UNIVERSITI PUTRA MALAYSIA

MULTI DEPOT DYNAMIC VEHICLE ROUTING PROBLEM WITH STOCHASTIC ROAD CAPACITY FOR EMERGENCY MEDICAL SUPPLY DELIVERY IN HUMANITARIAN LOGISTICS

WADI KHALID BIN ANUAR

IPM 2022 11

MULTI DEPOT DYNAMIC VEHICLE ROUTING PROBLEM WITH STOCHASTIC ROAD CAPACITY FOR EMERGENCY MEDICAL SUPPLY DELIVERY IN HUMANITARIAN LOGISTICS

By

WADI KHALID BIN ANUAR

Thesis Submitted to the School of Graduate Studies, Universiti Putra Malaysia, in Fulfillment of the Requirements for the Degree of Doctor of Philosophy

June 2022

COPYRIGHT

Abstract of thesis presented to the Senate of Universiti Putra Malaysia in fulfillment of the requirement for the degree of Doctor of Philosophy

MULTI DEPOT DYNAMIC VEHICLE ROUTING PROBLEM WITH STOCHASTIC ROAD CAPACITY FOR EMERGENCY MEDICAL SUPPLY DELIVERY IN HUMANITARIAN LOGISTICS

By

WADI KHALID BIN ANUAR

June 2022

Chairman : **Lee Lai Soon, PhD**
Institute : **Mathematical Research**

As part of humanitarian logistics research for emergency medical supply delivery during disaster, the modelling and solution of a Multi Depot Dynamic Vehicle Routing Problem with Stochastic Road Capacity (MDDVRPSRC) is presented. Based on the chaotic setting from a disaster event, the model and solution proposed are validated and analysed through a Decision Support System (DSS), MDDVRPSRC DSS. The model proposed is based on Markov Decision Processes (MDP) modelling framework as part of reinforcement learning (RL) solution approach. Through this model, multi objectives, multi depot, multi trip and split delivery among homogeneous fleet of vehicle are addressed. Moreover a stochastic road capacity distribution where its mean deteriorates over time is also depicted in the problem.

To solve the proposed model, an Approximate Dynamic Programming (ADP) approach is applied focusing on the lookahead approach. Specifically a PDS - Rollout Algorithm (PDS-RA) is adopted. Five variants of constructive base heuristics, Teach Based Insertion Heuristic (TBIH-1 - TBIH-5) are proposed complementing the PDS-RA when dealing with the lookahead rollout involving stochastic road capacity.

The computational results obtained are compared with a matheuristic approach. From this novel method, exact CPLEX computation is executed at every rollout decision epoch, based on two proposed reduced 2-stage stochastic programming ILP models (MDVRPSRC-2S1 and MDVRPSRC-2S2). These two models are derived from the Multi Depot Vehicle Routing Problem with Stochastic Road Capacity (MDVRPSRC-2S) which is proposed next to the deterministic model of Multi Depot

Vehicle Routing Problem with Road Capacity (D-MDVRPRC) stemming from the preliminary research, prior to the development of MDDVRPSRC model. Results indicate comparable quality of solution from the proposed heuristics to that of CPLEX in random setting of the problem instances. In addition, the proposed heuristics are especially superior in computation time.

The contributions of the research are as follows: (1) The MDP model for MDDVRPSRC, deterministic D-MDVRPRC as well as 2 stage stochastic ILP MDVRPSRC-2S models are respectively developed and presented; (2) based on the MDVRPSRC-2S it is shown how 2-stage stochastic programming model can be applied through CPLEX execution during each of Monte Carlo simulated PDS - RA by proposing another two models as two reduced version (MDVRPSRC-2S1 and MDVRPSRC-2S2) of the MDVRPSRC-2S; (3) the radial tremor disaster dispersion from a single or multiple epicentres and the corresponding deterioration of road capacity distribution mean and travel time are proposed; (4) the solution algorithm: TBIH-1, and it's 4 variants (TBIH-2, TBIH-3, TBIH-4 and TBIH-5) are presented; (5) test dataset is developed consists of simulated road networks and the damage unit of each roads due to the earthquake for experimentation and simulation purposes, and finally (6) a decision support system (MDDVRPSRC DSS) for simulating online delivery operation during disaster is designed. All of these could be applied to all types of dynamic vehicle routing problem involving any types of disaster and it's inherent stochastic road capacity as well as increased delayed travel time.

Abstrak tesis yang dikemukakan kepada Senat Universiti Putra Malaysia sebagai memenuhi keperluan untuk ijazah Doktor Falsafah

PERMASALAHAN NAVIGASI KENDERAAN DINAMIK PELBAGAI GUDANG MELIBATKAN KAPASITI JALAN YANG RAWAK UNTUK PENGHANTARAN BEKALAN PERUBATAN KECEMASAN DI DALAM LOGISTIK KEMANUSIAN

Oleh

WADI KHALID BIN ANUAR

Jun 2022

Pengerusi : **Lee Lai Soon, PhD**
Institut : **Penyelidikan Matematik**

Sebagai sebahagian daripada kajian logistik kemanusiaan untuk penghantaran bekalan perubatan kecemasan ketika bencana alam, model and penyelesaian kepada Masalah Navigasi Kenderaan Berbilang-Gudang Dinamik melibatkan Kapasiti Jalan yang Rawak telah dikemukakan. Berdasarkan keadaan gawat ketika musibah alam, model dan penyelesaian yang dikemukakan telah dianalisa dan dibuktikan melalui satu sistem sokongan keputusan, MDDVRPSRC DSS. Model yang dikemukakan adalah berdasarkan teori kerangka model Proses Keputusan Markov (MDP) sebagai sebahagian daripada penyelesaian pembelajaran pengukuhan. Melalui model ini, masalah berbilang objektif, berbilang gudang, berbilang perjalanan dan penghantaran secara berasingan oleh kumpulan kenderaan yang serupa telah diperhalusi. Di samping itu taburan kapasiti jalan di mana purata taburan tersebut berkurangan sepanjang tempoh operasi telah digambarkan di dalam masalah ini.

Untuk menyelesaikan model yang dicadangkan ini, kaedah Anggaran Pengatucaraan Dinamik (ADP) telah digunapakai memberi fokus kepada kaedah memandang ke hadapan. Secara khusus, algorithma Tebaran Keadaan Selepas Keputusan (PDS-Rollout (PDS-RA)) telah diadaptasikan. Lima varian daripada heuristik binaan, Heuristik Binaan Berdasarkan Pengajaran (TBIH-1 - TBIH-5) telah dikemukakan yang sesuai digunapakai bersama kaedah PDS-RA apabila diaplikasikan ketika menjalankan kaedah tebaran memandang ke hadapan melibatkan kapasiti jalan yang rawak.

Keputusan komputasi yang terhasil telah dibandingkan dengan kaedah matheuris-

tic. Dari kaedah yang novel ini, pengiraan CPLEX dilaksanakan pada setiap titik arahan tebaran berdasarkan kaedah permodelan 2 Tahap Pengaturcaraan Linear Integer Rawak (ILP) (MDVRPSRC-2S1 dan MDVRPSRC-2S2). Kedua-dua model ini dihasilkan daripada Masalah Navigasi Kenderaan Berbilang-Gudang melibatkan Kapasiti Jalan yang Rawak (MDVRPSRC-2S) yang diperkenalkan seperti juga model deterministik oleh Masalah Navigasi Kenderaan Berbilang-Gudang melibatkan Kapasiti Jalan (D-MDVRPRC) yang dihasilkan daripada kajian awal sebelum pembentukan model MDDVRPSRC. Keputusan yang terhasil menunjukan perbandingan yang lebih kurang setara diantara heuristik-heuristik yang diperkenalkan dan matheuristic tersebut di dalam keadaan rawak yang terdapat pada koleksi ujian. Tambahan lagi, heuristik-heuristik yang diperkenalkan adalah lebih baik berdasarkan masa pengiraan pengkomputeran yang diperlukan.

Sumbangan kajian ini adalah seperti berikut: (1) MDP model untuk MDDVRPSRC, deterministik D-MDVRPRC dan juga 2 Tahap Rawak ILP MDVRPSRC-2S model telah diperkenalkan; (2) berdasarkan MDVRPSRC-2S yang diperkenalkan, kajian menunjukkan bagaimana CPLEX boleh dilancarkan pada setiap simulasi Monte Carlo di dalam PDS-RA dengan memperkenalkan pula dua varian model (MDVRPSRC-2S1 dan MDVRPSRC-2S2) daripada model MDVRPSRC-2S; (3) taburan bulatan gegaran daripada satu atau lebih punca gempa bumi dan pengurangan purata taburan kapasiti jalan raya serta masa perjalanan telah diperkenalkan; (4) algorithma penyelesaian: TBIH-1 dan empat variannya (TBIH-2, TBIH-3, TBIH-4 dan TBIH-5) telah diperkenalkan; (5) data ujian telah dibangunkan terdiri daripada simulasi rangkaian jalan raya dan unit kerosakan jalan berpunca daripada gempa bumi untuk tujuan experimen dan simulasi; dan akhirnya (6) sebuah sistem sokongan keputusan (MDDVRPSRC DSS) untuk tujuan simulasi penghantaran bekalan ketika bencana telah dibangunkan. Kesemua sumbangan ini boleh diaplikasikan untuk semua jenis permasalahan navigasi kenderaan yang dinamik melibatkan bencana dan kapasiti jalan yang rawak yang terkesan serta kelewatan perjalanan masa yang bertambah.

ACKNOWLEDGEMENTS

My sincere and utmost gratitude to the All Mighty, Most Compassionate, Most Merciful Allah SWT Who has always been supporting me against the insurmountable odds, Who has always been bestowing me with knowledge, comprehension and ideas to keep me steady throughout my research path.

My family especially my father and mother and also my wife who has been there supporting me and accompanying me along this journey with their prayers and positive encouragements. My sincere appreciation to my research supervisor Prof. Dr. Lee Lai Soon, who has been guiding me throughout the research and publications while also managing the research funds, providing research tools for me as well as inviting me in his research circle. Not forgetting Prof. Dr. Stefan Pickle, Prof. Dr. Maximilian Moll, the late Assoc. Prof. Dr. Bakri Adam as well as Prof. Dr. Leong Wah June for their supports in my research.

My appreciation must also without doubt goes to Universiti Utara Malaysia (UUM) in supporting my career as a tutor while pursuing my PhD degree. In the same manner Universiti Putra Malaysia has also been aiding me with my research from day one with all the facilities and supports as well as knowledge that it has to offer as a public learning institution. Especially the Geran Putra-Inisiatif Putra Siswazah (GP-IPS/2017/9579500) which has been funding me with my research. I also extend my heartful appreciation to Institute for Mathematical Research (INSPEM), lead by Prof. Dr. Muhammad Rezal Kamel Ariffin, as the institution for my research with Dr. Nur Sumirah Mohd. Dom as my PhD coordinator. Last but not least Pn. Noor Suhana Mohd Azahari, Pn. Zahratun Nur Yosminar and all INSPEM staffs, researchers and fellow research students who has been supporting me.

In the spirit of honouring the MoU between Universiti Putra Malaysia (UPM) and Universität der Bundeswehr München (UniBw Munich), Prof. Dr. Lee Lai Soon (representing INSPEM), Prof. Dr. Stefan Pickl from UBw Munich as well as myself (representing both Institute for Mathematical Research (INSPEM, UPM) and Universiti Utara Malaysia (UUM)) set out to collaborate on research looking into the Machine Learning approach in dealing with the multi depot dynamic vehicle routing problem with stochastic road capacity for the application of Humanitarian Logistics. Our many discussions leads to the idea of applying for the German Academic Exchange Service (DAAD) Research Grants - Bi-nationally Supervised Doctoral Degrees, (2018/19 (57381333)) in enabling research attachment for me as an exchange postgraduate student in Munich. Without the crucial funding for the research collaboration, this collaboration would not have been possible.

I certify that a Thesis Examination Committee has met on 17 June 2022 to conduct the final examination of WADI KHALID BIN ANUAR on his thesis entitled ”MULTI DEPOT DYNAMIC VEHICLE ROUTING PROBLEM WITH STOCHASTIC ROAD CAPACITY FOR EMERGENCY MEDICAL SUPPLY DELIVERY IN HUMANITARIAN LOGISTICS” in accordance with the Universities and University Colleges Act 1971 and the Constitution of the Universiti Putra Malaysia [P.U.(A) 106] 15 March 1998. The Committee recommends that the student be awarded the Doctor of Philosophy.

Members of the Thesis Examination Committee were as follows:

Norazak bin Senu, Ph.D.
Associate Professor
Institute for Mathematical Research (INSPEM)
Universiti Putra Malaysia
(Chairperson)

Norihan binti Md Arifin, Ph.D.
Professor
Institute for Mathematical Research (INSPEM)
Universiti Putra Malaysia
(Internal Examiner)

Zanariah binti Abdul Majid, Ph.D.
Professor
Institute for Mathematical Research (INSPEM)
Universiti Putra Malaysia
(Internal Examiner)

Wolfgang Bein, Ph.D.
Professor
Department of Computer Science
University of Nevada, Las Vegas, USA
(External Examiner)

ZURIATI AHMAD ZULKARNAIN, Ph.D.
Professor and Deputy Dean
School of Graduate Studies
Universiti Putra Malaysia

Date:

This thesis was submitted to the Senate of Universiti Putra Malaysia and has been accepted as fulfilment of the requirement for the degree of Doctor of Philosophy. The members of the Supervisory Committee were as follows:

Lee Lai Soon, PhD
Professor
Faculty of Science
Universiti Putra Malaysia
(Chairperson)

Leong Wah June, PhD
Professor
Faculty of Science
Universiti Putra Malaysia
(Member)

Mohd Bakri Adam, PhD
Associate Professor
Faculty of Science
Universiti Putra Malaysia
(Member)

Stefan Pickl, PhD
Professor
Fakultät für Informatik
Universität der Bundeswehr München
Germany
(Member)

ZALILAH MOHD SHARIFF, PhD
Professor and Dean
School of Graduate Studies
Universiti Putra Malaysia

Date: 11 August 2022

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ALGORITHMS

LIST OF ABBREVIATIONS

VRP	Vehicle Routing Problem
RVRP	Rich VRP
HL	Humanitarian Logistics
DM	Disaster Management
DSS	Decision Support System
ML	Machine Learning
ADP	Approximate Dynamic Programming
MDDVRPSRC	Multi Depot Dynamic Vehicle Routing Problem with Stochastic Road Capacity
MDP	Markov Decision Processes
MILP	Mixed Integer Linear Programming
SP	Stochastic Programming
SILP	Stochastic Integer Linear Programming
D-MDVRPRC	Deterministic Multi Depot VRP with Road Capacity
MDVRPSRC-2S	Two-Stage SILP Multi Depot VRP with Stochastic Road Capacity
MDVRPSRC-2S1	Reduced Model 1 (Shelter)
MDVRPSRC-2S2	Reduced Model 2 (Depot)
D	Deterministic Model, D-MDVRPRC (Offline)
S	Stochastic Model, MDVRPSRC-2S (Offline)
SDY	Stochastic & Dynamic Model (Online)
RA	Rollout Algorithm
PRE	Pre Decision State
PDS	Post Decision State
PDS-RA	Post Decision State Rollout Algorithm
TBIH	Teach Based Insertion Heuristic
TBIH-1 – TBIH-5	TBIH variants
TP	Teaching Part
SP	Seeking Part

SIH(I1)	Sequential Insertion Heuristic Type (I1)
CW	Clarke and Wright
DSIH	Dynamic SIH
DCW	Dynamic CW
DLASIH	Dynamic Lookahead SIH
DLACW	Dynamic Lookahead CW
DF	Damage File

CHAPTER 1

INTRODUCTION

1.1 Overview

The occurrence of disasters throughout the world has lead to various negative impacts including famine, outbreak of diseases, poverty and loss of life among many others. Additionally, the trends towards higher frequency of disasters and the economic effects demands an immediate attention towards efficient relief operation (Motoyama, 2017). Moreover, the impact of disasters in underdeveloped and developing countries would be worse, as they are often without sufficient resources when dealing with the aftermath of disasters. For cases such as Nepal which is also

a landlocked country that recently suffered from a 7.8 magnitude Earthquake on the 25th April 2015 and lost nearly nine thousand lives due to the initial shock alone, the need for an efficient relief aid operation is even more dire (Zhao, 2016). This includes especially, urgent medical supply to the rural area and districts during the aftermath of the disaster as reflected by Sharma (2015) and Neupane (2015).

Tracing back on the details of the event it is clear that the medical supplies in landlocked country such as Nepal are mainly depended on international aids (Chauhan and Chopra, 2017). Furthermore, there were strong demands for medical supply such as vaccination that was lacking during the post disaster event. Wang et al. (2016a) stated that during the event almost 90% of local medical centres and facilities were inoperative and that surgical dressing as well as debridement were among the supplies needed. Moreover, Chauhan and Chopra (2017) also cautioned that preparations for infectious diseases outbreak should be made during the aid operation following the collapses of hygiene and sanitation facilities during the disaster. This further stressed the importance of medical supply during the disaster for landlocked countries such as Nepal. Radianti et al. (2016a) through their data collection based on twitters worldwide also mentioned the difficulties of delivering medical supply due to the regular customs that was imposed and the lack of transportations to disseminate supplies. Moreover, whichever transportation that was available had to operate under the lack of fuel supply due to the blockade of fuel by India (Hall et al., 2017). This further illustrates the strong dependency towards neighbouring countries for a landlocked country such as Nepal, which is devoid of access of sea for crude resources such as fuel. Hall et al. (2017) further stated that most of the existing medical supplies were not used as most of them were trapped under the rubbles due to the earthquakes.

In addition, Nepal as a landlocked country also suffers logistical nightmare as all international aid rushed towards the Tribhuvan Airport acting as the only local logistical hub which also worsen the distribution of the medical supply in Nepal. The much needed upgrade of the airport were unable to host over 400 humanitarian organization supplies causing a bottleneck at the airport (Menth and Stamm, 2015). The incompatibility of the airport to cater for much larger airfreights was also among the cause of delay for medical supply in Nepal (Radianti et al., 2016a). As a result, much of the airfreight carrying significant amount of medical supplies were rerouted to neighbouring countries.

Additionally, despite large supply resources due to global aid from various nations that flooded the Kathmandu Airport during the disaster, efficient distribution of supplies are hindered due to landslides disrupting the transportation network especially the parts that lead to the rural area where medical supplies are scarce ((Radianti et al., 2016a) and (Chiaro et al., 2015)). Hall et al. (2017) for example reported on the problem of accessing remote area due to the landslides and road damages. More similar cases were mentioned such as in Okamura et al. (2015) where heavy damages sustained by the Kathmandu – Bhaktapur Road of Araniko Highway caused major disruption in medical supply. Further reports on lack of coordination in delivering aid supply would also be a strong justification for the proposed research as efficient medical supply routing would also need to integrate coordination factor into its model and solution (Hall et al., 2017). Hence the problem observed includes limited transport vehicles, congested single point of supplies collection, and uncertainties in route options as well as road facility sustaining continuous damages (Radianti et al., 2016a).

From this particular aspect, it is seen that multi depots could be the means to alleviate the bottleneck problem in Kathmandu Airport where these depots could represent a number of alternative airports. Due to the limited transport vehicle, it is hypothesized that multi trip and split delivery between vehicles when delivering to specific destination could be useful. Similarly there is an obvious correlation between the disaster strike zone and the ongoing disruption of road network which necessitate dynamic update as a means to tackle the uncertainties.

Motivated by these observations, a Multi Depot Dynamic Vehicle Routing Problem with Stochastic Road Capacity (MDDVRPSRC) is proposed and presented. Furthermore prior investigation to the development of MDDVRPSRC has lead to the development of the Deterministic Multi Depot Vehicle Routing Problem with Road Capacity (D-MDVRPRC) model, the two-stage Stochastic Integer Linear Programming (SILP) Multi Depot Vehicle Routing with Stochastic Road Capacity (MDVRPSRC-2S) model, as well as two reduced models of MDVRPSRC-2S: MDVRPSRC-2S1 and MDVRPSRC-2S2 that can be applied iteratively to solve the MDDVRPSRC in a dynamic manner. As a result, a novel matheuristic rollout is also presented.

In addition, inspired by the work of Goodson et al. (2013), Novoa and Storer (2009) and many others whose application is more geared towards industry and commercial purposes, this research aims to expand further from the conventional Machine Learning (ML) approach in VRP of clustering the vehicle first, as part of the measure in dealing with the vast action space. Fresh approach is needed for problem where the clustering of vehicles might not be applicable due to the stochastic road capacity within the transport network. Furthermore, with the entry level machine available for the research, the Monte Carlo simulations and the lookahead horizon needed must be limited yet a good decision must be computed. Normally, this is a contradiction within the field of VRP adopting Rollout Algorithm (RA) in the scope of Reinforcement Learning (RL). This contradiction however is addressed by proposing individual computation of decision for each vehicle as opposed to computing decision collectively.

