Proposal for Master’s Thesis

Use of Traveling-Sales-Person Tour Length Estimators to Find Solutions to the Vehicle Routing Problem

Author: Stephen Feldman

Contact info: [feldms5@rpi.edu](mailto:feldms5@rpi.edu) | [sfeldman23@hotmail.com](mailto:sfeldman23@hotmail.com) | (732)-354-5379

Sponsors: Professor Cavdar, Professor Bailey

Contents

[Introduction 2](#_Toc207806178)

[1. Problem Statement 2](#_Toc207806179)

[2. Objective: 3](#_Toc207806180)

[3. Preliminary Literature Review 3](#_Toc207806181)

[3.1 A distribution-free TSP tour length estimation model for random graphs. *Bahar Çavdar, Joel Sokol.* 3](#_Toc207806182)

[3.2 PROBABILITY DISTRIBUTION OF THE LENGTH OF THE SHORTEST TOUR BETWEEN A FEW RANDOM POINTS: A SIMULATION STUDY. *Alexander Vinel & Daniel F. Silva.* 4](#_Toc207806183)

[3.3 Heuristics for Vehicle Routing Problem: A Survey and Recent Advances. *Fei Liu, Chengyu Lu, Lin Gui, Qingfu Zhang, Xialiang Tong, Mingxuan Yuan* 4](#_Toc207806184)

[3.4 Worst-Case Analysis of a New Heuristic for the Travelling Salesman Problem. *Nicos Christofides.* 6](#_Toc207806185)

[4. Methodology: 6](#_Toc207806186)

[4.1 Estimator Creation 6](#_Toc207806187)

[4.2 Dataset and Instance Generation 6](#_Toc207806188)

[4.3 Heuristic Creation 7](#_Toc207806189)

[5. Timeline: 7](#_Toc207806190)

[Appendix A: References/Reading List Resources 8](#_Toc207806191)

# Introduction

This research project is focused on optimizing last stage delivery. With delivery services becoming the norm, costs due to suboptimal route choices negatively impact businesses and their customers. Suboptimal routes cost more time and fuel increasing costs, harming the environment, and decreasing customer satisfaction. Currently, optimization models readily obtain solutions for instances of up to 100 customers, however, for instances with relatively few distribution centers servicing many customers it becomes computationally expensive to find optimal solutions. This research seeks to investigate a novel approach to solving vehicle routing problems. By utilizing breakthroughs in traveling-sales-person tour length estimation it may be possible to reduce the computational requirements for obtaining good vehicle routing solutions.

# 1. Problem Statement

The efficiency of last‑mile delivery operations is dependent on solving Vehicle Routing Problems (VRPs) quickly and accurately. In practice, many VRP solvers require solving numerous small Traveling Salesperson Problem (TSP) subproblems during local search or route evaluation. While exact TSP solvers can handle small instances optimally, repeatedly invoking them in high‑frequency VRP heuristics incurs significant computational cost, especially in large‑scale delivery networks with many customers and few depots.

One strategy to reduce this cost is to estimate the tour length without solving for the actual tour. Current estimator approaches often rely on knowledge of the distribution that was used to generate the nodes. The advent of distribution free TSP estimators means that it is possible to calculate tour length without first fitting a distribution.

There is therefore a need for a fast, accurate, and generalizable TSP length estimator that can be embedded into VRP heuristics without sacrificing solution quality. More specifically, the following research questions arise:

1. What features of a route (e.g., MST properties, spatial dispersion, convex hull metrics) most strongly predict its optimal TSP length? What are existing distribution free tour length estimators and how do they perform on Euclidean VRP problems?
2. How does integrating such an estimator into a VRP heuristic affect runtime, solution quality, and robustness compared to exact TSP evaluation?
3. Can the estimator maintain accuracy across different VRP instance classes (uniform, clustered, mixed) without retraining?
4. What other opportunities are presented by the reduced computational complexity of calculating the estimator? Is it possible to create a more intelligent search method based-on increasing the frequency of estimator calls?

Addressing these questions will enable the development of a computationally efficient VRP heuristic that leverages learned TSP estimators to achieve good solutions quicker.

# 2. Objective:

The objective of this research is to develop a heuristic to obtain good solutions for VRPs with reduced computational complexity by leveraging TSP estimators.

To develop this objective, the study will:

* Test the performance of existing TSP estimators with respect to true TSP route length.
* Acquire/develop a dataset of VRP instances with known solutions using existing approaches.
* Design an a search heuristic that is able to find good VRP solutions.
* Implement and tune the heuristics on the existing dataset.
* Design an experiment to benchmark heuristic performance against existing solution approaches.

The results of this study are valuable to enterprises and researchers pursuing improved vehicle routing.

