From generalisation to segmentation: Douglas-Peucker-Ramer and movement data

A. D. Miller*[1], A. B. Moore[1] and G. H. Leonard[1]

[1]National School of Surveying, University of Otago, Dunedin, New Zealand
*Email: aubrey.miller@otago.ac.nz

Abstract

With an explosion in the sensorisation of moving objects (people, cars, wildlife) that has come from the ease of use and acquisition of low-cost GNSS receivers, the volume of tracking data presents challenges for efficient data processing. For example, in analysis and modelling, tracks often need to be segmented based on particular characteristics (e.g., moving vs. stationary). Here we adapt a commonly used polyline generalisation algorithm (Douglas-Peucker-Ramer) to segment tracking data into uphill and downhill portions of the track based on vertex coordinates in terms of their elevation and distance between neighbouring vertices along a track (relative to start of the track). This adaptation supports robust segmentation even for tracks in complex terrain in order to best capture real-world conditions. We present our adapted algorithm and a case study with volunteered backcountry skier tracking data to demonstrate the performance of the algorithm. We conclude with thoughts on future development of tracking data segmentation algorithms.

Keywords: Generalisation, Douglas-Peucker, GPS, tracking, movement, recreation

1. Introduction

The field of cartography is rich with examples of generalisation techniques (e.g., simplification and smoothing) to ensure vector feature detail is appropriate to map scale. Semi-automated generalisation algorithms based on polyline vertex coordinates have been in use since the 1960s to generalise large polyline datasets (Tobler, 1964). Typically, generalisation algorithms work in two-dimensional space, smoothing, removing, collapsing, rectifying, or fractalising based on vertex xy coordinates (Shea and McMaster, 1989; McMaster, 1987).

With the development and widespread use of low-cost GNSS receivers there is a need for simple but powerful algorithms for segmenting GPS tracking data based on spatial and non-spatial attributes. The ways in which people move through a landscape, for example, depend on myriad factors, including the mode of travel (e.g., pedestrian, cycle, car, train), whether they are moving through an urban area or bush, whether they are constrained to a route network such as a road, the time in which the movement occurs, the familiarity with the area, and others. Track segmentation involves isolating portions of the track based on spatial (e.g. off or on a known route) and non-spatial (e.g., activity type such as cycling or walking) attributes to improve specificity in analysing the tracking data. Segmentation is also used for thinning datasets based on accuracy needs before further analysis.

As tracking datasets increase in size and ubiquity, a need for effective segmentation of tracks has arisen to better-characterise complex movement patterns and further movement modelling. GNSS technology in low-cost receivers such as mobile phones continue to improve in precision and accuracy reinforcing the need for simple algorithms for track segmentation.

Here we show how an adapted Douglas-Peucker-Ramer generalisation algorithm (Douglas and Pecuker, 1973; Ramer, 1972) can be used to segment GPS tracks into uphill and downhill portions of a track. Real-world conditions make this particular segmentation difficult because it is highly scale-dependent and characteristics such as movement speed fail to effectively segment tracks in large datasets with large variation in speed. We detail how the algorithm works (Section 2), present a case study with backcountry skier data from Colorado, USA to demonstrate the algorithm (Section 3) and conclude with ideas for future development and applications (Section 4).

2. The algorithm

A modification is needed to the standard Douglas-Peucker-Ramer algorithm (as formally expressed in Saalfeld, 1999) to change the standard (x, y) coordinate for any given vertex v that is part of a line to be simplified, into a (d, z) coordinate, where:

$d_h = \sqrt{(x_h - x_{h-1})^2 + (y_h - y_{h-1})^2} + d_{h-1}, h = 1,2, \ldots, n-1, n$ vertices in the line; and

z_h = the corresponding vertical component of the coordinate (elevation).

This gives the cumulative interpoint xy distance relative to the start of the track, which, when paired with the elevation value creates a single profile for the entire track (Figure 1).

