Durham University

Durham E-Theses

Online Optimisation of Casualty Processing in Major Incident Response

Wilson, Duncan

How to cite:

Wilson, Duncan (2015) *Online Optimisation of Casualty Processing in Major Incident Response*, Durham theses, Durham University. Available at Durham E-Theses Online: http://etheses.dur.ac.uk/10961/

Use policy

The full-text may be used and/or reproduced, and given to third parties in any format or medium, without prior permission or charge, for personal research or study, educational, or not-for-profit purposes provided that:

- a full bibliographic reference is made to the original source
- a link is made to the metadata record in Durham E-Theses
- the full-text is not changed in any way

The full-text must not be sold in any format or medium without the formal permission of the copyright holders.

Please consult the full Durham E-Theses policy for further details.

Academic Support Office, The Palatine Centre, Durham University, Stockton Road, Durham, DH1 3LE
e-mail: e-theses.admin@durham.ac.uk Tel: +44 0191 334 6107
http://etheses.dur.ac.uk

Online Optimisation of Casualty Processing in Major Incident Response

Duncan Thomas Wilson

A Dissertation
Presented to Durham University in Candidacy for the Degree of Doctor of Philosophy

Recommended for Acceptance
by the School of
Engineering & Computing Sciences
Adviser: Dr Graham Coates

May 2014

The copyright of this thesis rests with the author. No quotation from it should be published without the author's prior written consent and information derived from it should be acknowledged.

Abstract

Recent emergency response operations to Mass Casualty Incidents (MCIs) have been criticised for a lack of coordination, implying that there is clear potential for response operations to be improved and for corresponding benefits in terms of the health and well-being of those affected by such incidents. In this thesis, the use of mathematical modelling, and in particular optimisation, is considered as a means with which to help improve the coordination of MCI response.

Upon reviewing the nature of decision making in MCIs and other disaster response operations in practice, this work demonstrates through an in-depth review of the available academic literature that an important problem has yet to be modelled and solved using an optimisation methodology. This thesis involves the development of such a model, identifying an appropriate task scheduling formulation of the decision problem and a number of objective functions corresponding to the goals of the MCI response decision makers. Efficient solution methodologies are developed to allow for solutions to the model, and therefore to the MCI response operation, to be found in a timely manner.

Following on from the development of the optimisation model, the dynamic and uncertain nature of the MCI response environment is considered in detail. Highlighting the lack of relevant research considering this important aspect of the problem, the optimisation model is extended to allow for its use in real-time. In order to allow for the utility of the model to be thoroughly examined, a complementary simulation is developed and an interface allowing for its communication with the optimisation model specified. Extensive computational experiments are reported, demonstrating both the danger of developing and applying optimisation models under a set of unrealistic assumptions, and the potential for the model developed in this work to deliver improvements in MCI response operations.

Acknowledgements

I would like to express my gratitude to my primary supervisor, Graham Coates. Consistently generous with his time, encouragement, and patience, he helped me not only in completing this piece of work but in preparing for the next stages of my academic career. Thanks also to my secondary supervisor, Roger Crouch, for the many stimulating discussions in our Friday morning meetings.

Particular thanks goes to my colleague and friend Glenn Hawe, who, over countless discussions both in the office and the Victoria, taught me a great deal about research and how to hold your drink. His expertise, particularly in programming and optimisation, was a valuable resource throughout my work, and one he was always happy to share.

Finally, I would like to thank my brilliant bride-to-be, Amy Clow, for looking after me when I was busy and giving me some gentle 'encouragement' when my motivation was lacking.

Contents

List of Tables

List of Figures

Nomenclature

τ	Time (minutes)
$\mathcal{C}$	Set of all casualties
c_i	Casualty i
n_c	Total number of casualties
c_i^s	Stability of the health of casualty i
c_i^e	Extrication requirements of casualty i
c_i^t	Health of casualty i
$\mathcal{H}$	Set of all hospitals
h_i	Hospital i
n_h	Total number of hospitals
cap_h	Initial capacity of hospital h
max_h	Maximum capacity of hospital h
$rate_h$	Resource increase rate of hospital h
$\mathcal{R}$	Set of all response units
r_i	Response unit i
n_r	Total number of response units
$\mathcal{T}$	Set of all tasks
$t_{i,j}$	The jth task associated with casualty i
$t_{i,j}^r$	Response unit assigned to task $t_{i,j}$
$t_{i,j}^p$	The priority level assigned to task $t_{i,j}$
$t_{i,j}^h$	The hospital assigned to task $t_{i,j}$
$\mathcal{S}$	Set of all solutions
s	Solution
d	Distance of a path
$\hat{m}(d)$	Median travel time of distance d
f_1	Objective 1, fatalities
f_2	Objective 2, hospital arrival time
f_3	Objective 3, hospital quality
f_4	Objective 4, responder idle time
f_5	Objective 5, makespan
$g_1(= f_1)$	Composite objective 1, fatalities
g_2	Composite objective 2, suffering
g_3	Composite objective 3, efficiency
L	Set of triage states
τ_i^C	Time of arrival at CCS for casualty i
$p_i^T(\tau)$	Probability of casualty i being in triage state $T \in L$ at time τ
T	Triage state
w_T	Weight associated with triage level T
sp^{prob}	Probability of a priority 3 casualty self-presenting
$sp^{attract}$	Parameter describing how the attractiveness of a hospital varies with distance

sp^{wait}	Length of the interval over which self-presenters' waiting times will be uniformly distributed
ν	Specific injury type
$k\nu, T$	Hospital capability penalty for injury type ν and health T
β_c	Hospital capability penalty for casualty c
w_i	Weight of objective f_i when composed with other objectives
$o_i(t)$	ith task selection criteria, a function of task t
P	Set of all priority neighbours
$p_{t,b}(s)$	Priority neighbourhood operation
IN	Set of all insertion neighbours
$in_{t,\hat{t}(s)}$	Insert neighbourhood operation
H	Set of all hospital neighbours
$h_{t,h}(s)$	Hospital neighbourhood operation
SW	Set of all swap neighbours
$sw_{t_1,t_2}(s)$	Swap neighbourhood operation
k	Inter-neighbourhood search iteration in VND
k_{max}	Maximum inter-neighbourhood search iteration in VND
τ^*	Real-time threshold for local search termination
M_{off}	The offline model
M_{on}	The online model
τ_i^*	Simulated discovery time of casualty c_i
λ^{cas}	Casualty discovery rate
λ^{sp}	Self-presentation rate
e^{tri}	Error in triage assessment
$\theta_{j,i}$	True duration of task t_i at site j
μ_j	Mean of normal sampling distribution for $\theta_{j,i}$
σ_j^2	Variance of normal sampling distribution for $\theta_{j,i}$
$e_{j,i}$	Task duration assessment error
s^2	Variance of task duration assessment error
v_c	Cruise speed of responder units
a	Acceleration of responder units
ν, ξ	Parameters for the log-normal distribution for simulating travel times
G	Graphical representation of travel network
λ^{dist}	Travel network disruption
ς	Variance parameter for autonomous routing simulations
α	Average improvement of autonomous routing
β	Variance of improvement of autonomous routing
λ^{del}	Parameter governing the simulation of communication delays
Ψ	Confidence parameter in estimating task durations

γ	Exponential smoothing parameter for travel time estimation
λ_1	Experimental design parameter - error in the triage classification process
λ_2	Experimental design parameter - time between triage assessment
λ_3	Experimental design parameter - variation in task durations
λ_4	Experimental design parameter - error in task duration estimates
λ_5	Experimental design parameter - task duration confidence level
λ_6	Experimental design parameter - communication delay
λ_7	Experimental design parameter - network disruption

Chapter 1

Introduction

1.1 Motivation

The emergency response to recent large scale disasters such as the London bombings of July 7th 2005 have been criticised for a lack or coordination [81]. Coordination could be improved through better decision making, but the large scale emergency environment is a challenging one in this respect. Decision makers are subject to significant stress [73], information upon which to base decisions may be scarce, volatile or erroneous, and the decisions to be made are of a complex nature with many inter-dependant components.

One potential means with which to improve decision making is through the use of decision support systems [120]. Optimisation, using a mathematical model of the real problem and an appropriate algorithm to automatically generate solutions to that model, represents one such form of decision support. By explicitly encoding the problem in a quantitative manner, mathematical modelling allows solutions to be examined and compared in a systematic way. While an individual or team of people would rely on heuristics and basic rules-of-thumb to break down large problems into manageable quotients to be solved individually, an optimisation approach can consider a much larger problem at once. This can in turn lead to solutions which are of a much higher quality. Success of applying optimisation through mathematical models can be seen in a wide variety of similar contexts where decision problems are complex and unlikely to be solved to a high degree by an individual. Examples include inventory management, vehicle routing, task scheduling, facility locating and network flow problems.

The scope for optimisation and mathematical modelling to improve decision making in large scale emergency response provides the motivation for the work described in this thesis. At the most basic level, this thesis aims to answer the following question:

> To what extent can the coordination of large scale emergency response operations be improved through the use of optimisation?

This question is clearly wide ranging, encompassing a broad selection of problems and approaches to solve these problems. What is meant by a 'large scale emergency response operation'? Indeed, what constitutes an emergency? What make it large scale? What is a response operation? In addition to these questions regarding the problem to be addressed, questions also arise regarding the methodology to be applied to the problem. In particular, what type of optimisation methodology should be considered? Answers to these questions will be provided throughout Chapters 1, 2 and 3 of this thesis, with the motivating question evolving in a corresponding manner.

1.2 Aims and objectives

The aim of this thesis is to answer the question above. While we may not be capable of providing a definitive answer in this body of work, identifying a number of clear objectives to be pursued will help ensure that important and useful insights are gained. These objectives are:

1. identify a particular decision problem faced in large scale emergency response which shows potential to be amenable to optimisation based decision support;

2. develop and implement a mathematical model of this decision problem;

3. implement an appropriate solution methodology such that high quality solutions to the mathematical model can be found in a timely manner;

4. develop and implement a stochastic simulation of the large scale emergency response environment to be run in conjunction with the model;

5. evaluate the proposed model and solution algorithm through a Monte Carlo analysis using the simulation.

1.3 Scope

The scope of this work is limited to considering the response component of the disaster management cycle. The cycle, consisting of four stages stages of response, recovery, preparedness and mitigation, will be discussed in Chapter 2. Moreover, within the response stage the work will focus on the first few hours of the response operation.

This work will further focus on *large-scale* emergencies, to be distinguished from the routine incidents to which the emergency services must respond on a day-to-day basis. Further discussion of how the scale of an emergency is defined and interpreted can again be found in Chapter 2.

1.4 Importance

The field of disaster management is relatively young, but emerging in its importance [116]. Evidence of this may be seen both within the academic community, where an ever-increasing selection of conferences and journals provide outlets for academic research, but also in the wider public conscience. For example, the Financial

Times magazine cited disaster management as one of ten key research fields within science in 2011 [27].

As will be discussed in more detail in Chapter 3, a substantial amount of research has been conducted into the application of mathematical modelling and optimisation to problems faced within disaster management. However, due to the young age of the field, many key problems have yet to be tackled in this manner. As such, further research in this field which focusses on an as yet untreated problem will make an important contribution in terms of understanding the potential for mathematical modelling to be used effectively in this environment.

As the body of research in this area grows, and the understanding of the potential of optimisation and related techniques within disaster management improves, the likelihood of sophisticated decision support systems being implemented and used in the field also grows. Given the potential for such a system to lead to improved response operations, resulting in improvements in terms of the number of fatalities or the amount of suffering endured, this underscores the importance of this research. Moreover, by thoroughly assessing the limitations of such systems in specific settings, we can hope to avoid repeating previous mistakes where decision support systems have been implemented and subsequently led to a worsening of performance (for example, the London Ambulance Service Computer-Aided Despatch system (LASCAD) [50]).