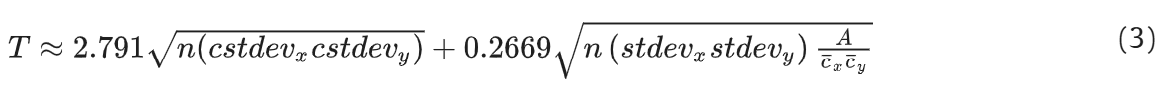
# 3. Preliminary Literature Review

This section reviews research regarding traveling-sales-person problem estimation techniques and existing vehicle routing heuristic algorithms.

## 3.1 A distribution-free TSP tour length estimation model for random graphs. *Bahar Çavdar, Joel Sokol.*

<https://doi.org/10.1016/j.ejor.2014.12.020>

This paper prompted the investigation into finding novel uses for TSP estimators. Cavdar and Sokol developed a function to estimate TSP tour length on random graphs that work for large and very large instances. The paper discusses the main function and an adjustment metric that is required when working with smaller routes. The key advantage is not needing to know the distribution that was used to generate the nodes when creating this estimate. This makes this estimator more widely applicable compared to previous iterations.

* Large instances (1000 nodes per route or greater)
  + Equation (3) is based on the intuition that dispersion of the nodes within the area is related to the optimal tour length.
  + Equation (3) returns solutions within 5% of optimal in O(n) for large instances).
* Medium Instances: (100 to 1000 nodes per route)
  + Equation (4) is an adjustment factor for equation (3) that is fitted to enhance performance on smaller instances.
  + This adjustment enabled superior performance compared to other tested approaches at that size. However, performance declines compared to the larger instances.

## 3.2 PROBABILITY DISTRIBUTION OF THE LENGTH OF THE SHORTEST TOUR BETWEEN A FEW RANDOM POINTS: A SIMULATION STUDY. *Alexander Vinel & Daniel F. Silva.*

<https://informs-sim.org/wsc18papers/includes/files/280.pdf>

The paper studies focuses on an estimation strategy introduced by Beardwood et al (1959) which shows that where is length of optimal tour, and is a constant. Various analyses are examined where simulation models were used to improve the value of the constant to generate more accurate performance. This presents an opportunity to compute optimal tour lengths in O(1) which would radically improve performance. The study attempts to approximate the TSP length by analytical determining if the tour lengths are normally distributed. For instances, where the normality assumption holds, it is possible to improve upon Beardwood et al and achieve a more consistent estimator that is equal in complexity.

The key limitation of this study is the dependance of the estimator on using square or circular regions to compute statistics. Poor performance on rectangular instances which are common in VRP problems leads to an estimator that cannot readily adapt to changing circumstances. It may be possible to utilize the intuition behind this approach in a closely supervised manner but not broadly applicable as estimator for small instances.

This study also concludes that the normality assumption fails on anisotropic regions or regions not uniformly distributed at tour sizes of n<10. This is a common problem with most estimators, however, small node sizes the computation time gained by using an estimator compared to the precision loss is not a worthwhile trade off. It is simpler and more efficient to use traditional heuristics such as Lin-Kernigan or Held-Karp to find tour lengths for such instances.

## 3.3 Heuristics for Vehicle Routing Problem: A Survey and Recent Advances. *Fei Liu, Chengyu Lu, Lin Gui, Qingfu Zhang, Xialiang Tong, Mingxuan Yuan*

<https://arxiv.org/pdf/2303.04147>

This paper provides a structured tour of how heuristics have evolved to tackle Vehicle Routing Problems (VRPs), highlights the major algorithmic building blocks, and points toward new directions. For a VRP variant that does not create routes until the last step several insights and tools emerge:

1. Three Pillars of VRP Heuristics
   * Constructive Heuristics: Build routes from scratch by following simple rules (nearest-neighbor, savings, insertion, sweep). These can quickly generate a feasible assignment which can be improved by an improvement heuristic. These can also be used to break free from local optima.
   * Improvement Heuristics (Local Search): Intra-route moves (relocate, exchange, 2-opt/3-opt, OR-exchange, GENI) tweak node order within a single route. Inter-route moves (cross, λ-interchange, swap*, 2-opt*) shuffle customers or segments between different routes. These are not as to the specific problem as these moves are based on sequence being a factor. As no tours are being created the sequence is not a factor when it comes to the solution.
   * Metaheuristics: High-level frameworks—Simulated Annealing, Tabu Search, Iterated Local Search, Large Neighborhood Search, Genetic Algorithms, Ant Colony Optimization, Memetic Algorithms—wrap local moves in strategies that balance exploration and exploitation. These serve as the driver which guide the moves based on variety of factors.
2. A General Framework for State-of-the-Art (SOTA) Methods Most top-performing VRP solvers follow this four-phase loop:
   1. Initialize a solution (often via multiple constructive heuristics).
   2. Perturb or recombine solutions (crossover, ruin-and-recreate, shaking).
   3. Improve each candidate using local moves.
   4. Select or replace solutions based on quality and diversity.