This research is differentiated further through the presentation of models that adapts to the changing of road capacity. Here, different base heuristics are proposed and applied in the Post Decision State Rollout Algorithm (PDS-RA) advocated by Goodson et al. (2013) that could be more suitable to tackle specific problem of deteriorating and stochastic road capacity. The base heuristics presented are then compared with the matheuristic rollout through which a new test dataset is introduced. Moreover, the tremor of the earthquake is also simulated through which the deteriorating damage and capacity of roads are anticipated and considered. Finally, an MDDVRPSRC Decision Support System (DSS) is presented as a simulation platform to support decision making for the case of MDDVRPSRC while also aiding the simulated data collection for validation and performance benchmark.

1.2 Research Background

This section served as a quick introduction to the core components of the research. The field of VRP and its inherent solution algorithm is vast and is considered known to the reader. To those uninitiated in VRP, the textbooks by Toth and Vigo (2002) and Toth and Vigo (2014) are referred. Additionally, to those interested to learn the basics of heuristic and metaheuristic, the textbooks by Labadie et al. (2016) and Talbi (2009) are also referred.

Full focus is given instead on Machine Learning (ML), Reinforcement Learning (RL) the solution of RL with regards to Approximate Dynamic Programming (ADP), as well as the lookahead approach in ADP as these fields represents the core of this research where every contributions are made towards modelling and solving the MDDVRPSRC for emergency delivery as part of the humanitarian operations in HL.

This chapter starts off with the general overview of ML and later zoom into the RL through the description of Markov Decision Processes (MDP), the solution methodology of MDP, the curse of dimensionality as well as the extended framework of MDP modelling involving Post Decision State (PDS) structure. The discussion on the curse of dimensionality leads to the topic of ADP. From ADP, the focus quickly narrows down into the lookahead approach where the Rollout Algorithm (RA) is described in detail. Finally the extended version of RA (PDS-RA) which is the core solution of MDDVRPSRC is explained.

1.2.1 Machine Learning, Reinforcement Learning and Markov Decision Processes

It is quite rare to talk about optimization or OR together with ML as they are treated as different entities branching on their own fields. However, recent trend in the literatures proves that this might no longer be the case. Various OR problems including VRPs are already addressed based on the ML approach specifically in RL. Future trend seems to point towards hybridization of ML and other classical optimization approaches such as metaheuristic, heuristic, and exact solution algorithms.

ML is the subfield of artificial intelligence (Figure 1.1) which is the subfield of computer science (Theobald, 2017). The aim of ML is to allow machine in the form of programs or codes to learn by itself the task that is required of them instead of the programmer having to tell them what to do. In fact, we now see programmers collecting data and feeding them into the learning model and algorithm leaving the ML code to organize and act based on the prediction made through its learning process (Alpaydin, 2016).

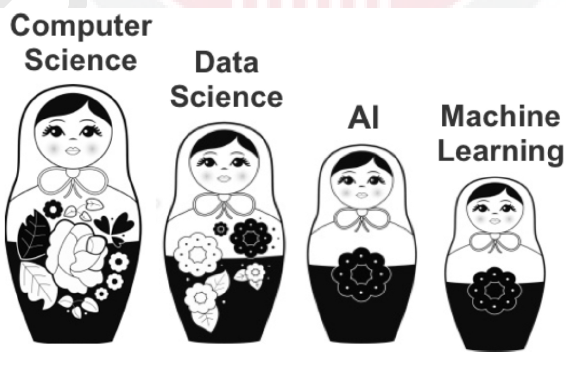


Figure 1.1: ML as the Subfield of Computer Science, Data Science and AI
[Source: Theobald, 2017]

Depending on the nature of learning, the ML is divided further into three categories (Theobald, 2017):

1. Supervised Learning,
2. Unsupervised Learning, and
3. Reinforcement Learning.

Despite the fact that there are three categories in ML, it is interesting to note that RL is adopted most for solving Complex Optimisation Problems (COPs). Such literatures includes Nadi and Edrisi (2017) and Keneally et al. (2016) where complex VRPs are addressed in times of disasters. On the other hand, few literatures are also seen adopting unsupervised learning through the Fuzzy C Means (FCM) variants such as in Ruan et al. (2016). Recently James et al. (2019) applied deep neural network, a supervised learning approach, in generating vehicle route online, while Zhang et al. (2020) applied multi agent RL in dealing with the VRP.

1.2.2 Markov Decision Processes

The core of this research is the Markov Decision Processes (MDPs) which is the prerequisite modelling framework to RL. It provides the modelling framework that represents RL given in:

- State, s_k in the state space S:
  - comprised of initial state s_0, the end or termination state s_K.
  - each state observable by the agent at decision epoch k at time t.
- Decision of action, a_k in the action space A:
  - at each decision epoch k, selection of action in action space is possible: $a_k \in A(s_k) \subset A$ where A is an action space consist of all available actions for all states available.
- Reward model, $R(s_k, a_k)$.
- Transition model:
  - deterministic transition model: $s_{k+1} = S^M(s_k, a_k)$.
  - stochastic transition model: $s_{k+1} = S^M(s_k, a_k, W_{k+1})$.

In this modelling framework, the state s_k represents a snapshot at decision epoch k of what is observed by an agent of a certain environment as opposed to how a human being would perceived an environment. It is deemed as limited as normally what is observable to a machine is limited to their sensory inputs and data transfer size. Different sensory inputs that are accessible to an agent are represented by different state variables. Therefore, an observation (state) made by an agent consist of a multidimensional vectors of variable values depending on what is observed by the agents. This observation is made by the agent at time t or decision point k, where agent is require to make a decision or take an action a_k. For example, in a specific point of time t where an agent needs to make a decision in a VRP, a state s_k might be represented by a number of unserved customers c, capacity status of each delivery vehicles q and demands of unserved customers d as it state's variable.

What is observed by the agent is then:

$$s_k = [c, \mathbf{q}, \mathbf{d}], \tag{1.1}$$

which is a multidimensional vector of specific state that could be represented as in Figure 1.2. The state space is then the space confined by the range of all variable axis $c, \mathbf{q}$ and $\mathbf{d}$ where all different points of s_k exists. A simple state is assumed by having c as an integer variable with $\mathbf{d}$ and $\mathbf{q}$ as a vector of n customer demands $d \in \{0,1\}^n$ and capacity of vehicles $\mathbf{q} \in \{0,50\}^{|M|}$ for vehicle $m \in M$. However, each value of variable may yet be another vector which extends the axis further enlarging the state space by the length of the axis. Action a_k represents a limited or selected scope of decision or action that an agent is permitted to make or perform at a decision epoch k based on what is observed during the triggered decision epoch. Making a decision changes the environment or process and this in turns is observed by the agent further as the next state of the environment. If the environment is deterministic, where no uncertainty is concerned, the next state observed by the agent due to the action taken is denoted as:

$$s_{k+1} = S^M(s_k, a_k). \tag{1.2}$$

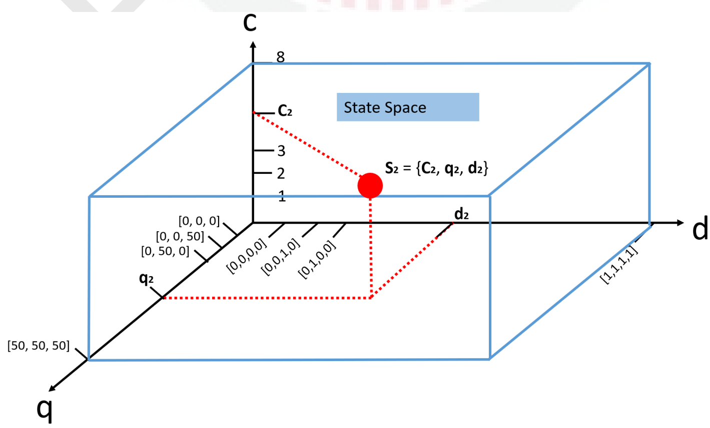


Figure 1.2: Illustration of State Space and its Variables

In MDP this is known as the deterministic transition model. When an agent has knowledge on both reward and transition model ($s_{k+1} = S^M(s_k, a_k)$), the MDP model is known as model based MDP. Alternatively a popular approach of MDP represent the environment as a black box where the agent, having no prior knowledge on rewards and transition model, could only perform an action and obtained reward in return of its action. Based only on this interaction with the black box (environment), an agent is required to compute an optimal policy by learning the transition model and reward model instinctively through series of action - reward interactions. This type of MDP model is known as the model free MDP seen applied in popular RL application such as the Atari (Mnih et al., 2013) and the Alphago Zero (Silver et al., 2017).

In a more realistic situation or environment, transition of states observed by the agents is rarely deterministic in nature. For example, if an agent make a decision to drop a leaf. In a deterministic world devoid of uncertainty elements, the agent would perceived based on the deterministic transition state of the MDP model that the leaf would fall in a vertical trajectory. However, in a realistic stochastic world, random uncertain gust of wind might influence the position of the leaf as it falls. As a result, the leaf trajectory might no longer be following a vertical line (Figure 1.3). In such a case an agent can only instead, ***expect*** a certain trajectory of the leaf falling based on random distribution of wind in the vicinity. An agent could only make prediction and expect such motion if the system model takes into account random information W.

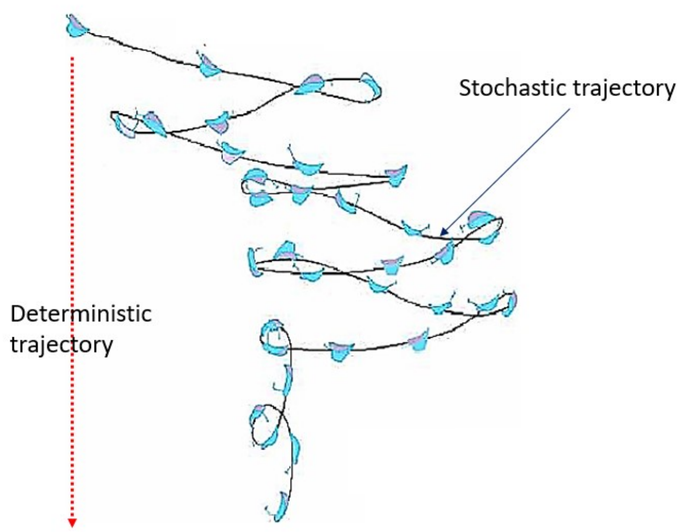


Figure 1.3: Deterministic Trajectory and Stochastic Trajectory
[Source: Haiyan, 2010]

Thus in a stochastic world, the transition model are now defined as:

$$s_{k+1} = S^M(s_k, a_k, W_{k+1}). \tag{1.3}$$

If one to reflect back on the model free MDP, such transition model is not specifically known to an agent. Instead, agent would need to sample series of an episode or realization ω by randomly performing an action and observed the transitioning state as well as the rewards obtained. An agent would need to perform random sampling each time creating a trajectory or episode ω that depends on random changes (denoted at the moment as $\hat{p}$) that occurred at decision epoch k at time t. As such, the random information W is depicted as $W_k(\omega_m(k))$ at decision epoch k or at time t as $W_t(\omega_m(t))$. An MDP model could consider more than one random variables hence the variable $\hat{p}$ could instead be a vector consist of all random variable observed by the agent.

A clear example of random information could be seen in Table 1.1:

Table 1.1: Example of Trajectories of Transitions Enabled by Sampling Random Information $W_k(\omega_m(t))$

	$t=1$	$t=2$	$t=3$	$t=4$	$t=5$	$t=6$	$t=7$	$t=8$
ω	$\hat{p}_1$	$\hat{p}_2$	$\hat{p}_3$	$\hat{p}_4$	$\hat{p}_5$	$\hat{p}_6$	$\hat{p}_7$	$\hat{p}_8$
1	(-14, -26)	(7, 16)	(-10, 20)	(41, 7)	(15, -11)	(8, -29)	(-6, -50)	(13, -36)
2	(11, 33)	(-36, -50)	(-23, 0)	(3, -50)	(7, -23)	(12, -10)	(35, 46)	(-18, -10)
3	(-50, -12)	(-2, -18)	(-15, 12)	(-31, 3)	(5, -2)	(-24, 33)	(10, 2)	(38, -19)
4	(1, -13)	(41, 20)	(-32, -2)	(3, -4)	(-46, 34)	(-15, 14)	(38, 16)	(48, -49)
5	(2, -34)	(-24, 34)	(-9, 25)	(-19, -40)	(1, -28)	(34, -7)	(36, -3)	(46, -35)
6	(41, -49)	(-24, -33)	(-5, -25)	(16, 36)	(8, 47)	(-17, -4)	(-29, 45)	(-48, -30)
7	(44, 37)	(7, -19)	(49, -40)	(-13, 5)	(38, 37)	(-30, 45)	(-48, -47)	(19, 41)
8	(-19, 37)	(-50, -35)	(-28, 32)	(-13, -17)	(2, -2)	(-10, -22)	(-2, 47)	(2, -24)
9	(-13, 48)	(-48, 25)	(-37, 39)	(-2, 30)	(-28, 33)	(-35, -49)	(-44, -13)	(-6, -18)
10	(48, 5)	(37, -39)	(43, 34)	(-13, -6)	(28, -37)	(-47, -12)	(13, 28)	(26, -35)

[Source: Powell, 2007]

From the table it could be seen that 10 samples of realization or episodes from decision epoch 1 ($k = 1$) to decision epoch 8 ($k = 8$), where each time random transition occurred and random information W_k is known. For example in the 10th sample or episode, at decision point $k = 5$ two random parameters are known ($\hat{p}_5 = \{28, -37\}$) and are then observed by the agent. Hence, at that decision point ($k = 5$) of the 10th sample random information $W_5(\omega_{10}(5)) = \hat{p}_5 = \{28, -37\}$ is obtained by sampling. The agent next observed from the environment a new state (s_{k+1}) (as modelled by Equation (1.3)) based on it's previous observation (s_k) and it's respective action (a_k) towards the observation as well as incoming random information (W_k).

With all the important components of MDPs being introduced, an MDPs tree depicting clear deterministic and stochastic transition are shown in Figure 1.4

and Figure 1.5, respectively. The concept of Post Decision State (PDS) is not yet discussed although it is shown in the figures as circles. At this point, the illustration of PDS is needed to mark the point of when random information is received. It could be seen from both Figure 1.4 and Figure 1.5 that agent performing action causes deterministic transition and incoming random information are observed next changing the trajectory of the transition into an uncertain stochastic transition (dashed lines).

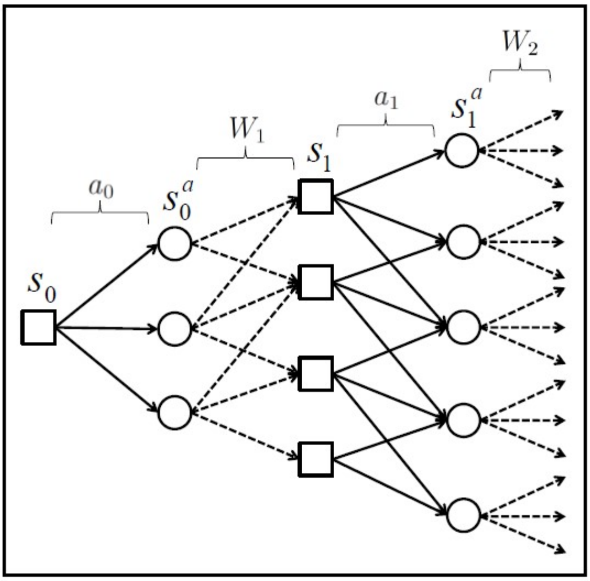


Figure 1.4: A Portion of MDP Tree Involving Deterministic and Stochastic Transitions
[Source: Goodson, 2010]

Reward on the other hand is a feedback obtained by the agent after performing an action or making a decision. Reward usually reflects good or bad decision or action performed by the agent. This is akin to a child getting a praise for saying thank you instead of a scolding for saying anything rude. A complex and effective learning process such as MDP considers long term collected rewards as opposed to immediate short term rewards to ensure that good planning over a longer horizon or time frame.

In RL the agent might already have a knowledge of what is good and what is bad (or specifically the reward of performing an action) given a state that it observed. This is usually specified by a human via lookup table of action - reward or state-reward mapping. This concept of RL could be further understood by the following in Figure 1.6 that shows the relationship between the components of MDPs or RL. In MDPs

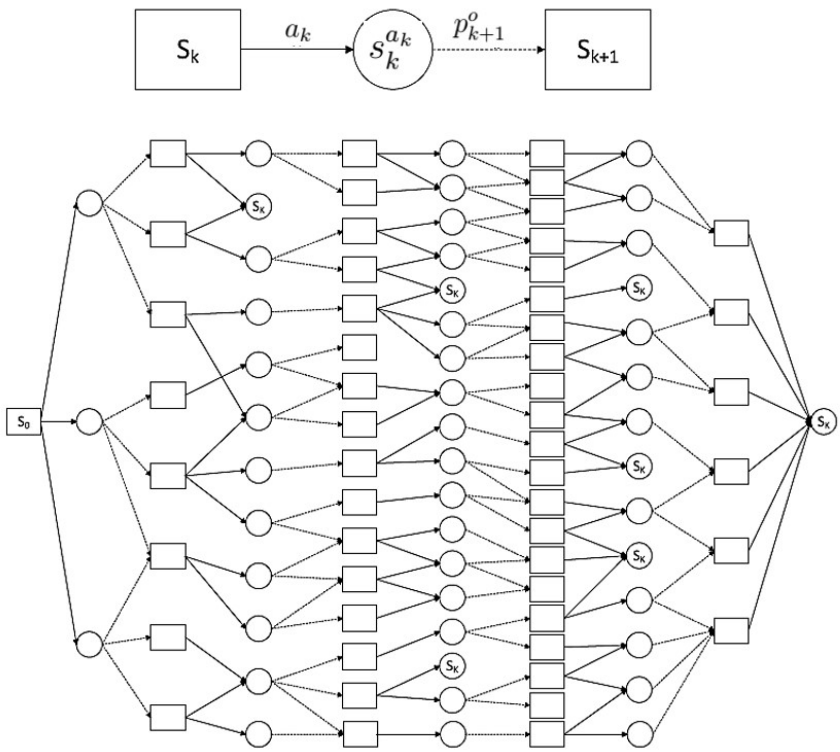


Figure 1.5: MDP Complete Tree Involving Deterministic and Stochastic Transitions
[Source: Ulmer, 2017]

or RL, an agent learns by observing a state of an environment. It then performs supposedly intelligent (computed offline or online) action. This action influence the environment and therefore changes the earlier state into a new observable state via state transition. Furthermore by performing an action, it receives a reward or incurred a cost in the form of a certain value.

The aim for an agent is to either maximize the collective rewards or to reduce the collective cost. By doing so, an agent is said to have optimized an operation (the environment model being the operation). It thus follows that to obtain the maximum sum of rewards as an indication of an optimized operation in RL, an agent would need to take the best or optimal action at each state that it observed where decision making is needed. This is achievable if an agent has a lookup table that maps each state to an optimal action it should take. This lookup table is what is known as the optimal decision rule based on an optimal policy in RL problem. Therefore, solving MDP model or problem means searching for an optimal policy to obtain an optimal decision rule.

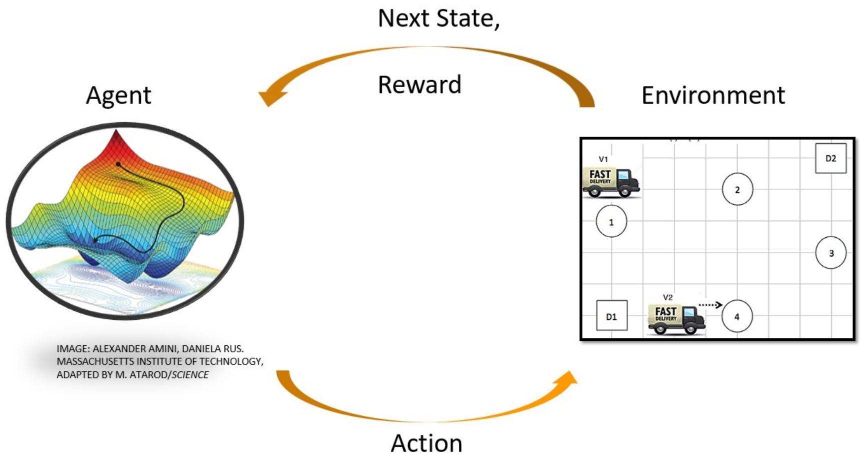


Figure 1.6: MDP Components and the Interaction with the Environment

1.2.3 Solving Markov Decision Processes

In Section 1.2.2, a brief concept of RL followed by MDPs is introduced. It is also understood that in order to solve a COPs through ML or specifically RL approach, one would have to solve for the optimal policy that would dictate an optimal decision rule.