1.5 Context - the REScUE project

This research forms part of a larger EPSRC funded project entitled 'Adaptive Coordinated Emergency Response to Rapidly Evolving Large-Scale Unprecedented Events (REScUE)' [24]. In parallel to the work described in this thesis, research has been conducted in the design and implementation of an agent-based simulation (ABS) of large-scale emergency response. Details of the resulting software and it's use can be found in, e.g., [62]. Collaboration between the two work streams has occurred throughout the project, particularly when developing an understanding of the problems faced in large-scale emergency response. However, the modelling of these problems has been carried out using different methodologies and with different goals. While the aim of the ABS research was to develop an appropriately realistic simulation of the problem environment and associated response operation as it would occur in reality, the focus of the work presented in this thesis is to explore to what extent optimisation techniques may be applied in these problems and how the utility of any such approach depends on key problem characteristics.

The REScUE project included a number of partners from local emergency planning offices. These included representatives of Cleveland Emergency Planning Unit, Tyne & Wear Emergency Planning Unit, Co. Durham & Darlington Civil Contingencies Unit and Government Office for the North East. Furthermore, a number of emergency response practitioners also contributed to the guidance of the project, including representatives of Co. Durham & Darlington Fire and Rescue Service, Cleveland Fire and Rescue Service, North East Ambulance Service, Northumbria Police, and Tyne & Wear Fire and Rescue Service. Discussions were held with project partners at six-month intervals throughout the project's life in the form of steering group meetings, where developments in the research were presented and feedback was obtained. In addition to these regular meetings, several visits to emergency planning and response offices were organised in order to gain further detailed feedback and input.

The simulation software developed as part of the REScUE project, entitled STORMI (Simulating Tactical and Operational Response to Major Incidents in the UK), includes a GUI enabled program which facilitates the defining of hypothetical major incidents, the environment in which they take place and the resources available to respond to them. This aspect of the software has been employed in this work stream. The future potential for the simulation capabilities of STORMI to be coupled with the optimisation modelling described in this thesis will be discussed in Chapter 8.

1.6 Contribution of the thesis

As discussed in Section 1.1, the research described in this thesis has been motivated by the following question -

> To what extent can the coordination of large scale emergency response operations be improved through the use of optimisation?

In answering the above, several significant and original contributions to knowledge have arisen. Casualty processing, a significant yet previously untreated problem, has been identified and defined through consultation with both emergency response practitioners and the academic literature. A multi-objective combinatorial optimisation model equipped with a real-time problem interface and solution methodology is proposed, capable of online use which allows for its application to the dynamic and uncertain problems encountered in this area. A detailed Monte Carlo simulation of casualty processing is also provided, with extensive computational analysis providing guidance regarding the utility of the centralized decision support for casualty processing over a broad range of problem characteristics and system constraints. Further details regarding each aspect of this contribution follow.

1.6.1 Identification of casualty processing as an important decision problem

The problem of coordination in disaster response is a large, varied and unstructured one. This thesis presents a discussion of the general problem environment, based on consultation with both emergency planning/response practitioners and associated practitioner documentation, identifying key features to be considered as decision variables, constraints or objectives in a novel mathematical formulation of the problem. Previous approaches into the design and implementation of decision support systems for use in disaster management are evaluated in this thesis, with work considering decision problems relating to casualty rescue, casualty treatment and hospital allocation, both individually and in isolation, reviewed. This thesis describes the first research to consider all elements together in a single casualty processing formulation. We hypothesise that such an integrated approach will maximise the potential for an optimisation-based approach to deliver increased performance, in terms of key metrics such as the numbers of fatalities and the amount of suffering endured by the casualties of the MCI.

1.6.2 A multi-objective combinatorial model of casualty processing

We propose combining the multiple objectives present within emergency response using a hierarchical framework, where lexicographic ordering is used to combine the high-level goals of minimising fatalities, minimising suffering and maximising efficiency. Each of these goals is formed of individual objective measures, combined using the weighted metric method. We provide the first model which acknowledges together fatalities, the time to arrival at hospital, the stochastic nature of casualty health, hospital resource levels and over-subscription, the role of specialist treatment facilities, autonomous self-presentation, and the spatial nature of the problem. Furthermore, the model accounts for the uncertain nature of key parameters, providing methods with which to forecsat their value over the course of the MCI response operation.

1.6.3 An interface between the model and real-time problems

Previous research in this domain has allowed for dynamic modelling in only a limited sense, with model parameters being updated to reflect changes in the problem environment at infrequent intervals of the order of hours. In this thesis we describe an interface between the model and the problem environment which allows for information to be passed in both directions with arbitrary frequency.

1.6.4 A disaster response simulation

Given the inherently dynamic nature of the response problem, a real-time disaster response simulation has been developed. This allows for the simulation of multi-site incidents occurring over time, the gradual accrual of known casualties due to search, the staggered arrival of responders through mutual aid, and the revision of all predicted quantities including transportation time, task durations, self-presentation and casualty health evolution. Together, this enables simulations of realistic situations with minimal assumptions.

In addition to simulating the problem environment, a heuristic algorithm controlling the decision making process in a manner reflecting current practice is described. Employing this algorithm allows for Monte Carlo simulations of the response to any given problem instance, the results of which provide a baseline comparison with which to judge any alternative decision making approach such as that enabled by employing the proposed optimisation model.

1.6.5 A solution method

In order to exploit the power of the proposed model and interface, a local search optimisation framework of the problem is described, including specification of a number of neighbourhood structures and a metaheuristic search strategy to enable an efficient optimisation process capable of finding high quality solutions to the problem. The optimisation process is closely linked to the real-time interface, allowing for the unification of the usually conflicting goals of finding a solution of high quality and finding a solution in a timely manner.

1.6.6 Evaluation of the model across problems and system constraints

The model, simulation, interface and solution method discussed in this thesis allow for extensive Monte Carlo computational experiments to be performed in order to analyse the implications of current and future policy in disaster response. Accordingly, a number of studies are presented which explore questions including: Can centralized decision making deliver benefits over decentralized? Can an optimisation model deliver benefit in evolving, dynamic problems? And, what underlying factors have significant predictive relationships with the utility of the optimisation model?

1.7 Overview of the thesis

In Chapter 2, the problems faced in large scale emergency response are discussed. Through exploring the definition of a large scale emergency in terms of both its cause and its effects, a sufficiently narrow focus is developed in terms of the specific problem of Mass Casualty Incidents (MCI). Decision making in MCIs is reviewed in order to provide the basis for development of a mathematical representation. In Chapter 3, the academic literature on decision support systems for large scale emergencies is surveyed. Through this process, a particular decision problem which has yet to be addressed in the literature is identified, and common modelling techniques used in addressing some of the fundamental characteristics of large scale emergencies are reviewed.

Moving on to the development of novel research, Chapter 4 describes the multi-objective combinatorial model designed for use in MCI response. This includes descriptions of both the decision space and the objective space. Algorithmic approaches to generating solutions to this model are then discussed in Chapter 5, which includes experimental results regarding the parametrisation of these algorithms. In Chapter 6, a simulation of the MCI environment is described. This simulation is designed to be coupled with the optimisation model via an appropriate interface and allow for more realistic scenarios to be considered, effectively reducing the number of assumptions necessary to use the model. This feature is then made use of throughout Chapter 7, where a number of computational experiments are described. Following this evaluation, the work is summarised and critically appraised in Chapter 8, where directions for future research are also suggested.

Chapter 2

Problem

Decision making in large-scale emergencies

2.1 Introduction

The term 'large scale emergency response' is a broad one, encompassing a wide range of scenarios and the decision making which takes place within them. In order to proceed with the design and implementation of an optimisation model for these problems, it is necessary to first define a suitable focus and scope of the work. In this chapter the range of problems encountered in large scale emergency response will be discussed. In particular, potential problem scenarios are broken down by their causes, their effects, and their size, in order to better understand their relation to each other and enable a suitable set of scenarios to be defined as the focus of this research. In the process, the notion of what constitutes an emergency 'response' will be defined, placing it in the context of the larger disaster management cycle.

Having discussed the range of potential problems, a focus on a specific class, Mass Casualty Incidents (MCIs), will be introduced. Following this, the decision problems encountered in MCI response will be discussed in order to identify potential decision problems which could be supported through mathematical modelling and optimisation. This discussion of decision making will include relevant background regarding key variables in the MCI response environment, such as the tasks which must be completed and the resources available to assist in their completion. Finally, current use of decision support systems in MCI response is reviewed.

2.2 Large scale emergencies

As outlined in Chapter 1, the focus of the research described in this thesis is the development of an optimisation model for 'large scale emergency response'. In order to proceed, we must first understand exactly what is meant by these three terms. That is,

- What is an emergency?
- Conditional on this, what then defines a large scale emergency?

- Finally, what constitutes the response to a large scale emergency?

These questions will be explored in turn. Firstly, we introduce response as part of a larger recognized disaster planning cycle. Following this, the different causes and effects of emergencies will be reviewed. Finally, the question of defining scale will be addressed. Throughout, the focus of this research will narrow as specific types of emergencies of a particular magnitude are highlighted and chosen as the primary motivation for the remainder of the thesis.

2.2.1 Response and the disaster management cycle

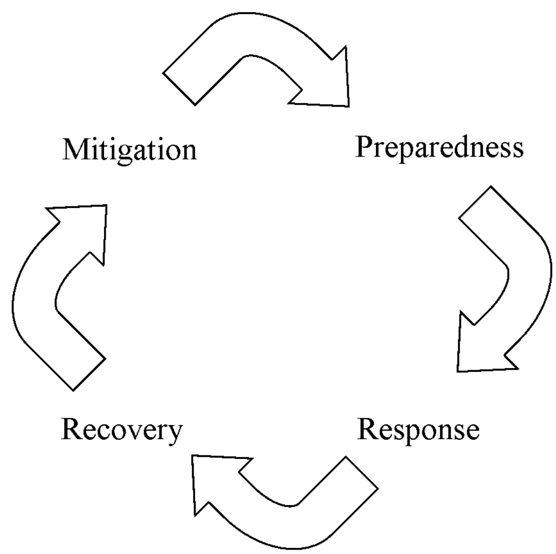

Figure 2.1: The four phases of disaster management.

Disaster management is an ongoing process, for which response, the focus of this thesis, is one of four key stages described in [28] and illustrated in Figure 2.1. However, the stages are not easily divisible into discrete and independent parts, either temporally or by considering those individuals or organizations involved. As such, in order to clearly define the scope of this work, it is important to introduce the whole disaster management process. The first stage, mitigation, involves taking steps to reduce the risk of disasters happening in the first place. Examples of this include:

- Installing flood levees to prevent a rise in sea level causing a flood in a populated area.
- Improving building design to better withstand the effect of earthquakes.

In preparedness, the subsequent stage, efforts are focused on doing what we can prior to the disaster occurring to aid performance in the subsequent response phase. While mitigation aims to reduce the chance of the disaster occurring in the first instance, preparedness aims to limit the effect of said disaster, assuming it will in fact happen at some point. Examples of preparedness measures include:

- Providing civilian training in how to cope, e.g. first aid and evacuation.
- Maintaining supplies of necessary resources, such as medical supplies, in key strategic locations.
- Running live, simulated training exercises for emergency response personnel.

