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

Koen Stevens

April 25, 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 heuristic (Lin and Kernighan, 1973). 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, H., G. Laporte, Y. Liu, and T. 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	2.2049	0.0701
groningen	Binnenstad	2.0410	0.0734
groningen	Oosterpoort	2.0627	0.0604
groningen	Rivierenbuurt	1.7826	0.0497
groningen	De Wijert	1.7755	0.0638
groningen	Oosterparkwijk	1.7838	0.0516
groningen	De Hoogte	1.6922	0.0587
groningen	Korrewegwijk	2.0751	0.0537
groningen	Schildersbuurt	2.2203	0.0489
groningen	Paddepoel	1.6792	0.0481
groningen	Oranjewijk	1.9282	0.0545
groningen	Tuinwijk	2.9039	0.0469
groningen	Selwerd	1.5247	0.0478
groningen	Vinkhuizen	1.4803	0.0435
groningen	Hoogkerk-zuid	1.5621	0.0590
groningen	Gravenburg	1.2985	0.0853
groningen	De Held	1.9223	0.0409
groningen	Reitdiep	1.6478	0.0530
groningen	Hoornse Meer	1.5935	0.0447
groningen	Corpus den Hoorn	1.6088	0.0508
groningen	Eemspoort	1.7165	0.0448
groningen	Euvelgunne	1.9353	0.1067
groningen	Driebond	1.9220	0.0599
groningen	Winschoterdiep	1.9630	0.0513
groningen	Eemskanaal	1.7551	0.0418
groningen	Helpman	2.0721	0.0618
groningen	Lewenborg	1.9303	0.0521
groningen	Beijum	1.7897	0.0429
groningen	Maarsveld	1.6844	0.0393
noord holland	Schrijverswijk	1.7493	0.0424
noord holland	Stad van de Zon	1.4790	0.1543
noord holland	Stadshart	1.4533	0.0663
noord holland	Jordaan	1.8471	0.0591
noord holland	Slotervaart	1.7456	0.0437
noord holland	IJburg	1.3349	0.0506
noord holland	Oostelijke Eilanden	1.6866	0.0424
noord holland	Oostelijk Havengebied	1.7129	0.0464
noord holland	Frederik Hendrikbuurt	2.3020	0.0582
noord holland	Van Lennepbuurt	1.7810	0.0596
noord holland	Da Costabuurt	2.4887	0.0585
noord holland	Kinkerbuurt	1.9679	0.0585
noord holland	Kersenboogerd	1.6571	0.0526
noord holland	Pax	2.1825	0.0415
noord holland	Graan voor Visch	2.2300	0.0531
noord holland	Vrijschot-Noord	2.4427	0.0556

Province	Neighborhood	β	MAE
noord holland	Toolenburg	1.2971	0.0443
noord holland	Floriande	1.9124	0.0485
noord holland	Overbos	1.7727	0.0419
noord holland	Bornholm	1.8024	0.0399
noord holland	Beukenhorst-Oost	1.7106	0.0833
noord holland	De Hoek	2.5441	0.0554
noord holland	West	2.0281	0.0489
noord holland	Zuid	1.6462	0.0527
noord holland	Oost	1.8809	0.0560
noord holland	Noord	1.6555	0.0485
noord holland	De President	1.5015	0.1269
noord holland	Graan voor Visch-Zuid	1.7029	0.0549
noord holland	Zuidwijk	1.4000	0.0406
noord holland	Buitenveldert-West	1.1975	0.0671
noord holland	Buitenveldert	1.1503	0.0710
noord holland	Apollobuurt	1.7468	0.0617
noord holland	Stadionbuurt	1.5084	0.0638
noord holland	Prinses Irenebuurt e.o.	1.8819	0.0428
noord holland	Hoofddorppleinbuurt	1.6903	0.0704
noord holland	Willemspark	1.9422	0.0576
noord holland	Schinkelbuurt	1.6424	0.0814
noord holland	Vondelparkbuurt	1.2941	0.0521
noord holland	Helmersbuurt	1.8039	0.0540
noord holland	Overtoomse Sluis	1.9497	0.0560
noord holland	Museumkwartier	1.7178	0.0597
noord holland	Rivierenbuurt	1.6822	0.0507
noord holland	IJselbuurt	1.5712	0.0496
noord holland	Scheldebuurt	1.3686	0.0565
noord holland	Rijnbuurt	1.6555	0.0510
noord holland	De Baarsjes	1.7849	0.0578
noord holland	Landlust	1.7685	0.0518
noord holland	Staatsliedenbuurt	1.8338	0.0520
noord holland	Spaarndammerbuurt	2.2421	0.0540
noord holland	De Pijp	2.2745	0.0521
noord holland	Grachtengordel	1.7832	0.0561
noord holland	Oud-Zuid	1.4443	0.0577