The question would then follows:

1. *How does an agent compute the best action taken for each state that it observed ?*
2. *What measurement or parameter serves as a basis in choosing the optimal action ?*

In fact, the answer to these two questions have long been answered by Richard E. Bellman. Rightly known as the father of dynamic programming, Richard E. Bellman not only came out with the concept of solving problem through learning processes (Dreyfus, 2002), but also developed the modelling framework that we now know as MDPs (Bellman, 1957). Furthermore, he theorized the solution principle known as the Principal of Optimality (Bellman et al., 1954). All of the aforementioned contributions of Richard E. Bellman are applied in this research. The solution principal then leads to another of his contribution known as the *Bellman Equation*. Incidentally he also coined the term *"the curse of dimensionality"* which is discussed in Section 1.2.3.1 paving way to the Approximate Dynamic Programming (ADP) (Section 1.2.4) as well as the Post Decision State (PDS) (Section 1.2.3.2).

Following Bellman's Principal of Optimality, it is stated that an action taken from the initial state onwards needs to be a consistently optimal action for one to form an optimal trajectory (initial state to the end state) or episodes. This is akin to saying that for a process to be an optimal process, every decision made throughout the process from the start towards the end needs to be a series of consecutively optimal decisions. This is none other the aforementioned optimal decision rule. Another important point from this principle is that solving the optimisation problem can now be done recursively via backward dynamic programming which lands Richard Bellman the nickname, father of the dynamic programming.

The Principal of Optimality leads to both question 1 and 2 above. These questions are then answered through the Bellman Equation presented in Bellman et al. (1954). The idea of this equation is to lead the agent into choosing an action based on a value assigned to each state associated by the action taken. Conventionally, a brute force algorithm is designed for the agent to compute the values for all possible state in the state space of the environment. Once this is computed, an agent needs only to navigate itself from the initial state to the end state, each time choosing an action that it now know would lead to the next state holding the highest state's value at each decision epoch. The resulting transition of state - action - next state cycle in the navigation path until the end state, is the optimal trajectories. An objective of an agent guided by this optimal policy is thus defined as:

$$\max_{\pi\in\Pi}(\mathbb{E}\{\sum_{k=0}^{K} R(s_k, A^{\pi}(s_k))\}). \tag{1.4}$$

It needs to be noted here that some literatures define policy as the decision rule itself. Although the definition of policy and decision rule are in the literature commonly intertwined with one another and sometimes even used interchangeably, their difference is made clear in this research. In this research, the policy is regarded as the theme in which a decision rule is based on. For example, capitalism could be a policy in making a business decision in general but do not dictate the decision rule explicitly. More specifically, in this research, a policy could be an approach of using a specific unknown theme for choosing an action, however the function mapping the state to (a supposedly good) action is not the theme itself.

Although a good decision making involves considering long term rewards, sometimes one would also need to put emphasize on a shorter horizon of planning by giving more priority to immediate rewards as opposed to delayed rewards in the subsequent transitions. One may consider the problem of an agent navigating its way to an end goal. However, if the agent discovers that going through a longer path leads to more rewards collected, it might not be in a hurry to reach the goal. In fact, it might never reach the goal. To balance the long term rewards and the objective of the problem, a decay rate $\gamma \in [0,1]$ is then introduced:

$$\max_{\pi \in \Pi}(\mathbb{E}\{\sum_{k=0}^{K} \gamma^k R(s_k, A^{\pi}(s_k))\}). \tag{1.5}$$

The core idea of solving MDP is to compute values for states involved at least in the transitions experienced by the agent. Ideally an agent would have to compute all values for all possible states that could exist in the state space which is known as the exact computation of the Bellman Equation.

The principal of optimality leads to the objective function and introduces the concept of computing value of states:

$$V_k(s_k) = R(s_k, a_k) + \gamma \mathbb{E}\{\sum_{k=k+1}^{K} R(s_k, a_k)\}. \tag{1.6}$$

It is important to note that the role of expectation ($\mathbb{E}$) computation in the equations as an agent is considering a stochastic planning.

Furthermore, the equation can be further derived into a standard Bellman Equation that give birth to the concept of dynamic programming (Equation (1.7)):

$$V_k(s_k) = R(s_k, a_k) + \gamma \mathbb{E}\{V_{k+1}(s_{k+1})|s_k\}, \tag{1.7}$$

which is solved exactly by looping over all possible next states $s_{k+1} = s'$ from the current state s_k solving the expectation as in Equation (1.8):

$$V_k(s_k) = R(s_k, a_k) + \gamma \sum_{s' \in S} p(s'|s_k, ak) V_{k+1}(s')|s_k. \tag{1.8}$$

Different action a_k taken leads to a different computed value. To guide the action into transitioning to the next best state as to maximize the rewards, an agent would have to compute the highest values possible for a state s_k in Equation (1.7) over all possible action a_k that can be made at decision epoch k. This is shown in Equation (1.9):

$$V_k^{\star}(s_k) = \max_{a_k \in A(s_k)} (R(s_k, a_k) + \gamma \mathbb{E}\{V_{k+1}(s_{k+1})|s_k)\}. \tag{1.9}$$

However, the objective for an agent is of course to compute the best decision rule $\rho(s_k)$ following the best policy $\pi^{\star}$ such that $\rho^{\pi^{\star}} : s_k \rightarrow a_k^{\star}$. The optimal action $a_k^{\star}$ can then be determined from the Bellman Equation by looping over all available actions at decision epoch k as shown in Equation (1.10):

$$a_k^\star = \underset{a_k \in A(s_k)}{\arg\max} \left(R(s_k, a_k) + \gamma \mathbb{E}\{V_{k+1}(s_{k+1}) | s_k)\right\}, \tag{1.10}$$

hence solving MDP would mean to either solve for maximum value of each states as in Equation (1.9) as to guide the search for optimal action indirectly, or to solve for an optimal decision rule by computing directly the optimal action as in Equation (1.10) at each decision epoch until terminal state is reached. Both Equation (1.10) and (1.19) are the core equations that solves MDP exactly.

Meanwhile, the conventional mean of solving MDP through basic MDP framework (without PDS) is illustrated in the Figure 1.7 - 1.9 below where the random parameter $\hat{p}_k$ is depicted through its possible random values p_k^f at decision epoch k where $f = \{1, 2\}$:

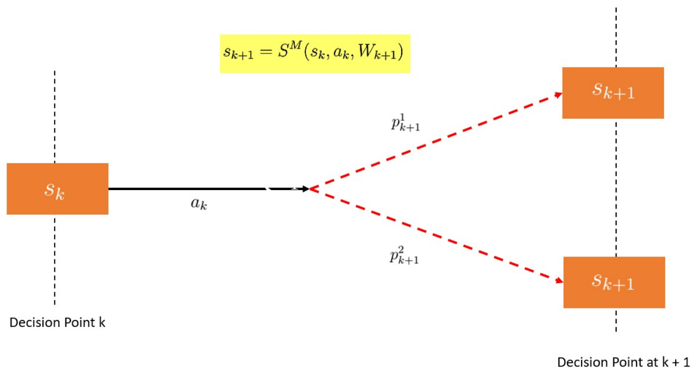


Figure 1.7: Conventional MDP structure: Current state & next state 1

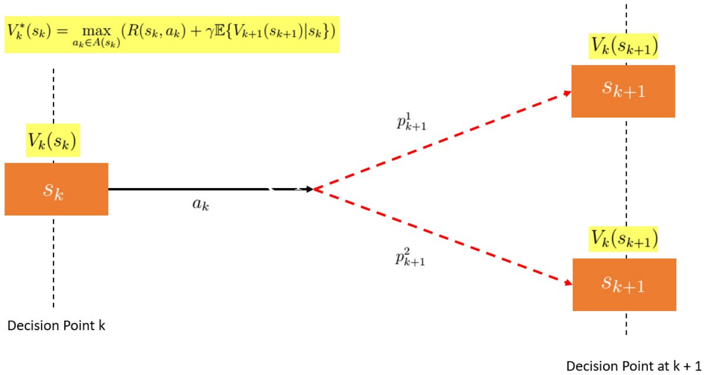


Figure 1.8: Conventional MDP structure: Current state & next state 2

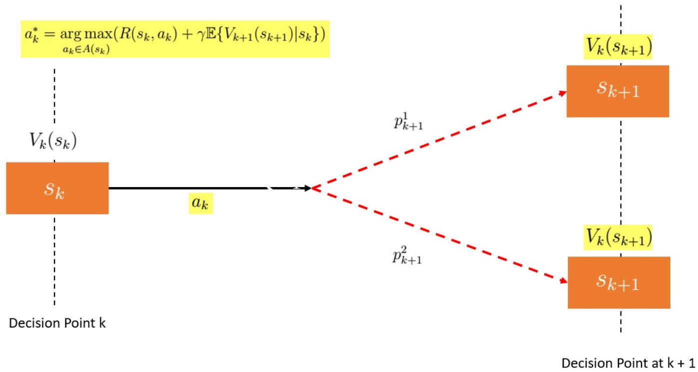


Figure 1.9: Conventional MDP structure: Current state & next state 3

1.2.3.1 The Curse of Dimensionality

Powell (2007) listed three curses of dimensionality in the dynamic programming which could be identified from the previous equation originated from the followings:

1. state space,
2. action space,
3. outcome space.

Without any approximation method, these three curses of dimensionality often would render the RL solution approach to be prohibitive. Most ADP approaches can be seen as solutions to solve any combination of these three curses. In short, ADP is a solution to approximate values computed to solve for MDP. For a simple

MDP problem, the three curses are manageable more so if it is a deterministic problem. However, a complex MDP problem leads to exhaustive computation due to explosion of state, action and outcome spaces.

A simple demonstration related to this research could illustrates the problem. For a complex MDP problem that revolves around 4 depots, 8 connecting nodes and 3 shelters connected by 58 edges with 3 vehicles for delivery operations, a stochastic road capacity is considered: low capacity, medium capacity, and high capacity. For the repetitive process of decision making, an agent is made to observe the following:

1. Current location of each vehicle, $l_k \in \{1,2,3,4,5,6,7,8,9,10,11,12,13,14,15\}^{|M|}$.
2. Arrival time to destination of each vehicle, $t_k \in [1..300]^{|M|} \subset \mathbb{Z}$.
3. Capacity status of each vehicle, $q_k \in \{0,50\}^{|M|}$.
4. Current occupancy of each roads associated with each vehicles, $e_k \in \{0,1\}^{174}$.
5. Demand status of each shelter (connecting nodes and depots set to zero), $d_k \in \{0,50,100,150,200,250,300\}^{|D|}$.
6. Road capacity, $r_k \in \{0,1,2\}^{58}$.

Therefore, the state observe at each decision epoch changes according the these variables:

$$s_k = [l_k, t_k, q_k, e_k, d_k, r_k]. \tag{1.11}$$

From this problem, we see that vehicle $m \in M$ with $|M|$ and nodes, n (depots, shelters and connecting nodes) $|N| = 15$. Furthermore, there are 58 edges available that made up to $(58 \times 3 = 174)$ edges associated to each vehicles. From here, it is computed that there are roughly:

$$(15^3) \times (300^3) \times (2^3) \times ((2)^{(58\times3)}) \times (|D|^{15}) \times (3^{58}) = 3.9 \times 10^{104} = |\mathbb{S}|,$$

possible states in the state space for this problem alone.

Moreover, since the agent need to constantly assigned location for each vehicle at every decision epoch k, there are $15^3 = 3375$ actions that need to be considered by the agent at each decision epoch and each action is associated further with 3 uncertainties: low, medium and high road capacity.

For an exact solution to this MDP problem, an agent would have to compute optimal action at decision epoch k through Equation (1.10) over all the values of 3.9×10^{104} states in the state space that involve iterative looping over all 3375 actions each featuring looping over all 3 potential outcomes when also computing the expectation term.

This means that solving stochastic MDP exactly for a realistic problem is prohibitive in practice. Due to this, many solution providers would avoid adopting RL solution approach. However, there is a real interest to solve for stochastic and dynamic problem in addressing real complex problem. The advancement of computation technology sparks a new life into the adoption of RL as a solution approach for many complex problems. Special consideration however are given to the three curses of dimensionality. Hence, the proposed structure of PDS and the birth of ADP.

1.2.3.2 The Post Decision State (PDS)

As the name suggest, the idea of PDS is quite simple and intuitive. A state is the observation of the environment made by an agent during a decision epoch based on the sensory inputs available to the agent. An agent then make a decision based on the observed state and in return observed a new state. In the proposed structure of PDS, an agent is assumed to be able to make observation just as a new random information W_{k+1} arrived after making decision a_k at decision epoch k. Thus a new state (PDS), $s_k^{a_k}$ is introduced to represent an agent's observation just after a current state s_k is observed and a decision a_k is made. This state observed by the agent right before a new state s_{k+1} is observed as the new information W_{k+1} is received.

The concept of PDS went naturally right after the introduction of MDP and dynamic programming by Bellman although no one made any official mentioned on the term PDS (Powell, 2007). Instead, the first to ever use the term PDS is Bertsekas in 1997 (Van Roy et al., 1997). It was later in 2004 that PDS is described as a solution strategy by Powell and Van Roy (Powell et al., 2005). The computational benefits and detailed explanation of PDS in regards to ADP are addressed in Powell (2007). A specific reference on how rollout algorithm could be computationally improved through the introduction of PDS is then shown in Goodson (2010).

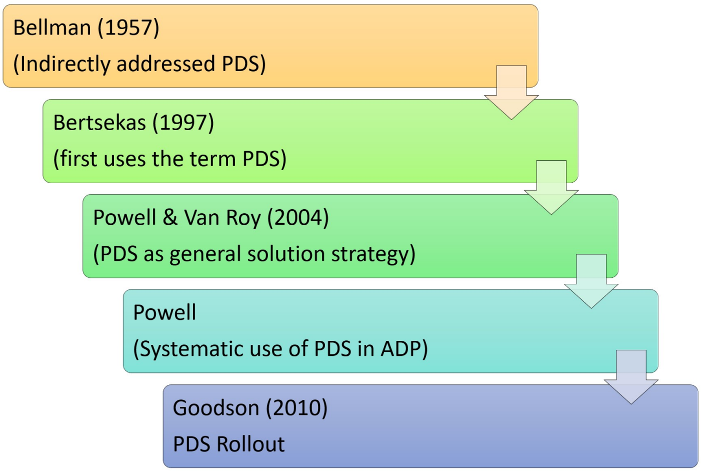


Figure 1.10: Brief History of Post Decision State in Relation of this Research
[Source: powell, 2007]

1.2.3.3 Post Decision State Structure in Markov Decision Processes

The concept of PDS is naturally embedded in MDP by dividing the transition of observation made by the agents into two: deterministic or stochastic transitions, which leads to the two observations made by the agent respectively:

1. A post decision state $s_k^{a_k}$ observed after making a decision or action a_k as a result of observing state s_k.
2. Next state s_{k+1} after random information W_{k+1} is received by the agent.

This is illustrated by the following Figure 1.4 and 1.5.

The process occurred dynamically at every decision epoch $k = 0,1,2,3,..,K$ consist of subsequent possible actions space $A(s_k) \subset \mathbb{A}$ and random information W_{k+1} followed by all possible outcomes p_k^o of random parameter $\hat{p}$ that could occur in the environment.

1.2.3.4 Solving Markov Decision Processes through Post Decision State Structure

In the PDS structure as explained in Powell (2007) and seen in Figure 1.5, the transition is divided into deterministic transition:

$$s_k^{a_k} = S^{(M,a)}(s_k, a_k), \tag{1.12}$$

and random transition as opposed to conventional MDP transition illustrated in Figure 1.8:

$$s_{k+1} = S^{(M,W)}(s_k^{a_k}, W_{k+1}). \tag{1.13}$$

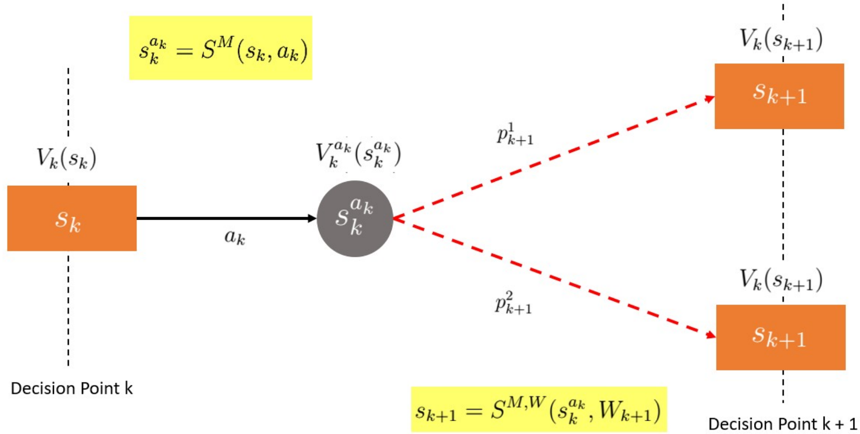


Figure 1.11: PDS in MDP Structure

Section 1.2.3.1 presented how the curse of dimensionality in MDP model leads to prohibitive computation of values according to a standard stochastic Bellman Equation (Equation (1.9)) due to the computation of expectation term:

$$\mathbb{E}\{V_{k+1}(s')|s_k\} = \sum_{s' \in S} p(s'|s_k, a_k) V_{k+1}(s'), \tag{1.14}$$

where $s_{k+1} = s'$. The computation in Equation (1.14) could be avoided thereby reducing the impact of outcome space as one of the curse of dimensionality.

The evolving Bellman's value function due to the PDS avoids computation of the expectation term. This is based on the definition of the value of the PDS which is also the expected value of the next state observed at $k+1$ decision epoch as shown in Equation (1.15):

$$V_k^{a_k}(s_k^{a_k}) = \mathbb{E}\{V_{k+1}(s_{k+1})|s_k^{a_k}\}, \tag{1.15}$$

which is further explained in Figure 1.12.

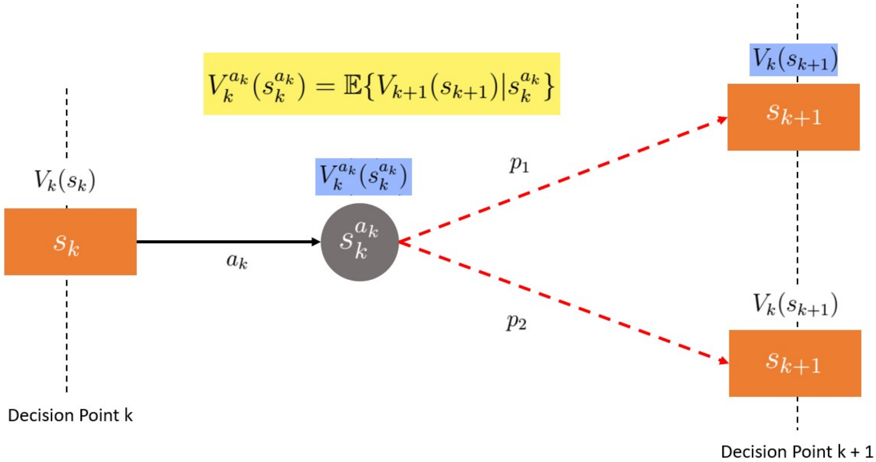


Figure 1.12: Computing Value of PDS

By substituting Equation (1.15) into the standard stochastic Bellman Equation (1.9) the expectation of next value at decision epoch $k+1$ is then replaced by the value of resulting PDS $s_k^{a_k}$ after making decision a_k based on the observation s_k made at k. This is seen in Equation (1.16):

$$V_k^*(s_k) = \max_{a_k \in A(s_k)} (R(s_k, a_k) + \gamma V_k^{a_k}(s_k^{a_k})). \tag{1.16}$$

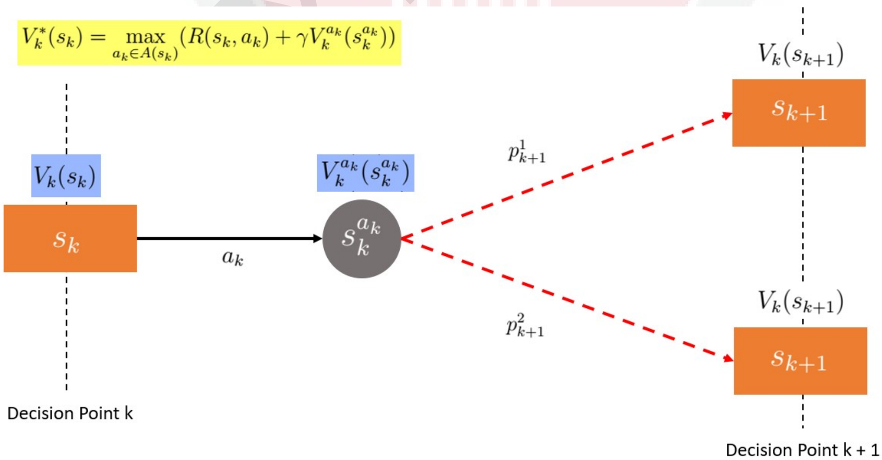


Figure 1.13: Value Computation Based on Value of PDS

In theory, the computation of the resulting PDS value $s_k^{a_k}$ also leads to the expectation computation as can be seen in Equation (1.15). However, this is only true for the conventional approach of solving MDP where all values need to be computed exactly. If the value of PDS $V_k^{a_k}(s_k^{a_k})$ could be approximated $\overline{V_k^{a_k}}(s_k^{a_k})$ instead of applying Equation (1.15), the computation burden in computing the expectation term could be avoided in both Equation (1.16) (as shown in Equation (1.17)) and Equation (1.15).

$$V_k^*(s_k) = \max_{a_k \in A(s_k)} (R(s_k, a_k) + \gamma \overline{V_k^{a_k}}(s_k^{a_k})). \tag{1.17}$$

Although one might argue that the value of the next state $V_{k+1}(s_{k+1})$ could also be approximated as Equation 1.18 below:

$$V_k^*(s_k) = \max_{a_k \in A(s_k)} (R(s_k, a_k) + \gamma \mathbb{E}\{\overline{V}_{k+1}(s_{k+1})|s_k\}), \tag{1.18}$$

one would still have to deal with computing the expectation even though the values computed is now approximated. As such, the computation involving the outcome space is not avoided.