Response operations are initiated upon acknowledgement of the disaster and comprise of all actions subsequently taken to limit the impact. As such, response is closely related to preparedness in that it addresses the same problem, with preparedness the focus before the event and response the focus during the event and in the short term afterwards. Examples of response activities include:

- The search for and rescue of civilians trapped in dangerous environments following an earthquake.
- The treatment of casualties, by emergency services or others such as trained civilians or voluntary organizations.
- The management of hospitals to free up more resources such as beds and operating rooms for the urgent cases resulting from the disaster.

Finally, in the longer term following the response to a disaster, recovery operations are carried out. During this phase, priorities will have shifted from those held during response, from focusing on saving lives and preventing suffering to carrying out work to return the affected area to its previous functioning state, or indeed an improved state. This reflects the link between the recovery and mitigation stages as indicated in Figure 2.1, as recovery efforts could include, for example, the strengthening of flood levees damaged in the disaster. Examples of recovery work include:

- Repairing damaged infrastructure such as transport networks and energy supply lines.
- Assisting displaced survivors in contacting relatives.
- Delivering essential supplies, such as food and water, to areas cut off from usual supply routes.

As is clear from the examples given, the disaster management cycle can not be cleanly divided into independent steps. Rather, it will be common for efforts to be undertaken by multiple organizations in several parts of the cycle at a single point in time. As such, while the stated focus of this thesis is coordination during emergency response, it is important to remain aware of the other related and possible dependant activities which will be carried out at the same time by the same organization. The key factor which will constrain our focus is that of time, in that all decision making to be considered will be carried out no earlier than upon notification of the disaster, and no later than the point where all that can be done to alleviate physical suffering has been completed. It is in this respect in which we define 'response'.

2.2.2 Causes and effects

Causes

In the U.K., the Cabinet Office [129] list fourteen types of emergency, arranged into three different groups:

- Natural events: Severe weather; Coastal flooding; Inland flooding; Pandemic human disease; Non-pandemic human disease; Animal disease.
- Major accidents: Major industrial accidents; Major transport accidents.
- Malicious attacks: Attacks on crowded places; Attacks on critical infrastructure; Attacks on transport systems; Non-conventional attacks; Cyber-attacks.

For each type of emergency, a different level of preparedness work will be deemed appropriate. This is determined through assessing the risk of the event, that is (i) the predicted impact together with (ii) the estimated probability of the event taking place. The UK National Risk Register [129] plots emergencies on these two dimensions to communicate these risks. As both impact and probability of occurrence will vary according to location, a wide range of preparedness levels are exhibited across the UK when all emergency types are considered. Indeed, the national risk register is refined at the local level via the community risk register to better reflect local conditions.

As can be seen from the list above, the number of events considered to be emergencies is clearly of a dynamic, evolving nature. In recent years, some events such as attacks on crowded places have risen in frequency and prominence, including such major examples as the World Trade Center attacks [25], the London bombings [81] and the Madrid bombings [123, 122, 33, 53]. The nature of many natural events is evolving in accordance with the changing environment. However, to some extent, the list of potential emergencies can be consolidated by focusing on the *effect* as opposed to the *cause*. As noted in [37], *"Disasters do not cause effects. The effects are what we call a disaster."*

Effects

Much of the primary motivation for this research arose from the London bombings and the associated response. The cause of this disaster was an 'attack on a crowded place'. The effect was largely characterized by the death and suffering caused as an immediate consequence of the blasts, and by the damage to infrastructure (in this case, the underground train network). It is the former effect which will drive the remainder of this research.

In terms of causing suffering, the London bombing led to approximately 700 injuries [80], with 340 of these being serious enough to require transport to hospital. These casualties were in addition to the 56 fatalities which resulted from the attack, including the four suicide bombers themselves. As a result, the emergency services were tasked with delivering treatment to this large group of injured people, a task which was complicated by the fact that much of the incident occurred in the enclosed environment of the underground train network. Furthermore, the extraction of those casualties unable to walk required the services of Fire and Rescue. Following their extraction, the Ambulance and the Health services were required to deliver treatment, which took place both at the incident sites and in nearby hospitals. The hospitals themselves were also required to enact emergency plans, in order to increase their capacity to deliver this treatment.

In addition to extracting and treating the injuries of the wounded, all fatalities required the attention and efforts of available response services. Specifically, fatalities had to be pronounced dead by qualified health service personnel before being transferred to an appropriate hospital. Many civilians who did not suffer significant injuries also required attention, which was delivered through survivor reception areas.

While this real-life incident provides a key motivation for this work, we do not wish to define our focus as 'terrorist attacks' or 'attacks on a crowded place'. Rather, we note that this incident was also classified as a 'Mass Casualty Incident'. Many other events with a different cause could also be classified as an MCI, exhibiting many of the same characteristics as those already described. For example, an earthquake could also lead to a large number of casualties, fatalities and survivors, many of whom may be trapped in a dangerous environment following a structural collapse. By focusing on MCIs, then, an appropriately broad view is taken which will encompass a large number of potential disasters in the UK.

2.2.3 Scale

The causes, effects and risks of large scale emergencies outlined in Section 2.2.2 provide some limited means to classify events. Considering more dimensions of variation will lead to a more detailed classification system, such as that proposed in [9]. Here, the authors propose using the following five factors when describing and classifying large scale emergencies:

i. type, i.e. natural or man-made,

ii. duration,

iii. degree of personal impact,

iv. potential for occupancy,

v. control over future.

Similar systems have been proposed in [32] and [54]. In the classification given in [9], dimensions (ii), (iii) and (v) all correspond to the size or the extent of the event. In addition to these measures, another intuitive factor which corresponds to the scale of an emergency would be the geographic area it covers. This is indeed one of the measures employed in [54], where emergencies are classified according to both the size of area affected and the number of people affected. A classification is imposed on this two dimensional scale, with labels 'small', 'medium', 'large', 'enormous' and 'gargantuan'. The classification is done in such as way that an increase in only one dimension is sufficient to progress to the next label. That is, an increase in geographic area affected (or, equivalently, number of people affected) is enough to proceed to the final label of gargantuan. In reality, we would expect some degree of automatic correlation between the two measures, i.e. as the area affected by an event increases so will the number of people affected increase. For example, considering the case of terrorist attacks, an event will affect a geographical area through affecting the people in that geographical area. Exceptions to this would include those events which affect largely unpopulated areas, such as the recent bush fires occurring in Australia [17].

A similar scale and classification is employed within the UK, and is set out in the UK 'Concept of Operations' document [128]. Geographic area affected is used as one of two dimensions, with the other defined as 'impact'. However, the term 'impact' is not entirely explained, and so use of the scale involves some degree of interpretation. Example classifications are given, where the London bombings, the 2007 UK floods and the 2009 avian H1N1 flu pandemic are all denoted 'serious'. In contrast to the classification of [54], the Concept of Operations scale requires an event to increase in both dimensions (area affected and impact) in order to progress to the next label of severity. The three labels used are 'significant', 'serious' and 'catastrophic', to which we add an initial label 'local response' as illustrated in Figure 2.2.

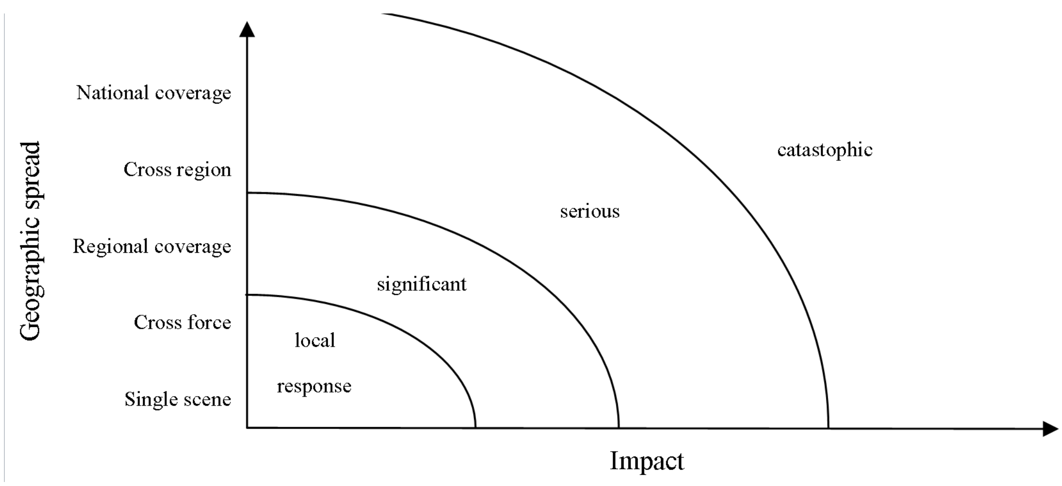

Figure 2.2: Classification of emergencies by scale, according to the UK Concept of Operations [128].

Having focused upon Mass Casualty Incidents in Section 2.2.2, we may now further define our scope considering the dimensions used in the Concept of Operations. Considering the geographic area, we intend to focus on a similar area of London as was affected by the 2005 bombings. In terms of impact, a lower limit of 200 casualties will be considered throughout. This figure will ensure that the results of this research will indeed be applicable to the large scale events which provide its motivation. Casualties are defined as those who have received injuries severe enough to require treatment at hospital.

To summarise, we have explicitly identified a range of scenarios, henceforth refereed to as 'Mass Casualty Incidents', which will provide the focus for our subsequent work. Specifically, we will consider problems where over 200 people are affected (that is, injured or immediately killed) in a densely populated urban area of approximately $50km^2$ size, and where the response is entirely focused on these casualties as opposed to other environmental hazards such as fire or CBRN damage.

2.3 Decision making in MCI response

2.3.1 Command and Control

The coordination of multiple emergency response agencies during the fast-paced environment of disaster response has been noted to be a challenging task [44, 26]. In order to better cope with this challenge, decision making in the response to an MCI is carried out within a structure common to all large scale emergencies, known as command and control. Under this structure, decision making is divided into three levels, each corresponding to a further level of granularity. The structure is of a hierarchical nature, as illustrated in Figure 2.3.

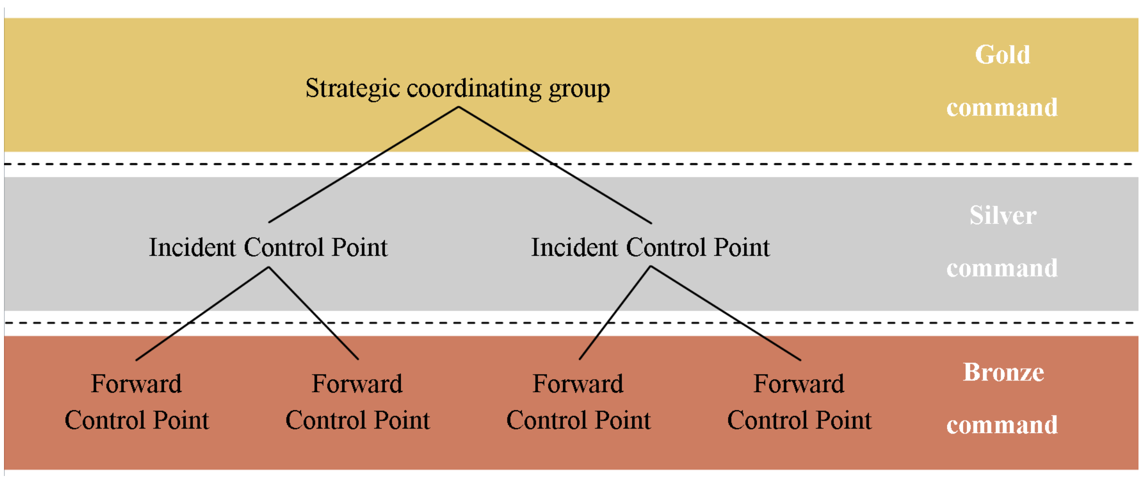

Figure 2.3: The hierarchical structure of Command and Control in the UK.

