The balancing traveling salesman problem: application to warehouse order picking

Article *in* Top · May 2020

DOI: 10.1007/s11750-020-00557-y

Rajan Batta
University at Buffalo, State University of New York

Mark Karwan
University at Buffalo, State University of New York

The Balancing Traveling Salesman Problem: Application to Warehouse Order Picking

Atieh Madani, Rajan Batta*, Mark Karwan

342 Bell Hall, Department of Industrial and Systems Engineering,
University at Buffalo, Buffalo, NY 14260, USA

*Corresponding author; batta@buffalo.edu

Abstract

This paper discusses the problem of predicting the length of Traveling Salesman Problem (TSP) tour in dynamic and uncertain environments. Our findings, which are guided by extensive simulation runs, provide statistical estimations for the tour length under different scenarios. One of the applications that can benefit from these estimates includes warehouse order picking, which has seen increased importance due to online shopping. The utility of statistical estimates for TSP tour length for order picking is demonstrated on a common warehouse layout.

Keywords: Traveling Salesman Problem, Warehousing Systems, Routing

1 Introduction

The traveling salesman problem (TSP) finds the shortest route to visit all the necessary nodes in a network and then return to the starting point. This problem has always been an interesting topic for many scientists. Even though there exists extensive literature regarding this topic, since the problem is NP-hard finding an optimal solution is likely to be computationally expensive. A good solution–which is optimal or close to optimal and fast to reach–which is not computationally expensive is therefore valuable. Some of the applications for this problem includes finding best routes for emergency vehicles, mobile services, and warehouse order-picking. TSP is considered an important problem because it either has a direct application or it appears as a subproblem in many application domains–including vehicle routing and airline scheduling. TSP has been shown to be NP-complete, Karp (1972) and Papadimitriou (1977), and that can explain the large variety of heuristic approaches that have been developed towards solving it.

TSP was first considered in a mathematical way in the 1930s by Merrill Flood who was interested in solving a school bus routing problem. Later in the 1950s and 1960s the problem became popular especially after the RAND Corporation in California proposed prizes for solving the problem. George Dantzig, Delbert Fulkerson and Selmer Johnson made significant contributions in investigating this problem, by modeling it as an Integer Program and developing a cutting plane solution algorithm. They wrote a seminal paper on this matter and found the optimal solution for a problem as large as 49 cities. Later on, many scientists and mathematicians studied this problem. For example, in the 1960s a new perspective started forming for this problem. Instead of looking for the optimal solution directly, one can find a reasonable range of solutions by creating lower bounds for the problem, Lawer et al. (1985).

There have been several approaches in the literature regarding how to solve the TSP, these include solving the problem either by itself (solo) or embedded in a larger problem. In the embedded cases, TSP is part of a more complex and larger problem. Examples for approaches towards solo problems are genetic algorithms (Grefenstette et al. (1985)), tabu search (Gendreau et al. (1994)), simulated annealing (Kirkpatrick et al. (1983)), swarm optimization (Kennedy (2011)), harmony search (Geem et al. (2001)), and many more.

In some cases of the TSP problem, finding or estimating the length of the optimal path becomes the key problem compared to finding the path itself. As an example, in public transportation, all the stops in a city are needed to be allocated to buses such that each bus has approximately the same amount of distance to cover. In this case, one will only need the length of TSP tour to solve the allocation problem. Beardwood et al. (1959) suggested that the length of the shortest path in a planar region with the area of A–in most cases–is asymptotically relative to $\sqrt{NA}$ (for large value of N) where N is the number of nodes, and tour length $T = b\sqrt{NA}$. This relativity holds for any bounded convex region where the appearance of locations are independent. They also presented bounds for k-dimensional Euclidean space. Later, Stein estimated b using Monte Carlo simulation experiments to be 0.75 in Stein (1978a) and 0.765 in Stein (1978b). There have been other works regarding the estimation of the TSP tour length. Daganzo (1984) proposed a formula for the length of the TSP tour by partitioning the area into clusters of nodes. Kwon et al. (1995) developed a model using regression and neural networks to estimate the length of the TSP tour for up to 80 nodes. They used Beardwood et al. (1959) as their master model. They suggested values for constant b and added parameters to improve the fitness of the model. Instead of estimating a universal value for b, some scientist decided to solve the problem by setting an upper bound for b. This is mainly due to the NP-hard nature of the problem and amount of time consumption to estimate b. For example, Ong and Huang (1989) used a heuristic approach to set 0.74 as an upper bound for b. Çavdar and Sokol (2015) developed a new method to estimate the length of TSP and outperformed previous algorithms in both accuracy and speed. They estimated the tour length using parameters such as location of all nodes, standard deviation of coordinates and standard deviation of absolute distances of all nodes from the central horizontal and vertical axes. Brown and Guiffrida (2017) studied a delivery fleet's planning using Beardwood et al. (1959)'s work in a stochastic model to identify the best course of action for the fleet. Franceschetti et al. (2017) presented a very thorough literature review on continuous approximation models in the application of freight distribution management. Nicola et al. (2019) developed regression models to estimate the travel distance in a TSP, the capacitated vehicle routing problem with Time Windows (CVRP-TW), and the multi-region multi-depot pickup and delivery problem (MR-MDPDP). They introduce some new parameters to the model such as sum and variance of nodes and depot distances, minimum, maximum, and sum and variance between nodes. Vinel and Silva (2018) explored the hypothesis that TSP tour length follows a normal distribution, for an isotope area with uniform distribution for the location of nodes and number of nodes greater than 10. They used Beardwood et al. (1959)'s equation to estimate the sample mean. For the standard deviation estimation, they proposed $\sigma = c + \frac{d}{\sqrt{n}}$ where $c = 0.4985$ and $d = 0.1826$.

In addition to the length of the TSP tour, we are also interested in the element of time over which arrivals occur. In most cases the goal is to minimize the travel cost or travel distance and the TSP is considered deterministic. However, in reality, routing can occur in a dynamic environment. To be more precise, time could have an influence on the number of nodes in the TSP network. Ignoring this aspect of the problem would cause one to underestimate the cost and sensitivity of other factors. To overcome this problem, researchers created new algorithms to include the element of time in the analysis using probabilistic modeling. Examples of models that can have a dynamic environment in this subject are the dynamic traveling salesman problem, traveling repairman problem, and dial-a-ride problem. These algorithms inform us on how to react to the appearance of an unexpected set of nodes in a given region, while considering objectives such as customer waiting time or travel distance. Another approach to this problem is the model by Bertsimas and Van Ryzin (1991) that uses queuing theory to reduce the waiting time in a vehicle routing problem in which the customers are waiting at each node to be served. Bertsimas (1992) considered a situation in which the demand is not fixed and each demand could be a potential node in the vehicle routing problem (VRP). He suggested creating a pre-fixed route and strategies to handle unexpected demand along the route. For further information on this topic, refer to Ghiani et al. (2003).

Most of the methods stated above are trying to estimate TSP tour length using Euclidean distance, for an existing system of nodes, whereas our method is about estimating TSP tour length to accommodate an unprecedented node/nodes (where the number of nodes is a random variable with a probability distribution) using both Euclidean and Manhattan distances. We decided to use Manhattan distance because of its practicality of applications in real world scenarios. Our approach helps the decision maker to react appropriately to an impromptu change in the TSP. Therefore, where other methods will give the length of the tour, our method will provide the expected percentage of change (increase) in tour length. Our problem focuses on both distance and time in the TSP. One of the examples of use of the balancing TSP is a scenario in which there are limited traversing vehicles (salesmen) and we are faced with the question of whether to wait longer to have more destinations versus starting the route with the current set of destinations. Considering that our objective function is to minimize both routing/traveling cost and lateness cost, we would need to decide at the beginning of each period to go on tour or wait for more nodes to be added to the system. By waiting longer, there will be a chance of having more nodes per TSP tour which means an overall reduction in the total cost (e.g. one tour vs. multiple tours). This means if the system receives 10 total nodes to visit, if we can arrange two tours to pick up all 10 versus three tours, it can save us some traveling cost since there will be less total traversed distance. However, if we wait longer we may not be able to visit all the destinations due to time constraints or we may face tardiness penalties. So the question of finding the best trade off point arises.

Practical scenarios include a flower shop that has a delivery service and would want to know whether or not to wait longer for more orders to come in (to reduce total transportation cost) or start delivering to improve service time and customer satisfaction; a mobile mechanic that has a service center and travels to the customers to provide services; or an order picker in a warehouse waiting for more orders to arrive.

The rest of this paper is organized as follows. Section 2 presents the problem description. The design of experiments and results are covered in Section 3. Instances with high rate of node arrival is discussed in Section 4. An analysis of a special case of this problem suitable for a warehouse picking situation is provided in Section 5. Finally, Section 6 contains our summary and conclusions.

2 Problem Description

The balancing traveling salesman problem is described as follows. There are N time periods. For every period there is a certain set of nodes that a salesman needs to visit. Area A includes all the potential nodes to form a TSP problem. Each arrival could lead to a new extra node in the TSP route. A node could be a customer who is waiting for service, an item to be picked up for shipment, etc. We assume the expected number of arrivals during each period is λ. We also know the number of orders already on hand. This leads to our main question: " What is the expected amount of time or distance the next period's node arrivals would add to the tour?". Our objective is to find the expected length of the tour for future periods. This will enable us to improve and determine the average duration of traversing the routes and the total transportation cost and predict any tardiness penalties. We have chosen simulated annealing method to solve the TSP part of our problem based on the Antosiewicz et al. (2013)'s study, as they demonstrated overall superiority of simulated annealing compared to many other algorithms. The decision to use a heuristic method over an exact solution algorithm to find the TSP tour length comes from the NP-hard nature of the TSP and the computational expenses that we face. In the following sections, one can see that considering the number of simulations in our experiments, it is not feasible to use the optimal approach. To fully explore the potential of using an exact approach, we used the Concorde solver to find an optimal solution to the TSP. Additionally, we implemented a sequential procedure to reach 0.1 relative error and confidence level of 95 percent, Law et al. (2000) to calculate the number of replications. If we used the Concorde solver the simulation experiments for some of the cases would take up to months of computational effort to calculate. We therefore selected the heuristic method for our experiments.