Following the same line of approach, it is also now possible through the introduction of PDS structure to compute for optimal action $a_k^\star$ guiding towards optimal decision rule $\rho^{\pi^\star} : s_k \rightarrow a_k^\star$ based on optimal policy applied:

$$a_k^* = arg \max_{a_k \in A(s_k)} (R(s_k, a_k) + \gamma V_k^{a_k}(s_k^{a_k})), \tag{1.19}$$

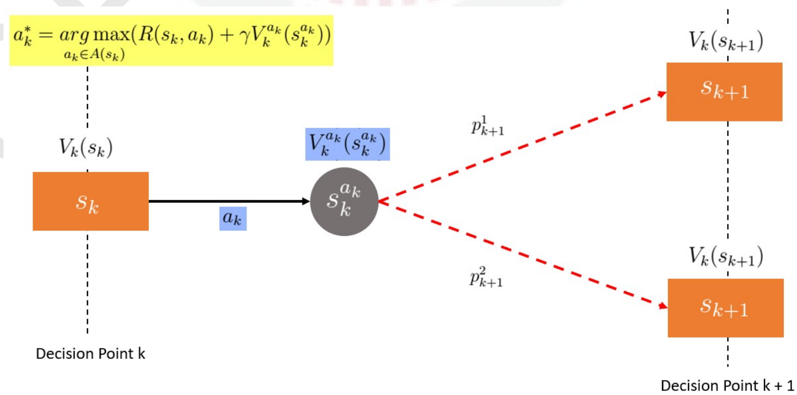


Figure 1.14: Optimal Value Computation through PDS

as opposed to the equation below (Figure 1.9):

$$a_k^\star = \underset{a_k \in A(s_k)}{\arg\max} \left(R(s_k, a_k) + \mathbb{E}\{\gamma^k V_{k+1}(s_{k+1})|s_k)\}\right). \tag{1.20}$$

Although the PDS structure provide a way to avoid computing the expectation in Bellman Equation thus relieving some of the computation burden in solving a complex MDP, one would still need to compute the value of PDS which therein lies another expectation term as is shown in Equation (1.15). In this case, ADP plays an important role of approximating the value of PDS rather than computing the value exactly through numerous method under the umbrella of ADP.

This research looks into the application of Rollout Algorithm or specifically the PDS - Rollout Algorithm (PDS-RA) as proposed by Goodson (2010) which is one of the lookahead approach in ADP.

1.2.4 Approximate Dynamic Programming

The ADP compliment the dynamic programming as a mean to deal with the curse of dimensionality that prohibit the computation of the Bellman Equation in order to solve the MDP problem. The sole idea of ADP are:

1. to avoid computing the value in Bellman Equation exactly thus avoiding oneself to deal with the exponentially vast state, action and outcome space (the three curses of dimensionality).
2. to serve as an algorithmic strategy for model free MDP.

These are achieved by moving/ stepping forward in time based on the following requirement of ADP (Powell, 2007):

- Ability to generate random samples/ episode/ trajectories:
  - Generating samples/ episodes by stepping forward in time.
  - Based on the samples, agent must learn the unknown transition model.
  - In generating episodes, random information is needed to allow for state transitions.
  - In ADP, this random information is sampled at random.
  - This is known as Monte Carlo simulation/ sampling.
- A way to make a decision:
  - Stepping forward in time involves making decision/ action as to allow forward transition from the current state onwards.

- Agent must at least has an initial rough plan on which action to take or decision to make as to create a complete trajectory (until the end state).
- In classical approach, decision is guided by value of states computed exactly by the agent.
- In ADP, these values could be approximated for obvious reason that exact computation is not possible.
- In advance and expensive method, agent is trained to recognize which state is most advantageous to move forward in time Silver et al. (2017).

By resorting to approximating values (hence the term ADP), imperfect solution is to be expected meaning that the solution would not be as good as an exact solution. However, the trade off would be the ability to solve for large and complex MDP problem. This does not mean that the approximation could be done half heartedly. In fact, this is the challenge of ADP: to approximate the value as to produce good and practical solution to the problem at hand.

The basic algorithm of ADP is given in Powell (2007):

Step 0. Initialization.

Step 0a. Initialize $\overline{V}_t^0(S_t)$ for all states S_t.

Step 0b. Choose an initial state S_0^1.

Step 0c. Set $n = 1$.

Step 1. Choose a sample path ω^n.

Step 2. For $t = 0, 1, 2, \ldots, T$ do:

Step 2a. Solve

$$\hat{v}_t^n = \max_{a_t \in \mathcal{A}_t^n} \left(C_t(S_t^n, a_t) + \gamma \sum_{s' \in \mathcal{S}} \mathbb{P}(s'|S_t^n, a_t)\overline{V}_{t+1}^{n-1}(s') \right),$$

and let a_t^n be the value of a_t that solves the maximization problem.

Step 2b. Update $\overline{V}_t^{n-1}(S_t)$ using

$$\overline{V}_t^n(S_t) = \begin{cases} \hat{v}_t^n, & S_t = S_t^n, \\ \overline{V}_t^{n-1}(S_t), & \text{otherwise.} \end{cases}$$

Step 2c. Compute $S_{t+1}^n = S^M(S_t^n, a_t^n, W_{t+1}(\omega^n))$.

Step 3. Let $n = n+1$. If $n<N$, go to step 1.

Figure 1.15: Basic ADP Algorithm
[Source: Powell, 2007]

where the similarity of the algorithm to backward dynamic programming is noted. However in this basic ADP algorithm, the computation of value is done by stepping forward in time in Step 2c (Figure 1.15). The forward transition is done through Monte Carlo sampling by randomly choosing one random information at each transition. When moving forward from the current state, an agent is said to look into the future as far as T decision point. When T is reached, one complete trajectory or one Monte Carlo simulation is said to be performed. In Figure 1.15, N Monte Carlo simulation is done to approximate the values. Prior to each trajectory or episode n, a sample path ω^n is pre-sampled in step 1 for all decision point t until T in the trajectory.

More importantly, the computation of values at Step 2a is done by the approximated value of the next state $V_{t+1}^{n-1}(s')$. It is important to note that the algorithm require only computation of state value that is transitioned as opposed to computing all state value in the state space. This substantially reduce the computation burden that is seen in the classical approach of solving MDP although it may presents the exploitation vs exploration issue. Step 2a also requires computation of expectation that maintain the burden in dealing with the outcome space. Additionally the curse of action space is also not avoided. All of these limitations is noted as the basic algorithm only served as to give an overview concept of ADP. In a more sophisticated approach of ADP, more improvements could be seen in terms of mechanism proposed in dealing with the curses of dimensionality. The output of the algorithm in Figure 1.15 is a lookup table that features computed values involve in the sampled episodes. The state which is not visited during the sampled episodes retained their initial value at Step 0a. A more generic ADP algorithm is depicted below in Figure 1.16:

Step 0. Given an initial state S_0^1 and value function approximations $\overline{V}_t^0(S_t)$ for all S_t and t, set $n = 1$.

Step 1. Choose a sample path ω^n.

Step 2. For $t = 0, 1, 2, \ldots, T$ do:

  Step 2a. Optimization: Compute a decision $a_t^n = A_t^\pi(S_t)$ and find the post-decision state $S_t^{a,n} = S^{M,a}(S_t^n, a_t^n)$.

  Step 2b. Simulation: Find the next pre-decision state using $S_t^n = S^M(S_t^n, a_t^n, W_{t+1}(\omega^n))$.

Step 3. Update the value function approximation to obtain $\overline{V}_t^n(S_t)$ for all t.

Step 4. If we have not met our stopping rule, let $n = n+1$ and go to step 1.

Figure 1.16: Generic ADP Algorithm
[Source: Powell, 2007]

According to Powell (2007) there are three strategies that can be used to approximate the values:

1. Lookup table approach.

2. Parametric models of value.

3. Non-parametric models of value.

This is closely related to the class of broad ADP solution approaches (Powell, 2007):

- Lookahead Approach.
- Policy Function Approximation (PFA).
- Value Function Approximation (VFA).

However in Ulmer (2017), another type of VFA is considered as a new class of solution approach:

- Lookahead Approach.
- Value Function Approximation (VFA).
- Approximate Value Iteration (AVI).

Lookahead approach such as rollout algorithm is often applied in the dynamic and stochastic MDP. The lookahead approach emphasized on the dynamic nature of the problem as the rolling horizon could be done at each decision epoch but at the cost of computation which normally needs some form of decomposition (Goodson, 2010).

On the other hand Value Function Approximation (VFA) are more common in solving complex MDP through a parametric model which is usually trained by a supervised learning as can be seen in Silver et al. (2017). The concept of VFA is to approximate the true value based on multiple basis functions based on multiple selected features or attributes important to a state. Here the attributes are akin to what is observed (normally important parameters) by the agents. A basis function is then define as $\phi_f(s_k)$. Each different features $f \in \mathscr{F}$ forms different basis functions. This features on the other hands are weighted with θ_f where the approximation of value of a state at decision epoch k can then be solved through:

$$\overline{V(s_k)} = \sum_{f \in \mathscr{F}} \theta_f \phi_f(s_k). \tag{1.21}$$

An accurate approximation might vary with the complexity of the problem. A linear regression might need less training as compared to neural network and could produce satisfying results (Ulmer, 2017). However both approaches need to be trained offline which is not desirable when one tries to execute dynamic/ online algorithm.

In the same manner, the simplest example of Policy Function Approximation could be seen in Silver et al. (2017) where effective approximation could only be done with the state of the arts computing equipment involving countless training for the complex problem it is solving.

On the other hand, Approximate Value Iteration (AVI) make used of Monte Carlo simulation in order to generate random episodes. Once this is done, the value is updated via incremental means based on the learning rate/ step size defined by the user for the agent. The approach are similar to the general ADP algorithm presented in Figure 1.16.

1.2.4.1 Lookahead Approach - Rollout Algorithm

In relation to earlier Section 1.2.4, the lookahead approach is adopted for this research considering its suitability in terms of the requirement of live simulation as well as limitation of the hardware. Such approach allows for flexibility independent of any types of dynamic complex problem. This means that no prior training is needed to deploy the algorithm. Such is not the case with PFA or VFA that opted for parametric model approach which is normally trained offline explicitly according to the problem at hand (Silver et al., 2017). In other words, the lookahead approach could be considered as an online ADP approach as opposed to PFA and VFA which are offline ADP.

The idea of lookahead approach is to approximate the value of the state or PDS, from which the lookahead starts until the end of the lookahead horizon. Ideally the limited lookahead horizon could be expanded until the end state. However, that is normally not the case in the usual practice as the computation involved is considerably high due to the details of the problem presented in the limited time frame (Ulmer, 2017). Furthermore, decision needs to be made fast, hence computation needs to take into account how long an agent is allowed to think (compute by considering future simulations or lookahead) before making decision at each decision epoch. In the case of Alphago zero (Silver et al., 2017) for example, the agent cannot think too long before making a move in a Go game.

An extreme version of lookahead would be the tree search that considers all possible actions and all possible outcomes (Powell, 2007). When all possible outcomes are reduced to Monte Carlo sampling during the computation, the sparse sampling tree search leads to a more manageable computation. The concept is then further evolved into Rollout Algorithm which limits the tree search based on one decision and one

Monte Carlo sampling of random information per each decision epoch following a certain base policy and decision rule until the lookahead horizon ended.

The rollout is first presented in Bertsekas et al. (1997) and later applied in scheduling problem in Bertsekas and Castanon (1999). From then on further variants of rollout algorithm were introduced by Goodson (2010). Among them is the PDS - Rollout (PDS-RA) which is the solution approach applied in this research.

Conventional Rollout

The aim of rollout algorithm is to identify the optimal action or decision $a_k^\star$ at decision epoch k based on the return value $\overline{V}_{k+1}(s_{k+1})$ approximated by the rollout return $B(\pi_{\mathscr{B}(s_{k+1})}, k+1)$ where $\pi_{\mathscr{B}(s_{k+1})}$ is a certain base policy applied at a certain state s_{k+1} at decision epoch $k+1$. The standard solution to Bellman in determining optimal action $a_k^\star$ in Equation (1.10) is approximated through the approximated next state value:

$$a_k^\star = \underset{a \in A(s_k)}{\arg\max} (R(s_k, a_k) + \lambda^k \mathbb{E}\{\overline{V}_{k+1}(s_{k+1})\}), \tag{1.22}$$

where $\overline{V}_{k+1}(s_{k+1})$ is approximated through rollout algorithm as the average of return of $B(\pi_{\mathscr{B}(s_{k+1})}, k+1, j)$ from decision epoch $k+1$ towards j following the decision rule $\rho^{\pi_{\mathscr{B}(s_{k+1})}}$ based on policy $\pi_{\mathscr{B}(s_{k+1})}$ that is to apply base heuristic $\mathscr{B}(s_{k+1})$:

$$B(\pi_{\mathscr{B}(s_{k+1})}, k+1, j) = \mathbb{E}^{\pi_{\mathscr{B}(s_{k+1})}}\{\sum_{i=k+1}^{j-1} \lambda^{i-k+1} R(s_i, A^{\pi_{\mathscr{B}(s_{k+1})}}(s_i))|s_{k+1}\}. \tag{1.23}$$

In this research the rollout is applied until the decision epoch j where $j = K$. Therefore the return is computed as:

$$B(\pi_{\mathscr{B}(s_{k+1})}, k+1) = \mathbb{E}^{\pi_{\mathscr{B}(s_{k+1})}}\{\sum_{i=k+1}^{K-1} \lambda^{i-k+1} R(s_i, A^{\pi_{\mathscr{B}(s_{k+1})}}(s_i))|s_{k+1}\}, \tag{1.24}$$

leaving out the j parameter in the base heuristic notation.

The rollout is executed over many Monte Carlo simulations such that different trajectories of lookahead episodes are generated. Eventually, the average return of the lookaheads performed is computed as the approximated value of the next state

$\overline{V}_{k+1}(s_{k+1})$ where, the optimal decision can be computed based on Equation (1.22).

The act of approximating value through the conventional rollout algorithm leads to a more tractable computation. However, the term expectation in Equation (1.22) still introduces extra computation based on the outcome space of the model.

By taking advantage of the PDS structure in MDP, one could further reduce the computational burden by approximating the value of respective PDS rather than the next state value via PDS-RA.

Post Decision State - Rollout Algorithm (PDS-RA)

From the previous section it can be seen that by separating the transition into deterministic (Equation (1.12)) and stochastic transition (Equation (1.13) with the value of PDS solved through Equation (1.15), one would be able to avoid computing expectation term in the standard Bellman equation by replacing the value of PDS instead of the expectations of value of the next state as in Equation (1.19).

Similarly, through ADP the exact computation PDS value is avoided by approximating it in Equation (1.25):

$$a_k^\star = \underset{a_k \in A(s_k)}{\arg\max} (R(s_k, a_k) + \lambda^k \overline{V_k^{a_k}}(s_k^{a_k})), \tag{1.25}$$

with the averaged rollout return after N number of Monte Carlo Simulation for each PDS following this basic pseudo Algorithm 1.1 (Goodson, 2010):

Algorithm 1.1 Compute $\overline{V_k^a}(s_k^a)$

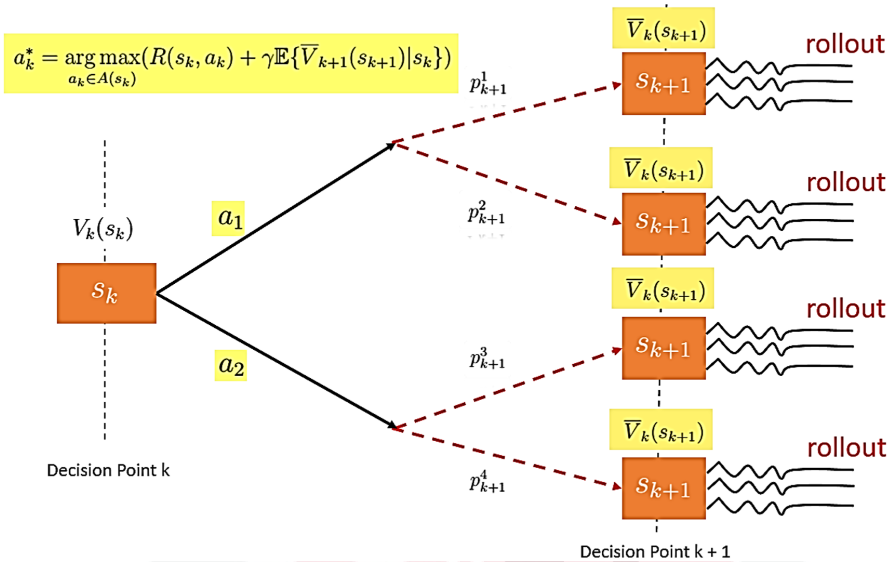
Input: s_k, λ, a_k
Output: $\overline{V_k^a}(s_k^a)$
1: Initialize $n, k, R(s_k, a_k), B^n$
2: $s_k^{a_k} = S^{M,a}(s_k, a_k)$
3: $\pi_{\mathscr{B}(s_k^a)} \leftarrow \mathscr{B}(s^a)$
4: **while** $n \leq N$ **do**
5: $\quad s^a \leftarrow s_k^a$
6: $\quad$ **while** $k \neq K$ **do**
7: $\quad\quad R(s_k, a_k) = R(s_k, a_k) + \lambda^k R(s_k, a_k)$
8: $\quad\quad s_k = S^{M,W}(s^a, W_{k+1}(\omega(k+1)))$
9: $\quad\quad a_k \leftarrow A^{\pi_{\mathscr{B}(s_k^a)}}(s_k)$
10: $\quad\quad s^a = S^{M,a}(s_k, a_k)$
11: $\quad\quad k = k+1$
12: $\quad$ **end while**
13: $\quad \widehat{B}^n(\pi_{\mathscr{B}(s_k^a)}, k+1) \leftarrow R(s_k, a_k)$
14: $\quad \overline{V_k^a}^n(s_k^a) = \overline{V_k^a}^{n-1}(s_k^a) + \frac{1}{n}\left(\widehat{B}^n(\pi_{\mathscr{B}(s_k^a)}, k+1) - \overline{V_k^a}^{n-1}(s_k^a)\right)$
15: $\quad n = n+1$
16: **end while**
17: return $\overline{V_k^a}^N(s_k^a)$

Similar to the conventional rollout, an expectation term could be seen in the Equation (1.26) below when computing the return during rollout phase:

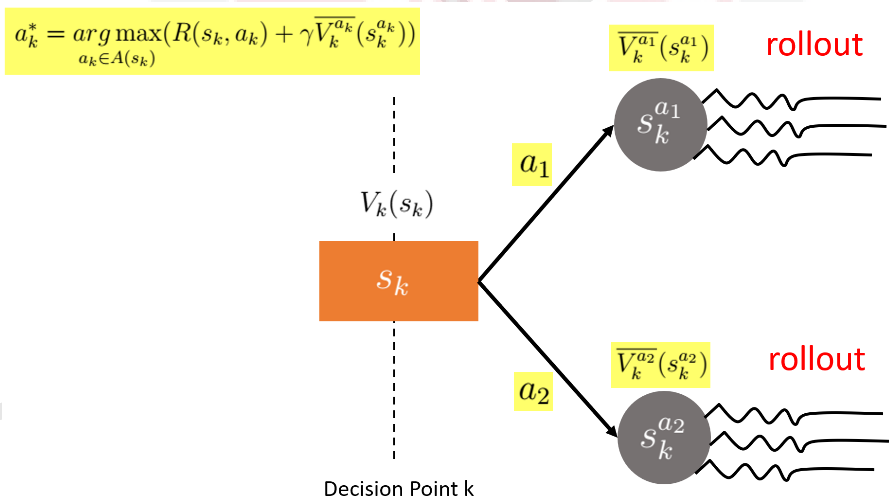
$$B(\pi_{\mathscr{B}(s_k^{a_k})}, k+1) = \mathbb{E}^{\pi_{\mathscr{B}(s_k^{a_k})}}\{\sum_{i=k+1}^{K-1} \lambda^{i-k+1} R(s_i, a_i | s_{k+1}) | s_{k+1}\}, \tag{1.26}$$

However, this is only true when exact computation of the return is done by taking into consideration the large outcome space that exist in the rollout transitions. In the lookahead approach, the common approach to execute the rollout is instead to sample the stochastic variable $p_{\hat{k+1}} = W_{k+1}(\omega(k+1))$ at each decision epoch $k+1$ as to allow full deterministic (through the base heuristic resulting decision rule) and stochastic (via Monte Carlo sampling) transition to generate an episode (from when the rollout is invoked to the end state s_K). This could be seen in line 8 of the Algorithm 1.1. Hence the expectation term (outcome space) is avoided when computing the Bellman Equation.