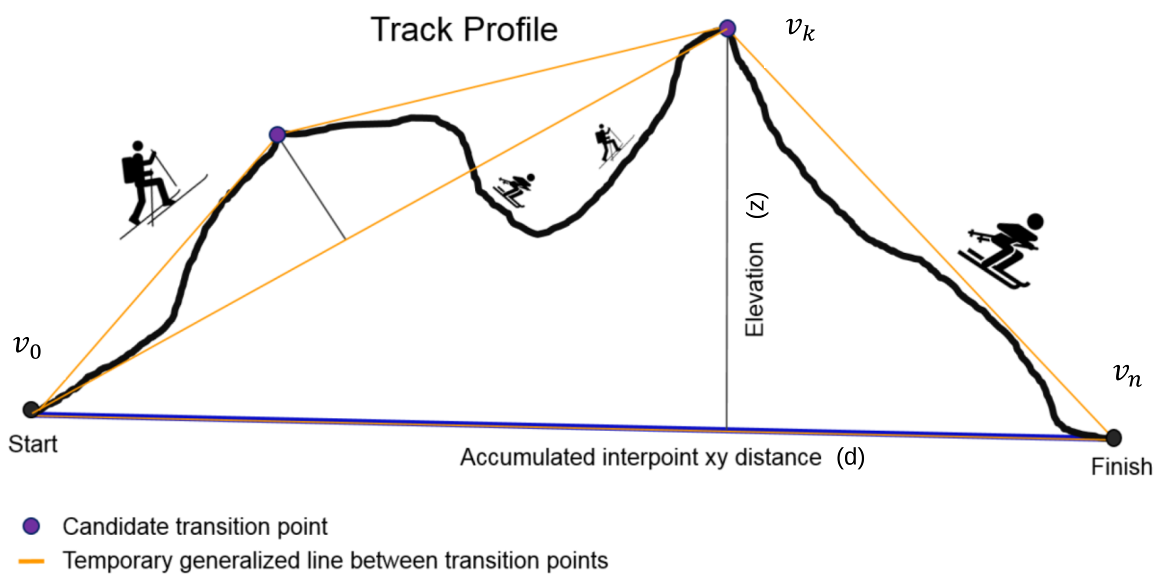

Figure 1: Concept of adapted Douglas-Peucker-Ramer algorithm to 3D change in tracking data.

The profile coordinates are processed using the standard algorithm. First, the entire track is passed to a Simplify function (Saalfeld, 1999):

Initialize output polyline vertex set V to $\{v_0, v_n\}$;
Simplify (P_{0n}, ε);
Output V;

where $\{v_0, v_1, \dots, v_{n-1}, v_n\}$ is an ordered set of (d,z) vertices; P_{0n} is the entire track polyline; and $\varepsilon > 0$ is the threshold distance that guides how much the simplified line can deviate from the original polyline.

The Simplify function calculates the maximum perpendicular distance of the polyline from the edge connecting the first and last vertices (initially v_0 and v_n), using the vertex at the maximal point to split the polyline. Each polyline subset is then recursively simplified (Figure 1). The Simplify function is set out below (Saalfeld, 1999):

```
Simplify (P_ij, ε){
    If j > i + 1, then {
        Find k ∈ (i, j) with δ(v_k, e_ij) ≥ δ(v_s, e_ij), s ∈ (i, j);
        If δ(v_k, e_ij) > ε, then {
            Add v_k to V;
            Simplify (P_ik, ε);        //recursively process left sub-polyline
            Simplify (P_kj, ε);        // recursively process right sub-polyline
        }
    }
}
```

where P_{ij} is the sub-polyline currently being processed (i is the start vertex address, j is the end vertex address); e_{ij} is the edge connecting vertex v_i and vertex v_j; k is the address of the vertex with the maximum perpendicular distance from e_{ij}; $\delta(v_k, e_{ij})$ is that perpendicular distance value from vertex v_k to edge e_{ij}; P_{ik} is the left sub-polyline from vertex v_i to vertex v_k; P_{kj} is the right sub-polyline from vertex v_k to vertex v_j.

Finally, each generalised line section formed from the ordered output vertex set V is classified as uphill, downhill or flat sections:

If $z_h > z_{h+1}$, then gradient type g_h= downhill

If $z_h < z_{h+1}$, then gradient type g_h= uphill

If $z_h = z_{h+1}$, then gradient type g_h= flat

where $h = 0,1, \dots, m-2, m-1$ vertices in the simplified line.