Table 1 gives a snapshot of how this paper compares to other work that has already been published on this topic.

Table 1: Comparison of Models that Estimate TSP Tour Length

Article	**Direct Prediction of TSP Tour Length**	**No Restriction on # of Nodes**	**Nodes Arrival Random Var**	**No Locations Info Needed for Estimation**
Beardwood et al. (1959)	✓	✓	×	✓
Çavdar and Sokol (2015)	×	✓	×	×
Vinel and Silva (2018)	✓	×	×	✓
Our Model	✓	✓	✓	✓

3 Design of Experiments and Results

In this section the effects of arrival rate of nodes on TSP tour length are investigated under different scenarios which include both static and dynamic environments. Since our purpose is to find the estimation for TSP tour length, our experiments involve one tour at a time or per simulation. We studied three cases: **Single Arrival**, **Poisson Arrival**, **Poisson Arrivals with Prior Knowledge of Node on Hand**. **Single Arrival** investigates the effect of adding one additional node on the TSP length in a static environment with no uncertainty. **Poisson Arrival** studies the change in TSP length per period when the arrival of a node follows a Poisson distribution. Lastly, **Poisson Arrivals with Prior Knowledge of Node**

on Hand considers a Poisson distribution for the arrival of nodes, which is conditional on knowing the exact number of nodes on hand. The difference between the second and third cases is related to the assumption of the number of nodes on hand. In the second case we assume that there are initially no nodes in the system, whereas in the third case we assume that there are one or more nodes initially in the system. The following sections provide detailed explanations of design of experiments for each of these cases.

3.1 Single Arrival

Let us assume we have $n-1$ nodes for which we already know the TSP tour length. We would like to estimate how much one extra node would add to the existing tour length. Our purpose in this section is to find a distribution or a fit for this change in the length of the TSP tour, as a function of the total number of nodes n in the system (including the newly generated one). Assume that area A contains all the original nodes plus the newly generated node. Also assume that node locations are independent and uniformly distributed in A. In our experiments, we want to ensure that we produce enough replications so that the empirical results are statistically reliable and unbiased. We randomly locate the first $n-1$ nodes $30 * 2^{n-1}$ times and for each instance we create 100 replications for generating locations of the n^{th} node. This results in the total of $3000 * 2^{n-1}$ simulation runs. For example, when $n = 4$, for the first three nodes we ran $30 * 2^3$ instances and for the 4^{th} node we ran 100 instances. We assume that the highest number that n can take is 15, for which we replicate the first 14 nodes $30 * 2^{14}$ times and for each replication we generate the 15^{th} node's location 100 times. We consider both square and rectangle shape area as well as Euclidean and Manhattan distances, recognizing that the Manhattan distance has application in warehouse order picking and in mobile services that use city street layouts.

Our results indicate that the trend for increased distance when adding an extra node follows a power equation with positive coefficient and negative power. In our experiments, we took the mean of all the runs for each n^{th} node and show the result for the fitted line. Figure 1 presents the average percentage increase in total Manhattan distance of a square area by adding an n^{th} node. The corresponding equation for the dashed line has an R^2 of 0.81, which indicates a good fit. The line follows $y = 0.23(x)^{-0.77}$ where $x \in \{2, 3, ..., n\}$. The results for a square area with Euclidean travel (square-Euclidean), a rectangular area with Euclidean travel (rectangle-Euclidean), and a rectangular area with Manhattan travel (rectangle-Manhattan) are, respectively, $y = 0.23(x)^{-0.79}$, $y = 0.62(x)^{-1.32}$, and $y = 0.29(x)^{-1.07}$ with R^2 values ranging between 0.82 and 0.91.

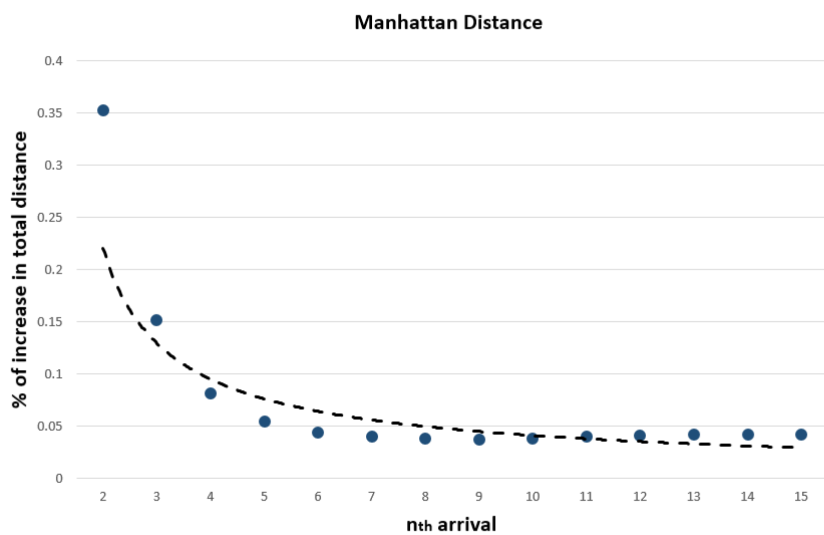

Figure 1: Results of Single Arrivals' Tour Length Increase for a Square Area with Manhattan Distance (Square-Manhattan)

The standard deviation of additional length as a result of one extra node n is calculated for all $n \in \{1, 2, 3, ..., 15\}$. Using these standard deviations a line is fitted to estimate the variance for the change in TSP tour length by adding an additional extra node. The fitted line for square-Manhattan model has $R^2 = 0.88$ and follows $y = 0.23x^{-0.79}$ where x signifies the number of nodes added ($x \in \{1, 2, 3, ..., 15\}$) and y represents the standard deviation of change in tour length. We found $y = 0.23x^{-0.79}$ to be the fitted line for the square-Euclidean model with $R^2 = 0.88$ as well. The lines for rectangle-Euclidean and rectangle-Manhattan models are in order $y = 1.03x^{-1.45}$ and $y = 0.69x^{-1.18}$ respectively with the R^2 of 0.95 and 0.98 where x signifies the number of nodes added and y represents the standard deviation of change in tour length. The standard deviation of results when the second node is added is the highest. This is expected, because by having more nodes on hand, the probability of facing a significant increase in tour length by adding an extra node decreases due to saturation of the area. For example, assuming the area is a square, when adding a second node to the route, the two nodes can be located as far as the length of the area's diagonal. This is in contrast to the case of adding a third node, where the possible longest distance between the third node and the other two nodes (assuming both of them are not located in one point) is one side of the square. In other words, the upper bound for the possible additional distances from the new node to existing nodes decreases as the number of existing nodes increases. There will be a point in the process when the existing nodes cover the entire area and adding another node has minimal impact on the length of the route (full saturation of the area). We observe this phenomenon in our experiments. As the number of nodes increases the variation in results reduces. There is a slight bump at the end of the Figure 1 that can be explained by the fact that we used a heuristic solution method (simulated annealing) to solve for the best TSP route as opposed to an optimal scheme. After a certain point (in this example after the 7^{th} node) the rate of increase in TSP tour length decreases. However, since we are using simulated annealing to solve the TSP, there are more opportunities to improve the route and this causes slightly higher variation.

3.1.1 Comparison of single arrival vs Beardwood et al. (1959) method vs Çavdar and Sokol (2015) method

Beardwood et al. (1959) and Çavdar and Sokol (2015) approaches are compared to the single arrival case. We chose these two approaches after viewing the comparison table presented in Çavdar and Sokol (2015). Çavdar and Sokol showed their model outperforms any other existing model in this area. Even though both methods perform well with a high number of nodes, we want to show this comparison to provide a perspective about our model versus existing methods. We ran 100 simulations of rectangle-Euclidean model. Each simulation is a TSP problem in which random nodes were sequentially added to the network up to a total of 10 nodes (excluding the depot). For comparison purposes we only added up to 10 nodes since we observed a saturation effect in the regression line. Considering that the duration to solve a TSP with 10 nodes vs 15 nodes increases significantly and that the change in tour length slows down considerably after 10 nodes, we decided to run the simulation for this section up to 10 nodes. After each additional node we calculated the TSP tour using simulated annealing and mixed integer programming (Miller-Tucker-Zemlin formulation, Miller et al. (1960)). After each new node was added we also captured the changed in tour length. We used Beardwood et al. (1959)'s formula and Cook (2011) estimation for the constant in the formula which is $b = 0.712$ and Çavdar and Sokol (2015)'s method to estimate the tour length and the change in tour length after each additional node is introduced to the system. The results show that the single arrival regression has a better estimation than two other methods for the estimation of both tour length and tour length change. The R^2 for tour estimation using single arrival regression is 0.96 whereas for the other two methods they have negative R^2 with the ratio of estimated length to optimal length of 0.64 for Beardwood et al. (1959) and 0.65 for Çavdar and Sokol (2015). Comparing the estimated increase in tour length, single arrival regression gives R^2 of 0.79 versus Beardwood et al. (1959) and Çavdar and Sokol (2015) give R^2 of 0.69 and 0.68 respectively. Çavdar and Sokol (2015) showed that for number of nodes less than 10 their model does not provide a good estimation.

3.2 Poisson Arrivals

The actual number of nodes added during a period is a random variable. We selected a Poisson random variable with the rate of λ for testing purposes, since this random variable closely models situations such as warehouse order picking (an intended application of our model). The experiment itself is similar to 3.1 with the exception of the number of added nodes per period. This means the locations are still independently and uniformly distributed. The assumption of node locations to be independent and uniformly distributed is held in this case as well. For the first $t-1$ periods we have simulated $30 * 2^{t-1}$ arrivals; an arrival contains number of node(s) and their location(s), and for each instance the t^{th} period is then replicated 100 times.

