mean	0.1668
df	135.78
95% conf int.	0.0493
95% conf int.	0.0618
H₀	Rejected

Table 3: Results of Welch-test to compare the two methods

References

- [1] R. AITKEN, *Wilderness Areas in Scotland*, PhD thesis, unpublished Ph.D. Thesis. University of Aberdeen, Aberdeen, 1977.
- [2] M. S. GOH, D. F. SHEN, AND S. H. HONG, *Processing of gps data with difference hdop in guide robot for the visually impaired*, International Journal of Computer Science and Network Security, 7 (2007), pp. 90–97.
- [3] M. GULLIKSSON AND P. WEDIN, *Modifying the qr-decomposition to constrained and weighted linear least squares*, SIAM Journal on Matrix Analysis and Applications, 13 (1992), pp. 1298–1313.
- [4] C. HIRT, M. S. FILMER, AND W. E. FEATHERSTONE, *Comparison and validation of the recent freely available aster-gdem ver1, srtm ver4.1 and geodata dem-9s ver3 digital elevation models over australia*, Australian Journal of Earth Sciences, 57 (2010), pp. 337–347.
- [5] E. LANGMUIR, *Mountaincraft and Leadership*, SportsScotland, 3 ed., 1995.
- [6] W. NAISMITH, *Notes and queries*, Scottish Mountaineering Club Journal, 2 (1892), p. 133.
- [7] A. PITMAN, M. ZANKER, J. GAMPER, AND P. ANDRITSOS, *Individualized hiking time estimation*, in 2012 23rd International Workshop on Database and Expert Systems Applications, Sept 2012, pp. 101–105.
- [8] F. V. RENATUS, *The Military Institutions of the Romans (De Re Militari), Book I: The Selection and Training of New Leviess*, english translation by John Clarke,, 3 ed., 1767.
- [9] A. K. SKIDMORE, *A comparison of techniques for calculating gradient and aspect from a gridded digital elevation model*, International Journal of Geographical Information Systems, 3 (1989), pp. 323–334.
- [10] E. R. SWANSON, *Geometric dilution of precision*, Navigation, 25 (1978), pp. 425–429.
- [11] W. R. TOBLER AND N. C. FOR GEOGRAPHIC INFORMATION & ANALYSIS (U.S.), *Three Presentations on Geographical Analysis and Modeling: Non-isotropic Geographic Modeling; Speculations on the Geometry of Geography; and Global Spatial Analysis*, vol. 93-1 of NCGIA technical report, NCGIA, University of California, 1993.