Once the attribute of uphill, downhill or flat is added to each vertex in a track (reaching the starting vertex ID again), the next track is processed automatically, necessary for datasets with many tracks.

3. Case study: Recreation GPS tracking data, Colorado, USA.

We used GPS data from a winter recreation study (2011-2013) in Colorado, USA to test our uphill/downhill segmentation algorithm. Backcountry skiers carried a small GPS receiver (Qstarz BT-

Q1300, positional accuracy < 10m, logging frequency 5-sec) for their trip duration. Attributes recorded by the GPS receiver included elevation, speed, heading, logging time, and track ID. For more details on study design and methodology, see Olson et al. (2017).

We started by manually segmenting four individual test tracks based on expert knowledge of the terrain. The heuristic we looked to match with the automated segmentation was whether a skier would take off or put on climbing skins at an uphill-downhill transition point. The tracks (Figure 2) represented a range of terrain and skier characteristics, including steep and relatively flat terrain, short (~5km) and long (~18km) total trip length, a range of speeds, as well as the presence of data gaps where the GPS receiver did not meet satellite fix requirements to log a point. We then applied the algorithm to these four test tracks using 10m, 25m, 100m, 200m threshold distance parameters and inspected the results for any points segmented differently than the test tracks.

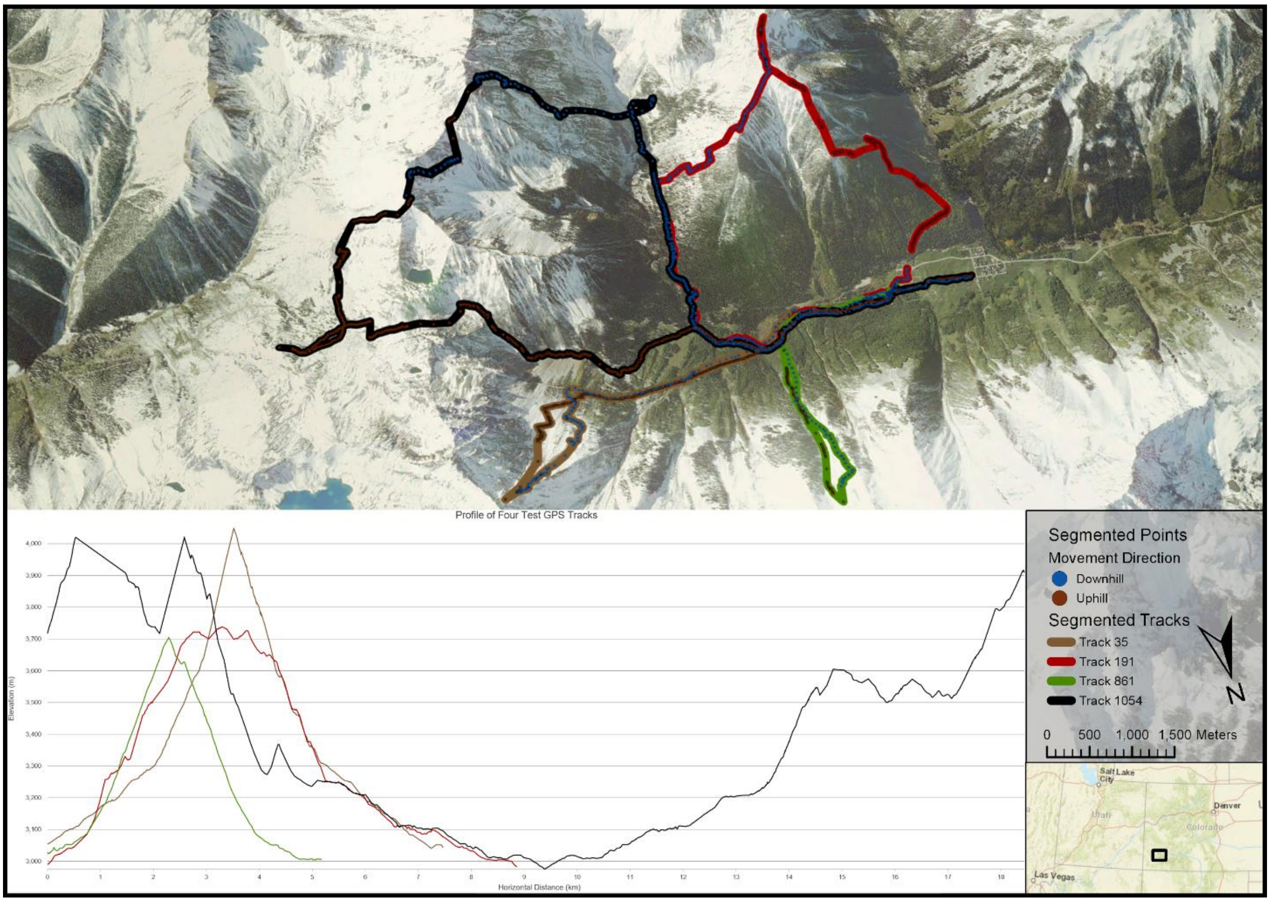

Figure 2: Scene showing four test tracks for segmentation of skier movement data in Colorado, USA along with profile for each track. Segmented points are shown on top of the tracks. Note, scale bar represents linear distance in the centre of the scene.

We found the 25m threshold distance to be best at capturing the change in direction of travel that best mirrors real-world conditions (i.e., climbing skins on or off at a transition). We then segmented a larger dataset of 687 individual tracks (comprising 433,555 points/vertices) into uphill and downhill portions of the track using the threshold distance of 25m, and spot-checked 10 individual tracks to identify any mis-segmented points. We found the 25m threshold distance to work well in all 10 cases with less than 1% of total points found to be mis-segmented. The results of the automated segmentation based on the 25m threshold are shown in Figure 3.

Segmented skier data support closer inspection of movement characteristics, allowing for more useful analysis of dataset-level terrain selection and movement patterns. For example, 67% of the points were uphill and 32% were downhill, suggesting that skiers spent on average two thirds of their moving time traveling uphill compared with one third move downhill. We also measured differences in the mean moving speed of uphill points (2.7kph) and downhill points (7.16kph).