Our main purpose is to observe system behavior by introducing a new dynamic element, which is the number of nodes per period. Towards the end we created a statistical measurement. In our experiments, the rate of the Poisson distribution λ is an independent variable that takes different values, $\lambda \in \{0.25, 0.5, 0.75, 1, 1.25\}$. These values are largely chosen around 1 in order to be able to compare the results with the single arrival case.

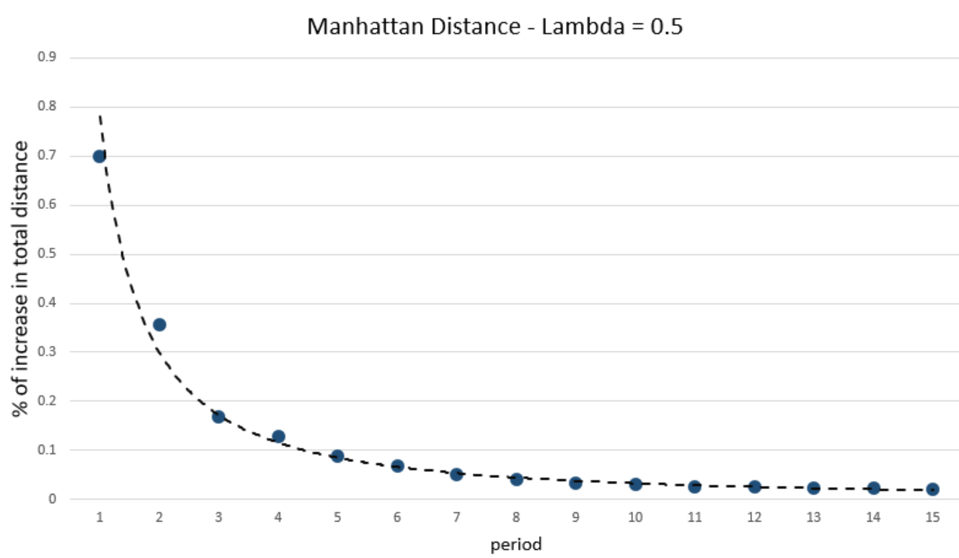

Figure 2: Results of Poisson Arrivals' Tour Length Increase for a Square Area with Manhattan Distance (Square-Manhattan)

The result shows a power equation with the same coefficients and the power signs as Section 3.1. The values of these parameters depend upon the value of λ. Figure 2 illustrates the trend line for the percentage change in tour length using a Poisson distribution with $\lambda = 0.5$. The corresponding equation for the fitted line is $y = 0.78(x)^{-1.38}$, where period $x \in \{1, 2, 3, .., 15\}$, and y represents the percentage of change in tour length and the R^2 is

0.99.

Table 2 shows equation parameters for each experimented λ and their R^2 goodness of fit for Manhattan distance.

Table 2: Parameters of Fitted Lines for Poisson Arrivals with Different Rates

Square Shape Area				Rectangle Shape Area			
Lambda	**Coefficient**	**Power**	R^2	**Lambda**	**Coefficient**	**Power**	R^2
0.25	1.17	-1.43	0.99	**0.25**	0.06	-0.46	0.91
0.5	0.78	-1.38	0.99	**0.5**	0.07	-0.38	0.88
0.75	0.47	-1.12	0.96	**0.75**	0.10	-0.39	0.82
1.0	0.37	-0.95	0.90	**1.0**	0.11	-0.31	0.78
1.25	0.24	-0.7	0.88	**1.25**	0.12	-0.22	0.64

One can determine parameters and therefore the equation for the increase in total tour length when the number of additional nodes is based on a Poisson distribution with a rate of λ. The fitted lines for both coefficient and power are polynomial. The fitted lines for these two parameters of the square shape area are $y_c = 0.83\lambda^2 - 2.16\lambda + 1.65$ and $y_p = 0.35\lambda^2 + 0.22\lambda - 1.53$ respectively with R^2 equal to 0.99 for both, for positive values of λ. The rectangle shape area results for coefficient and power parameter fitted lines are $y_c = -0.04\lambda^2 + 0.11\lambda - 0.42$ and $y_p = 0.02\lambda^2 + 0.05\lambda + 0.05$ respectively with the R^2 of 0.99 for the former and 0.88 for the latter.

The standard deviations for all the fitted lines in this part of the study were also calculated. These equations help us estimate uncertainty in predicted tour length when the arrivals follow a Poisson distribution. Figure 3 shows the standard deviation for square-Manhattan model when $\lambda = 0.75$ with the equation of $y = 0.51x - 1.19$ and $R^2 = 0.98$ where x signifies period $x \in \{1, 2, 3, ..., 14\}$ and y represents the standard deviation of change in tour length. Based on the results of all standard deviations, statistical equations are developed to help us find the variation for different λ values. Table 3 shows the standard deviation's parameter values as a function of λ. The fitted lines for square shape area for the power and coefficient of standard deviations are $y_p = 1.27\lambda^2 - 1.86\lambda - 0.32$ and $y_c = 1.8e^{-0.99\lambda}$ respectively with $R^2 = 0.79$

for the former and $R^2 = 0.71$ for the latter. Similarly, the results for rectangle shape area for the power and coefficient are $y_p = 3.2\lambda^2 - 4.97\lambda + 0.91$ and $y_c = -3.33\lambda^2 + 4.84\lambda - 0.62$ respectively with the $R^2 = 0.98$ for the former and $R^2 = 0.88$ for the latter. As expected, the Poisson arrival case has greater variability than the single arrival case discussed in 3.1. This is due to increased uncertainty associated with the number of nodes being generated per period.

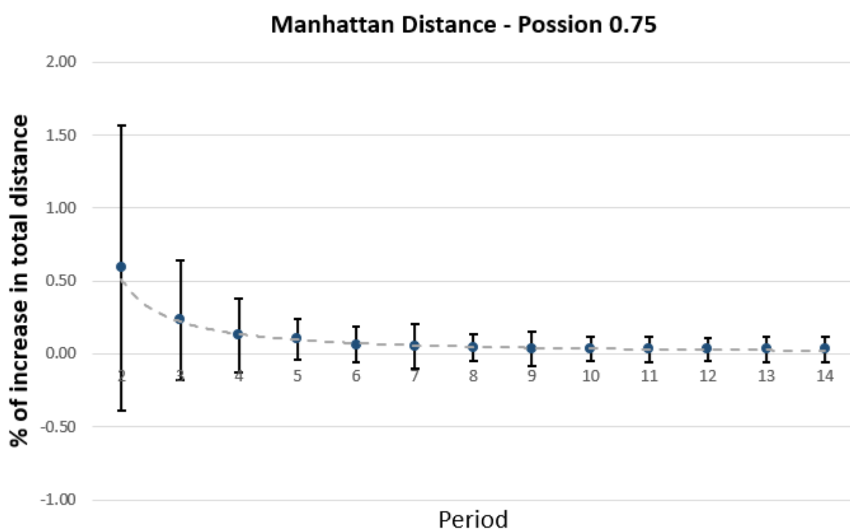

Figure 3: Results of Poisson Arrivals' Tour Length Increase Standard Deviation for a Square Area with Manhattan Distance (Square-Manhattan)

Table 3: Parameters of Standard Deviation Trend Lines of Poisson Arrivals with Different Rate

Square Shape Area				Rectangle Shape Area			
Lambda	**Coefficient**	**Power**	R^2	**Lambda**	**Coefficient**	**Power**	R^2
0.25	1.12	-0.66	0.89	**0.25**	0.43	-0.16	0.61
0.5	1.61	-1.05	0.90	**0.5**	0.79	-0.70	0.77
0.75	0.73	-0.92	0.93	**0.75**	1.35	-1.08	0.78
1.0	0.74	-0.90	0.91	**1.0**	0.79	-0.85	0.83
1.25	0.48	-0.68	0.75	**1.25**	0.25	-0.30	0.87

3.3 Poisson Arrivals with Prior Knowledge of Node on Hand

In this part of our study we investigate the change in tour length per period due to additional nodes generated from a Poisson distribution, given that there is a known number of nodes in the system. This problem is a combination of the material in Sections 3.1 and 3.2. In Section 3.1 the focus was on how adding the n^{th} node to the system changes tour length, whereas in Section 3.2 it was the change in tour length due to Poisson arrival of nodes per period. We combine these two perspectives to address a case that has a known number of nodes in the system and there are additional nodes generated by a Poisson distribution. Figure 4 represents the results from the set of experiments that were performed which were conditional on having one node on hand and Poisson arrivals with $\lambda \in \{0.25, 0.5, 0.75, 1, 1.25\}$.

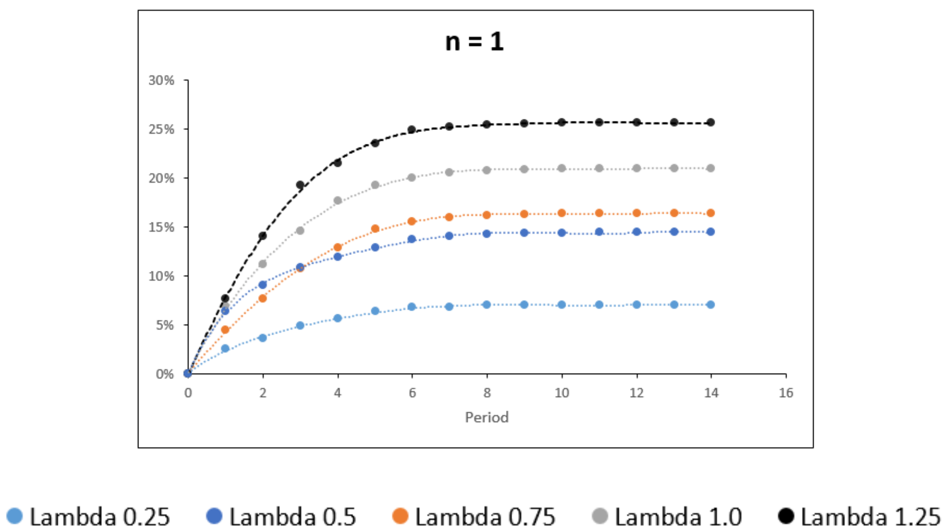

Figure 4: Results of Poisson Arrivals with Prior Knowledge of Node on Hand' Tour Length Increase for a Square Area

The rate of increase in tour length does not have a direct relationship with the value of λ. As one can see from Figure 4, in the first period the rate of change for $\lambda = 0.5$ is faster than that for $\lambda = 0.75$. The reason behind this behavior can be explained by saturation effect of the area. This Figure represents the percentage of change and not the tour length itself. When $\lambda = 0.5$ it takes longer to reach saturation and as a result there is more increase in the percentage of change at the beginning versus when $\lambda = 0.75$. On the contrary, this effect may not be observed if the value of λ is high enough to the point that even with the saturation effect, the change in TSP tour length would be more for higher arrival rates (e.g. λ 0.75 compare to 1.0). A phenomenon that is common for all arrival rates very close to zero, is that the increase in length goes up with an increase in λ value.