Command and control is of a collaborative nature, occurring across and within each of the main emergency response organizations. Each emergency service place their own commanders at each level, who coordinate with their counterparts as appropriate. As set out in [128] and [125], the three levels of command and control are summarized as follows:

- **Operational** (Bronze) command will usually be located at an on-scene Forward Control Point or close to the incident. The operational commander will concentrate on the tasks within their area of responsibility. For example, an operational commander in the police may focus on the setting up of cordons around the disaster area. At this level, first responders will take immediate measures to save lives before going on to making an assessment of the situation. Following this assessment, the operational commander will decide whether or not a silver command level is warranted.

- **Tactical** (Silver) command ensures coordination at the operational level. A tactical coordinating group (composed of silver commanders from each responding agency) will be set up to determine priorities for the allocation of resources to bronze command teams. Furthermore, this group will plan how and when tasks will be undertaken, obtain additional resources as necessary, and assess the level of risk. Other responsibilities include deciding where to locate key functions and facilities, such as ambulance loading points, helicopter landing sites, marshalling areas for the assembly of equipment, and body holding areas. The tactical coordinating group will operate from an incident control point, preferably a permanent office although a mobile command unit may be used if necessary. In incidents with multiple incident sites, multiple incident control points will be set up. Incident control points should be located nearby or adjacent to the incident site, with an alternative location identified as a back-up. The tactical coordinating group will assess the situation and decide whether or not an additional level of command (i.e. strategic) is appropriate.

- **Strategic** (Gold) command is carried out by a strategic coordinating group, composed of strategic commanders from the various responding agencies. Located at a pre-planned location, remote from the incident, the strategic coordinating group will develop clear strategic aims and objectives and prioritizes the allocation of resources to the tactical command level.

Following the discussion of scale given in Section 2.2.3, the types of MCIs to be considered in this research would invoke command and control at each stage of the hierarchy. Regarding the potential applicability of an optimisation model, there is potential for impact at each level, although particular attention should be given to the tactical command level. The types of decisions described at this level, namely resource distribution and facility location, are problems which have been studied in a generic form by the operational research community for decades [90, 83]. Another decision which should be highlighted is that of routing, occurring at the operational level, which again has received much attention in its generic form [39].

2.3.2 Objectives

If effective decision support is to be given in MCI response, it is vital to understand the objectives held by the relevant response organizations. In the UK, there are thirteen objectives common to all organizations and agencies involved with the local emergency response [128]. These are:

(i) saving and protecting human life

(ii) relieving suffering

(iii) protecting property

(iv) providing the public with information

(v) containing the emergency - limiting its escalation or speed

(vi) maintaining critical services

(vii) maintaining normal services at an appropriate level

(viii) protecting the health and safety of personnel

(ix) safeguarding the environment

(x) facilitating investigations and inquiries

(xi) promoting self-help and recovery

(xii) restoring normality as soon as possible

(xiii) evaluating the response and identifying lessons to be learned

Given any problem guided simultaneously by several objectives, it is not a trivial task to combine them to provide some coherent, single direction to base decisions upon. In terms of the objectives listed it is noted in Concept of Operations that "their relative priority may shift as the emergency develops", and that government ministers "advise on the appropriate balance to strike in light of the circumstances" [128].

The list of objectives is designed to cover the full scope of potential emergencies, as discussed in Section 2.2. As such, the list may be filtered in order to better reflect the focus of MCIs set out in Section 2.2.2. Given that the incidents considered do not involve fire or chemical elements, the objectives of protecting property (iii) and safeguarding the environment (ix) do not apply. Furthermore, objectives (x) and (xiii) will be disregarded as they will only come into play after the response operation has been completed. The objective list can be further simplified by noting its hierarchical structure. Although the list is given with no specified priorities attached, it is clear that in any MCI the objectives of saving and protecting human life (i) and of relieving suffering (ii) will dominate all others. Objectives such as (iv), providing the public with information, can be viewed as operating at a lower level, where work contributing in this direction will ultimately help in terms of the two higher level objectives. This is also the case for objectives (v) - (viii), (xi) and (xii).

The incorporation of relevant objectives within decision support models proposed through the research community will be discussed in Chapter 3. For now, it is worth noting that these practical objectives are often omitted. For example, the model proposed in [14] aims only to keep total time taken to complete all response operations as low as possible. On the other hand, the model proposed in [124] incorporates several

Table 2.1: Tasks carried out in MCI response, grouped by type.

Type	Tasks
Logistics support:	collecting supplies from supply points; delivering supplies to points of demand; setting up distribution centers; setting up equipment marshalling areas; assembling vehicles and equipment.
Gathering information:	performing risk assessments; searching for casualties; assessing the injury level of casualties (i.e. performing triage); searching for secondary devices.
Dealing with the healthy	setting up and running survivor centers; setting up and managing cordons; re-designation of traffic flow; evacuating areas.
Dealing with the wounded	setting up casualty clearing stations; designating ambulance loading points; transporting the wounded; rescuing trapped casualties; administering on-site treatment; designating helicopter landing points.
Tackling the disaster environment	pumping out excessive water; safely removing contaminated water; fighting fires; clearing obstructions; stabilizing damaged structures; bomb disposal.
Fatality management	certifying death; setting up body holding areas; transporting fatalities.

objectives, including financial cost, total time of operations and a notion of fairness. However, this still falls short of explicitly accounting for the two key objectives of protecting life and relieving suffering.

2.3.3 Tasks

Through reviewing relevant documentation for emergency planners and responders in the UK [127, 126, 87, 128, 125] a wide range of tasks which may require completion during an MCI response operation can be identified. As the design of an optimisation model will depend heavily on which of these tasks are included, we set out in Table 2.1 a list of these tasks grouped into a number of categories. This list is not comprehensive; rather, it is presented in order to provide the reader with a clear impression of the breadth of potential decision problems which could potentially be included in this work.

As in the case of objectives, the set of all possible tasks in an MCI may be further refined to meet the specific scope set out in Section 2.2. Those tasks within the category "tackling the disaster environment" will not be considered further in this thesis, as their impact in terms of those objectives of most importance described in Section 2.3.2 will be limited. It is also useful to distinguish tasks in terms of the frequency with which they must be carried out. In particular, a natural division emerges, seperating tasks into those of a one-off nature occurring during the setup of a response incident and tasks which are repeatedly carried out for the remainder.

Table 2.2: Tasks carried out in the setup and execution of MCI response.

Setup	$\rightarrow$	Execution
Distribution centers		Collection and delivery
Marshalling areas		Assembling equipment
Survivor centers		Risk assessments
Cordons		Search and rescue
Casualty clearing stations		Triage
Ambulance loading points		Treatment
Helicopter landing points		Transportation to hospital
Body holding area		Certifying death

Table 2.3: Resources used in MCI response, grouped by type.

Type	Resources
Personnel	police officer; paramedic; general practitioner; specialist medical staff; firefighters (wholetime and retained); local authority emergency officer.
Equipment	rescue dogs; firearms; mobile medical equipment; urban search and rescue pods; high volume pumps; temporary shelters; ladders and aerial platforms.
Vehicles	police cars; ambulances; fire appliances.
Consumables	food and water; medical supplies; clothing; blankets.

Accordingly, this structure suggests a decomposition of the entire decision process in two stages, *setup* and *execution*. These stages are summarized in Table 2.2.

Tasks pertaining to the setup of the response are primarily of a *facility location* type of decision problem. The choice of location for all the listed features could impact the performance of the remainder of the response operation. In the execution phase, each task must be carried out repeatedly and so, while any individual decision relating to a single task will have limited impact upon performance, a strategy which together dictates how all such tasks should be completed will have a significant impact.

2.3.4 Resources

Associated with the set of potential tasks discussed in Section 2.3.3 are the resources which are employed in their execution. By resources, we include all vehicles, equipment, personnel and consumable commodities such as blankets and food. As with tasks, we present a non-exhaustive list of resources which are involved in large scale emergency response in Table 2.3

Again, this set of resources may be refined to reflect the specific nature of MCI response. In particular, consumable supplies of food and water, clothing and blankets will be important when considering response operations lasting a significant length of time, but will be of less importance in cases where the response operation is completed

Table 2.4: Environmental factors in MCI response, grouped by type.

Type	Environmental factors
People	health level; demographic information (students, ethnicity, faiths, vulnerable); civilian local knowledge; responder local knowledge; autonomous civilian actions (self-help, response activities); 'emotional' factors (modesty, contrariness).
Natural	meteorological data; geographical features (rivers, sea, hills).
Structural	telecommunications networks and their reliability; energy networks and their reliability; transportation networks and subnetworks.
Disasters	explosives; fires; hazardous materials; collapsed or near-collapse buildings; pre-event risk assessments.

in 2 - 3 hours.

2.3.5 Environment

Information about the environment in which the disaster occurs is important when making decisions. Details associated with the environment can include natural (e.g. time of day, weather), structural (transportation networks, notable buildings like schools) and human (demographic data, civilian behaviors). All environmental details are external and not under full control of emergency responders.

The classification of the health of people in particular requires elaboration. In MCI response in the UK a triage system is employed when measuring the health of casualties. The purpose of triage (derived from the French word 'trier', to sort) is to partition casualties into a number of categories which reflect the urgency with which they require treatment. The resulting information can then be used when deciding how to allocate scarce resources to a large number of casualties, prioritizing those who are likely to benefit most.

Two triage systems, namely triage sieve and triage sort, are used in UK MCI response, each working at a different level of granularity. Triage sieve is carried out immediately following an MCI and must be completed before any treatment can take place. The outcome is the classification of each casualty into one of four categories as described in Table 2.5 [1]. A physical label, color coded according to the triage category, is affixed to the casualty to allow for rapid recognition by other responders during the remainder of the response operation.

These categories are used in the first stage of the MCI response, where casualties must be extracted from the incident site and taken to a designated safe area, close to the incident site, where basic treatment can be administered in order to stabilize casualties and prepare them for transportation to an appropriate hospital. This area is known as the Casualty Clearing Station (CCS). At the CCS, a further, more detailed triage operation is carried out: triage sort. Triage sort allows for further detailed discrimination within the T1 class. A 0-12 integer scale is used, where 0 corresponds

Table 2.5: Triage levels assigned to casualties [1].

Category	Description	Explanation
T1	Immediate	Require immediate life-saving procedure
T2	Urgent	Require surgical or medical intervention within 2-4 hours
T3	Delayed	Less serious cases whose treatment can safely be delayed beyond 4 hours
Dead		

to Dead, 1-10 denote varying levels of severity within T1, and 11 and 12 denote T2 and T3 respectively. The triage sort score is used to prioritize casualties as they are transported from the CCS to a hospital.

The categorization of casualties according to triage sieve and triage sort provides a natural representation of casualty health to be used in any mathematical model of MCI response. However, we note two aspects of this representation which pose challenges to its successful use. Firstly, despite the extensive training of Ambulance Service staff, the outcome of any triage operation is subject to error, both from measurement and from bias. For example, there is a documented tendency to assign children to triage states of higher severity than is actually warranted [52]. When providing decision support based on triage information, the effect of such errors should be monitored closely. In addition, we note that the health of casualties is typically dynamic, and in particular is likely to deteriorate over time. This is reflected by current practice in triage sort, where casualties are re-assessed every 15 minutes in order to monitor any changes in their health. Predicting such dynamic behavior is a significant challenge in its own right; again, the sensitivity of any proposed model to inaccuracies in such predictions should be assessed.