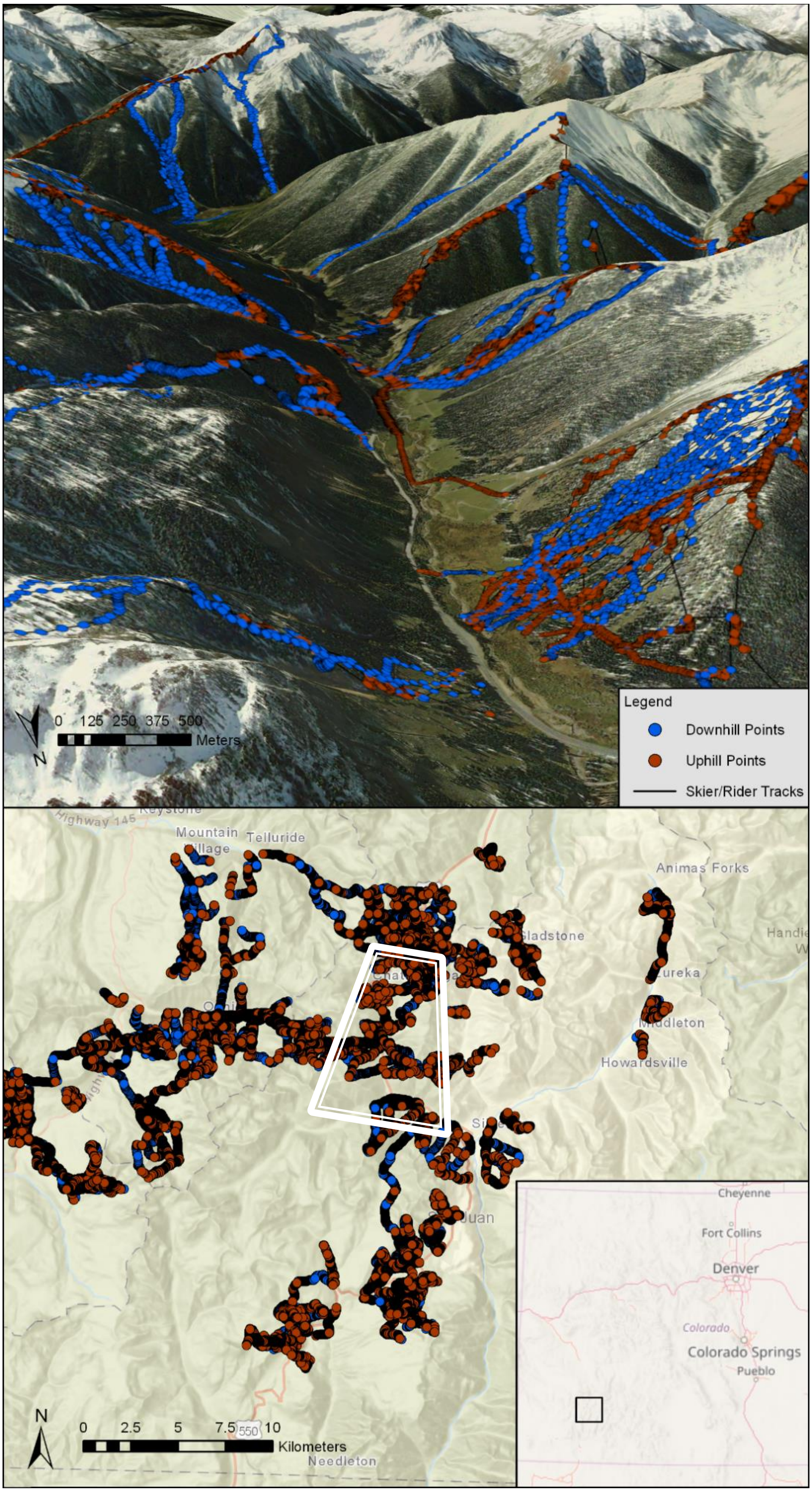

Figure 3: Recreation data from Colorado, USA used to scale-up adapted Douglas-Peucker-Ramer algorithm for segmentation into uphill and downhill portions of track.

4. Conclusion

Our adaptation of the Douglas-Peucker-Ramer generalisation algorithm to segment movement data will be useful for researchers wanting better specificity with regards to characterising movement patterns, especially with an explosion in the collection of tracking data. For example, as more and more people and vehicles are equipped with low-cost GNSS receivers (autonomous vehicles, public transit, fitness and recreation tracking) the need for efficient data processing prior to further modelling will be essential.

One area for future development of this adapted algorithm for uphill/downhill travel is the development of a simple algorithm to identify the most suitable threshold distance for an individual track based on its length, elevation change and speed to avoid the need for a global parameter and user input. Also, additional parameterisation from real-world conditions specific to the mode of travel (for example slope angle for bicycles) will be necessary as this algorithm is applied to other tracking datasets.

5. Acknowledgements