Table 4 shows the results for change in tour length from all our simulations. The x axis represents the period and the y axis represents the number of node(s) on hand. The table presents the results for the following arrival rates, $\lambda = 0.25, 0.5, 0.75, 1.0, 1.25$. The number of nodes on hand, n, is a set of integers $\in \{0, 1, 2, ..., 10\}$ and the total number of periods, t, is 14. Each cell shows how much increase in TSP tour length is expected for the corresponding λ, period, and number of nodes on hand. Due to heuristic nature of simulated annealing method that we use to solve the TSP problem, we may see a negative change in tour length specially with a smaller λ. However this phenomenon has not been observed often and it can be neglected since it has a minimum effect on fitting our regression.

We use multiple-regression to develop a statistical model to be able to produce the expected change in tour length. Results from Table 4 provide the inputs to this model. The independent variables in the multiple-regression are the number of nodes on hand, arrival rate λ and the period in which the system is simulated. The dependent variable is the expected change in the TSP tour length. The output of regression for each arrival rate is illustrated in Table 5. The results show that the goodness of fit improves with an increase in arrival rate. The regression equations are quadratic polynomial. In order to be able to find the parameters for different λ values, we fitted a line to each parameter of the multiple-regression equations, see Table 6. Using these equations, one can obtain the model's parameters for any λ value, and form an equation to calculate the expected change in tour length by period and number of nodes on hand. It is interesting to note that for both square and rectangle

Table 4: Results of Poisson Arrivals with Prior Knowledge of Node on Hand' Tour length Increase for a Square Area

0.25	t=1	t=2	t=3	t=4	t=5	t=6	t=7	t=8	t=9	t=10	t=11	t=12	t=13	t=14
n=1	2.5%	3.6%	4.9%	5.6%	6.3%	6.8%	6.8%	7.0%	7.0%	7.0%	7.0%	7.0%	7.0%	7.0%
n=2	1.1%	1.2%	3.0%	3.9%	4.3%	4.6%	4.7%	4.7%	4.7%	4.7%	4.7%	4.7%	4.7%	4.7%
n=3	0.9%	1.3%	1.6%	2.1%	2.1%	2.1%	2.2%	2.2%	2.2%	2.2%	2.2%	2.2%	2.2%	2.2%
n=4	1.1%	3.1%	3.6%	3.9%	4.2%	4.3%	4.3%	4.3%	4.3%	4.3%	4.3%	4.3%	4.3%	4.3%
n=5	0.8%	-0.3%	-0.6%	0.0%	0.1%	0.2%	0.2%	0.3%	0.3%	0.3%	0.3%	0.3%	0.3%	0.3%
n=6	1.6%	3.3%	4.2%	4.6%	4.8%	4.9%	5.0%	5.0%	5.0%	5.0%	5.0%	5.0%	5.0%	5.0%
n=7	0.8%	5.2%	5.1%	5.3%	5.4%	5.4%	5.4%	5.5%	5.5%	5.5%	5.5%	5.5%	5.5%	5.5%
n=8	0.1%	0.3%	0.6%	0.9%	1.2%	1.3%	1.3%	1.3%	1.3%	1.3%	1.3%	1.3%	1.3%	1.3%
n=9	0.0%	0.2%	0.8%	1.1%	1.3%	1.3%	1.3%	1.3%	1.3%	1.3%	1.3%	1.3%	1.3%	1.3%
n=10	-0.1%	1.2%	1.3%	1.4%	1.4%	1.4%	1.4%	1.4%	1.4%	1.4%	1.4%	1.4%	1.4%	1.4%

0.5	t=1	t=2	t=3	t=4	t=5	t=6	t=7	t=8	t=9	t=10	t=11	t=12	t=13	t=14
n=1	6.4%	9.1%	10.8%	11.9%	12.9%	13.7%	14.1%	14.2%	14.4%	14.4%	14.4%	14.4%	14.4%	14.4%
n=2	2.5%	4.9%	6.6%	8.3%	9.0%	9.4%	9.7%	9.8%	9.8%	9.8%	9.8%	9.8%	9.8%	9.8%
n=3	2.0%	2.1%	3.0%	4.8%	5.5%	6.0%	6.1%	6.2%	6.2%	6.2%	6.2%	6.2%	6.2%	6.2%
n=4	1.8%	4.1%	5.7%	6.8%	7.3%	7.6%	7.7%	7.7%	7.7%	7.7%	7.7%	7.7%	7.7%	7.7%
n=5	1.6%	3.5%	3.7%	4.4%	4.6%	4.7%	4.8%	4.9%	4.9%	4.9%	4.9%	4.9%	4.9%	4.9%
n=6	1.0%	3.3%	4.4%	4.7%	4.8%	4.9%	4.9%	4.9%	4.9%	4.9%	4.9%	4.9%	4.9%	4.9%
n=7	1.8%	3.0%	4.4%	4.8%	5.1%	5.2%	5.3%	5.3%	5.3%	5.3%	5.3%	5.3%	5.3%	5.3%
n=8	1.0%	1.2%	2.6%	3.3%	3.5%	3.6%	3.7%	3.7%	3.7%	3.7%	3.7%	3.7%	3.7%	3.7%
n=9	3.1%	4.3%	5.2%	6.2%	6.4%	6.3%	6.3%	6.3%	6.3%	6.3%	6.3%	6.3%	6.3%	6.3%
n=10	2.3%	3.5%	4.8%	5.3%	5.4%	5.4%	5.4%	5.4%	5.4%	5.4%	5.4%	5.4%	5.4%	5.4%

0.75	t=1	t=2	t=3	t=4	t=5	t=6	t=7	t=8	t=9	t=10	t=11	t=12	t=13	t=14
n=1	4.5%	7.7%	10.7%	12.9%	14.8%	15.5%	16.0%	16.2%	16.3%	16.3%	16.3%	16.3%	16.3%	16.3%
n=2	3.3%	6.2%	8.0%	8.5%	9.3%	9.9%	10.1%	10.2%	10.3%	10.3%	10.3%	10.3%	10.3%	10.3%
n=3	3.0%	5.8%	8.7%	10.6%	11.4%	11.8%	12.0%	12.1%	12.2%	12.2%	12.2%	12.2%	12.2%	12.2%
n=4	2.4%	3.8%	5.1%	6.7%	7.0%	7.4%	7.5%	7.5%	7.5%	7.6%	7.6%	7.6%	7.6%	7.6%
n=5	4.0%	8.3%	9.8%	11.0%	11.5%	11.7%	11.9%	11.9%	11.9%	11.9%	11.9%	11.9%	11.9%	11.9%
n=6	2.9%	4.1%	6.3%	7.4%	8.0%	8.1%	8.2%	8.3%	8.3%	8.3%	8.3%	8.3%	8.3%	8.3%
n=7	3.6%	5.9%	7.5%	8.0%	8.3%	8.4%	8.5%	8.5%	8.5%	8.5%	8.5%	8.5%	8.5%	8.5%
n=8	1.6%	3.7%	5.1%	5.4%	5.5%	5.7%	5.8%	5.8%	5.8%	5.8%	5.8%	5.8%	5.8%	5.8%
n=9	1.7%	2.4%	3.5%	3.9%	4.0%	4.1%	4.1%	4.1%	4.1%	4.1%	4.1%	4.1%	4.1%	4.1%
n=10	2.2%	3.7%	4.7%	5.0%	5.6%	5.6%	5.6%	5.6%	5.6%	5.6%	5.6%	5.6%	5.6%	5.6%

1.00	t=1	t=2	t=3	t=4	t=5	t=6	t=7	t=8	t=9	t=10	t=11	t=12	t=13	t=14
n=1	6.9%	11.1%	14.6%	17.6%	19.3%	20.0%	20.5%	20.7%	20.9%	20.9%	20.9%	20.9%	20.9%	20.9%
n=2	5.0%	9.3%	11.3%	13.8%	15.4%	16.0%	16.2%	16.3%	16.4%	16.4%	16.4%	16.4%	16.4%	16.4%
n=3	3.9%	6.8%	9.7%	11.5%	12.5%	13.0%	13.2%	13.4%	13.4%	13.4%	13.4%	13.4%	13.4%	13.4%
n=4	4.1%	6.5%	9.3%	11.0%	11.7%	12.1%	12.3%	12.4%	12.5%	12.5%	12.5%	12.5%	12.5%	12.5%
n=5	2.7%	7.2%	9.9%	10.8%	11.4%	11.8%	11.9%	12.0%	12.1%	12.1%	12.1%	12.1%	12.1%	12.1%
n=6	2.0%	4.7%	6.7%	7.6%	8.3%	8.5%	8.6%	8.7%	8.7%	8.7%	8.7%	8.7%	8.7%	8.7%
n=7	2.2%	5.5%	7.5%	8.3%	8.9%	9.2%	9.3%	9.3%	9.3%	9.3%	9.3%	9.3%	9.3%	9.3%
n=8	4.8%	8.1%	10.2%	11.2%	11.6%	11.9%	12.1%	12.1%	12.1%	12.1%	12.1%	12.1%	12.1%	12.1%
n=9	7.0%	9.6%	11.5%	11.6%	12.0%	12.2%	12.2%	12.2%	12.2%	12.2%	12.2%	12.2%	12.2%	12.2%
n=10	2.2%	4.0%	4.9%	5.7%	6.0%	6.0%	6.0%	6.0%	6.0%	6.0%	6.0%	6.0%	6.0%	6.0%