In short, the number of rollout one would have to performed is reduced when PDS-RA is applied as opposed to the conventional rollout algorithm. This is further illustrated in Figure 1.17. Meanwhile, a graphical progressions of PDS-RA are shown in Appendix A.



(a) Conventional RA



(b) PDS-RA

Figure 1.17: Comparison between PDS-RA (b) and Conventional Rollout (a)

1.3 Scope of the Research

This research can be divided into several research scopes depending on the field of which they are addressing. In this case, the research addressed simultaneously the field of Operations Research (OR), Disaster Management (DM), as well as ML. In the field of OR, the research looks into the field of humanitarian operation, a subset of Humanitarian Logistics (HL), specifically the medical supply delivery. The case of 2015 Nepal Earthquake is taken as a motivating example where the rich model MDDVRPSRC is considered and developed. The difficulties of delivering medical supply to many temporary shelters and temporary depots as an alternative to elevate the bottleneck at Kathmandu Airport is especially addressed in this research. Furthermore, the route towards these shelters could be obstructed due to the damage sustained during the Earthquake. In addressing this, a stochastic road capacity problem with progressive decrease (dynamic) of road capacity mean in the road capacity distribution depending on the tremor of the earthquakes is proposed. To this end, a simulation of earthquake tremor is also developed to model the tremor radially dispersed where the interceptions of the roads (edges) and the radial tremor lines could be translated as the damage unit sustained on the respective roads. Since both time and cost are of importance in this humanitarian problem, a multi-objectives problem is thus considered. Moreover, split deliveries and multi trip operation are also depicted in this model to ease the delivery attempt to shelters considering the limited transportation available as well as the uncertainties that plagued the delivery operation.

The DM consists of four phases namely, the preparedness, mitigation, response, and recovery phase are addressed in the research. Here the scope is narrowed down into the preparedness and response phase of DM. In the response phase, the medical supply is delivered efficiently through the proposed model and solution taking into the account of all the aforementioned factors in the previous paragraphs. Furthermore, this research also indirectly addressed the preparedness whereby the MDDVRPSRC DSS could be the tool for decision makers in generating simulation for many possible instances and prepare the respective routing plan in advance prior to any hazardous disaster.

Among the key components of MDDVRPSRC DSS is the agent interacting with the user and taking part in the simulation by making decision for the medical delivery operation in MDDVRPSRC based on the ML approach. In this case, the ML field is narrowed further into RL where agents is allowed to learn in making decision based on rewards instead of training the agent through supervised or unsupervised learning. Furthermore, since the model is complex which incurred the curse of dimensionality of ML, the solution approach is further specified by computing decision based on a lookahead policy. This lookahead policy leads to an approximate optimal decision for each state the agent is in, in real time by performing simulation of possible future episodes based on Monte Carlo sampling instead of exactly solving the problem through a dynamic programming. Through the ADP solution approach, a one step approximate policy iteration or RA is performed at each Post Decision State (PDS) (known as PDS-RA algorithm) to compute an approximated value of a PDS associated to a potential decision. This involves performing lookahead into the future scenario where hypothetical decision is determined by the presented base heuristics which are compared with the novel matheuristic rollout. Finally, by weighing the computed values of PDSs associated with each potential decision, the near optimal decision is selected and applied by the agent when making real time decision.

In Figure 1.18, the specified scope of different fields covered by this research is shown.

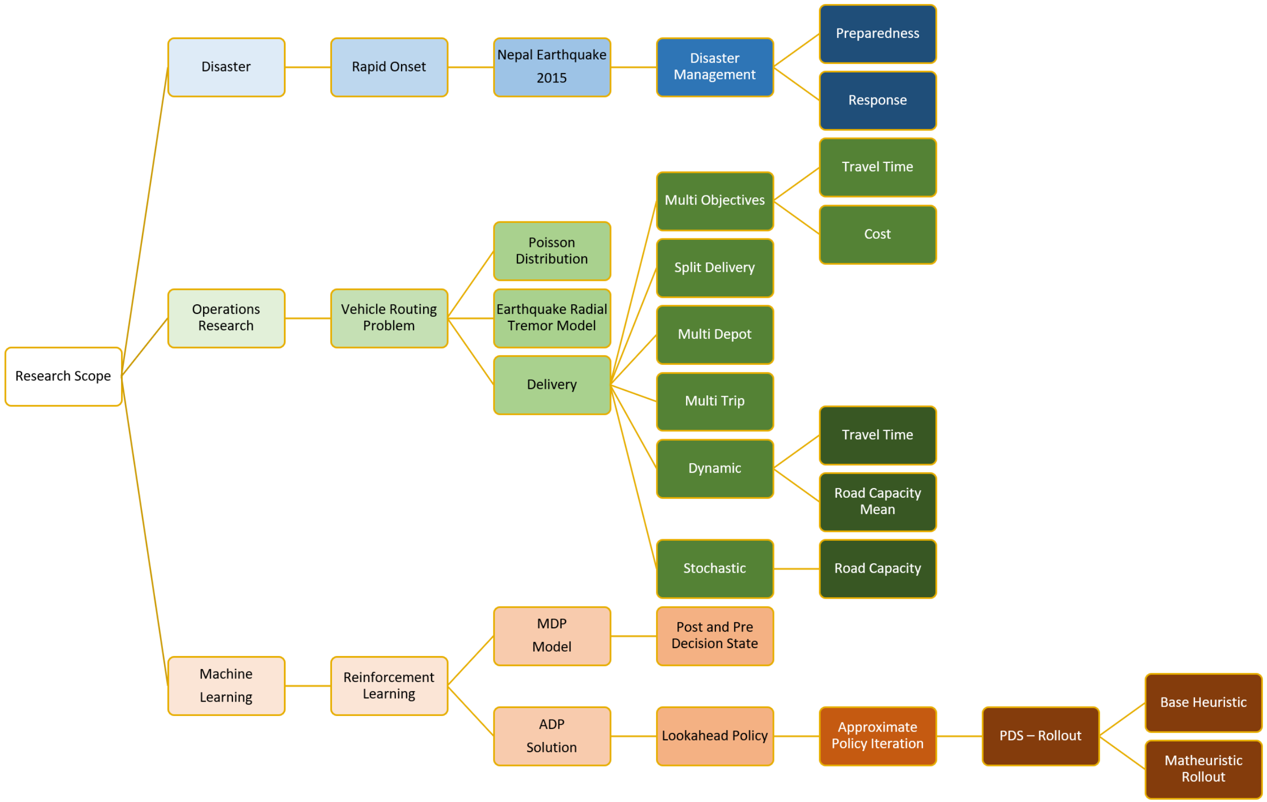


Figure 1.18: Research Scope

1.4 Problem Statements

Based on the scope of this research, the problem statement could be formulated as the followings:

1. *The vast majority of literatures that address the problem of VRP based on the approach of RL try to alleviate the curses of dimensionality through the extended framework of MDP model and ADP solution approach. The richer and more complex problem of VRP in humanitarian operations however is undoubtedly more difficult to solve which stresses further on the computation limitation at hand. Such research is therefore needed to fill the gap of VRP from the perspective of RL that address not only the application side of humanitarian operation instead of commercial problem, but also the computational challenge that comes with it.*

2. *Multiple model driven based DSS have been developed to aid with various humanitarian operation involving VRP that allow for simulation and optimization. However an intelligent DSS that react to simulated environment of such problem is still lacking. If an intelligent agent can defeat a human champion in the game of GO, it is logical for one to harness the same potential in developing an intelligent DSS that could react to the simulated environment and perform the humanitarian operation such that the results could be learned for future planning purposes.*

3. *Compared to the ever expanding solution and modelling methodology of VRP in regards to both commercial and humanitarian operations, the test dataset and benchmark development for complex and rich VRP such as those that deals with the stochastic road capacity problem is still far lacking. Most are rather outdated begging the question of their suitability in addressing modern VRPs that could also leads to the development of their respective DSS. As such, a suitable test dataset and benchmark method need to not only account for complex problem it is addressing, but also holistically flexible in ways that it could be used for the application of a DSS.*

1.5 Research Questions

Deriving from problem statements that have been formulated in Section 1.4, the research questions are listed as follows:

1. *What is the best way to optimize the medical supply delivery in a chaotic and dynamic environment that addressed the issue of stochastic road capacity and road damage in relation to the radial tremor dispersing from the disaster points of origin ?*

2. *How to model the road capacity distributions in relation to the disaster point*

of origin and the damage that the roads sustained based on their respective effects towards the operation in a realistic manner ?

3. *How to apply the ADP solution approach into a stochastic and dynamic VRP such that the curse of dimensionality is reduced and split delivery is allowed as opposed to common practices ?*

4. *How to Ireduced Monte Carlo simulations and lookahead horizon due to hardware limitation and yet able to compute good decision ?*

5. *How to perform lookahead in rollout algorithm with stochastic road capacity that render previous computed routes invalid ?*

6. *Are there any suitable test dataset and benchmark methods or solutions that can be compared with the proposed solutions ?*

7. *How to monitor the progression of the delivery operations through graphical version of the problem through simulation in order to validate the model and investigate the performance of the solutions proposed through sufficient numbers of simulation ?*

8. *Can numerous simulation be performed through a suitable interactive DSS ?*

1.6 Research Gaps

Based on these research gaps, the research questions and objectives are formed that shaped the research. These gaps could be grouped in terms of modelling aspects and solution aspects as listed below:

1. To date, very few researches address the VRP with regards to stochastic road capacity within the road network. Thus a suitable test dataset to investigate and validate this type of problem is lacking.

2. Existing VRPs that are addressed by MDP are typically less comprehensive such that most are covering problems involving only a single vehicle.

3. Models based on MDP are scarcely applied when describing rich VRP problem (RVRP) as compared to mathematical programming model such as Mixed Integer Linear Programming (MILP), Integer Linear Programming (ILP) and Stochastic Programming (SP) models.

4. Only few researches could be seen employing ML solution approach as compared to exact solution and metaheuristic solution approaches when solving the RVRP.

5. Only a number of RVRP are modelled through the structure of Pre Decision State (PRE) and Post Decision State (PDS) when modelled based on MDP model.

6. Among the researches that solved the stochastic or dynamic VRP through the ML solution approach, most applied clustering approach as a method in dealing with the curse of dimensionality of ML. To date, there is no alternative that try to approach such problem without performing clustering prior to computing the route for the case of multi vehicle operation.

7. A very limited number of RVRP are solved based on lookahead solution approach within the branch of ADP.

8. Very few within the field addressed the different base heuristics in the application of rollout algorithm.

9. It is yet to be seen the application of exact solution within the rollout as part of the dynamic base policy.

10. Only a handful of researches describe deteriorating facilities in their model when addressing VRPs in the application of humanitarian operation.

11. Among those, very few actually model the source of disruption explicitly such as the progression tremor of earthquake that leads to continuous deterioration of critical facilities such as roads and highways.

These research gaps are elaborated more in Chapter 2.

1.7 Research Objectives

This research could be separated into **four main objectives** along with their sub-objectives:

Objective 1 : to identify the rich model components of the MDDVRPSRC and to apply them in the preliminary investigation and comparison of the deterministic and stochastic version of the problem.

1.1 to identify the rich MDDVRPSRC model components.

1.2 to model the stochastic road capacity distribution that can be accessed by the MDDVRPSRC model and the 2-stage stochastic version of the model.

1.3 to model the radial tremor dispersion of an earthquake or any other disaster from an origin point.

Objective 2 : to enable agent to learn through RL through the development of the MDP model and to deal with the 3 curses of dimensionality based on the Extended standard framework of MDP.

2.1 to develop an MDP MDDVRPSRC model through the extended MDP structure that reduces the curse of dimensionality.

Objective 3 : **to solve the MDDVRPSRC MDP model based on the solution approach that alleviates the 3 curses of dimensionality which also accounts for road capacity problem and the limited computation power available and to develop the corresponding benchmarking method to the solution proposed.**

3.1 to develop an ADP solution based on the lookahead approach such that the curse of dimensionality is addressed.

3.2 to introduce a computing mechanism that allows for good decision computation with less Monte Carlo simulation and limited lookahead horizon.

3.3 to develop PDS Rollout base heuristics based on the dynamic changes of the problem and stochastic road capacity that render fixed route heuristic impractical.

3.4 to develop a new benchmark solution method using iterative CPLEX through a novel matheuristic rollout approach suitable for benchmarking in filling in the gap of MDDVRPSRC.

Objective 4 : **to develop an MDDVRPSRC DSS with suitable features for monitoring and analysis.**

4.1 to develop a DSS that could verify the proposed models and solutions based on a simulated test dataset.

4.2 to develop a DSS with an online monitoring visualization as a platform for DSS.

4.3 to develop a test dataset where simulated data could be compiled and comparative analysis could be performed between proposed base heuristics and the benchmark solution method.

Additionally the research objectives could also be organized based on the three core components of the research listed as follows:

- Modelling
  1. Model components identification.
  2. Model development.
- Solution
  1. Solve the specified MDDVRPSRC.
  2. Comparing performance of proposed solutions.
- Decision Support System
  1. MDDVRPSRC DSS development.
  2. Simulation and verification.

These research objectives based on the core components of the research are illustrated in Figure 1.19, Figure 1.20 and Figure 1.21 respectively.

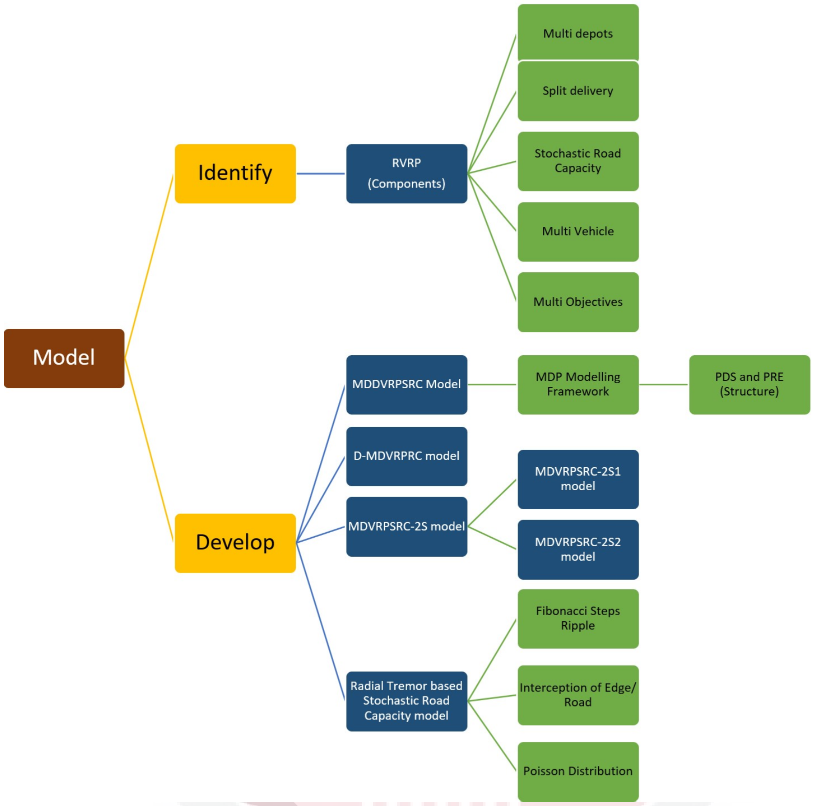


Figure 1.19: Research Objectives on Modelling Aspect

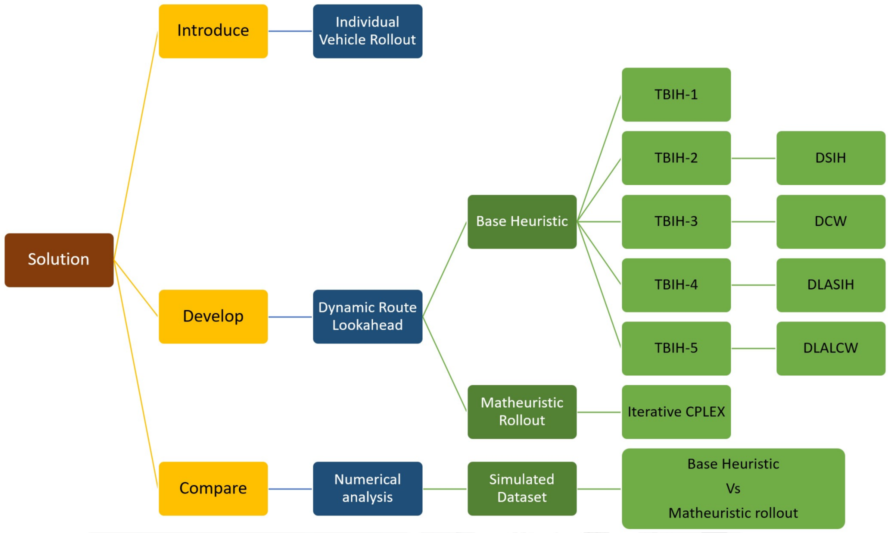


Figure 1.20: Research Objectives on Solution Aspect

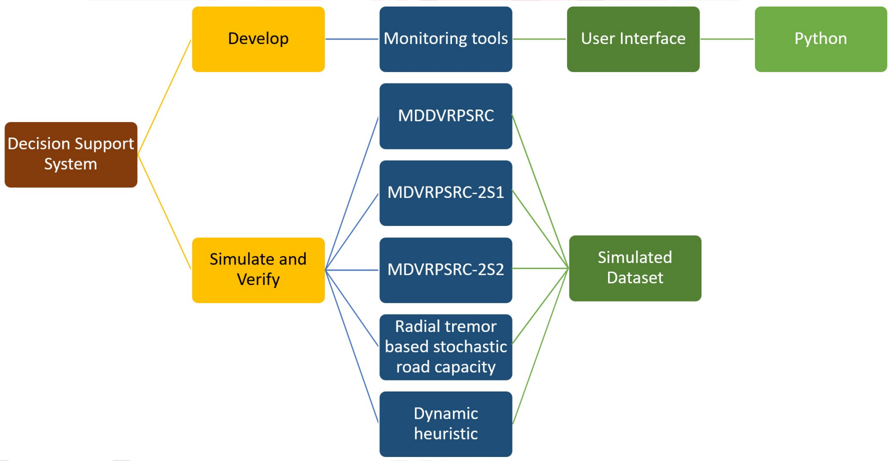


Figure 1.21: Research Objectives on Decision Support System Aspect

1.8 Research Limitations

Through this research, several limitations are encountered. Thus, the research objectives are achieved subjected to these limitations:

1. Hardware limitation
    - This research is conducted using a laptop computer running on Intel^R Core^{TM} i7-7500U CPU at 2.70–2.90 GHz with 20 GB RAM. Although the RAM is upgraded to a sufficient level and the processors are running on an adequate speed, the physical core of dual processors is limiting the numbers of Monte Carlo simulations performed in the rollout algorithm.
    - Due to the core duo processors available, no parallel (parallel threading) Monte Carlo simulations could be performed that would potentially improve the computation time.
    - Furthermore, the matheuristic rollout proposed are only limited to a certain numbers of nodes despite the reduced models proposed.
    - Additionally, the number of vehicles performed by the matheuristic rollout is also limited in a certain test instance for benchmarking due to computing time that is no longer reasonable. Such case is deemed to be prohibitive as the computing time requires for the proposed base heuristics is clearly far superior.
2. Data limitation
    - The test dataset that are simulated to emulate the scenario of 2015 Nepal earthquake in terms of the location and placement of the nodes in the road network. The real data during the event are difficult to obtained due to the sensitivity of the data.

