Assessment of an approximation method for TSP path length on road networks

Koen Stevens

April 21, 2025

Bachelor's Thesis Econometrics

Supervisor: dr. N.D. van Foreest

Second assessor:

Assessment of an approximation method for TSP path length on road networks

Koen Stevens

Abstract

1 Introduction

The Traveling Salesman Problem (TSP) is an important problem in operations research. It is particularly relevant for last-mile carriers and other logistics companies where efficient routing directly impacts cost, time and service quality. Since the number of parcels worldwide has increased between 2013 and 2022 and is expected to keep increasing Statista (2025), the need for fast, scalable route planning methods becomes ever more pressing.

The TSP is an NP-hard problem, it is computationally intensive to find the exact solution for large instances. In many real-world scenarios, the exact optimal routes may not be needed, but instead a rough, reliable estimate of the optimal route length. For instance, consider a postal delivery company. This firm may need to assign a certain amount of deliveries or a certain area to each postman. Reliable estimates for the route length can provide valuable information for making such decisions.

Efficient approximation methods provide a solution for such practical applications where exact solutions are too computationally intensive to conduct or not feasible due to insufficient data. These methods aim at approximating the expected optimal total travel time or distance, while using minimal data and computational effort.

There is extensive research on such approximation methods and how they perform in the Euclidean plane. Consider n uniformly drawn locations from some area in $\mathbb{R}^2$ with area A. Beardwood, Halton, and Hammersley (1959) prove the relation:

$$L \to \beta\sqrt{nA}, \quad \text{as } n \to \infty \tag{1.1}$$

as an estimation for the length of the shortest TSP path measured by Euclidean distance through these random locations, where β is some proportionality constant. This formula is a very elegant result, and it requires very little data. However, its assumptions, uniform random locations and euclidean space differ from real-world applications, which are defined by complex geographic features, such as road networks.

This research investigates how well this approximation method performs when considering real road networks. Using OpenStreetMap data, TSP instances are simulated in a wide variety of different urban areas in the Netherlands, then solve these for the actual shortest paths using the Lin–Kernighan Lin and Kernighan (1973) heuristic. Then the β from equation 1.1 is estimated and the performance of this formula is analyzed. Additionally, the results for β and the performance across the selected areas is compared.

In section 2 a deep dive in the context and previous research in this field is provided. In section 3 the experimental design is documented.

2 Literature Review

In this section the existing literature on the Beardwood formula and some applications, and on the Lin-Kernighan heuristic and its implementations is reviewed.

2.1 Applications of the Beardwood formula

This research concerns the performance of formula 1.1 for reasonable amounts of locations a delivery person can visit in a workday, say $10 \leq n \leq 90$. Lei, Laporte, Liu, and Zhang (2015) estimates the values of β for a selection of values for n. In their research, the points were generated uniformly and the L_2 distance metric was used. Table 1 lists the results.

Table 1: Empirical estimates of β as a function of n, $20 \leq n \leq 90$ (Lei et al. (2015))

n	$\beta(n)$
20	0.8584265
30	0.8269698
40	0.8129900
50	0.7994125
60	0.7908632
70	0.7817751
80	0.7775367
90	0.7773827

Figliozzi (2008) is the first research to apply approximation formulas to real-world instances of TSPs (and VRPs (Vehicle Routing Problems)). An extension of formula 1.1 that works for VRPs is assessed in a real-world setting. It is found that this model has an R^2 of 0.99 and MAPE (Mean Absolute Prediction Error) of 4.2%. This prediction error is slightly higher than when it is applied to a setting where Euclidean distances are considered (3.0%), but the formula still performs well (Figliozzi (2008)).