1.25	t=1	t=2	t=3	t=4	t=5	t=6	t=7	t=8	t=9	t=10	t=11	t=12	t=13	t=14
n=1	7.6%	14.0%	19.2%	21.5%	23.5%	24.8%	25.2%	25.4%	25.6%	25.6%	25.6%	25.6%	25.6%	25.6%
n=2	7.0%	12.5%	15.9%	19.6%	21.3%	22.2%	22.6%	22.8%	22.9%	22.9%	22.9%	22.9%	22.9%	22.9%
n=3	4.6%	10.1%	13.3%	15.4%	16.7%	17.3%	17.6%	17.7%	17.8%	17.8%	17.8%	17.8%	17.8%	17.8%
n=4	5.5%	12.2%	16.0%	17.8%	18.8%	19.4%	19.6%	19.7%	19.7%	19.7%	19.7%	19.7%	19.7%	19.7%
n=5	4.9%	7.8%	11.3%	13.2%	13.9%	14.4%	14.7%	14.7%	14.8%	14.8%	14.8%	14.8%	14.8%	14.8%
n=6	3.4%	6.2%	8.1%	9.1%	9.8%	10.2%	10.3%	10.4%	10.4%	10.4%	10.4%	10.4%	10.4%	10.4%
n=7	5.2%	10.7%	13.0%	13.9%	14.6%	14.9%	15.0%	15.1%	15.1%	15.1%	15.1%	15.1%	15.1%	15.1%
n=8	7.0%	9.7%	10.6%	11.6%	12.1%	12.3%	12.4%	12.4%	12.4%	12.4%	12.4%	12.4%	12.4%	12.4%
n=9	4.8%	7.4%	9.0%	10.3%	10.7%	10.8%	10.8%	10.8%	10.8%	10.8%	10.8%	10.8%	10.8%	10.8%
n=10	3.3%	8.3%	9.7%	10.9%	11.3%	11.3%	11.3%	11.3%	11.3%	11.3%	11.3%	11.3%	11.3%	11.3%

areas, the type of curve which were fitted to a parameter is the same. For example, the intercept follows a linear equation.

Table 5: Parameters of Fitted Lines for Poisson Arrivals with Prior Knowledge of Node on Hand with Different Rate

Square Shape Area

Lambda	**Intercept**	n^2	t^2	n	t	R^2
0.25	0.01432	0.00002	0.00011	-0.00057	-0.00238	0.34
0.50	0.03293	0.00017	0.00025	-0.00241	-0.00529	0.70
0.75	0.03936	0.00004	0.00033	-0.00114	-0.00693	0.82
1.00	0.05500	0.00012	0.00047	-0.00212	-0.00977	0.82
1.25	0.07297	0.00014	0.00064	-0.00271	-0.01310	0.86

Rectangle Shape Area

Lambda	**Intercept**	n^2	t^2	n	t	R^2
0.25	0.02776	0.00010	0.00030	-0.00170	-0.00526	0.57
0.50	0.04661	0.00010	0.00060	-0.00224	-0.00963	0.78
0.75	0.07142	0.00003	0.00096	-0.00201	-0.01561	0.81
1.00	0.09491	0.00009	0.00119	-0.00360	-0.01994	0.86
1.25	0.11444	0.00007	0.00179	-0.00313	-0.02709	0.85

Table 6: Fitted Lines of Multi-regression for the Parameters of Poisson Arrivals with Prior Knowledge of Node on Hand (where x is Rate of Arrival λ)

Square Shape Area

Variable	**Equation**	R^2
Intercept	$y = 0.056x + 0.001$	0.98
n^2	$y = 0.00003e^{1.31x}$	0.34
t^2	$y = 0.0005x - 0.00002$	0.99
n	$y = -0.001\ln(x) - 0.002$	0.52
t	$y = -0.006\ln(x) - 0.01$	0.92

Rectangle Shape Area

Variable	**Equation**	R^2
Intercept	$y = 0.84x + 0.01$	0.99
n^2	$y = 0.00005x^2 - 0.0002x + 0.0002$	0.81
t^2	$y = 0.002x - 0.0004$	0.99
n	$y = -0.0004x^2 + 0.0008x - 0.003$	0.93
t	$y = -0.03x + 0.0003$	0.99

4 Higher Poisson Arrival Rate

So far in our experiments our λ's were chosen around 1 so we can compare the results of Section 3.2 and 3.3 with the single arrival in Section 3.1. However, we also would like to be able to use this method for higher arrival rates. For this purpose, we ran all the previous experiments for rectangle shape area for both Euclidean and Manhattan distances where λ is equal to 3.5 and 5.5 to demonstrate the performance of this approach.

4.1 Poisson Arrival

The goodness of fit for both arrival rates shows desirable results. The fitted lines that can be utilized to obtain the expected change in TSP tour length for Manhattan and Euclidean distances when λ is 3.5 are $y = -0.08\ln(x) + 0.29$ and $y = -0.08\ln(x) + 0.30$, respectively, with the $R^2 = 0.93$ for both cases. Here x signifies period and y represents the percentage of change in tour length. Similarly for $\lambda = 5.5$, the fitted lines for Manhattan and Euclidean distances are $y = -0.15\ln(x) + 0.45$ and $y = -0.17\ln(x) + 0.48$, respectively, with the $R^2 = 0.98$ for both distances.

4.2 Poisson Arrivals with Prior Knowledge of Node on Hand

Table 7 represents the parameters of fitted lines for each arrival rate by type of distance. These fitted lines are the regression models that we can utilize to get the expected change in TSP tour length when the rate of arrival is either 3.5 or 5.5. The R^2 values for all the lines indicate a well fitted curve. This shows that this approach is capable of handling higher values of λ as well.

Table 7: Parameters of Fitted Lines for Poisson Arrivals with Prior Knowledge of Node on Hand with Higher Arrival Rate λ

Distance Type	Lambda	Intercept	n^2	t^2	n	t	R^2
Manhattan Distance	3.5	0.3105	-0.0001	0.0066	-0.0039	-0.0878	0.90
	5.5	0.4663	0.0003	0.0101	-0.0110	-0.1329	0.90
Euclidean Distance	3.5	0.2970	0.0001	0.0053	-0.0053	-0.0782	0.90
	5.5	0.4419	0.0001	0.0087	-0.0073	-0.1216	0.86

5 Warehouse Picking Case - Triangular Segmented Area

One of the assumptions in previous sections was that the shape of the area where nodes were generated was rectangular/square. However, in many real world cases, the area is not a rectangle shape. For example, in many warehouses, the area has been divided into different segments based on their demand. This means within an area of interest, arrival rate of nodes is not uniformly distributed and differs segment by segment. Figure 5 shows one of many typical warehouse layouts. In this layout, area I contains only 10% of the items but supports 70% of the demand, area II has 20% of demand and contains 20% of items, and area III has 70% of the items and supports the remaining demand which is 10%. The reason behind the segmentation is to make the retrieval and return of orders easier. Popular items are being stored in area I which is the closest to the Input/Output point and easier to access. Even though these popular items are only 10% of the total items, they are the most popular ones. This means the rate of arrival for items (future nodes) in area I is greater than area II, and area II greater than III. In all the experiments performed in Section 2 the probability of a node appearing at a given space for the entire area follows a uniform distribution. Whereas, with the segmented layout that is not the case with the segmented layout. The motivation behind designing a segmented experiment is to study the implications of our work in a warehousing order picking situation.

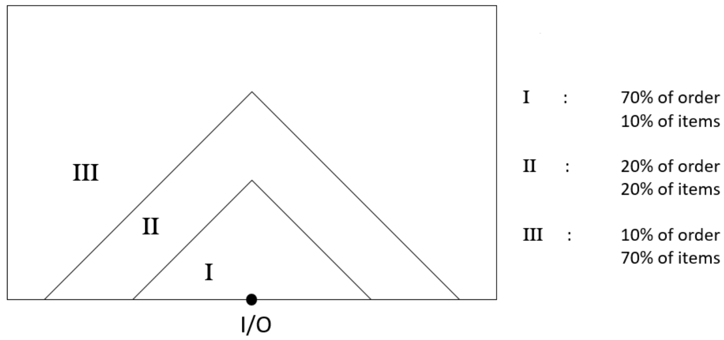

Figure 5: Orders Distribution in a Common Warehouse Layout

5.1 Single Arrival

This section is very similar to Section 3.1 with two differences - one being the shape of the area and the other being that the probability of node appearance is not uniformly distributed. The layout of the area is a triangular shape with different Poisson arrival rates; see Figure 5. As a part of experimental setting, the same rectangle from Section 3 is assumed for the total area and shape. The objective here is to find the expected change in the tour length by adding one node to the triangular shape area. The remaining parameters are the same as Section 3.1 since they are controlled variables.

From the result the fitted line of Manhattan distance for the net percentage change in tour length follows a power line with a coefficient of 0.475 and power of -0.81. The R^2 value of 0.88 shows that this is a well fitted line. The results for Euclidean distance shows similar outcomes with a power line with a coefficient of 0.48, power of -0.82 and R^2 value of 0.86. The fitted lines for "Manhattan and Euclidean distance standard deviations" of change in tour length follow $y = 0.46x - 0.78$ with $R^2 = 0.80$, where x signifies the number of nodes added.