1.9 Research Contributions

The contributions stemmed from this research could be viewed from four key aspects of the research as depicted in Figure 1.22.

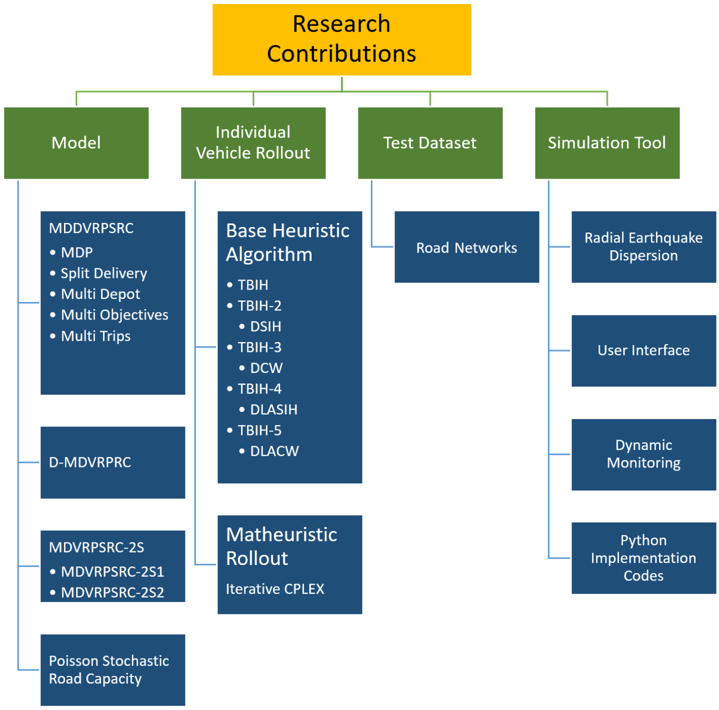


Figure 1.22: Research Contributions based on Research Components

From this research four models are developed and proposed prior to developing MD-DVRPSRC as a preliminary study:

1. D-MDVRPRC model for the case of deterministic road capacity.
2. Stochastic Integer Linear Programming (SILP) MDVRPSRC-2S model which derived further two reduced SILP models through which iterative decision rule is computed during rollout lookahead:
   (a) MDVRPSRC-2S1 SILP model.
   (b) MDVRPSRC-2S2 SILP model.
3. MDDVRPSRC MDP model.

Besides the MDVRPSRC-2S1 and MDVRPSRC-2S2 models, the D-MDVRPRC, MDVRPSRC-2S, and MDVRPSRC-2S incorporate a multi trip, multi objectives, multi depots, and split delivery operation.

Apart from these models, a radial dispersion earthquake lines is created to mimic the earthquake tremor that helps in determining road damages and the respective deteriorating road capacity mean. Random road capacity is then sampled from road capacity distribution based on the Poisson distribution for each road in the network with the deteriorating road capacity mean as its parameter.

From the solutions of MDDVRPSRC perspective, few contributions are noted. First, the limitation of performing rollout with less Monte Carlo simulation is addressed by performing rollout for each vehicle individually instead of collectively as opposed to the conventional approach. Pivoting from this are the five presented base heuristics where TBIH-1 is first developed as the framework to differentiate the Teaching Part (TP) and the Seeking Part (SP) mechanism in determining a decision. Here SP consist of pure random selection of the next destination. Next, SIH (I1) is adopted in the SP mechanism forming TBIH-2 with embedded DSIH. The same is done by adopting CW forming TBIH-3 with embedded DCW. Deriving from these two each, are TBIH-4 with embedded DLASIH and TBIH-5 with embedded DLACW where the SP mechanism is improved by introducing lookahead in choosing promising nodes. Through these approaches it is shown how CW and SIH could be applied in the dynamic setting and could be modified further by introducing lookahead. To benchmark these proposed heuristic, new solution methodology is presented based on iterative CPLEX computation throughout every lookahead episode which can be coined as matheuristic rollout.

From the simulation tool perspective, a contribution is made by proposing a MDDVRPSRC DSS platform consist of agent observing and computing the decision during the simulation, user interface, as well visualization showing the progression of the delivery operation simulation, all coded in Python 2.7. Through this DSS, the simulated data is compiled for comparative analysis.

1.10 Outline of Thesis

This thesis is organized beginning with Chapter 1 where the foundation of the research is elaborated including the research scope, research questions, problem statements, research objectives, research limitations, as well as the research contributions. Much can be discussed for the vast field of VRP and the respective solution algorithms. The author however, choose to focus instead on the concept of RL as this is the main theory that underpins the core of research. Chapter 2 elaborates on the literature review performed covering the narrowed down research scope discussed earlier in Chapter 1.

The methodology of the research in regards to conventional models, MDP models and the ADP solutions proposed as well as the MDDVRPSRC DSS are presented in Chapter 3, 4, 5 and 6 respectively.

Chapter 3 presents the deterministic and SILP models along with the road capacity distribution model based on the Poisson distribution. Here, the computation of the deteriorating road capacity mean based on the damage sustained by the road is also presented. In Chapter 4, the MDDVRPSRC model based on MDP with extended framework of Post Decision State (PDS) is presented while the solution methodology is introduced in Chapter 5. In this chapter, the matheuristic rollout as well as the TBIH-1 and its variants are introduced in a solution methodology based on ADP through the rollout implementation for each individual vehicle when needed. The MDDVRPSRC DSS is presented next in Chapter 6 detailing the Python classes, methods and threads as the building blocks of the DSS.

Meanwhile the computational results based on computational experiments performed in regards to models presented in Chapter 3 and Chapter 4 using the methodology described in Chapter 5 are presented in Chapter 7. Here, the development of test dataset is explained in detail. Furthermore, the analysis for respective computational experiments are also discussed in this Chapter. Finally, the research is summarized and concluded in Chapter 8.

REFERENCES

Abdulkader, M. M. S., Gajpal, Y., and ElMekkawy, T. Y. (2018). Vehicle routing problem in omni-channel retailing distribution systems. *International Journal of Production Economics*, 196:43–55.

Adhikari, I. M., Pyakurel, U., and Dhamala, T. N. (2020). An integrated solution approach for the time minimization evacuation planning problem. *International Journal of Operations Research*, 17(1):27–39.

Afsar, H. M., Prins, C., and Santos, A. C. (2014). Exact and heuristic algorithms for solving the generalized vehicle routing problem with flexible fleet size. *International Transactions in Operational Research*, 21(1):153–175.

Ahmadi, G., Tavakkoli-Moghaddam, R., Baboli, A., and Najafi, M. (2020). A decision support model for robust allocation and routing of search and rescue resources after earthquake: a case study. *Operational Research*, 20:1–43.

Al Theeb, N. and Murray, C. (2017). Vehicle routing and resource distribution in postdisaster humanitarian relief operations. *International Transactions in Operational Research*, 24(6):1253–1284.

Alexander, D. E. (2002). *Principles of emergency planning and management*. Terra Publishing.

Alinaghian, M., Aghaie, M., and Sabbagh, M. S. (2019). A mathematical model for location of temporary relief centers and dynamic routing of aerial rescue vehicles. *Computers & Industrial Engineering*, 131:227–241.

Allahviranloo, M., Chow, J. Y., and Recker, W. W. (2014). Selective vehicle routing problems under uncertainty without recourse. *Transportation Research Part E: Logistics and Transportation Review*, 62:68–88.

Alpaydin, E. (2016). *Machine Learning: The New AI*. MIT Press.

Apte, A., Heath, S. K., Pico, A., Tan, R., and Hui, Y. (2015). Evacuating people with mobility-challenges in a short-notice disaster. *Decision Sciences*, 46(4):731–754.

Archetti, C., Savelsbergh, M. W., and Speranza, M. G. (2006). Worst-case analysis for split delivery vehicle routing problems. *Transportation science*, 40(2):226–234.

Aric A. Hagberg, D. A. S. and Swart, P. J. (2008). Exploring network structure, dynamics, and function using networkx. In Gäel Varoquaux, T. V. and Millman, J., editors, *Proceedings of the 7th Python in Science Conference (SciPy2008)*, pages 11–15, Pasadena, CA USA. cited as requested by author `https://networkx.org/documentation/networkx-1.10/reference/citing.html`.

Artha, Y. E. B., Santoso, S., and Liputra, D. T. (2020). Usulan penentuan rute distribusi dengan multi trip, multi product, dan time windows menggunakan algoritma nearest addition heuristic. In *Seminar National Teknik Industri Universitas*

Gadjah Mada 2020, pages 105–110. Depertemen Teknik Mesin dan Industri FT UGM.

Azimi, S., Delavar, M., and Rajabifard, A. (2018). An optimized multi agent-based modeling of smart rescue operation. *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 42(3/W4).

Baharmand, H., Comes, T., and Lauras, M. (2017). Managing in-country transportation risks in humanitarian supply chains by logistics service providers: Insights from the 2015 Nepal earthquake. *International Journal of Disaster Risk Reduction*, 24:549–559.

Baou, E., Koutras, V. P., Zeimpekis, V., and Minis, I. (2018). Emergency evacuation planning in natural disasters under diverse population and fleet characteristics. *Journal of Humanitarian Logistics and Supply Chain Management*, 8:447–476.

Beazley, D. (2010). Understanding the python gil. In *PyCON Python Conference. Atlanta, Georgia*.

Bell, W. J., Dalberto, L. M., Fisher, M. L., Greenfield, A. J., Jaikumar, R., Kedia, P., Mack, R. G., and Prutzman, P. J. (1983). Improving the distribution of industrial gases with an on-line computerized routing and scheduling optimizer. *Interfaces*, 13(6):4–23.

Bellman, R. (1957). A markovian decision process. *Journal of Mathematics and Mechanics*, 6(5):679–684.

Bellman, R. et al. (1954). The theory of dynamic programming. *Bulletin of the American Mathematical Society*, 60(6):503–515.

Bellman, R. and Lee, E. (1984). History and development of dynamic programming. *IEEE Control Systems Magazine*, 4(4):24–28.

Bertazzi, L., Bosco, A., Guerriero, F., and Lagana, D. (2013). A stochastic inventory routing problem with stock-out. *Transportation Research Part C: Emerging Technologies*, 27:89–107.

Bertazzi, L., Bosco, A., and Laganà, D. (2015). Managing stochastic demand in an inventory routing problem with transportation procurement. *Omega*, 56:112–121.

Bertazzi, L. and Secomandi, N. (2018). Faster rollout search for the vehicle routing problem with stochastic demands and restocking. *European Journal of Operational Research*, 270(2):487–497.

Bertsekas, D. P. and Castanon, D. A. (1999). Rollout algorithms for stochastic scheduling problems. *Journal of Heuristics*, 5(1):89–108.

Bertsekas, D. P., Tsitsiklis, J. N., and Wu, C. (1997). Rollout algorithms for combinatorial optimization. *Journal of Heuristics*, 3(3):245–262.

Braekers, K., Ramaekers, K., and Van Nieuwenhuyse, I. (2016). The vehicle routing problem: State of the art classification and review. *Computers & Industrial Engineering*, 99:300–313.

Brown, G. G., Ellis, C. J., Graves, G. W., and Ronen, D. (1987). Real-time, wide area dispatch of mobil tank trucks. *Interfaces*, 17(1):107–120.

Bruni, M., Beraldi, P., and Khodaparasti, S. (2018). A fast heuristic for routing in post-disaster humanitarian relief logistics. *Transportation Research Procedia*, 30:304–313.

Buzón-Cantera, I. E., Mora-Vargas, J., Ruiz, A., and Soriano, P. (2015). A hybrid optimization model: An approach for the humanitarian aid distribution problem. *Applied Mathematical Sciences*, 9(127):6329–6346.

Caceres-Cruz, J., Arias, P., Guimarans, D., Riera, D., and Juan, A. A. (2014). Rich vehicle routing problem: Survey. *ACM Computing Surveys (CSUR)*, 47(2):1–28.

Cai, R., Hu, D., and Jin, Z. (2019). Research on commodities distribution and location routing problem based on time deprivation in emergency logistics. In *CICTP 2019*, pages 4758–4770. ASCE.

Cao, C., Li, C., Yang, Q., Liu, Y., and Qu, T. (2018a). A novel multi-objective programming model of relief distribution for sustainable disaster supply chain in large-scale natural disasters. *Journal of Cleaner Production*, 174:1422–1435.

Cao, J., Han, H., Jiang, Y.-P., and Wang, Y.-J. (2018b). Emergency rescue vehicle dispatch planning using a hybrid algorithm. *International Journal of Information Technology & Decision Making*, 17(06):1865–1890.

Chauhan, A. and Chopra, B. K. (2017). Deployment of medical relief teams of the indian army in the aftermath of the Nepal earthquake: Lessons learned. *Disaster Medicine and Public Health Preparedness*, 11(3):394–398.

Chauhan, C., Gupta, R., and Pathak, K. (2012). Survey of methods of solving tsp along with its implementation using dynamic programming approach. *International Journal of Computer Applications*, 52(4):(0975–8887.

Chen, A. Y. and Chu, J. C. (2016). TDVRP and BIM integrated approach for in-building emergency rescue routing. *Journal of Computing in Civil Engineering*, 30(5):C4015003.

Chen, A. Y., Peña-Mora, F., and Ouyang, Y. (2011). A collaborative GIS framework to support equipment distribution for civil engineering disaster response operations. *Automation in Construction*, 20(5):637–648.

Chen, D. and Yang, Z. Multiple depots vehicle routing problem in the context of total urban traffic equilibrium. *Journal of Advanced Transportation*, 2017:8524960.

Chen, L. and Miller-Hooks, E. (2012). Optimal team deployment in urban search and rescue. *Transportation Research Part B: Methodological*, 46(8):984–999.

Chen, R.-M. and Guo, J.-C. (2018). Optimal routing for split delivery vehicle routing problems. In *2018 7th International Congress on Advanced Applied Informatics (IIAI-AAI)*, pages 667–670. IEEE.

Chiaro, G., Kiyota, T., Pokhrel, R. M., Goda, K., Katagiri, T., and Sharma, K. (2015). Reconnaissance report on geotechnical and structural damage caused by the 2015 Gorkha earthquake, Nepal. *Soils and Foundations*, 55(5):1030–1043.

Chini, G., Poli, C., Oddi, G., Pietrabissa, A., and Grigioni, M. (2014). Receding horizon multi-vehicle routing for emergency scenarios. In *Control and Automation (MED), 2014 22nd Mediterranean Conference of*, pages 374–379. IEEE.

Çimen, M. and Soysal, M. (2017). Time-dependent green vehicle routing problem with stochastic vehicle speeds: an approximate dynamic programming algorithm. *Transportation Research Part D: Transport and Environment*, 54:82–98.

Clarke, G. and Wright, J. W. (1964). Scheduling of vehicles from a central depot to a number of delivery points. *Operations Research*, 12(4):568–581.

Clarke, J., Gascon, V., and Ferland, J. A. (2017). A capacitated vehicle routing problem with synchronized pick-ups and drop-offs: The case of medication delivery and supervision in the dr congo. *IEEE Transactions on Engineering Management*, 64(3):327–336.

Cook, A. D., Shrestha, M., and Htet, Z. B. (2016). International response to 2015 Nepal earthquake lessons and observations. Technical report, Department of Roads Nepal.

Dantzig, G. B. and Ramser, J. H. (1959). The truck dispatching problem. *Management Science*, 6(1):80–91. Informs.

Daud, M. S. M., Hussein, M., Nasir, M., Abdullah, R., Kassim, R., Suliman, M., and Salu-din, M. (2016). Humanitarian logistics and its challenges: The literature review. *International Journal of Supply Chain Management*, 5(3):107–110.

de León, A. D., Lalla-Ruiz, E., Melián-Batista, B., and Moreno-Vega, J. M. (2019). A simulated annealing-based approach for aid distribution in post-disaster scenarios. In *International Conference on Computer Aided Systems Theory*, pages 335–343. Springer.

de Vries, H. and Van Wassenhove, L. N. (2017). Evidence-based vehicle planning for humanitarian field operations. *INSEAD Working Paper No. 2017/62/TOM/Social Innovation Centre*.

de Weerdt, M. M., Stein, S., Gerding, E. H., Robu, V., and Jennings, N. R. (2016). Intention-aware routing of electric vehicles. *IEEE Transactions on Intelligent Transportation Systems*, 17(5):1472–1482.

Deus, R. P. G. (2018). *Estimating the Efficacy of Mass Rescue Operations in Ocean Areas with Vehicle Routing Models and Heuristics*. PhD thesis, Universidade de Lisboa (Portugal).

do C. Martins, L., Hirsch, P., and Juan, A. A. (2021). Agile optimization of a two-echelon vehicle routing problem with pickup and delivery. *International Transactions in Operational Research*, 28(1):201–221.

Dreyfus, S. (2002). Richard bellman on the birth of dynamic programming. *Operations Research*, 50(1):48–51.

Dror, M. (1993). Modeling vehicle routing with uncertain demands as a stochastic program: properties of the corresponding solution. *European Journal of Operational Research*, 64(3):432–441.

Dror, M., Laporte, G., and Trudeau, P. (1989). Vehicle routing with stochastic demands: Properties and solution frameworks. *Transportation Science*, 23(3):166–176.

Du, M. and Yi, H. (2013). Multi-objective emergency logistics vehicle routing problem: Road congestion, unilateralism time window. In *LISS 2012*, pages 43–48. Springer.

Du, R. and Zhou, H. (2018). Location-routing problem under emergency with simultaneous relief delivery and victim evacuation. In *2018 International Conference on Computer Modeling, Simulation and Algorithm (CMSA 2018)*. Atlantis Press.

Dubois, F., Renaud-Goud, P., and Stolf, P. (2019). Capacitated vehicle routing problem under deadlines. In *2019 International Conference on Information and Communication Technologies for Disaster Management (ICT-DM)*, pages 1–8. IEEE.

Eksioglu, B., Vural, A. V., and Reisman, A. (2009). The vehicle routing problem: A taxonomic review. *Computers & Industrial Engineering*, 57(4):1472–1483.

Eom, H. B. and Lee, S. M. (1990). Decision support systems applications research: a bibliography (1971–1988). *European Journal of Operational Research*, 46(3):333–342.

Eom, S. and Kim, E. (2006). A survey of decision support system applications (1995–2001). *Journal of the Operational Research Society*, 57(11):1264–1278.

Eom, S. B., Lee, S. M., Kim, E. B., and Somarajan, C. (1998). A survey of decision support system applications (1988–1994). *Journal of the Operational Research Society*, 49(2):109–120.

Escribano Macias, J. J., Angeloudis, P., and Ochieng, W. (2018). Integrated trajectory-location-routing for rapid humanitarian deliveries using unmanned aerial vehicles. In *2018 Aviation Technology, Integration, and Operations Conference*, page 3045.

et al., A. (1995). Capacitated vrp instances. `https://neo.lcc.uma.es/vrp/vrp-instances/capacitated-vrp-instances/`. Accessed: 2020-10-18.

Faiz, T. I., Vogiatzis, C., and Noor-E-Alam, M. (2019). A column generation algorithm for vehicle scheduling and routing problems. *Computers & Industrial Engineering*, 130:222–236.

Fan, J., Wang, X., and Ning, H. (2006). A multiple vehicles routing problem algorithm with stochastic demand. In *2006 6th World Congress on Intelligent Control and Automation*, volume 1, pages 1688–1692. IEEE.

Farahani, R. Z., Lotfi, M., Baghaian, A., Ruiz, R., and Rezapour, S. (2020). Mass casualty management in disaster scene: A systematic review of OR & MS research in humanitarian operations. *European Journal of Operational Research*, 287(3):787–819.

Farwell, D. C. and Farwell, T. (1982). Decision support system for ski area design. *Journal of Systems Management*, 33(3):32–37.

Ferrer, J. M., Martín-Campo, F. J., Ortuño, M. T., Pedraza-Martínez, A. J., Tirado, G., and Vitoriano, B. (2018). Multi-criteria optimization for last mile distribution of disaster relief aid: Test cases and applications. *European Journal of Operational Research*, 269(2):501–515.