Merchán and Winkenbach (2019) use circuity factors to measure the relative detour incurred for traveling in a road network, compared to the Euclidean distance. This circuity factor is defined as, where p and q are locations:

$$c = \frac{d_c(p,q)}{d_{L_2}(p,q)} \tag{2.1}$$

By construction, c is greater or equal to 1, a value closer to 1 indicates a more efficient network. Then, β_c is estimated by $\beta_c = c\beta$. This value c, is estimated for three different areas in São Paulo, for which the results are listed in table 2. These values indicate real travel distances are on average 2.76 times longer in area 1 compared to the L_2 metric. These values were obtained by uniformly generating n locations (for n ranging from 3 to 250), computing near-optimal tour lengths under the Euclidean metric, and solving for β, then scaling by the empirical circuity factor. It is important to note,

Table 2: Estimates of the circuity factor c and its corresponding β_c (Merchán and Winkenbach (2019))

	Area 1	Area 2	Area 3
c	2.76	2.34	1.82
β_c	2.48	2.10	1.64

however, that the assumptions in this study may limit the generality of the findings. In particular, the use of uniformly distributed locations does not accurately reflect the spatial distribution of delivery points in real urban environments, where locations tend to cluster in residential, commercial, or industrial zones. Additionally, within small urban areas, high-rise buildings and single-family homes may coexist in the same neighborhoods, further challenging the assumption of uniformly distributed delivery points. Furthermore, the circuity factor c can vary significantly within a single city, depending on local street patterns, infrastructure, and topography. These variations suggest that a fixed circuity factor may oversimplify the complexity of real-world delivery contexts, especially when applied to smaller sub-regions or neighborhoods.

2.2 Lin-Kernighan Heuristic

To be able to efficiently solve many TSPs, to find a good estimate for β, a fast and reliable solution algorithm is needed. The Lin-Kernighan Lin and Kernighan (1973) heuristic provides outcome, it is generally considered to be one of the most effective methods of generating (near) optimal solutions for the TSP. In this research a modified implementation of the heuristic is used Helsgaun (2000). The run times of both heuristics increase by approximately $n^{2.2}$, but the modified heuristic is much more effective. It is able to find optimal solutions to large instances in reasonable times Helsgaun (2000).

PARAGRAPH ABOUT HOW THE HEURISTIC WORKS

3 Experimental design

3.1 Data

3.2 Generation and solving of TSPs

4 Results

5 Discussion

6 Conclusion

References

Beardwood, Jillian, John H Halton, and John Michael Hammersley (1959). The shortest path through many points. In *Mathematical proceedings of the Cambridge philosophical society*, Volume 55, pp. 299–327. Cambridge University Press.

Figliozzi, Miguel Andres (2008). Planning approximations to the average length of vehicle routing problems with varying customer demands and routing constraints. *Transportation Research Record 2089*(1), 1–8.

Helsgaun, Keld (2000). An effective implementation of the lin–kernighan traveling salesman heuristic. *European journal of operational research 126*(1), 106–130.

Lei, Hongtao, Gilbert Laporte, Yajie Liu, and Tao Zhang (2015). Dynamic design of sales territories. *Computers & Operations Research 56*, 84–92.

Lin, Shen and Brian W Kernighan (1973). An effective heuristic algorithm for the traveling-salesman problem. *Operations research 21*(2), 498–516.

Merchán, Daniel and Matthias Winkenbach (2019). An empirical validation and data-driven extension of continuum approximation approaches for urban route distances. *Networks 73*(4), 418–433.

Statista (2025). Global parcel shipping volume between 2013 and 2027 (in billion parcels)*.

7 Appendix

Table 3: Empirical estimates for β in selected neighborhoods.