2.4 Decision support in MCI response

2.4.1 Potential benefits of decision support

The potential for decision support systems to assist in MCI, and other major incident, response has been noted by [76]. Several key points where benefits could make themselves felt can be identified.

Firstly, it is known that decision makers in MCI response environments are placed under a significant amount of stress [94, 73]. Providing automatic suggestions regarding the many small decisions to be made could relieve some of this stress, in turn allowing for decision makers to invest more of their time and attention to those aspects of the response operation which are not amenable to assistance via optimisation models.

Decision support systems are also capable of providing a centralized view of the entire problem and associated decision space, conditional on sufficient data being available. This has been well documented in other problem domains, where humans faced with a large, complex decision space can struggle to find the best possible option. Rather than systematically identifying each option and comparing them all, heuristics are employed in order to arrive at a sensible solution quickly. Employing methods which can search a vast array of options in a systematic way could lead to better overall response plans than would be arrived at otherwise.

It should be noted that the context of MCI response is one which would be very sensitive to the performance of a decision support system. Given the high importance attached to the objectives of saving and protecting human life, it is clear that any system which could lead to modest benefits in these respects would be of considerable value. Equally, however, any errors resulting from the use of a decision support system, as quantified in this manner, would be considered intolerable.

2.4.2 Current decision support systems

Currently, decision support systems are rarely employed by the emergency services in the response phase. Information systems, on the other hand, are widely used. The purpose of information systems is to help organize, filter and present information which could be useful to decision makers as they move through the decision making process. Examples of information systems include those of a Geographic Information System (GIS) nature used to record and forecast flood coverage. GIS is also used in conjunction with GPS location devices attached to vehicles in order to track in real-time the position of resources. Computer Aided Dispatch (CAD) systems have been implemented successfully, although not without issue as demonstrated by the initial high-profile 'crash' of the London Ambulance Computer-Aided Dispatch system [50]. The Command Support System™ [131] is designed specifically for use in large scale emergencies, but again the focus is on assisting the decision making process through better information management and facilitating collaboration, as opposed to automatically locating and evaluating courses of action and suggesting these to the decision makers.

2.4.3 Challenges of decision support development

Given the preceding review of the MCI response environment and associated decision problems, a number of factors which challenge the design and implementation of a decision support system which is both safe and useful may be identified. The structure imposed on the problem by the command and control structure, with well-defined roles and responsibilities existing for all emergency response agencies, will help in the modelling of the decision problem. However, in addition to formulating the decision problem another key component of any decision support system is a mathematical means with which to evaluate and compare any two alternative solutions. This requirement is particularly challenging. Three key characteristics should be borne in mind:

- Speed: Response operations will not be delayed in order to wait for decision support to be generated, and so for any system to be of pracitcal use it is essential that support can be offered in a timely manner.
- Uncertainty: Many parameters will need to be estimated in any mathematical model, and in many cases it will be unrealistic to assume that such estimation will always be completely accurate. The need to include such uncertainty in any evaluation of proposed solutions presents a significant challenge.
- Dynamicity: Even in cases where parameters are estimated with complete accuracy, it may be the case that they are likely to evolve over the course of the response operation. A successful model should therefore require the ability to continuously update its parameters based on information gathered in real time.

The latter two challenges could be sidestepped through the introduction of associated assumptions. For example, assuming that the health of all casualties in an MCI will remain constant over time would allow for a model to be developed without the need to continually update casualty health information. When considering any such assumption, the benefits it brings in terms of allowing for a model to be realistically implemented must be contrasted with the risks associated with it. Throughout this thesis, a key theme will be the removal of many assumptions which are frequently made in similar approaches to decision support, accompanied by extensive assessment of the benefit afforded through their removal.

2.5 Summary

The set of problems which fall under the banner of 'large scale emergency response' is large. Through consideration of the causes, effects and scales of such problems, a focus has been defined for consideration in the remainder of this thesis. Specifically, Mass Casualty Incidents and the decision problems found in associated response operations will be studied. Key parameters to consider have been outlined, including the objectives held in such a response operation, the tasks which are required to be completed, and the resources available to complete them. The potential for decision support systems to assist in such problems has been argued, and the lack of current systems in use noted. Having reviewed the problem from the perspective of the practitioner, in Chapter 3 we proceed to review the academic contribution made through the design and implementation of decision support systems.

Chapter 3

Review

Optimisation models for emergency response

In Chapter 2, a host of decision problems faced in large scale emergencies were identified. A focus on the *response* phase of *Mass Casualty Incidents* (MCI) was specified, which will be maintained throughout the model development and evaluation described in Chapters 4 through 7. In this chapter we seek to review the academic literature relevant to this focus. In particular, we aim to cover the history and the state of the art of research into the design and implementation of optimisation-based decision support models in disaster response. Note that the problems considered in this review (problems of *disaster* response) represent a larger space than the aforementioned MCI response problems. This widened scope is appropriate due to many shared characteristics which decision problems in disaster response share. For example, research focused on the distribution of medical supplies to locations of demand following an earthquake, taking a commodity flow form, must consider factors such as the modelling of the health of casualties and the effect of damage to the transport network. These same considerations will also be important in research designing a model of casualty evacuation, taking the form of a Vehicle Routing Problem.

This review is structured in the following manner. Firstly, two recent papers [5, 55] highlighting common modelling assumptions and assessing their validity are introduced in Section 3.1. These assumptions provide a means with which to classify the remainder of work under review. Secondly, in Section 3.2 a detailed discussion of key papers is given, with the aim of identifying gaps in the research, in terms of the decision problem addressed, to be pursued further in this thesis. Following this, in Section 3.3 a number of key problem features are considered, with discussions provided of how these features have previously been modeled. Similarly, methodologies employed in order to address the uncertainty which characterizes the MCI response environment are reviewed in Section 3.4, and finally, in Section 3.5, solution methodologies employed are surveyed. Throughout, the aim of this review is to both identify good modelling practice which can be built upon in the developments of Chapters 4 - 7, whilst identifying key gaps in the literature which may be filled by novel developments.

3.1 Common assumptions

Some common assumptions made in the design of operational research models for disaster operations management are identified in [55] and reproduced in Table 3.1. In the paper, each assumption was classified according to the degree of realism it affords. Three labels were used for this purpose: reasonable, limited, or unrealistic. Each assumption is labelled to enable easy referral throughout this Chapter, where we use the prefix 'GB' in reference to the authors of the paper.

Further common assumptions covering the more general area of disaster planning are listed in [5], reproduced in Table 3.2. In this case the author did not explicitly label the extent to which each assumption is deemed reasonable, and so the corresponding labels have been added.

These assumptions will be referred to throughout this chapter, as previous research into the application of mathematical modelling for the purpose of decision support in disaster response is reviewed.

3.2 Decision Support Models for disaster response

While there has been a substantial amount of research published in the area of decision support models and optimisation tools for *emergency* response as a whole (see [116] for a review on the subject), only a small subset of it has been concerned with *large scale* emergency response. Where an emergency can include the routine calls made to the fire, ambulance and police emergency services on a day-to-day basis, large scale emergencies involve highly unlikely events with potentially devastating consequences. These specific problems have unique attributes as discussed in Chapter 2, such as inherent uncertainty, large problem size (in terms of geographical area, numbers of responders/casualties/supplies) and involve many highly inter-dependent decisions. Their analysis therefore requires specialist treatment quite apart from 'everyday' emergencies, although clearly synergies exist between the two areas.

Much work has been carried out in the areas of mitigation and preparedness [2], where the objective is to prevent disasters occurring in the first instance (e.g. barrier construction to avoid flooding) and in effectively preparing the community for when one does (e.g. development of communication systems). A survey of OR/MS research in disaster management [2] identified that the area of response, where the objective is to distribute available resources (e.g. medical commodities, personnel, specialist equipment) across the disaster area in a fast, efficient manner, has received less attention by the research community.

Decision problems found in the large scale emergency environment are typically 'large' in a mathematical sense, in that there are many decision variables and associated constraints. This can have implications in terms of the tractability of the models, which in turn can affect the time required for a solution to be produced. In the areas of mitigation and preparedness there is no urgent demand for decisions to be made, and therefore when designing decision support models the time it takes to arrive at a solution is not subject to temporal pressures. However, the response environment

Table 3.1: Common assumptions made when applying OR/MS methodology to problems in disaster response, as identified in [55] .

ID	Assumption	Type
GB1	Immediate availability of supplies and other resource after the occurrence of the disaster	Limited
GB2	Deterministic and given post-disaster travel times, costs and demand. If scenarios are used, these parameters are assumed deterministic and given within each scenario	Unrealistic
GB3	A given set of scenarios whose probabilities and behaviour are based on experts opinions and historic data	Limited
GB4	Static parameters (demand, travel time, costs, etc.)	Unrealistic
GB5	Resistant buildings and network infrastructures	Unrealistic
GB6	For distribution of relief goods and humanitarian supply chains, total resource availability generally considered enough to cope with total demand (otherwise, a penalty for unmet demand is considered as an input)	Limited
GB7	Network topology is given for transportation problems	Reasonable
GB8	Set of candidate locations considered as an input for location-allocation problems	Reasonable
GB9	Statistical independence of events	Limited
GB10	Information about disaster impact (demand, network status, etc.) available immediately after the disaster	Unrealistic
GB11	Perfect information of evacuees about road network and traffic conditions	Unrealistic
GB12	No time-windows delimiting delivery times in relief goods distribution	Unrealistic
GB13	Population divided into categories according to their urgency of attention	Reasonable
GB14	Post-disaster coverage models based on pre-disaster computed distances regardless of infrastructure vulnerability	Limited
GB15	Evacuation area divided into zones, where each zone has a given destination	Reasonable
GB16	In staged evacuations, evacuees from the same area evacuate at the same time	Reasonable
GB17	Poisson distribution for describing the process of disaster occurrence	Limited
GB18	Capacitated roads or links	Reasonable
GB19	Evacuation guidance performed ideally	Unrealistic
GB20	Evacuee's exercise free-will for selecting which shelter to evacuate to and which route to take	Limited

is short lived and dynamic by nature. Consequently, for any decision support model to be of use it is imperative that it provides candidate solutions in a timely manner.

Table 3.2: Common assumptions made in disaster planning, as identified in [5].

ID	Assumption	Type
dH1	Dispatchers will hear of the disaster and send emergency units to the scene, with no units self-dispatching	Limited
dH2	Trained emergency personnel, as opposed to other survivors, will carry out field search and rescue	Unrealistic
dH3	Trained EMS personnel will carry out triage, provide first aid or stabilizing medical care, and - if necessary - decontaminate all casualties before they are transported	Limited
dH4	Casualties will be transported to hospitals by ambulance	Unrealistic
dH5	Casualties will be transported to hospitals appropriate for their needs and in such a manner that no hospitals receive a disproportionate number	Unrealistic
dH6	Authorities in the field will ensure that area hospitals are promptly notified of the disaster and the numbers, types and severities of casualties to be transported to them	Limited
dH7	The most serious casualties will be the first to be transported to hospitals	Reasonable

This factor of computation time may go some way to explain why there has been relatively little work done on disaster response decision models since the technology has not been capable of delivering high quality results fast enough, even when implementing state-of-the-art algorithms. With the advancements made in computing and in the science of algorithms over recent years, it is starting to become possible to develop decision support models capable of operating in real-time, connected with the emergency environment by being embedded in a larger decision support system, using new information as it becomes available to update its proposed solutions.