(2) is challenged significantly by the presence of non-uniform capacity customers. It can be hard to ruin/recreate a solution while preserving the solution properties. Other metaheuristic approaches can help eliminate the need for perturbation and instead allow traversing solutions with feasible local moves.

1. Emerging Trends and Their Relevance:
   * Unified Heuristics: One solver framework handles many VRP variants by plugging in attribute-specific modules (time windows, backhauls, heterogeneous fleets).
   * Automatic Heuristic Design: Auto-tuning parameters (e.g., the ruin/recreate weights in ALNS), algorithm selection (picking the best heuristic family per instance), or even composing algorithmic components on the fly from a library. A trained model could pick, at each step, which move operator (relocate vs. swap vs. cross-exchange) yields the largest estimator-predicted gain.
   * Machine-Learning-Assisted Heuristics: Instead of end-to-end learning, use ML to guide building blocks the search. Predict which move will improve a route most. This can also be applied to initial solution generation.

By reviewing existing tools this paper provides the foundation for building a VRP heuristic. It also summarizes various approaches which can be benchmarked against going forward.

## 3.4 Worst-Case Analysis of a New Heuristic for the Travelling Salesman Problem. *Nicos Christofides.*

<https://doi.org/10.1007/s43069-021-00101-z>

Christofides revisits the classical double‐tree heuristic for the metric Traveling Salesman Problem (TSP), in which one begins by computing a minimum spanning tree (MST), doubling its edges to form an Eulerian multigraph, and then shortcutting repeated visits under the triangle inequality to obtain a Hamiltonian tour. He proves that this procedure—while guaranteed never to produce a tour shorter than the MST—yields a tour whose length does not exceed 1.5 times the length of an optimal TSP tour.

While not a new estimator this approach reveals that the computation of MST O(n^2) can reliably bound the optimal length. This implies that it can be used to tighten the estimator performance. However, this also led to an additional intuition. If the MST is tied to the TSP optimal tour length then perhaps it would be possible to estimate the TSP optimal tour length (without computing the sequence) from the length of the MST.

# 4. Methodology:

The project seeks to obtain reliable estimators for TSP tour lengths across Euclidean capacitated vehicle routing problems. Then implement a heuristic that uses those estimators to obtain solutions to the problem. Finally, benchmark the heuristic against existing approaches.

### 4.1 Estimator Creation

The first step in the process is to create a reliable objective function that can be used to approximate the tsp optimal tour length across all instances. The complexity of this function should be lower than the complexity of computing the complete TSP as the length of the tour increases. For very small tours, it may be reasonable to compute the TSP using existing approaches due to negligible performance gain. To create this function existing research will be leveraged and tested to see where estimation strategies perform the best. A hybrid selection of these strategies will be combined to create a single function which will be tested against vehicle routing problems with known solutions. The objective of a good estimator is to closely model the cost of the vehicle routing problem. Internal benchmarking is required as most estimators were benchmarked on TSP instances meaning their errors were assessed on a per case basis. In this instance, the error in a VRP will be compounded by the number of vehicles. This will be considered during the objective function construction phase.

### 4.2 Dataset and Instance Generation

The objective function will need to be tested against existing VRP instances. This means constructing or procuring a dataset of VRPs and their optimal solutions. Furthermore, efficient TSP solver such as Concorde or LKH-3 would be required to transform the assignments created by the future heuristic to complete solutions. The first dataset for which the training pipeline will be implemented is the XML100 dataset which contains 10,000 instances and solutions.

### 4.3 Heuristic Creation

Once the objective function has been created and a dataset/evaluation pipeline is created. The core of the system, the heuristic itself can be created. This will require exploring new swapping strategies due to the lack of sequence involved. Initial testing will occur on simpler vehicle routing problems with uniform demand allowing the swap mechanisms to be simpler and straightforward. When sufficient performance is achieved on simpler instances, more swap mechanisms will be added to effectively traverse the solution space of complicated CVRPs.

After the heuristic performs well on XML100, the heuristic will be tested against new larger CVRP instances. In this case, an optimal solution may not be known, and the performance will be checked by other heuristic approaches.

# 5. Timeline:

Research will take place over the course of 12 weeks, 20 hours a week for a total of 6 credits. Professor Cavdar and Professor Bailey will be advising this research. Review meetings are scheduled for Wednesdays at 3:00 PM, to encourage the open flow of ideas and to discuss new findings.