Ferrer, J. M., Ortuño, M. T., and Tirado, G. (2020). A new ant colony-based methodology for disaster relief. *Mathematics*, 8(4):518.

Fikar, C., Gronalt, M., and Hirsch, P. (2016). A decision support system for coordinated disaster relief distribution. *Expert Systems with Applications*, 57:104–116.

Flores, I., Ortuño, M. T., Tirado, G., and Vitoriano, B. (2020). Supported evacuation for disaster relief through lexicographic goal programming. *Mathematics*, 8(4):648.

Fragkos, M. E., Zeimpekis, V., Koutras, V., and Minis, I. (2022). Supply planning for shelters and emergency management crews. *Operational Research*, 12:57–73.

Gao, X., Nayeem, M. K., and Hezam, I. M. (2019). A robust two-stage transit-based evacuation model for large-scale disaster response. *Measurement*, 145:713–723.

Gautam, M. G., Aryal, D., Pokharel, S., Dangol, S., Malla, P., Shrestha, G. M., and Talchabhandari, P. (2018). Strategic road network of Nepal pie chart showing various pavement types of strategic road of Nepal stacked bar chart showing strategic road length in various development regions. Technical report, Department of Roads Nepal.

Gendreau, M., Laporte, G., and Séguin, R. (1996). Stochastic vehicle routing. *European Journal of Operational Research*, 88(1):3–12.

Gholami-Zanjani, S. M., Jafari-Marandi, R., Pishvaee, M. S., and Klibi, W. (2019). Dynamic vehicle routing problem with cooperative strategy in disaster relief. *International Journal of Shipping and Transport Logistics*, 11(6):455–475.

Goerigk, M., Grün, B., and Heßler, P. (2013). Branch and bound algorithms for the bus evacuation problem. *Computers & Operations Research*, 40(12):3010–3020.

Gökçe, M. A. and Ercan, E. (2019). Multi-period vehicle routing & replenishment problem of neighbourhood disaster stations for pre-disaster humanitarian relief logistics. *IFAC-PapersOnLine*, 52(13):2614–2619.

Goldberg, M. L. (2015). Nepal earthquake and facts and figures. `https://www.undispatch.com/nepal-earthquake-facts-and-figures/`.

Golden, B. L., Kovacs, A. A., and Wasil, E. A. (2014). *Vehicle Routing Applications in Disaster Relief.*, chapter 14, pages 409–436. In Toth and Vigo (2014), 2nd edition.

Golden, B. L. and Wasil, E. A. (1987). OR practice—computerized vehicle routing in the soft drink industry. *Operations Research*, 35(1):6–17.

Goodson, J. C. (2010). *Solution methodologies for vehicle routing problems with stochastic demand*. PhD thesis, University of Iowa.

Goodson, J. C., Ohlmann, J. W., and Thomas, B. W. (2013). Rollout policies for dynamic solutions to the multivehicle routing problem with stochastic demand and duration limits. *Operations Research*, 61(1):138–154.

Goodson, J. C., Thomas, B. W., and Ohlmann, J. W. (2016). Restocking-based rollout policies for the vehicle routing problem with stochastic demand and duration limits. *Transportation Science*, 50(2):591–607.

Gorry, G. A. and Morton, M. S. (1989). A framework for management information systems. *Sloan Management Review*, 30(3):49–61.

Grass, E. and Fischer, K. (2016). Two-stage stochastic programming in disaster management: A literature survey. *Surveys in Operations Research and Management Science*, 21(2):85–100.

Güner, A. R., Murat, A., and Chinnam, R. B. (2017). Dynamic routing for milk-run tours with time windows in stochastic time-dependent networks. *Transportation Research Part E: Logistics and Transportation Review*, 97:251–267.

Haghani, A. and Oh, S.-C. (1996). Formulation and solution of a multi-commodity, multi-modal network flow model for disaster relief operations. *Transportation Research Part A: Policy and Practice*, 30(3):231–250.

Hall, M., Lee, A., Cartwright, C., Maharatta, S., Karki, J., and Simkhada, P. (2017). The 2015 Nepal earthquake disaster: lessons learned one year on. *Public Health*, 145:39–44.

Han, Y.-q., Li, J.-q., Liu, Z., Liu, C., and Tian, J. (2020). Metaheuristic algorithm for solving the multi-objective vehicle routing problem with time window and drones. *International Journal of Advanced Robotic Systems*, 17(2):1729881420920031.

Hanum, F., Hadi, M., Aman, A., and Bakhtiar, T. (2019). Vehicle routing problems in rice-for-the-poor distribution. *Decision Science Letters*, 8(3):323–338.

He, Y., Wen, J., and Huang, M. (2015). Study on emergency relief VRP based on clustering and PSO. In *Computational Intelligence and Security (CIS), 2015 11th International Conference on*, pages 43–47. IEEE.

Heckmann, I., Comes, T., and Nickel, S. (2015). A critical review on supply chain risk–definition, measure and modeling. *Omega*, 52:119–132.

Hu, Z. (2018). Model and algorithm for the large material distribution problem in maritime transportation. *Journal of Coastal Research*, 82:294–306.

Huang, M., Smilowitz, K., and Balcik, B. (2012). Models for relief routing: Equity, efficiency and efficacy. *Transportation Research Part E: Logistics and Transportation Review*, 48(1):2–18.

Huang, X. and Song, L. (2016). An emergency logistics distribution routing model for unexpected events. *Annals of Operations Research*, pages 1–17.

Igbaria, M., Sprague Jr, R. H., Basnet, C., and Foulds, L. (1996). The impact and benefits of a DSS: The case of fleet manager. *Information & Management*, 31(4):215–225.

Issac, P. and Campbell, A. M. (2017). Shortest path problem with arc failure scenarios. *EURO Journal on Transportation and Logistics*, 6(2):139–163.

James, J., Yu, W., and Gu, J. (2019). Online vehicle routing with neural combinatorial optimization and deep reinforcement learning. *IEEE Transactions on Intelligent Transportation Systems*, 20(10):3806–3817.

Jansen, B. J. (1998). The graphical user interface. *ACM SIGCHI Bulletin*, 30(2):22–26.

Ji, Y.-S., Hu, D.-W., and Li, S.-H. (2018). The research of location-routing problem in emergency logistics based on uncertainties. In *CICTP 2017: Transportation Reform and Change—Equity, Inclusiveness, Sharing, and Innovation*, pages 1396–1405. American Society of Civil Engineers Reston, VA.

Jiang, J., Li, Q., Wu, L., and Tu, W. (2017). Multi-objective emergency material vehicle dispatching and routing under dynamic constraints in an earthquake disaster environment. *ISPRS International Journal of Geo-Information*, 6(5):142–142.

Jiang, Y., Bian, B., and Liu, Y. (2021). Integrated multi-item packaging and vehicle routing with split delivery problem for fresh agri-product emergency supply at large-scale epidemic disease context. *Journal of Traffic and Transportation Engineering (English Edition)*, 8(2):196–208.

Jitt-Aer, K. (2018). *The integration of Geographic Information Systems and Capacitated Vehicle Routing Problem for Humanitarian Logistics: A Case Study of Preparedness for a Tsunami in Phuket, Thailand*. PhD thesis, University of Portsmouth.

Joe, W. and Lau, H. C. (2020). Deep reinforcement learning approach to solve dynamic vehicle routing problem with stochastic customers. In *Proceedings of the International Conference on Automated Planning and Scheduling*, volume 30, pages 394–402.

Kashyap, A., Ghose, D., Menon, P. P., Sujit, P., and Das, K. (2019). UAV aided dynamic routing of resources in a flood scenario. In *2019 International Conference on Unmanned Aircraft Systems (ICUAS)*, pages 328–335. IEEE.

Katsuma, R. and Yoshida, S. (2018). Dynamic routing for emergency vehicle by collecting real-time road conditions. *International Journal of Communications, Network and System Sciences*, 11(2):27–44.

Kearby, J. A., Winz, R. D., Hodgson, T. J., Kay, M. G., King, R. E., and McConnell, B. M. (2020). Modeling and transportation planning for US noncombatant evacuation operations in South Korea. *Journal of Defense Analytics and Logistics*, 4(1):41–69.

Keenan, P. B. (1998). Spatial decision support systems for vehicle routing. *Decision Support Systems*, 22(1):65–71.

Keneally, S. K., Robbins, M. J., and Lunday, B. J. (2016). A markov decision process model for the optimal dispatch of military medical evacuation assets. *Health care management science*, 19(2):111–129.

Keogh, E. and Mueen, A. (2011). Curse of dimensionality. In *Encyclopedia of Machine Learning*, pages 257–258. Springer.

Khare, A., Batta, R., and Kang, J. E. (2021). On the analysis of last-mile relief delivery on a tree network: Application to the 2015 Nepal earthquake. *Journal of the Operational Research Society*, 72(4):727–743.

Khorsi, M., Chaharsooghi, S. K., Bozorgi-Amiri, A., and Kashana, A. H. (2020). A multi-objective multi-period model for humanitarian relief logistics with split delivery and multiple uses of vehicles. *Journal of Systems Science and Systems Engineering*, 29(3):360–378.

Khoukhi, S., El Yaakoubi, O., Bojji, C., and Bensouda, Y. (2019). A genetic algorithm for solving a multi-trip vehicle routing problem with time windows and simultaneous pick-up and delivery in a hospital complex. In *Proceedings of the 3rd International Conference On Machine Learning and Soft Computing*, pages 76–80.

Klingman, D., Phillips, N., Steiger, D., and Young, W. (1987). The successful deployment of management science throughout citgo petroleum corporation. *Interfaces*, 17(1):4–25.

Knott, R. (1988). Vehicle scheduling for emergency relief management: A knowledge-based approach. *Disasters*, 12(4):285–293.

Kokaji, K. and Kainuma, Y. (2018). Development of a disaster relief logistics model minimizing the range of delivery time. *Operations and Supply Chain Management*, 11(2):66–72.

Krasko, V. and Rebennack, S. (2017). Two-stage stochastic mixed-integer nonlinear programming model for post-wildfire debris flow hazard management: Mitigation and emergency evacuation. *European Journal of Operational Research*, 263(1):265–282.

Labadie, N., Prins, C., and Prodhon, C. (2016). *Metaheuristics for vehicle routing problems*. John Wiley & Sons.

Lahyani, R., Khemakhem, M., and Semet, F. Taxonomy for rich vehicle routing problems. In *Proceedings of the 1st International Conference on Logistics Operations Management*, pages 00–00.

Lamos Díaz, H., Aguilar Imitola, K., Barreto Robles, M. A., Niño Niño, P. N., and Martínez Quezada, D. O. (2018). A memetic algorithm for location-routing problem with time windows for the attention of seismic disasters a case study from bucaramanga, colombia. *INGE CUC*, 14(1):75–86.

Laporte, G. (2009). Fifty years of vehicle routing. *Transportation Science*, 43(4):408–416.

Larson, R. C. and Odini, A. R. (1999). Urban operations research. Massachusetts. `http://web.mit.edu/urban_or_book/www/book/chapter6/6.4.12.html`.

Lee, K., Chae, J., Song, B., and Choi, D. (2020). A model for sustainable courier services: vehicle routing with exclusive lanes. *Sustainability*, 12(3):1077.

Li, B., Patankar, S., Moridian, B., and Mahmoudian, N. (2018a). Planning large-scale search and rescue using team of UAVs and charging stations. In *2018 IEEE International Symposium on Safety, Security, and Rescue Robotics (SSRR)*, pages 1–8. IEEE.

Li, D. and Zhou, H. (2018). Robust optimization for vehicle routing problem under uncertainty in disaster response. In *2018 15th International Conference on Service Systems and Service Management (ICSSSM)*, pages 1–5. IEEE.

Li, H., Peng, J., Li, S., and Su, C. (2017). Dispatching medical supplies in emergency events via uncertain programming. *Journal of Intelligent Manufacturing*, 28(3):549–558.

Li, J., Ma, Y., Gao, R., Cao, Z., Lim, A., Song, W., and Zhang, J. (2021). Deep reinforcement learning for solving the heterogeneous capacitated vehicle routing problem. *IEEE Transactions on Cybernetics*, :1–14.

Li, Q., Tu, W., and Zhuo, L. (2018b). Reliable rescue routing optimization for urban emergency logistics under travel time uncertainty. *ISPRS International Journal of Geo-Information*, 7(2):77.

Li, W. and Cassandras, C. G. (2006). A cooperative receding horizon controller for multivehicle uncertain environments. *IEEE Transactions on Automatic Control*, 51(2):242–257.

Li, W., Wu, Z., and Song, L. (2018c). Single-level LRP problem of emergency logistics system considering demand priority. *Earth and Environmental Science*, 189(6):062020.

Li, Y. and Chung, S. H. (2019). Disaster relief routing under uncertainty: A robust optimization approach. *IISE Transactions*, 51(8):869–886.

Li, Z., Yu, H., Chen, X., Zhang, G., and Ma, D. (2019). Tsunami-induced traffic evacuation strategy optimization. *Transportation Research Part D: Transport and Environment*, 77:535–559.

Lin, B., Ghaddar, B., and Nathwani, J. (2021). Deep reinforcement learning for the electric vehicle routing problem with time windows. *IEEE Transactions on Intelligent Transportation Systems*, :1–11.

Liu, B., Sheu, J.-B., Zhao, X., Chen, Y., and Zhang, W. (2020). Decision making on post-disaster rescue routing problems from the rescue efficiency perspective. *European Journal of Operational Research*, 286(1):321–335.

Liu, C., Kou, G., Peng, Y., and Alsaadi, F. E. (2019a). Location-routing problem for relief distribution in the early post-earthquake stage from the perspective of fairness. *Sustainability*, 11(12):3420.

Liu, J., Liu, J., Xie, K., and Xie, K. (2017). Emergency materials transportation model in disasters based on dynamic programming and ant colony optimization. *Kybernetes*, 46(4):656–671.

Liu, X., Zhao, Y., and Qiao, C. (2019b). Autonomous vehicle dispatching for person evacuation. In *2019 IEEE Global Communications Conference (GLOBECOM)*, pages 1–6. IEEE.

Loughran, B. and Cocks, P. (1977). Airline programme planning in british airways european division. *Interfaces*, 7(2):21–36.

Luis, E., Dolinskaya, I. S., and Smilowitz, K. R. (2012). Disaster relief routing: Integrating research and practice. *Socio-economic Planning Sciences*, 46(1):88–97.

Ma, Y. and Wu, N. (2018). Research on seismic site emergency rescue traffic path analysis system based on public image information. *Journal of Information Hiding and Multimedia Signal Processing*, 9(3):577–585.

Maghfiroh, M. F. and Hanaoka, S. (2018). Dynamic truck and trailer routing problem for last mile distribution in disaster response. *Journal of Humanitarian Logistics and Supply Chain Management*, 8(2):252–278.

Masaki, J. E., AbdelRazig, Y. A., and Kobelo, D. C. (2019). An evolutionary algorithm for the autonomous vehicle routing problem under emergency evacuation. *Journal of Multidisciplinary Engineering Science Studies*, 5:2811–2817.

Menth, M. and Stamm, J. L. H. (2015). An agent-based modeling approach to improve coordination between humanitarian relief providers. In *Winter Simulation Conference (WSC), 2015*, pages 3116–3117. IEEE.

Miller, D. M. (1987). An interactive, computer-aided ship scheduling system. *European Journal of Operational Research*, 32(3):363–379.

Mills, A. F., Argon, N. T., and Ziya, S. (2015). Dynamic distribution of casualties to medical facilities in the aftermath of a disaster. Technical report, Working paper, Kelley School of Business, Indiana University. Available at: http://www.unc. edu/˜ ziya/dynamic-casualty-dispatch. pdf, accessed: February 15.

Mizuno, H. and Takahashi, S. (2016). Goods sharing problems for emergency supply. In *Advanced Applied Informatics (IIAI-AAI), 2016 5th IIAI International Congress on*, pages 1168–1173. IEEE.

Mnih, V., Kavukcuoglu, K., Silver, D., Graves, A., Antonoglou, I., Wierstra, D., and Riedmiller, M. (2013). Playing ATARI with deep reinforcement learning. *arXiv preprint arXiv:1312.5602*.

Moghaddas, Z. et al. (2018). A three-stage model for location problem under fuzzy environments. *International Journal of Industrial Engineering & Production Research*, 29(4):483–496.

Mohamed, R. E., Kosba, E., Mahar, K., and Mesbah, S. (2018). A framework for emergency-evacuation planning using gis and dss. In *Information Fusion and Intelligent Geographic Information Systems (IF&IGIS'17)*, pages 213–226. Springer.

Moin, N. H. and Halim, H. Z. A. (2018). Solving inventory routing problem with stochastic demand. In *AIP Conference Proceedings*, volume 1974, page 020104. AIP Publishing LLC.

Mole, R. and Jameson, S. (1976). A sequential route-building algorithm employing a generalised savings criterion. *Journal of the Operational Research Society*, 27(2):503–511.

Molina, J., López-Sánchez, A., Hernández-Díaz, A. G., and Martínez-Salazar, I. (2018). A multi-start algorithm with intelligent neighborhood selection for solving multi-objective humanitarian vehicle routing problems. *Journal of Heuristics*, 24(2):111–133.

Motoyama, T. (2017). Optimal disaster-preventive expenditure in a dynamic and stochastic model. *Journal of Macroeconomics*, 51:28–47.

Nadeem, I., Uduman, P. S., Iqbal, P., and Khaliq, R. (2019). A heuristic routing approach for flood evacuation. *International Journal of Management, Technology And Engineering*, 9(2):1735 – 1742.

Nadi, A. and Edrisi, A. (2017). Adaptive multi-agent relief assessment and emergency response. *International Journal of Disaster Risk Reduction*, 24:12–23.

Nazari, M., Oroojlooy, A., Snyder, L. V., and Takác, M. (2018). Deep reinforcement learning for solving the vehicle routing problem. *arXiv preprint arXiv:1802.04240*.

Neupane, S. P. (2015). Immediate lessons from the Nepal earthquake. *The Lancet*, 385(9982):2041–2042.

Ngueveu, S. U., Prins, C., and Calvo, R. W. (2010). An effective memetic algorithm for the cumulative capacitated vehicle routing problem. *Computers & Operations Research*, 37(11):1877–1885.

Ni, S., Yu, J., and Zhong, Q. (2014). Routing design for emergency supplies based on multi-objective programming. In *ICLEM 2014: System Planning, Supply Chain Management, and Safety*, pages 463–469. ASCE.

Novoa, C. and Storer, R. (2009). An approximate dynamic programming approach for the vehicle routing problem with stochastic demands. *European Journal of Operational Research*, 196(2):509–515.

Novoa, C. M. (2005). *Static and dynamic approaches for solving the vehicle routing problem with stochastic demands*. PhD thesis, Lehigh University.

Nowak, M. A. (2005). *The pickup and delivery problem with split loads*. PhD thesis, Georgia Institute of Technology.

Nugroho, S., Nafisah, L., Khannan, M., Mastrisiswadi, H., and Ramdhani, M. (2020). Vehicle routing problem with heterogeneous fleet, split delivery, multiple product, multiple trip, and time windows: A case study in fuel distribution. In *IOP Conference Series: Materials Science and Engineering*, volume 847, page 012066. IOP Publishing.

Nurprihatin, F. and Lestari, A. (2020). Waste collection vehicle routing problem model with multiple trips, time windows, split delivery, heterogeneous fleet and intermediate facility. *Engineering Journal*, 24(5):55–64.

OCHA, U. (2003). Glossary of humanitarian terms in relation to the protection of civilians in armed conflict. *New York: United Nations. UN OCHA (2012): OCHA on Message: Humanitarian Principles. New York: UN OCHA. UN OCHA (nedatovno): Consolidated Appeal Process: About the Process, United Nations Office for the Coordination of Humanitarian Affairs*, 15(11):2013.

Oh, B. H., Kim, K., Choi, H.-L., and Hwang, I. (2018). Cooperative multiple agent-based algorithm for evacuation planning for victims with different urgencies. *Journal of Aerospace Information Systems*, 15(6):382–395.

Okamura, M., Bhandary, N. P., Mori, S., Marasini, N., and Hazarika, H. (2015). Report on a reconnaissance survey of damage in kathmandu caused by the 2015 Gorkha Nepal earthquake. *Soils and Foundations*, 55(5):1015–1029.

Ozkapici, D. B., Ertem, M. A., and Aygüneş, H. (2016). Intermodal humanitarian logistics model based on maritime transportation in Istanbul. *Natural Hazards*, 83(1):345–364.

Paessens, H. (1988). The savings algorithm for the vehicle routing problem. *European Journal of Operational Research*, 34(3):336–344.