3.2.1 General Decision Support Models

Later in this review we will focus on optimisation-based decision support models, as such models will correspond with the aims of this thesis. Initially, it is of interest to consider more general decision support models.

An early paper describing decision support systems for use by emergency relief organizations is provided by [8], where four examples are given. As motivation for the use and continued development of such systems, the authors note that decision makers in a large scale emergency relief environment face unique challenges, particularly psychological aspects ('cognitive strain') and data which is dynamic and often uncertain. These decision makers could therefore benefit from a system which helps them make use of all available information and to arrive at solutions to the large, complex problems faced. As a general framework for a decision support system, it is noted that in addition to modelling capabilities they should also include a data bank, the ability to analyze the collected data and finally a means of displaying data to, and interfacing with, the decision maker. Of the four examples given, two involve decision support for the disaster response phase by using an optimisation model to help decide how to distribute limited resources effectively. In particular, a system developed for the Regional Emergency Medical Organization assists in coordinating the transportation of the injured with the distribution of medical commodities, while a system developed for incidents at nuclear power plants assists in deciding which schools should be used as shelters and which routes ambulances should take when transporting the wounded.

More early work in the field is given in [132], which builds upon this specification of a decision support system. It provides an early example of researchers adopting the Federal Emergency Management Agency (FEMA) division of large scale emergency relief into the phases of mitigation, preparedness, response and recovery and describes the information requirements of each. The authors go on to categorize the decisions made in each phase as operational, managerial (or tactical) and strategic, corresponding to how often the decision must be made. In an attempt to distinguish which are particularly amenable to computerized decision support, the authors note that a degree of 'structure' exists within each task, with the more structured tasks lending themselves to applications of decision support systems. For example, the task of damage assessment is identified as being structured due to its relatively predictable length, whereas the task of carrying out search and rescue is noted as unstructured. Along with [8] this paper maintains the importance of involving the decision makers and the organizations to which they belong in the process of developing a decision support system. They state that "The goal of this approach is to encode the facts that experts have about a problem domain and the methods of reasoning they use on a decision problem."

More recently, [121] presents the case for a new field of emergency management and engineering, incorporating the use of advanced communications and computing technologies along with decision support systems. The authors note that decomposing the whole problem faced in large scale emergency relief into two separate stages, pre-event tasks (mitigation, preparedness) and post-event tasks (response, recovery), solving both to optimality, can lead to a sub-optimal solution for the larger problem. It is argued that where possible the two stages should be somehow integrated in order to avoid this.

A decision support system for dealing with the large scale emergency of forest fires is given in [141]. Comprising of five components, the tool offers support to emergency personnel through weather monitoring, analysis of fire risk, offering firefighting advice, fire detection and fire modeling. The tool is designed for use in both the preparedness and response stage of a forest fire, helping emergency services limit the potential damage caused. Of the five components, all but the firefighting advisor are concerned with accepting input data from the environment and processing it into a form which is of use to the decision maker and then presenting it to them. The firefighting advisor component instead attempts to process the information available and use it to suggest possible operational plans. It is designed to help in the preparedness stage by assisting emergency services develop a plan which maximizes their level of preparedness, where a plan may include such activities as patrolling, imposing access restrictions of conducting local fire safety campaigns. However, little detail is given on the methodology used in this component of the tool.

It is noted in [146] that one of the main activities to be carried out in humanitarian assistance operations is the real-time coordination of relief actions. Observing that information is an essential resource when attempting to carry out such activities, the authors go on to present a knowledge management framework for supporting decision makers. Included in this framework is a decision making component which implements case-base reasoning (CBR), a technique designed to make use of similar past experiences. Here, CBR first gains an understanding of the problem by collecting values of certain attributes, before going on to compare with the corresponding values of previous cases and arriving at a conclusion. The system is capable of learning by accepting feedback from the outcome of operations, both successes and failures. The authors note that in order for the CBR system to operate effectively, an extensive body of knowledge of the specific disaster domain is needed.

The potential of using agent-based systems in decision support systems for large scale emergency response, both in providing a simulation environment in which to test policies and in coordinated decision making, is noted in [48]. In [139] the authors integrate an agent-based discrete event simulator with a geographical information system to form a tool designed to accurately mimic real-life large scale emergency response operations. With this model, possible response plans can be simulated quickly to evaluate their effectiveness and so through an iterative process a near-optimal plan can be found. The authors describe a simulation which uses a rule base, consisting of general response principles, to initiate an initial response once an event occurs and some data relating to it, such as the number of victims, has been received. The simulator incorporates optimisation and statistical techniques which can be used to further specify the commands given by the rule base in order to maximize the efficiency of the response operation, although details of these routines are not supplied. The simulation is run iteratively, where after each run the results are taken in and used to update the rule base so better decisions may be developed in the next run. After several iterations, a plan is settled on and presented to the decision makers. The simulation cycle can then be run again, using any new data that has become available.

Agent technology is also employed in [40] in a component of a general disaster management tool designed to offer support to the Emergency Operations Center (EOC) and the decisions made there. In this case, a multi-agent system is developed which contains *decision support agents*, who continually evaluate the disaster environment and offer advice to the human decision makers. Agents are developed for the support of the fire service, the ambulance service, search and rescue, reconnaissance and infrastructure. The environment observed by the agents consists of several simulators which interact to reproduce the dynamic aspects of the real-life disaster world and response operations, including an earthquake simulation to estimate building damage, a casualty simulator to model the health of the injured, and a fire simulator. Simulators of response personnel interact with these disaster environment simulators, observing the disaster world in a realistic manner. Specifically, simulated response personnel do not have access to all information regarding the disaster world, but only a limited selection corresponding to what would be available in a real incident. This paper also contains a discussion how computer based decision support may be integrated into emergency response operations, noting the potential danger of blind trust of the tool and the possible reluctance of users to accept the tool as a decision aid. It is noted that the tool can be used in two stages of the decision process, firstly in recognizing the problem and offering a selection of possible solutions for the decision maker to consider, and secondly in the evaluation of a candidate plan by using the simulation capabilities to predict it's outcome.

3.2.2 Optimisation Based Decision Support Systems

Having discussed decision support systems in general, we now proceed to review systems which feature an optimisation model at their center. Firstly, we shall review research which approaches the large scale emergency response problem from the angle of logistics. We will go on to include papers which explicitly incorporate the transportation of casualties in their models. Finally, we will describe work which addresses the problem in question through a model of resource allocation and task scheduling.

Logistical formulations

A popular view of the problems presented in large scale emergency response involves considering the logistical issues present [65]. In particular, it is proposed that response operations could be improved significantly by focusing on the distribution of resources, such as medical supplies or food, around the emergency environment. These approaches make use of the extensive amount of previous research into general vehicle routing and network flow problems.

An early example of such an approach is [71]. The specific problem addressed is that of large scale food distribution following a major disaster, and so involves the transportation of many different types of commodities of varying size and weight. The author notes that supplies are initially forwarded to distribution centers close to the areas of demand, after which each distribution center is responsible for delivering its supplies to those who need it. The research focuses on the second stage, i.e. how a distribution center should supply the various points of demand in its area. A model is described which uses the relevant data that is available to suggest which vehicles should be used to distribute which commodities and in what order. The information used includes knowledge of travel times and the terrain to be crossed, vehicle capacities and service intervals, and weights and volumes of commodities. Using this information in conjunction with a rule base developed with the assistance of practitioners, a heuristic recommends which vehicles should be used to distribute which commodities in such a way as to minimize overall fuel consumption.

The problem of routing emergency responders in a 3D sense is addressed in [74], modeling movement inside multi-story buildings as well as movement on the ground transportation system. The authors' approach to this problem involves developing an intelligent real-time 3D GIS to be used to help coordinate the emergency response operations and evacuation during a disaster. In particular, each building is modeled as a 3D network data structure using graph theory techniques, and each of these structures is connected to the 2D network representing the transportation system. Each data structure contains a variety of data including occupancy levels in each room, temperature, and details on smoke and fire. This data is proposed to be gained through the use of various sensors throughout the building, although the authors note this may not be a realistic assumption due to objections to excessive surveillance. Using this network structure, optimal paths for emergency responders to follow are found, taking into account blockages on both the transport network and within buildings (e.g. collapsed stairs) as well as multiple points of possible entry and exit.

A decision support model presented in [61] tackles the problem of transporting several different commodities over a network via several possible modes of transport, such as boat, truck or helicopter. The model allows supply and demand at various locations, changing over time. It also incorporates the ability to move commodities from one mode of transport to another at certain transshipment nodes. The problem is presented as a mixed integer linear program. The objective used is the minimization of total cost, where cost is contributed to by the flow of commodities over the network links, the cost of transferring commodities over modes, and the cost of carrying over supply from one time period to the next. The result of the model is a detailed description of how the available vehicles should move over the transportation network over time along with specifications of the flow of each type of commodity.

Further complexities inherent in a large scale emergency response situation have been modeled. In [6] the authors attempt to incorporate uncertainty into their model. They approach a problem similar to that tackled in [61], the distribution of supplies to locations of demand using multiple modes of transport. However, the authors divide the decision process into two stages and use stochastic programming techniques to arrive at an optimal solution given the expectation of future events. The model is designed to be used immediately after an earthquake has struck, when information on the epicenter and magnitude has been received but details of actual impact are still uncertain. Given the basic information, the model considers a set of possible impact scenarios and, considering how likely the realization of each scenario is, formulates a plan on how to initially move commodities bearing in mind that changes may have to be made. Following more information being received on the impact of the earthquake, the model can then further specify an optimal plan in order to satisfy demand. The stochastic elements of the problem include the levels of supply, demand and the capacities of the transportation links. The objective employed is that of minimizing the expected cost while satisfying the demand.

Another possible complication to the general commodity flow problem is the introduction of multiple objectives, a problem addressed in [124]. Whereas other models focus only on minimizing the cost of transportation, this paper acknowledges that the objectives of minimizing the time taken to deliver the commodities and of maximizing the fairness are also of importance. Here the notion of fairness is defined as the smallest level of satisfaction across all demand points, and is designed to avoid situations where critical but hard to reach locations are neglected in the pursuit of a more efficient overall operation. The problem faced consists of deciding at which locations some intermediate distribution center or 'transfer depots' should be set up, and then how commodities should be shipped from supply locations to transfer depots and then out to the demand locations. The model is designed to be dynamically updated as new information on the status of transport links becomes available, updating routing advice as necessary. In order to deal with the three objectives simultaneously, the model denotes the measure of fairness as the principle objective and implements the measures of cost and time as soft constraints. This results in the objective function being penalized as the cost of a solution grows and as the time taken to implement it increases. A further contribution of this paper in addition to the mathematical model is some detail on the collection of information to be used as input. It is noted that data collection tasks can be classified as 'pre-operation stage' or 'disaster information transmission', where the former denotes data that can be collected before a disaster occurs (e.g. the transport network) while the later must be gathered as the disaster develops (e.g. damage to roads).

Presented in [92] is a model for the inner city transportation of commodities using several modes of transportation following a disaster. This model acknowledges that several different types of vehicles will be used in this situation and that the initial locations of these vehicles need not be at specific depots already known to the decision maker. The model allows vehicles to be called on from every point in the network, allowing the model to replicate the use of civilian vehicles. The model has been designed to facilitate re-planning so as to be of use in the rapidly changing environment of disaster response. To avoid the computational strain often encountered in vehicle routing problems when keeping track of each individual vehicle in the model, this work treats vehicles as another commodity. This approach leads to a smaller formulation of the problem. Taking details of the transport network, including sub-networks for each transportation mode, along with forecasts of how supply, demand and vehicle availability are going to change over the planning horizon, the model specifies how commodities (including vehicles) move across each link of the transport network over time.