The following project milestones will include:

1. Investigation Phase (May – June 2025): is it possible to use an optimization model to obtain an optimal solution to a VRP based only on Professor Cavdar’s TSP estimation function.
2. Research Phase (July – August 2025)

* Setup training and testing pipelines
* Research existing estimator strategies
* Develop and implement an objective function
* Test the objective function on VRP instances to assess the overall estimator quality.

1. Heuristic Development (September 2025)
   * How do we know if these estimations are accurate, useful and provide value?
   * What correlations exist? Are there any relationships among factors?
2. Experimental Design and Analysis of Results (October 2025)
   * Validate the performance of the heuristic on XML100.
   * Prepare and execute an experiment comparing the performance of the new estimator driven heuristic against other existing strategies on the same XML100 dataset.
   * Setup a new pipeline for larger CVRP instances (far more than 100 nodes) and compare the performance of the new heuristic to existing strategies.
3. Technical Writing (1st half of November 2025)
4. Peer Review/ Finalization (November – December 2025)

|  |  |  |
| --- | --- | --- |
| Week Start Date | Week End Date | Description |
| 5/1/2025 | 5/7/2025 | Past |
| 5/8/2025 | 5/14/2025 | Past |
| 5/15/2025 | 5/21/2025 | Past |
| 5/22/2025 | 5/28/2025 | Past |
| 5/29/2025 | 6/4/2025 | Past |
| 6/5/2025 | 6/11/2025 | Past |
| 6/12/2025 | 6/18/2025 | Past |
| 6/19/2025 | 6/25/2025 | Past |
| 6/26/2025 | 7/2/2025 | Past |
| 7/3/2025 | 7/9/2025 | Past |
| 7/10/2025 | 7/16/2025 | Past |
| 7/17/2025 | 7/23/2025 | Past |
| 7/24/2025 | 7/30/2025 | Past |
| 7/31/2025 | 8/6/2025 | Past |
| 8/7/2025 | 8/13/2025 | Past |
| 8/14/2025 | 8/20/2025 | Past |
| 8/21/2025 | 8/27/2025 | Past |
| 8/28/2025 | 9/3/2025 | Assess estimator accuracy |
| 9/4/2025 | 9/10/2025 | Explore correlations |
| 9/11/2025 | 9/17/2025 | Adapt swap strategies |
| 9/18/2025 | 9/24/2025 | Integrate swap mechanisms |
| 9/25/2025 | 10/1/2025 | Validate heuristic on XML100 |
| 10/2/2025 | 10/8/2025 | Compare strategies on XML100 |
| 10/9/2025 | 10/15/2025 | Setup large CVRP pipeline |
| 10/16/2025 | 10/22/2025 | Compare large CVRP performance |
| 10/23/2025 | 10/29/2025 | Complete October analysis |
| 10/30/2025 | 11/5/2025 | Draft technical report |
| 11/6/2025 | 11/12/2025 | Continue technical writing |
| 11/13/2025 | 11/19/2025 | Peer review & revisions |
| 11/20/2025 | 11/26/2025 | Finalize writing & edits |
| 11/27/2025 | 12/3/2025 | Submit & finalize report |
| 12/4/2025 | 12/10/2025 | Buffer & wrap-up |
| 12/11/2025 | 12/17/2025 | Final polishing |

# Appendix A: References/Reading List Resources

[1] Çavdar B, Sokol J. A distribution-free TSP tour length estimation model for random graphs. European Journal of Operational Research. 2015;244(1):21–30. <https://doi.org/10.1016/j.ejor.2014.12.020>

[2] Vinel A, Silva DF. Probability distribution of the length of the shortest tour between a few random points: A simulation study. In: Proceedings of the 2018 Winter Simulation Conference; 2018. p. 3157–3168. <https://informs-sim.org/wsc18papers/includes/files/280.pdf>

[3] Liu F, Lu C, Gui L, Zhang Q, Tong X, Yuan M. Heuristics for Vehicle Routing Problem: A Survey and Recent Advances. arXiv. 2023;arXiv:2303.04147. <https://arxiv.org/pdf/2303.04147>

[4] Christofides N. Worst-Case Analysis of a New Heuristic for the Travelling Salesman Problem. Operations Research Forum. 2022;3:20. <https://doi.org/10.1007/s43069-021-00101-z>

[5] Varol T. NN-Estimators-For-Optimal-TSP-Tour-Lengths. GitHub repository. 2022. <https://github.com/tahavarol/NN-Estimators-For-Optimal-TSP-Tour-Lengths>

[6] Chen Y. Estimators for Optimal TSP Tour Lengths. GitHub Pages. 2022. <https://realcyh.github.io/yuhang-chen/tsp/>