Province	Neighborhood	β	MAE
groningen	Hortusbuurt	2.2017	0.0789
groningen	Binnenstad	2.0464	0.0811
groningen	Oosterpoort	2.0587	0.0690
groningen	Rivierenbuurt	1.7778	0.0558
groningen	De Wijert	1.7749	0.0745
groningen	Oosterparkwijk	1.7818	0.0612
groningen	De Hoogte	1.6913	0.0650
groningen	Korrewegwijk	2.0728	0.0601
groningen	Schildersbuurt	2.2188	0.0558
groningen	Paddepoel	1.6758	0.0554
groningen	Oranjewijk	1.9335	0.0611
groningen	Tuinwijk	2.9016	0.0625
groningen	Selwerd	1.5229	0.0556
groningen	Vinkhuizen	1.4801	0.0492
groningen	Hoogkerk-zuid	1.5613	0.0720
groningen	Gravenburg	1.2986	0.0950
groningen	De Held	1.9209	0.0496
groningen	Reitdiep	1.6469	0.0587
groningen	Hoornse Meer	1.5943	0.0482
groningen	Corpus den Hoorn	1.6080	0.0580
groningen	Eemspoort	1.7209	0.0571
groningen	Euvelgunne	1.9316	0.1281
groningen	Driebond	1.9184	0.0934
groningen	Winschoterdiep	1.9621	0.0585
groningen	Eemskanaal	1.7589	0.0511
groningen	Helpman	2.0775	0.0704
groningen	Lewenborg	1.9260	0.0576
groningen	Beijum	1.7935	0.0461
groningen	Maarsveld	1.6849	0.0481
noord holland	Schrijverswijk	1.7494	0.0473
noord holland	Stad van de Zon	1.4783	0.1775
noord holland	Stadshart	1.4551	0.0722
noord holland	Jordaan	1.8469	0.0651
noord holland	Slotervaart	1.7467	0.0490
noord holland	IJburg	1.3368	0.0559
noord holland	Oostelijke Eilanden	1.6894	0.0532
noord holland	Oostelijk Havengebied	1.7176	0.0599
noord holland	Frederik Hendrikbuurt	2.3015	0.0636
noord holland	Van Lennepbuurt	1.7840	0.0659
noord holland	Da Costabuurt	2.4961	0.0652
noord holland	Kinkerbuurt	1.9740	0.0663
noord holland	Kersenboogerd	1.6580	0.0604
noord holland	Pax	2.1805	0.0583
noord holland	Graan voor Visch	2.2276	0.0588
noord holland	Vrijschot-Noord	2.4317	0.0706

Province	Neighborhood	β	MAE
noord holland	Toolenburg	1.2983	0.0489
noord holland	Floriande	1.9099	0.0524
noord holland	Overbos	1.7756	0.0459
noord holland	Bornholm	1.8038	0.0450
noord holland	Beukenhorst-Oost	1.7033	0.0951
noord holland	De Hoek	2.5458	0.0620
noord holland	West	2.0234	0.0587
noord holland	Zuid	1.6443	0.0612
noord holland	Oost	1.8792	0.0630
noord holland	Noord	1.6545	0.0536
noord holland	De President	1.5065	0.1393
noord holland	Graan voor Visch-Zuid	1.7057	0.0630
noord holland	Zuidwijk	1.4003	0.0468
noord holland	Buitenveldert-West	1.1983	0.0752
noord holland	Buitenveldert	1.1466	0.0818
noord holland	Apollobuurt	1.7467	0.0691
noord holland	Stadionbuurt	1.5038	0.0735
noord holland	Prinses Irenebuurt e.o.	1.8808	0.0476
noord holland	Hoofddorppleinbuurt	1.6837	0.0790
noord holland	Willemspark	1.9420	0.0647
noord holland	Schinkelbuurt	1.6432	0.0929
noord holland	Vondelparkbuurt	1.2973	0.0847
noord holland	Helmersbuurt	1.8065	0.0593
noord holland	Overtoomse Sluis	1.9518	0.0632
noord holland	Museumkwartier	1.7210	0.0679
noord holland	Rivierenbuurt	1.6870	0.0562
noord holland	IJselbuurt	1.5689	0.0683
noord holland	Scheldebuurt	1.3635	0.0614
noord holland	Rijnbuurt	1.6563	0.0558
noord holland	De Baarsjes	1.7897	0.0661
noord holland	Landlust	1.7654	0.0572
noord holland	Staatsliedenbuurt	1.8320	0.0608
noord holland	Spaarndammerbuurt	2.2379	0.0616
noord holland	De Pijp	2.2742	0.0570
noord holland	Grachtengordel	1.7834	0.0634
noord holland	Oud-Zuid	1.4461	0.0659