In [63] a multiple objective model of the distribution of supplies from multiple supply points to a single demand point is presented. The model assumes that the demand exhibited at the emergency location can be predicted accurately, and does not consider the details of exactly which vehicles are used for transportation - only the flow of supplies is modeled. The cost of transporting supplies is deemed to be proportional to the traveled distance. The two objectives of minimizing cost and minimizing shortage are combined through the use of weights, into a single objective. The resulting linear optimisation model is reported to be solved within 1 second.

In the research presented in [22], the problem of distributing a single resource from multiple sources to multiple emergency points is considered. It is assumed that the levels of demand at emergency locations is known. A model is developed which attempts to find the optimal loadings of the single resource among the capacitated vehicles available, and details of the routes each vehicle should take. It is noted that the resulting model is similar to the classic Traveling Salesman Problem.

Incorporating casualty transportation

The focus of the work reviewed thus far has been on how to distribute emergency personnel, equipment and commodities to the disaster area in an efficient, coordinated way. However, during any urban disaster there is likely to be large numbers of civilians, both healthy and injured, that need to be acknowledged in any emergency response plans. There has been some effort to incorporate both the transportation of commodities and the collection of casualties into the same model. It is noted in [7] that helicopters can be used in disaster response operations to both deliver supplies and to transfer the wounded to hospitals, often combining the two tasks into one trip. The authors model this by treating casualties as another type of commodity and allowing locations in the model to act as both supply and demand nodes. The model itself is divided into two sub-models, separating tactical and operational decisions. On the tactical side the model suggests which helicopters from which bases should be used in the response operation. It also specifies which of the available pilots should operate which helicopter, and determines the number of trips each helicopter can make. At the operational level, given a helicopter with designated pilots and a set number of trips, the routes of each trip are specified along with instructions of which commodities to pick up or drop off where, and when re-fueling should be carried out. This results in two separate models with two separate, conflicting objectives, namely maximizing cost-efficiency for tactical decisions and minimizing total time for operational decisions.

The problem of coordinating logistics support and evacuation operations is also tackled in [144]. Their decision support model deals with the transportation of commodities and of casualties, but also incorporates the problem of locating temporary emergency centers from which supplies can be distributed and at which casualties can receive some limited treatment. This paper builds upon [92] and incorporates the same technique of treating vehicles as commodities, but improves the model through adding the facility location problem. The problem is decomposed into a mixed integer network flow problem where all commodity flows (including vehicles) are specified, following which the problem of specifying instructions for the individual vehicles is solved. In addition, [143] again examine the disaster response problem from a logistics perspective, dealing with the distribution of commodities to areas of demand whilst also evacuating the wounded to medical centers. Much of the model formulation is borrowed from [144], although the problem of locating temporary distribution centers is omitted. The novelty of this work lies in its solution method, where an ant colony optimisation algorithm is developed. The model incorporates time-varying supply and demand, prioritized classification of the wounded and prioritized commodities. Furthermore, it allows for a heterogeneous fleet of emergency vehicles, which can have different capacities.

Allocation and scheduling formulations

In this section we will examine research which views the problems of interest as some form of resource allocation or scheduling problem. [46] provides an example of such work. Here, the authors consider a problem involving the allocation of resources, including personnel and equipment, to an operational area after an earthquake has struck. The authors identify different types of operational area depending on the type of work that is to be carried out in them, be it search and rescue, stabilization work to avoid secondary disasters, or the repairing of essential networks. In total, six tasks and seven resources are identified and built into the model. The model also recognizes the role of uncertainty in a post-disaster environment and accepts as input the survival rates for trapped victims and for casualties who have been rescued but are awaiting treatment. Casualty health is modeled using a reduced version of the triage system descried in Section 2.3.5 and used in emergency response, where the model classifies casualties as either 'injured' or 'deceased'. It also quantifies probabilities of secondary disasters occurring in stabilizing areas, although the authors note that due to the unique nature of such areas these probability estimates need to be provided individually by experts at the scene. Taking this information, along with details of transportation times and the duration of the task to be carried out, the authors go on to develop a detailed objective function. The goal of the model is to minimize the number of fatalities, and the model accounts for fatalities occurring due to many factors. It includes those due to secondary disasters, the duration of the search and rescue operation, and any delays affecting the duration of transport and resulting in casualties not reaching a hospital in a timely manner. When assigning casualties to hospitals, it is noted that each hospital has a fixed capacity. Incorporating all these factors, the model then defines a work schedule for all resources which will minimize the expected number of fatalities.

The work in [47] addresses the same problem of resource allocation following a major earthquake. The decision support system developed is composed of three submodels: a simulation of the disaster environment; a simulation of the movement and operations of response resources; and a decision support model. Similar to the work of [40], the simulation of the disaster environment includes models to simulate the spread of fire, the damage to buildings caused by an earthquake and the health of casualties as time progresses. Simulators of resources which are relevant for search and rescue operations are included, such as a helicopter module, a crane module and a fire engine module. A decision support model is used to issue orders to the response resource simulators, which go on to carry out these tasks, interacting with the environment simulators as they do so. The decision support tool is modeled through the use of agents, with the objective of minimizing loss of life and property damage. The author describes testing the model with real world data of a city with 40, 000 inhabitants. Given a scenario involving 70 response resources, 10 blocked roads, eight buildings initially on fire and 250 casualties trapped inside buildings, a simulation of 72 hours of search and rescue activities took 3 hours on a Pentium 4 PC.

Similarly, [113] formulate a decision support system for the distribution of resources following a disaster. Here, the period considered is the first 3 days from the disaster, noting that it is during this time that is most critical to search and rescue operations. The model includes a capability to forecast the dynamic demand for relief resources across the disaster area. The distribution of relief resources is separated into two distinct stages: first, relief is moved from supply locations to relief distribution centers, after which they are in turn delivered to the affected areas. It is assumed that geographical information is available, in particular the number of affected areas and their geographical relations. In addition, it is assumed that updated information can be obtained as operations progress, and that the vehicles used to transport relief can carry multiple types of resources. A multiple objective optimisation routine is developed to decide how the several types of resources available at the relief distribution centers should be delivered to grouped affected areas, taking dynamic demand and supply into account. The two objectives used are the maximization of satisfying demand and the minimization of cost. Once it has been decided the amounts of each resources are to be transported from the distribution centers to the grouped-affected areas, the model then examines each individual resource and formulates a plan to transport them from the distribution centers to each individual affected area.

In [4] the authors identify the problem of ambulance dispatch and relocation as one which could be assisted by a decision support model. These two problems are treated separately but are tied together through the notion of preparedness, for which a new quantitative measure is presented. The response area is partitioned into zones and the n closest ambulances make a contribution to the preparedness based on how long it would take them to travel there. This sum is then weighted by the demand of the zone (i.e. the frequency of emergency calls), and so this measure of preparedness increases as ambulances move closer to the zone and decreases as demand increases. The model accepts pre-calculated travel times between zones, accounts for the frequency of calls historically received in each zone, and then decides which ambulance should be dispatched to a given call. Individual calls are assigned a priority level, which is also taken into consideration. As ambulances are dispatched to calls the relocation problem can present itself if any zone is left unprotected, i.e. if the preparedness level of that zone drops below a certain threshold. When this happens, the decision support model relocates the available ambulances across the response area in order to achieve maximum *global* preparedness.

A similar problem is tackled in [60]. Here the context of disaster response is explicitly taken into account, and the notion of casualty 'clusters' is presented. The authors suggest that following a disaster there will be areas with high concentrations of casualties and by dispatching ambulances to these clusters as opposed to individual casualties located far away from any other, a higher level of efficiency will be achieved since ambulances will be able to fill their capacity quickly in each trip. Much of the work presented focuses on how to model these clusters, in particular how to predict how they will grow or shrink over time depending on what level of service they are receiving. By considering only casualty clusters, the allocation model is significantly reduced in size and is therefore easier to solve in a timely manner. Another assumption which reduces the complexity of the problem is that the number of wounded people takes the form of a continuous number as opposed to an integer value. The tool is designed to be used regularly, reallocating ambulances as the disaster evolves and new clusters of casualties are formed. The authors assume that the plan can only be updated at certain intervals, e.g. every hour, and note that the decision of how long these intervals are can have a significant effect on the efficiency of the tool. No detailed suggestions of appropriate lengths are offered, although it is suggested that it is essential that solutions are found quickly. The objective used is that of minimizing the time until the last casualty cluster has been completely treated.

The notion of clusters of casualties is also implemented in [67]. Again the problem addressed is the dispatching of emergency vehicles to casualties in order to transport them to a hospital, although in this case the problem of routing is also addressed. Given a distribution over the response area of casualties, each with a level of priority assigned, the model suggests which of the available emergency vehicles should be sent to which casualty or cluster of casualties. Once a vehicle has picked up it's assigned casualties, the model then suggests which hospital it should deliver them to. Another contribution of this research was it's, albeit limited, use of data fusion techniques to incorporate uncertainty into the model. For example the fact that varying amounts of congestion or road damage could affect travel times is incorporated by assessing the probabilities of certain levels of disruption occurring.

While most research in the area of disaster response focuses on category one responders (e.g. blue light service), little attention has been paid to decision support for category two responders (e.g. utility companies). One such paper is [149], where a decision support model for the logistical problems faced by electric utility companies is presented. Similarly to [4], [60] and [67], the principal problem is the allocation of emergency vehicles to some job, in this case the allocation of emergency repair vehicles to repair tasks. Specifically, the model suggests a partitioning of the response zone into districts of responsibility. Once these zones are defined, each emergency repair vehicle is left to deal with repair tasks as they arise in their district. Furthermore, the model also advises on how many vehicles/districts are needed in order to meet a pre-defined service level, and on a dispatching policy for a vehicle with multiple calls waiting to be serviced (for example, first come first served). To solve the problem of deciding on district boundaries an unspecified optimisation procedure is used, whereas a simulation is used to evaluate dispatching policies. In making these decisions, the model acknowledges the transportation network, the priority level of each call and the expected durations of repair tasks.

A similar problem is considered in [112], namely the assignment of electric power repair crews and related resources to various damaged areas following a natural disaster. In addition, the proposed decision support system also finds optimal locations in the operational area at which to place depots containing a selection of resources. To model these decisions, the decision maker first divides the disaster area into a number of cells and, based on the level and type of damage observed in these cells, assigns values representing demand for the various resources available. In addition to this, the model also accepts input in the form of descriptions of any resources depots available for deployment, including the level of resources available in them. Using a mixed integer linear program, the model then finds the optimal locations for any depots and how resources should be distributed from them to damaged areas, with an aim of minimizing the total transport cost.

A problem of allocating personnel is also addressed in [14], although in this case the units are engineering battalions and the tasks consist of repair work to be carried out following an earthquake. Here the authors decompose the overall problem into tactical and operational levels. This model is designed to be run in real-time and has the ability to incorporate input from the decision maker, although it is noted that this rarely increases solution quality. The authors note that 'excessive logistical details' are omitted since the model is designed for use on a national scale and therefore details like transportation networks are of limited benefit. The objective is to complete all repairs as quickly as possible.