5.2 Poisson Arrivals

In this section, similar to Section 3.2, the number of nodes per period is established from a Poisson process. The area however is a rectangle with triangular segmentation, divided into three sub-areas with different arrival rates, see Figure 5. The rate of arrivals for items in area I is 70% , in area II is 20%, and in III is 10% of the total rate of arrivals in the area. To ensure statistical accuracy, the first $t-1$ periods are simulated $2^{S(t-1)} * 30$ times in which $S = 3$ is the number of segments in the area, and for each of these simulations the t^{th} period is replicated 100 times. This means, for example, that the total number of simulations for period 15 is $30*2^{42}*100$. Since the arrival rate in area I is the highest and most of the ordered items are located there, we chose λ_I to be $\in \{0.75, 1, 1.25, 1.5, 1.75, 2, 2.25\}$ and adjusted λ_{II} and λ_{III} accordingly to become 0.2 and 0.1 times of total $\lambda = \lambda_T = (\lambda_I + \lambda_{II} + \lambda_{III}) = \dfrac{\lambda_I}{0.7} = \dfrac{\lambda_{II}}{0.2} = \dfrac{\lambda_{III}}{0.1}$.

Figure 6 shows the net percentage change in total length of TSP tour when the distance is Manhattan and $\lambda_I = 1$. Corresponding to $\lambda_I = 1$, the other two arrival rates would be $\lambda_{II} = \dfrac{2}{7}\lambda_I = \dfrac{2}{7}$ and $\lambda_{III} = \dfrac{1}{7}\lambda_I = \dfrac{1}{7}$. The fitted line for net percentage of change in TSP tour length in Figure 6 follows $y = 0.51x - 0.73$ with $R^2 = 0.96$, where x as an input is the period number.

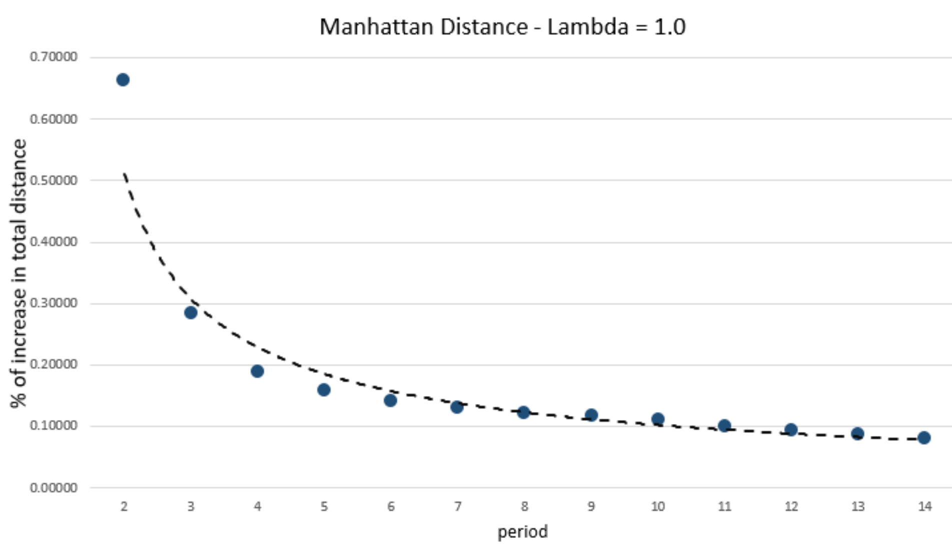

Figure 6: Results of Poisson Arrivals' Tour Length Increase for a Triangular Segmented Area with Manhattan Distance

Table 8 shows the fitted line parameters for different values of λ_I, λ_{II}, and λ_{III}. Using these results, we developed statistical equations to estimate the parameters of the model for tour length change with rate of arrival λ as an independent variable. The coefficient as one of the parameters follows $y_c = 0.01\lambda + 0.51$ with $R^2 = 0.61$ and similarly for the power parameter, the fitted line is $y_p = 0.05\ln(\lambda) - 0.78$ with $R^2 = 0.73$ where λ is the rate of node arrival. Comparing these results to Section 3.2, we observe that the correlations for power and coefficient are weaker. In Section 3.2 the values of R^2 are around 0.90, whereas R^2 in this case are between 0.61 and 0.73. The reason is that there is now uncertainty due to the placement of the order as well as uncertainty in the number of orders.

Table 8: Parameters of Fitted Lines for Poisson Arrivals with Different Rate where the Area is Triangular Segmented

Lambda	Coefficient	Power	R^2
0.75	0.53	-0.81	0.96
1.0	0.51	-0.73	0.96
1.25	0.53	-0.71	0.97
1.5	0.53	-0.70	0.99
1.75	0.56	-0.71	0.98
2.0	0.54	-0.68	0.98
2.25	0.59	-0.72	0.97

The standard deviations for the results of change in tour length is calculated. This enables us to estimate the risk or the variance of the expected change in TSP length. For example the standard deviations for Manhattan distance in Section 3.2 when the arrival rate is 0.75, follows $y = 0.53x - 0.81$ with $R^2 = 0.96$ where x signifies period $x \in 1, 2, 3, ..., 14$ and y represents the standard deviation of change in tour length. Due to more uncertainty in the segmented layout, the standard deviations go down by period at a lower rate than that in Section 3.2, see Figure 7.

Table 9 presents coefficients and powers for fitted lines of the standard deviations, for different λ values. We fitted two lines to each of these parameters so we could estimate

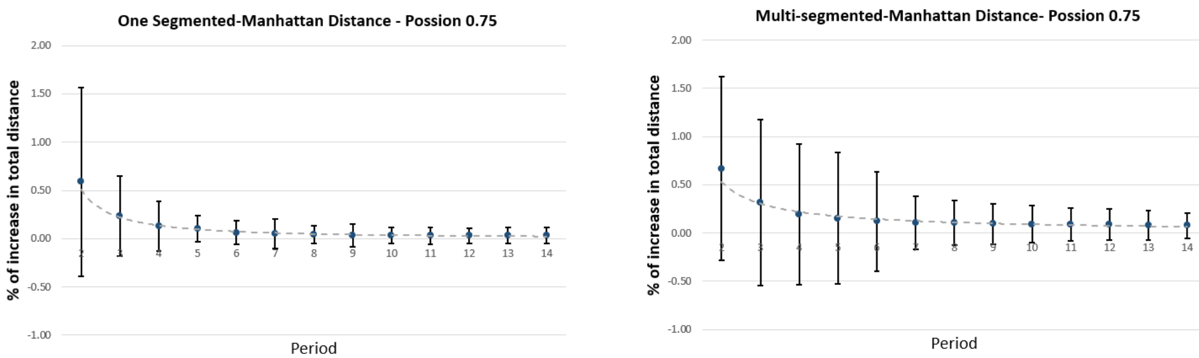

Figure 7: Results of Poisson Arrivals' Tour Length Increase Standard Deviation for a Square Area (One Segmented) and a Triangular Segmented area (Multi-Segmented) with Manhattan Distance

the standard deviation for any λ values by period. The coefficient line has R^2 of 0.59 and follows $y_c = 1.95e^{-0.31\lambda}$ where λ is the rate of node arrival. The line for the power equation is $y_p = 0.19\lambda^2 - 0.55\lambda - 0.58$ with R^2 of 0.36. The goodness of fit for the segmented layout standard deviation is weaker than for the simple layout case. The reason behind the change in goodness of fit is that the segmented system is more volatile and has more uncertainty. Even though, these regression lines do not have a very good R^2, if the decision makers want to involve risks to their analysis, they can utilize these models to their benefit.

Table 9: Parameters of Standard Deviation Trend Lines of Poisson Arrival with Different Rate where the Area is Triangular Segmented

Lambda	**Coefficient**	**Power**	R^2
0.75	1.44	-0.88	0.86
1.0	1.45	-0.95	0.99
1.25	1.44	-0.98	0.98
1.5	1.05	-0.86	0.99
1.75	1.45	-1.05	0.95
2.0	1.08	-0.92	0.98
2.25	0.86	-0.83	0.99

5.3 Poisson Arrivals with Prior Knowledge of Nodes on Hand

In this section we experiment with the case in Section 3.3 but with the triangular shape layout and multiple Poisson processes for each segment. Table 10 shows the results, with the x axis representing the period and the y axis representing the number of orders on hand, for λ from 0.75 to 2.25 with intervals of 0.25. When we have more nodes on hand, the expected increase in the total length of the tour decreases. This behavior can again be explained by the saturation of the area with nodes. A statistical model is developed to estimate the expected percentage change in tour length by different arrival rates, number of nodes on hand, and number of periods.