Thanks to the recreationists who volunteered to carry the GPS units and the research staff who collected the data.

6. References

Douglas, D.H. and Peucker, T.K. 1973. Algorithms for the reduction of the number of points required to represent a digitized line or its caricature. *Cartographica* **10**(2), pp.112-122.

Olson, L.E., Squires, J.R., Roberts, E.K., Miller, A.D., Ivan, J.S., and Hebblewhite, M. 2017. Modeling large-scale winter recreation terrain selection with implications for recreation management and wildlife. *Applied Geography*. **86**, pp.66–91.

McMaster, R.B. 1987. Automated line generalization. *Cartographica*. **24**(2), pp.74-111.

Ramer, U. 1972. An iterative procedure for the polygonal approximation of plane curves. *Computer Graphics and Image Processing*. **1**, pp.244-256.

Saalfeld, A. 1999. Topologically Consistent Line Simplification with the Douglas-Peucker Algorithm. *Cartography and Geographic Information Science*. **26**(1), pp.7-18.

Shea, K.S. and McMaster, R.B. 1989. Cartographic generalization in a digital environment: When and how to generalize. *Auto-carto 9. Proc. symposium*, Baltimore, MD, **9**, pp.56-67.

Tobler, W.R. 1964. An experiment in the computer generalization of maps. Technical Report No. 1, ONR Task No. 389-137.