Petrik, M. (2010). *Optimization-based approximate dynamic programming*. PhD thesis, University of Massachusetts Amherst.

Pico, A. and Tan, Y. H. (2012). Heuristics for solving problem of evacuating non-ambulatory people in a short-notice disaster. Technical report, NAVAL POST-GRADUATE SCHOOL MONTEREY CA.

Pillac, V., Van Hentenryck, P., and Even, C. (2016). A conflict-based path-generation heuristic for evacuation planning. *Transportation Research Part B: Methodological*, 83:136–150.

Pourrahmani, E., Delavar, M. R., Pahlavani, P., and Mostafavi, M. A. (2015). Dynamic evacuation routing plan after an earthquake. *Natural Hazards Review*, 16(4):04015006.

Powell, W. B. (2007). *Approximate Dynamic Programming: Solving the curses of dimensionality*, volume 703. John Wiley & Sons.

Powell, W. B., George, A., Bouzaiene-Ayari, B., and Simao, H. P. (2005). Approximate dynamic programming for high dimensional resource allocation problems. In *Proceedings. 2005 IEEE International Joint Conference on Neural Networks, 2005.*, volume 5, pages 2989–2994. IEEE.

Qin, J., Ye, Y., Cheng, B.-r., Zhao, X., and Ni, L. (2017). The emergency vehicle routing problem with uncertain demand under sustainability environments. *Sustainability*, 9(2):288.

QURESHI, A. G. and TANIGUCHI, E. (2017). A multi-period relief distribution model considering limited resources and decreasing resilience of affected population. *Journal of the Eastern Asia Society for Transportation Studies*, 12:57–73.

Qureshi, A. G. and Taniguchi, E. (2020). A multi-period humanitarian logistics model considering limited resources and network availability. *Transportation Research Procedia*, 46:212–219.

Rabbani, M., Aghabegloo, M., and Farrokhi-Asl, H. (2017). Solving a bi-objective mathematical programming model for blood mobiles location routing problem. *International Journal of Industrial Engineering Computations*, 8(1):19–32.

Radianti, J., Hiltz, S. R., and Labaka, L. (2016a). An overview of public concerns during the recovery period after a major earthquake: Nepal twitter analysis. In *System Sciences (HICSS), 2016 49th Hawaii International Conference on*, pages 136–145. IEEE.

Radianti, J., Hiltz, S. R., and Labaka, L. (2016b). An overview of public concerns during the recovery period after a major earthquake: Nepal twitter analysis. In *System Sciences (HICSS), 2016 49th Hawaii International Conference on*, pages 136–145. IEEE.

Rand, G. (2014). 50 years of the savings method for vehicle routing problems. *Operations Research Management Science Today*, 41(1):14–15.

Raziei, Z., Tavakkoli-Moghaddam, R., Rezaei-Malek, M., Bozorgi-Amiri, A., and Jolai, F. (2018). Postdisaster relief distribution network design under disruption risk: a tour covering location-routing approach. In *Integrating Disaster Science and Management*, pages 393–406. Elsevier.

Ren, G., Xu, L., and Duan, T. (2019). Modeling phased evacuation network optimization based on public transit. In *CICTP 2019*, pages 4389–4401.

Rey, D., Almi'ani, K., and Nair, D. J. (2018). Exact and heuristic algorithms for finding envy-free allocations in food rescue pickup and delivery logistics. *Transportation Research Part E: Logistics and Transportation Review*, 112:19–46.

Ritzinger, U., Puchinger, J., and Hartl, R. F. (2016). A survey on dynamic and stochastic vehicle routing problems. *International Journal of Production Research*, 54(1):215–231.

Rivera, J. C., Afsar, H. M., and Prins, C. (2015). A multistart iterated local search for the multitrip cumulative capacitated vehicle routing problem. *Computational Optimization and Applications*, 61(1):159–187.

Rivera, J. C., Afsar, H. M., and Prins, C. (2016). Mathematical formulations and exact algorithm for the multitrip cumulative capacitated single-vehicle routing problem. *European Journal of Operational Research*, 249(1):93–104.

Roseman, M. (2007). Tkdocs. `https://tkdocs.com/tutorial/index.html`. Official Tkinter Documentation.

Ruan, J., Wang, X., Chan, F., and Shi, Y. (2016). Optimizing the intermodal transportation of emergency medical supplies using balanced fuzzy clustering. *International Journal of Production Research*, 54(14):4368–4386.

Ruan, J., Wang, X., and Shi, Y. (2014). A two-stage approach for medical supplies intermodal transportation in large-scale disaster responses. *International journal of environmental research and public health*, 11(11):11081–11109.

Sabouhi, F., Bozorgi-Amiri, A., Moshref-Javadi, M., and Heydari, M. (2019). An integrated routing and scheduling model for evacuation and commodity distribution in large-scale disaster relief operations: A case study. *Annals of Operations Research*, 283(1-2):643–677.

Sabouhi, F., Heydari, M., and Bozorgi-Amiri, A. (2016). Multi-objective routing and scheduling for relief distribution with split delivery in post-disaster response. *Journal of Industrial and Systems Engineering*, 9(3):17–27.

Sahebjamnia, N., Torabi, S. A., and Mansouri, S. A. (2017). A hybrid decision support system for managing humanitarian relief chains. *Decision Support Systems*, 95:12–26.

Sakiani, R., Seifi, A., and Khorshiddoust, R. R. (2020). Inventory routing and dynamic redistribution of relief goods in post-disaster operations. *Computers & Industrial Engineering*, 140:106219.

Savelsbergh, M. (1990). A parallel insertion heuristic for vehicle routing with side constraints. *Statistica Neerlandica*, 44(3):139–148.

Scott Morton, M. S. (1971). Management decision systems: Computer based support for decision making. *Boston: Harvard University*.

Secomandi, N. (1998). *Exact and heuristic dynamic programming algorithms for the vehicle routing problem with stochastic demands*. PhD thesis, University of Houston.

Secomandi, N. (2000). Comparing neuro-dynamic programming algorithms for the vehicle routing problem with stochastic demands. *Computers & Operations Research*, 27(11-12):1201–1225.

Secomandi, N. (2001). A rollout policy for the vehicle routing problem with stochastic demands. *Operations Research*, 49(5):796–802.

Secomandi, N. (2003). Analysis of a rollout approach to sequencing problems with stochastic routing applications. *Journal of Heuristics*, 9(4):321–352.

Sepulveda, J. and Bull, J. (2019). A model-driven decision support system for aid in a natural disaster. In *International Conference on Human Systems Engineering and Design: Future Trends and Applications*, pages 523–528. Springer.

Sepúlveda, J. M., Arriagada, I. A., and Derpich, I. (2018). A decision support system for distribution of supplies in natural disaster situations. In *2018 7th International Conference on Computers Communications and Control (ICCCC)*, pages 295–301. IEEE.

Shahabi, K. and Wilson, J. P. (2018). Scalable evacuation routing in a dynamic environment. *Computers, Environment and Urban Systems*, 67:29–40.

Shahparvari, S., Abbasi, B., Chhetri, P., and Abareshi, A. (2019). Fleet routing and scheduling in bushfire emergency evacuation: A regional case study of the black saturday bushfires in Australia. *Transportation Research Part D: Transport and Environment*, 67:703–722.

Shapiro, S. S. and Wilk, M. B. (1965). An analysis of variance test for normality (complete samples). *Biometrika*, 52(3/4):591–611.

Sharda, R., Barr, S. H., and McDonnell, J. C. (1988). Decision support system effectiveness: A review and an empirical test. *Management Science*, 34(2):139–159.

Sharma, D. C. (2015). Nepal earthquake exposes gaps in disaster preparedness. *The Lancet*, 385(9980):1819–1820.

Shavarani, S. M. and Vizvari, B. (2018). Post-disaster transportation of seriously injured people to hospitals. *Journal of Humanitarian Logistics and Supply Chain Management*, 8(2):227–251.

Shen, L., Tao, F., Shi, Y., and Qin, R. (2019). Optimization of location-routing problem in emergency logistics considering carbon emissions. *International Journal of Environmental Research and Public Health*, 16(16):2982.

Shiri, D., Akbari, V., and Salman, F. S. (2020). Online routing and scheduling of search-and-rescue teams. *OR Spectrum*, pages 1–30.

Shiripour, S. and Mahdavi-Amiri, N. (2019). Optimal distribution of the injured in a multi-type transportation network with damage-dependent travel times: Two metaheuristic approaches. *Socio-Economic Planning Sciences*, 68:100660.

Silver, D., Hubert, T., Schrittwieser, J., Antonoglou, I., Lai, M., Guez, A., Lanctot, M., Sifre, L., Kumaran, D., Graepel, T., et al. (2017). Mastering chess and shogi by self-play with a general reinforcement learning algorithm. *arXiv preprint arXiv:1712.01815*.

Soeffker, N., Ulmer, M. W., and Mattfeld, D. C. (2016). Problem-specific state space partitioning for dynamic vehicle routing problems. *Proceedings of MKWI*, pages 229–240.

Solomon, M. M. (1987). Algorithms for the vehicle routing and scheduling problems with time window constraints. *Operations Research*, 35(2):254–265.

Sörensen, K. and Glover, F. (2013). Metaheuristics. *Encyclopedia of Operations Research and Management Science*, 62:960–970.

Sprague Jr, R. H. (1980). A framework for the development of decision support systems. *MIS Quarterly*, pages 1–26.

Stott Jr, K. L. and Douglas, B. W. (1981). A model-based decision support system for planning and scheduling ocean-borne transportation. *Interfaces*, 11(4):1–10.

Su, Z.-P., Jiang, J.-G., Liang, C.-Y., and Zhang, G.-F. (2011). Path selection in disaster response management based on Q-learning. *International Journal of Automation and Computing*, 8(1):100–106.

Takabatake, T., Fujisawa, K., Esteban, M., and Shibayama, T. (2020). Simulated effectiveness of a car evacuation from a tsunami. *International Journal of Disaster Risk Reduction*, 47:101532.

Talarico, L., Meisel, F., and Sörensen, K. (2015). Ambulance routing for disaster response with patient groups. *Computers & Operations Research*, 56:120–133.

Talbi, E.-G. (2009). *Metaheuristics: From Design to Implementation*, volume 74. John Wiley & Sons.

Tavakkoli-Moghaddam, R., Memari, P., and Talebi, E. (2018). A bi-objective location-allocation problem of temporary emergency stations and ambulance routing in a disaster situation. In *2018 4th International Conference on Optimization and Applications (ICOA)*, pages 1–4. IEEE.

Tavana, M., Abtahi, A.-R., Di Caprio, D., Hashemi, R., and Yousefi-Zenouz, R. (2018). An integrated location-inventory-routing humanitarian supply chain network with pre-and post-disaster management considerations. *Socio-Economic Planning Sciences*, 64:21–37.

Theobald, O. (2017). *Machine Learning for Absolute Beginners: A Plain English Introduction*. Scatterplot press.

Tian, R., Li, S., and Yang, G. (2018). Research on emergency vehicle routing planning based on short-term traffic flow prediction. *Wireless Personal Communications*, 102(2):1993–2010.

Tikani, H. and Setak, M. (2019). Ambulance routing in disaster response scenario considering different types of ambulances and semi soft time windows. *Journal of Industrial and Systems Engineering*, 12(1):95–128.

Tlili, T., Abidi, S., and Krichen, S. (2018). A mathematical model for efficient emergency transportation in a disaster situation. *The American Journal of Emergency Medicine*, 36(9):1585–1590.

Torabi, S. A. and Hassini, E. (2008). An interactive possibilistic programming approach for multiple objective supply chain master planning. *Fuzzy sets and systems*, 159(2):193–214.

Toth, P. and Vigo, D. (2002). *The Vehicle Routing Problem*. SIAM.

Toth, P. and Vigo, D. (2014). *Vehicle Routing: Problems, Methods, and Applications*. SIAM.

Ulmer, M. W. (2017). *Approximate dynamic programming for dynamic vehicle routing*, volume 61 of *Operations Research/Computer Science Interfaces Series*. Springer International Publishing, 1 edition.

Ulmer, M. W., Goodson, J. C., Mattfeld, D. C., and Hennig, M. (2019). Offline–online approximate dynamic programming for dynamic vehicle routing with stochastic requests. *Transportation Science*, 53(1):185–202.

Ulmer, M. W., Goodson, J. C., Mattfeld, D. C., and Thomas, B. W. (2017a). Route-based markov decision processes for dynamic vehicle routing problems. Technical report, Braunschweig.

Ulmer, M. W., Heilig, L., and Voß, S. (2017b). On the value and challenge of real-time information in dynamic dispatching of service vehicles. *Business & Information Systems Engineering*, 59(3):161–171.

Ulmer, M. W., Mattfeld, D. C., Hennig, M., and Goodson, J. C. (2016). A rollout algorithm for vehicle routing with stochastic customer requests. In *Logistics Management*, pages 217–227. Springer.

Ulmer, M. W., Mattfeld, D. C., and Köster, F. (2018). Budgeting time for dynamic vehicle routing with stochastic customer requests. *Transportation Science*, 52(1):20–37.

Vahdani, B., Veysmoradi, D., Noori, F., and Mansour, F. (2018a). Two-stage multi-objective location-routing-inventory model for humanitarian logistics network design under uncertainty. *International Journal of Disaster Risk Reduction*, 27:290–306.

Vahdani, B., Veysmoradi, D., Shekari, N., and Mousavi, S. M. (2018b). Multi-objective, multi-period location-routing model to distribute relief after earthquake by considering emergency roadway repair. *Neural Computing and Applications*, 30(3):835–854.

Van Roy, B., Bertsekas, D. P., Lee, Y., and Tsitsiklis, J. N. (1997). A neuro-dynamic programming approach to retailer inventory management. In *Proceedings of the 36th IEEE Conference on Decision and Control*, volume 4, pages 4052–4057. IEEE.

Vidal, T., Crainic, T. G., Gendreau, M., and Prins, C. (2013). Heuristics for multi-attribute vehicle routing problems: A survey and synthesis. *European Journal of Operational Research*, 231(1):1–21.

Vieira, Y. E., Bandeira, R. A., Lopes, L. A., Júnior, O. S. S., and Júnior, M. M. B. (2020). A procedure to support the distribution of drinking water for victims of drought: The case of the Brazilian semi-arid region. *Transportation Research Procedia*, 47:331–339.

Vogiatzis, C., Yoshida, R., Aviles-Spadoni, I., Imamoto, S., and Pardalos, P. M. (2013). Livestock evacuation planning for natural and man-made emergencies. *International Journal of Mass Emergencies and Disasters*, 31(1):25–37.

Wang, D., Zhu, J., Wei, X., Cheng, T., Yin, Y., and Wang, Y. (2019). Integrated production and multiple trips vehicle routing with time windows and uncertain travel times. *Computers & Operations Research*, 103:1–12.

Wang, J., Ding, H., Lv, Q., Chen, J.-h., Sun, Y.-f., Fan, H.-j., and Liu, Q.-l. (2016a). 2015 Nepal earthquake: analysis of child rescue and treatment by a field hospital. *Disaster Medicine and Public Health Preparedness*, pages 1–4. Cambridge Univ Press.

Wang, J., Ding, H., Lv, Q., Chen, J.-h., Sun, Y.-f., Fan, H.-j., and Liu, Q.-l. (2016b). 2015 Nepal earthquake: analysis of child rescue and treatment by a field hospital. *Disaster medicine and public health preparedness*, pages 1–4.

Wang, L. (2020). A two-stage stochastic programming framework for evacuation planning in disaster responses. *Computers & Industrial Engineering*, 145:106458.

Wang, S., Liu, F., Lian, L., Hong, Y., and Chen, H. Integrated post-disaster medical assistance team scheduling and relief supply distribution. *The International Journal of Logistics Management*, 29.

Wang, X., Choi, T.-M., Liu, H., and Yue, X. (2016c). A novel hybrid ant colony optimization algorithm for emergency transportation problems during post-disaster scenarios. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 48(4):545–556.

Widuch, J. (2020). Current and emerging formulations and models of real-life rich vehicle routing problems. In *Smart Delivery Systems*, pages 1–35. Elsevier.

Wu, Y., Pan, F., Li, S., Chen, Z., and Dong, M. (2020). Peer-induced fairness capacitated vehicle routing scheduling using a hybrid optimization ACO–VNS algorithm. *Soft Computing*, 24(3):2201–2213.

Xavier, I. R., Bandeira, R. A., Bandeira, A. P., Campos, V. B., and Silva, L. O. (2020). Planning the use of helicopters in distribution of supplies in response operations of natural disasters. *Transportation Research Procedia*, 47:633–640.

Xie, Q., Gaohu, L., Chen, H., Xu, C., and Feng, B. (2017). Seismic damage to road networks subjected to earthquakes in Nepal, 2015. *Earthquake Engineering and Engineering Vibration*, 16(3):649–670.

Yan, S., Chu, J. C., Hsiao, F.-Y., and Huang, H.-J. (2015). A planning model and solution algorithm for multi-trip split-delivery vehicle routing and scheduling problems with time windows. *Computers & Industrial Engineering*, 87:383–393.

Yang, X., Hao, W., and Lu, Y. (2018). Inventory slack routing application in emergency logistics and relief distributions. *Plos One*, 13(6):e0198443.

Yano, C. A., Chan, T. J., Richter, L. K., Cutler, T., Murty, K. G., and McGettigan, D. (1987). Vehicle routing at quality stores. *Interfaces*, 17(2):52–63.

Yi, J.-H., Wang, J., and Wang, G.-G. (2019). Using monarch butterfly optimization to solve the emergency vehicle routing problem with relief materials in sudden disasters. *Open Geosciences*, 11(1):391–413.

Yi, W. and Kumar, A. (2007). Ant colony optimization for disaster relief operations. *Transportation Research Part E: Logistics and Transportation Review*, 43(6):660–672.

Yin, R., Liu, Z., and Fu, X. (2019). Vehicle routing strategy for disaster relief operation: A case study. In *CICTP 2019*, pages 2694–2705. ASCE.

Yuan, Y., Zhou, X., and Yang, M. (2019). Emergency logistics vehicle routing optimization based on insufficient supply. In *E3S Web of Conferences*, volume 136, page 04068. EDP Sciences.

Zaiontz, C. (2015). Real statistics using excel.

Zanakis, S. H., Evans, J. R., and Vazacopoulos, A. A. (1989). Heuristic methods and applications: A categorized survey. *European Journal of Operational Research*, 43(1):88–110.

Zhan, S.-l., Chen, L., Chen, P.-K., and Ye, Y. (2019). A vehicle route planning method of two-phase large-scale crowd evacuation in typhoon relief activities. *Mathematical Problems in Engineering*, 2019:9539746.

Zhang, K., He, F., Zhang, Z., Lin, X., and Li, M. (2020). Multi-vehicle routing problems with soft time windows: A multi-agent reinforcement learning approach. *Transportation Research Part C: Emerging Technologies*, 121:102861.

Zhang, L. (2016). Real-time transit vehicle routing optimization in intermodal emergency evacuations. Technical report, Mississippi State University.

Zhang, Q. and Xiong, S. (2018). Routing optimization of emergency grain distribution vehicles using the Immune Ant Colony Optimization algorithm. *Applied Soft Computing*, 71:917–925.

Zhao, B. (2016). April 2015 Nepal earthquake: observations and reflections. *Natural Hazards*, 80(2):1405–1410.

Zhao, X., Qian, W.-w., and Huo, H. (2019). Multi-objective routing optimization for urban emergency rescue vehicles. In *CICTP 2019*, pages 3855–3866. ASCE.

Zhao, Y., Chen, X., Jia, Q.-S., Guan, X., Zhang, S., and Jiang, Y. (2010). Long-term scheduling for cascaded hydro energy systems with annual water consumption and release constraints. *IEEE Transactions on Automation Science and Engineering*, 7(4):969–976.

Zhen, L., Ma, C., Wang, K., Xiao, L., and Zhang, W. (2020). Multi-depot multi-trip vehicle routing problem with time windows and release dates. *Transportation Research Part E: Logistics and Transportation Review*, 135:101866.

Zheng, H. (2014). Optimization of bus routing strategies for evacuation. *Journal of Advanced Transportation*, 48(7):734–749.

Zhou, X., Tong, L., Mahmoudi, M., Zhuge, L., Yao, Y., Zhang, Y., Shang, P., Liu, J., and Shi, T. (2018). Open-source VRPLite package for vehicle routing with pickup and delivery: A path finding engine for scheduled transportation systems. *Urban Rail Transit*, 4(2):68–85.

Zhu, L., Gong, Y., Xu, Y., and Gu, J. (2019). Emergency relief routing models for injured victims considering equity and priority. *Annals of Operations Research*, 283(1-2):1573–1606.