Table 10: Results of Poisson Arrivals with Prior Knowledge of Node on Hand' Tour length Increase for a Triangular Segmented Area

0.75	t=2	t=3	t=4	t=5	t=6	t=7	t=8	t=9	t=10	t=11	t=12	t=13
n=1	29.2%	46.5%	59.0%	71.3%	82.1%	92.4%	102.3%	112.0%	121.2%	129.9%	138.1%	145.7%
n=2	19.4%	33.1%	44.0%	54.6%	65.1%	75.1%	84.7%	93.9%	102.8%	111.3%	119.1%	125.9%
n=3	12.0%	25.7%	36.6%	47.6%	57.8%	67.7%	77.1%	86.0%	94.5%	102.5%	109.8%	116.1%
n=4	12.0%	21.8%	32.4%	43.0%	53.1%	62.8%	71.9%	80.4%	88.6%	96.2%	102.8%	108.3%
n=5	10.9%	21.6%	32.4%	42.6%	52.4%	61.7%	70.5%	78.7%	85.8%	91.5%	95.3%	97.1%
n=6	10.5%	21.5%	32.6%	43.0%	52.6%	61.1%	69.2%	76.4%	82.1%	86.1%	87.9%	87.9%
n=7	12.3%	22.7%	33.1%	42.5%	51.7%	59.9%	66.3%	70.0%	70.6%	70.6%	70.6%	70.6%
n=8	11.0%	20.8%	30.7%	39.8%	48.2%	54.6%	57.8%	57.8%	57.8%	57.8%	57.8%	57.8%
n=9	11.2%	20.2%	29.2%	37.6%	43.0%	43.7%	43.7%	43.7%	43.7%	43.7%	43.7%	43.7%
n=10	10.0%	20.0%	28.8%	30.1%	30.1%	30.1%	30.1%	30.1%	30.1%	30.1%	30.1%	30.1%

1	t=2	t=3	t=4	t=5	t=6	t=7	t=8	t=9	t=10	t=11	t=12	t=13
n=1	33.2%	51.7%	66.7%	81.2%	94.9%	107.8%	120.0%	131.2%	141.6%	151.3%	159.7%	166.8%
n=2	25.4%	42.4%	57.5%	72.1%	85.7%	98.2%	109.8%	120.7%	130.8%	140.0%	148.5%	156.3%
n=3	16.9%	33.4%	47.8%	61.8%	74.3%	86.4%	97.7%	108.3%	117.9%	126.6%	134.1%	140.2%
n=4	14.3%	29.0%	42.6%	55.9%	68.0%	79.5%	90.1%	99.9%	109.1%	117.3%	124.3%	129.7%
n=5	14.9%	29.1%	42.3%	55.4%	67.5%	78.4%	88.6%	98.0%	106.2%	112.6%	117.0%	119.1%
n=6	15.0%	29.0%	41.6%	54.3%	66.0%	76.6%	86.4%	94.5%	100.3%	103.3%	103.3%	103.3%
n=7	13.3%	27.0%	39.9%	51.7%	62.7%	72.5%	80.0%	84.3%	84.6%	84.6%	84.6%	84.6%
n=8	14.9%	27.7%	40.1%	51.4%	61.9%	70.2%	74.9%	75.3%	75.3%	75.3%	75.3%	75.3%
n=9	12.6%	24.4%	35.6%	46.1%	52.9%	53.8%	53.8%	53.8%	53.8%	53.8%	53.8%	53.8%
n=10	12.2%	23.0%	34.3%	40.4%	40.4%	40.4%	40.4%	40.4%	40.4%	40.4%	40.4%	40.4%

1.25	t=2	t=3	t=4	t=5	t=6	t=7	t=8	t=9	t=10	t=11	t=12	t=13
n=1	30.9%	50.6%	69.0%	85.8%	102.2%	116.7%	130.3%	142.5%	153.6%	163.8%	172.9%	180.6%
n=2	24.9%	45.7%	63.2%	79.9%	95.4%	109.6%	122.2%	133.8%	144.5%	154.3%	162.8%	170.0%
n=3	20.3%	39.2%	56.2%	72.3%	87.5%	101.0%	113.4%	124.4%	134.6%	144.0%	152.6%	159.9%
n=4	20.5%	38.2%	55.4%	71.3%	85.4%	98.1%	110.0%	120.7%	130.5%	139.0%	146.0%	151.1%
n=5	19.0%	36.7%	54.1%	68.6%	81.8%	94.3%	105.4%	115.5%	124.3%	131.2%	136.0%	138.3%
n=6	16.8%	34.0%	50.2%	64.5%	77.2%	88.7%	99.1%	107.9%	114.3%	117.8%	118.3%	118.3%
n=7	17.3%	33.4%	49.1%	62.4%	74.6%	85.7%	94.5%	100.1%	101.7%	101.7%	101.7%	101.7%
n=8	20.2%	36.1%	50.1%	62.3%	73.7%	82.7%	88.2%	89.3%	89.3%	89.3%	89.3%	89.3%
n=9	16.0%	32.9%	45.7%	58.2%	68.1%	73.0%	73.0%	73.0%	73.0%	73.0%	73.0%	73.0%
n=10	14.7%	28.6%	41.9%	46.1%	46.1%	46.1%	46.1%	46.1%	46.1%	46.1%	46.1%	46.1%

1.50	t=2	t=3	t=4	t=5	t=6	t=7	t=8	t=9	t=10	t=11	t=12	t=13
n=1	33.1%	55.5%	77.0%	96.0%	114.3%	129.9%	144.0%	156.6%	168.1%	178.6%	188.0%	196.1%
n=2	23.4%	46.1%	66.5%	85.2%	102.2%	117.2%	130.7%	142.8%	153.8%	163.7%	172.8%	181.1%
n=3	19.6%	40.7%	61.3%	79.6%	96.0%	110.4%	123.4%	135.1%	145.6%	155.2%	163.5%	170.4%
n=4	22.0%	42.9%	62.5%	80.4%	95.8%	109.7%	122.0%	133.1%	143.5%	152.6%	160.4%	166.3%
n=5	24.5%	45.6%	65.3%	82.2%	97.1%	109.8%	121.5%	132.2%	141.5%	149.0%	154.2%	157.0%
n=6	21.4%	40.5%	58.5%	74.1%	88.0%	100.7%	112.1%	121.6%	128.5%	132.2%	132.4%	132.4%
n=7	23.7%	42.7%	59.4%	74.5%	87.5%	99.4%	109.1%	115.7%	118.7%	118.7%	118.7%	118.7%
n=8	20.8%	38.7%	55.7%	70.1%	83.0%	92.5%	97.1%	97.1%	97.1%	97.1%	97.1%	97.1%
n=9	20.3%	37.2%	52.2%	65.3%	73.0%	73.0%	73.0%	73.0%	73.0%	73.0%	73.0%	73.0%
n=10	19.3%	36.2%	51.5%	55.8%	55.8%	55.8%	55.8%	55.8%	55.8%	55.8%	55.8%	55.8%

1.75	t=2	t=3	t=4	t=5	t=6	t=7	t=8	t=9	t=10	t=11	t=12	t=13
n=1	40.0%	69.3%	94.5%	116.1%	135.6%	152.2%	166.8%	179.8%	191.5%	202.1%	211.8%	220.5%
n=2	26.6%	54.4%	77.3%	98.0%	115.8%	131.6%	145.5%	158.0%	169.2%	179.4%	188.8%	197.2%
n=3	30.6%	54.5%	78.1%	98.0%	114.8%	129.7%	142.8%	154.7%	165.3%	174.5%	181.8%	187.0%
n=4	28.4%	52.3%	73.0%	91.1%	107.7%	122.1%	135.0%	146.5%	157.0%	166.2%	173.7%	179.3%
n=5	24.7%	48.8%	69.9%	87.5%	102.6%	116.1%	128.5%	139.7%	149.2%	156.5%	161.2%	163.1%
n=6	27.6%	50.6%	69.8%	86.5%	101.5%	114.6%	126.5%	136.5%	144.1%	148.7%	150.1%	150.1%
n=7	24.3%	44.9%	62.8%	79.9%	94.1%	106.4%	115.5%	120.4%	120.4%	120.4%	120.4%	120.4%
n=8	25.4%	46.9%	64.5%	79.8%	93.5%	103.9%	109.3%	109.3%	109.3%	109.3%	109.3%	109.3%
n=9	21.1%	41.3%	58.5%	73.0%	80.1%	80.1%	80.1%	80.1%	80.1%	80.1%	80.1%	80.1%
n=10	19.9%	38.6%	53.9%	54.1%	54.1%	54.1%	54.1%	54.1%	54.1%	54.1%	54.1%	54.1%

2	t=2	t=3	t=4	t=5	t=6	t=7	t=8	t=9	t=10	t=11	t=12	t=13
n=1	37.7%	68.7%	94.2%	117.1%	137.2%	154.3%	169.0%	182.3%	194.1%	204.7%	214.4%	223.3%
n=2	36.8%	66.2%	92.1%	115.2%	134.0%	150.3%	164.5%	177.1%	188.4%	198.7%	208.2%	216.7%
n=3	29.3%	59.3%	83.6%	104.4%	122.2%	137.8%	151.4%	163.7%	174.7%	184.6%	193.2%	200.3%
n=4	30.2%	57.0%	80.6%	100.0%	116.9%	131.9%	144.9%	156.7%	167.3%	176.4%	183.8%	189.2%
n=5	28.0%	53.6%	76.3%	95.4%	112.3%	126.4%	139.3%	150.6%	159.9%	166.5%	170.1%	170.6%
n=6	29.0%	55.9%	76.3%	93.9%	109.7%	123.7%	136.3%	146.5%	153.9%	157.7%	157.8%	157.8%
n=7	27.3%	51.5%	71.2%	87.8%	102.3%	115.3%	125.5%	131.9%	134.1%	134.1%	134.1%	134.1%
n=8	25.6%	47.4%	65.7%	81.9%	96.4%	107.1%	112.5%	112.5%	112.5%	112.5%	112.5%	112.5%
n=9	27.4%	48.4%	66.7%	81.9%	89.4%	89.4%	89.4%	89.4%	89.4%	89.4%	89.4%	89.4%
n=10	25.7%	44.6%	62.5%	70.6%	70.6%	70.6%	70.6%	70.6%	70.6%	70.6%	70.6%	70.6%

2.25	t=2	t=3	t=4	t=5	t=6	t=7	t=8	t=9	t=10	t=11	t=12	t=13
n=1	38.2%	70.3%	99.2%	122.8%	143.0%	160.2%	175.3%	188.6%	200.5%	211.2%	220.9%	229.4%
n=2	41.9%	73.3%	100.6%	124.0%	143.6%	160.2%	174.8%	187.5%	199.0%	209.4%	218.9%	227.3%
n=3	36.5%	69.0%	93.6%	115.1%	133.6%	149.5%	163.4%	175.7%	186.8%	196.9%	205.9%	213.5%
n=4	30.9%	61.6%	85.9%	106.2%	124.0%	139.0%	152.6%	164.4%	175.3%	184.9%	193.0%	199.4%
n=5	33.4%	64.1%	86.9%	106.6%	123.9%	138.5%	151.4%	162.9%	172.6%	180.0%	184.8%	186.7%
n=6	32.6%	61.2%	84.3%	102.6%	118.8%	133.0%	145.6%	155.8%	163.0%	166.6%	166.6%	166.6%
n=7	31.8%	57.7%	78.2%	97.0%	112.1%	125.8%	137.0%	144.9%	148.9%	148.9%	148.9%	148.9%
n=8	29.7%	53.1%	74.1%	91.2%	106.0%	116.5%	120.9%	120.9%	120.9%	120.9%	120.9%	120.9%
n=9	28.0%	51.5%	71.2%	87.7%	96.3%	96.3%	96.3%	96.3%	96.3%	96.3%	96.3%	96.3%
n=10	26.6%	47.4%	65.7%	70.4%	70.4%	70.4%	70.4%	70.4%	70.4%	70.4%	70.4%	70.4%

Table 11a shows the results for quadratic polynomial regressions for different arrival rates. These results show that higher λ values provide a better fit for the regression line. An equation to estimate each parameter of the quadratic polynomial model is presented in Table 11b.

Table 11: Regression Parameters for Poisson Arrivals with Prior Knowledge of Node on Hand

(a) Parameters of Fitted Lines for Poisson Arrivals with Prior Knowledge of Node on Hand with Different Rate where the Area is Triangular Segmented

Lambda	Intercept	n^2	t^2	n	t	R^2
0.75	0.187	0.00047	0.000	-0.011	-0.012	0.583
1.0	0.228	0.00025	0.000	-0.011	-0.015	0.711
1.25	0.259	0.00003	0.001	-0.008	-0.020	0.901
1.5	0.289	0.00002	0.001	-0.008	-0.024	0.934
1.75	0.359	0.00007	0.001	-0.011	-0.035	0.925
2.0	0.390	-0.00019	0.002	-0.008	-0.044	0.975
2.25	0.427	-0.00065	0.002	-0.004	-0.055	0.979

(b) Fitted Lines of Multi-regression for the Parameters of Poisson Arrivals with Prior Knowledge of Node on Hand (where x is Rate of Arrival λ) for the Triangular Segmented Area

Variable	Equation	R^2
Intercept	$y = 0.1635x + 0.0602$	0.9896
n^2	$y = -0.0006x + 0.0009$	0.8390
t^2	$y = 0.0001e1.2897x$	0.9479
n	$y = 0.0021x^2 - 0.0032x - 0.0092$	0.4914
t	$y = -0.014x^2 + 0.0133x - 0.0144$	0.9969

5.4 Triangular Segmented Shape vs Rectangle Shape

In this part we want to provide a brief comparison between Section 5 with rectangle shape form and Section 3 with triangular shape form. In both cases the area is the same, however the node dispersion is different. In the rectangular shape case, Section 3, nodes are independently and uniformly distributed whereas in the triangular segmented case, Section 5, the occurrence of node appearance in the smaller triangular area is higher compared to the rest of the area. The regression lines to predict the change in TSP tour length for Single and Poisson arrivals of both Sections have the same type of functions with different values for parameters(coefficient and intercept). The parameters(coefficient and intercept) for the change of TSP tour length regression lines of Section 5 are higher than Section 3. For the case of "Poisson Arrivals with Prior Knowledge of Node on Hand", both Sections 5 and 3 show different functions for their regression lines to predict the change in TSP tour length.

6 Testing the Predictive Models

In this section we have created testing sets to evaluate the regression models that we have developed. To determine an appropriate number of incidents in each testing set, we use the criteria that the relative error reaches 0.1 with confidence level of 95 percent, Law et al. (2000). We have developed two separate testing sets, one for the Rectangle shape area with uniformly distributed node dispersion, and the other for the Triangular Segmented area with uneven node dispersion (e.g., higher probability of node arrival in inner triangle). For each of the incidents in these sets, rate of arrivals $\lambda \in [0, 3.5]$, number of node on hand $n \in \{0, 1, 2, ..., 15\}$, and the targeted period for which we need the prediction $p \in \{1, 2, ..., 15\}$, are all randomly chosen within their range. The results from the Rectangle shape area yield a mean absolute percent error(MAPE) of 10 percent and the results for the Triangular Segmented area give us a MAPE value of 15 percent. Even though the MAPE values are not as low as we desired them to be, they can still be utilized when one needs to predict the TSP tour length with unknown number of nodes. For example, one way to adjust these can be using the standard deviations to tailor the models to the appropriate application.

7 Conclusions and Summary

In this study we researched the possibility of finding patterns in the change of the length of a TSP tour in order to estimate the length of tour when nodes are added in both deterministic and probabilistic settings. These patterns are useful when one needs to predict a change in the tour length. There are many applications that can use these results to their benefit, such as warehouse order-picking, public transportation allocation, deliveries in e-commerce, and mobile service industries. We have shown results for different types of area (layout), distance measurement, arrival of nodes, and probability distribution of node appearance. The process in this paper is flexible and can be adjusted to different constraints. In our study we considered three different layouts, a simple rectangular layout, a simple square layout and a rectangular layout with triangle-shaped segmented parts. We then simulated the problem with deterministic and probabilistic arrival of nodes. We also examined a system which had no memory of the number of nodes in hand and also another case which had the memory of the number of nodes on hand. We have used different types of statistical fits to find these patterns, such as multiple, single, power, logistic, and polynomial regressions. Moreover, we showed the goodness of fit for each of these models. Using these results on TSP tour length prediction, one can save time and cost and help smooth the decision processes.

For future research in this area, we suggest the study of different probabilistic processes for the arrival rates. Also, we suggest investigating the impact of different layout shapes, to determine the correlation between shape, size and the outcomes.

Acknowledgment

The authors would like to thank two anonymous referees for their insightful comments on earlier versions of this paper. Addressing these comments has led to a much improved manuscript.

References

Antosiewicz, M., Koloch, G., and Kamiński, B. (2013). Choice of best possible metaheuristic algorithm for the travelling salesman problem with limited computational time: quality, uncertainty and speed. *Journal of Theoretical and Applied Computer Science*, 7(1):46–55.

Beardwood, J., Halton, J. H., and Hammersley, J. M. (1959). The shortest path through many points. In *Mathematical Proceedings of the Cambridge Philosophical Society*, volume 55, pages 299–327. Cambridge Univ Press.

Bertsimas, D. J. (1992). A vehicle routing problem with stochastic demand. *Operations Research*, 40(3):574–585.

Bertsimas, D. J. and Van Ryzin, G. (1991). A stochastic and dynamic vehicle routing problem in the euclidean plane. *Operations Research*, 39(4):601–615.

Brown, J. R. and Guiffrida, A. L. (2017). Stochastic modeling of the last mile problem for delivery fleet planning. In *Journal of the Transportation Research Forum*, volume 56.

Çavdar, B. and Sokol, J. (2015). A distribution-free tsp tour length estimation model for random graphs. *European Journal of Operational Research*, 243(2):588–598.

Cook, W. J. (2011). *In pursuit of the traveling salesman: mathematics at the limits of computation*. Princeton University Press.

Daganzo, C. F. (1984). The distance traveled to visit n points with a maximum of c stops per vehicle: An analytic model and an application. *Transportation Science*, 18(4):331–350.

Franceschetti, A., Jabali, O., and Laporte, G. (2017). Continuous approximation models in freight distribution management. *Top*, 25(3):413–433.

Geem, Z. W., Kim, J. H., and Loganathan, G. (2001). A new heuristic optimization algorithm: harmony search. *Simulation*, 76(2):60–68.

Gendreau, M., Hertz, A., and Laporte, G. (1994). A tabu search heuristic for the vehicle routing problem. *Management science*, 40(10):1276–1290.

Ghiani, G., Guerriero, F., Laporte, G., and Musmanno, R. (2003). Real-time vehicle routing: Solution concepts, algorithms and parallel computing strategies. *European Journal of Operational Research*, 151(1):1–11.

Grefenstette, J., Gopal, R., Rosmaita, B., and Van Gucht, D. (1985). Genetic algorithms for the traveling salesman problem. In *Proceedings of the first International Conference on Genetic Algorithms and their Applications*, pages 160–165.

Karp, R. M. (1972). Reducibility among combinatorial problems. In *Complexity of computer computations*, pages 85–103. Springer.

Kennedy, J. (2011). Particle swarm optimization. In *Encyclopedia of machine learning*, pages 760–766. Springer.

Kirkpatrick, S., Gelatt, C. D., Vecchi, M. P., et al. (1983). Optimization by simulated annealing. *science*, 220(4598):671–680.

Kwon, O., Golden, B., and Wasil, E. (1995). Estimating the length of the optimal tsp tour: An empirical study using regression and neural networks. *Computers & operations research*, 22(10):1039–1046.

Law, A. M., Kelton, W. D., and Kelton, W. D. (2000). *Simulation modeling and analysis*, volume 3. McGraw-Hill New York.

Lawer, E., Lenstra, J. K., Kan, A. R., and Shmoys, D. (1985). The traveling salesman problem: a guided tour of combinatorial optimization.

Miller, C. E., Tucker, A. W., and Zemlin, R. A. (1960). Integer programming formulation of traveling salesman problems. *Journal of the ACM (JACM)*, 7(4):326–329.

Nicola, D., Vetschera, R., and Dragomir, A. (2019). Total distance approximations for routing solutions. *Computers & Operations Research*, 102:67–74.

Ong, H. and Huang, H. (1989). Asymptotic expected performance of some tsp heuristics: an empirical evaluation. *European Journal of Operational Research*, 43(2):231–238.

Papadimitriou, C. H. (1977). The euclidean travelling salesman problem is np-complete. *Theoretical computer science*, 4(3):237–244.

Stein, D. M. (1978a). An asymptotic, probabilistic analysis of a routing problem. *Mathematics of Operations Research*, 3(2):89–101.

Stein, D. M. (1978b). Scheduling dial-a-ride transportation systems. *Transportation Science*, 12(3):232–249.

Vinel, A. and Silva, D. F. (2018). Probability distribution of the length of the shortest tour between a few random points: A simulation study. In *2018 Winter Simulation Conference (WSC)*, pages 3156–3167. IEEE.