More work on the allocation and deployment of resources was carried out in [137], where the problem of coordinating oil spill clean up operations is faced. The authors explicitly identify the tactical portion of the problem and focus their efforts on it. Specifically, they develop a model that accepts as input the locations and amounts of each type of clean up equipment and go on to suggest which components should be deployed over the time of the clean up operations and from which locations these components should be taken. It is noted that a response system which is capable of helping to clean up an oil spill is composed of a number of separate components, such as a pump, a skimmer or a barge. The problem faced then involves deciding on which response systems to deploy, and furthermore how each system should be composed i.e. where the components should be taken from and at which location should they be assembled (given constraints on the area available). From the several possible objectives that could be used, the minimization of total response time was noted as the most appropriate: "Social consciousness along with the penalties accessed in recent court cases have led to the objective of responding as quickly as possible, regardless of the cost of the response itself." Response time here is defined as being composed of the time taken to ship all necessary components to a staging area, the time taken to assemble the components into a response system and the time taken to deploy the system into operation. A novel approach is taken to reduce the dimensions of the problem in order to achieve faster computation times: instead of considering every possible type of response system, a heuristic method using graph theory techniques is used to select a subset of 'promising' systems for the given oil spill. These are then used as decision variables in a general integer program. To solve this model, two heuristic algorithms based on a linear relaxation of the problem are presented, which achieve fast run times (i.e. < 10 seconds) on problems based on real-world data.

An agent-based simulation is used in a decision support framework described in [140]. It is noted that the emergency response problem may be viewed as an assignment problem, and so the decision support framework advises on how agents, which represent emergency responders, should be assigned to response tasks. In total, three resources and three tasks are incorporated into the model. Details of the objective function used are omitted.

3.2.3 Summary

Several examples of models which give no explicit consideration to the processing of casualties exists in the literature. Such work has generally focused on either the distribution of emergency responder units to areas which require their attention, or on the distribution of some vital commodities such as food and medicine around the affected area. Of the former type [14, 46, 106, 135, 136], a varying degree of detail in the modeling of casualties is present. Only [46] considers casualties explicitly in their model, providing a means with which to forecast the number of fatalities resulting from any proposed responder assignment which they use as an objective function. The proposed method considers the overall changes on the entire casualty group incurred due to factors such as delayed rescuing or delayed transportation to hospital. In contrast, [14, 106, 135, 136] all employ objectives relating to how long the response operation takes and do not explicitly consider casualties. Due to the abstract nature of the tasks to which responders are assigned to, it may be possible to interpret them as the tasks required when processing casualties. However, no details regarding how this could be implemented are given.

Considering models focusing on the distribution of vital goods [61, 6, 92, 20, 124, 113, 85, 12, 79, 101, 147], common objectives used in the models include the minimization of the cost of transporting the goods in question, minimizing the time taken to distribute the goods, and the minimization of unsatisfied demand. The models described in [124] and [79] are notable for their inclusion of objectives designed to maximize the "fairness" of the distribution by examining the largest difference between the unsatisfied demand at all locations in their problem environment. In all of these models, casualties are at best present in an implicit manner, assumed to be generating demand for the goods in question at various points in the problem environment but not being modeled explicitly.

A further set of models which address the distribution of vital goods incorporate the transportation of casualties into the same model. That is, the same vehicles used to distribute emergency supplies are used to transport casualties to hospitals or other appropriate treatment facilities. The model proposed in [92] is extended in this fashion in [144, 143, 91]. These models consider casualties as another good or commodity which requires transportation from supply points to demand points, and as such the same commodity flow objectives of minimizing transportation cost and unsatisfied demand as used above are employed, albeit with weights used to differentiate between casualties and goods. In [7] the authors describe a model based upon the vehicle routing problem which includes the specification of the routes to be taken by response helicopters and at which point on these routes they should collect casualties to return them to base. In [21] the problem of evacuating civilians in an urban environment whilst simultaneously directing responders into the environment is modeled, where the objective is to minimize the total travel time with different groups being assigned different priorities. The problem of assigning ambulances to clusters of casualties is described in [60] and developed in [67], where a model for online (i.e. making decision sequentially rather than simultaneously) use is described. The model advises where an ambulance should be sent once it becomes free, and then to which hospital it should transport its charge. The model does not account for other parts of casualty processing, nor does it approach the problem in a holistic manner.

Only two pieces of work have been found to address the treatment of casualties in the major incident response environment [119, 30]. Survival time distributions are employed in [30] in a model designed to suggest from which of a number of health classes a casualty should be selected whenever an operating room becomes free, with the aim of minimizing the expected number of fatalities. In contrast, the model detailed in [119] considers the treatment of casualties taking place at the disaster scene, attempting to prescribe optimal sequences of patients to medical teams with the same aim of minimizing expected fatalities. However, in both cases the related decisions of how casualties should be rescued and how they should be transported to hospital are not incorporated.

Regarding some of the key modelling features discussed, Table 3.3 captures their presence in the work reviewed.

Table 3.3: A summary of the decision variables and objectives employed in models of disaster response. Decisions are denoted as a for allocation, s for sequencing.

	Decision variables							*Objectives*		
	transportation		treatment		rescue		hospital	fatalities	suffering	makespan
Model	a	s	a	s	a	s	a			
[14, 106, 135, 136, 147]	×	×	×	×	×	×	×	×	×	✓
[46]	×	×	×	×	✓	✓	×	✓	×	✓
[61, 6, 92, 20, 124, 113, 85, 12, 79, 101, 147]	×	×	×	×	×	×	×	×	✓	✓
[144, 143, 91, 21, 60]	✓	×	×	×	×	×	×	×	✓	✓
[7]	✓	×	×	×	×	×	×	×	×	✓
[67]	✓	×	×	×	×	×	✓	×	✓	✓
[119, 30]	×	×	✓	✓	×	×	×	✓	×	×

Whereas a number of models have been developed to give decision support to a tactical decision maker during disaster response, there has yet to be any comprehensive treatment of the entire casualty processing procedure. We hypothesize that a model which incorporates a high level of detail with regards to this area, allowing for control at the level of individual casualties and spanning the entire processing timeline, will lead to significant efficiencies in response operations and a corresponding contribution towards objectives (i) and (ii) as listed in Section 2.3.2.

3.3 Key sub-models

Although the work reviewed has encompassed a range of decision problems and objective measures, there are a number of problem characteristics which have been found to be incorporated in many cases. Specifically, in the work reviewed it is common for the health of casualties to be considered to some degree. Given the MCI focus of the research described in this thesis, as outlined in Section 2.2.2, we anticipate that the modelling of casualty health will have an important role in the design of the proposed optimisation model. As such, it is of interest to examine in further detail how casualty health has been represented in previous work.

Similarly, the modelling of hospitals is potentially an important component of the MCI response optimisation model. The modelling of hospitals, and the role they play in response decision making, has been included in previous work to varying degrees. This will be examined further in Section 3.3.2 in order to inform how hospitals should be represented in our proposed model. Finally, given the spatial nature of multi-site MCIs as discussed in Section 2.3.5, we also examine in detail how the routing of response vehicles has been represented in previous work.

3.3.1 Casualty health

Many decision support models have been proposed for use in disaster response problems where casualties are an important factor, with a wide range of approaches to the modelling of health exhibited. Where distinction between different health levels of casualties is made, not all models account for the possibility of a casualty's health changing from one level to another as the response operation progresses. Those that do typically fall into two categories: simulating health progression, or analytically predicting it. In the case of the former, the models aim to realistically replicate the dynamics of any given casualty's health. In the case of the latter, the model aims to provide probabilistic estimates of how a casualty's health is likely to evolve over time, enabling the prediction of quantities such as the number of fatalities which will occur over a period of time.

Several decision support models designed for use in disaster response which do not explicitly model casualty health are described in the literature. [6] describe a stochastic programming model designed to assist in deciding how first aid resources should be distributed across a disaster area. Whilst the health of casualties is not modeled explicitly, it could be argued that the demand for first aid resources at any one location within the disaster area includes implicit information regarding the health of the corresponding casualties. This is also the case for the resource distribution model proposed by [85], which aims to satisfy the demand for resources at a number of hospitals. Casualties are included explicitly by [7], who propose a model to determine how helicopters should be employed to transport casualties from a number of locations to hospital. However, no distinction between the health of any two casualties is included.

A task scheduling model is proposed by [106], with an objective function which minimizes a generic measure of cost. The model is extended by [135], where the authors again model the completion of relatively generic tasks and do not include any explicit modelling of casualties or their health. [136] note that the tasks discussed may have time-windows for their completion and that such a window could correspond to an expected survival time for a casualty, suggesting the authors envisage a possible application of their model where tasks are related to individual casualties. However, no explicit details of how such a casualty survival window could be determined are included.

Moving on from models which do not account for any detail in casualty health, some examples of its inclusion at a basic level also exist. [21] present a routing model describing the movement of evacuees out of a disaster area and the simultaneous movement of emergency responders into it. The model has the capacity to assign levels of priority to different groups of evacuees, which could be utilized to distinguish between groups with different average injury levels.

Casualties are modeled as 'clusters' in [60], to which ambulance responder units are to be assigned. Although the health of individual casualties is not modeled, the authors do allow for weights to be used to reflect the relative importance of each cluster of casualties. As in the work of [21], these weights could be used to reflect average injury levels. An example of including health information at the level of individual casualties can be found in the work of [144], where a model describing the simultaneous routing of disaster relief supplies and transportation of casualties to hospital is described. Each casualty in the model is assigned to a weighted injury category, where the weights used are noted to be 'subjective parameters'. The model does not account for the possible evolution of health, nor the potential uncertainty in its measurement.

In some cases the dynamic nature of casualty health is captured through simulation. In [67] the authors extend the work of [60] by incorporating a simulation of the response operations to run alongside the decision support model. Casualties are modeled as belonging to one of two injury classes, with the deterioration of health possible in the simulation. No analysis of the effect of dynamic health levels is presented, nor is a model for analytically predicting how health will change presented. Similarly, [47] discusses a decision support model designed to be used in conjunction with a simulation. Here, casualties may be in one of four health states and the evolution of health is affected by the casualties' environment. Health is assumed to be stable on arriving at a hospital.

Further examples of the simulation of casualty health can be found in the literature [111, 99, 133]. Casualties in the agent based simulation of [111] may take one of five discrete health levels, the transition between which is modeled via a series of Markov chains. The parameters of the discrete time Markov chain (that is, its transition probabilities) are dependent on both the environment a casualty is in and the type of treatment (if any) they are receiving.

A discrete event simulation of the ambulance service response to MCIs is described by [99], where the health of casualties is described on a 0-100 scale. Health is assumed to fluctuate as casualties wait for, and receive, treatment. Such fluctuations are presumed to be of a linear form, with the survival time of each casualty sampled from a uniform distribution. The simulation model of [133] incorporates two health state descriptions at different levels of granularity, similar to the triage sieve and sort systems described in Section 2.3.5. The logistic function of [108] is employed to estimate the probability of survival of a given casualty, parameterized by their health state.

The work discussed immediately above allows for the simulation of casualty health but not necessarily its prediction. This task is explicitly tackled by [46], where a detailed model predicting the number of fatalities to arise from a given response operation is presented. Casualties are assumed to belong to one of two injury classes, information which is employed to predict the number of deaths arising from several causes including a delay in being rescued and a delay in receiving treatment at hospital. An exponential survival function is used in these calculations.

In contrast, a Weibull distribution of survival times is employed in the model described by [30], which aims to assist in the allocation of casualties to operating rooms at a hospital following an MCI. The optimisation of treatment is also the goal in the work of [119], although here the problem environment is the incident scene itself. Casualties in the model are assigned to one of four injury levels, and deterioration from one to another is assumed to occur at known points in time. For example, a casualty of health state T3 will move to health state T2 after exactly 120 minutes.

Only one work has been found which allows for uncertainty in the assessment of health. In [67] the authors assume that several estimates of the health of a casualty, described as one of four possible health states, are available at the outset of the response operation. The authors propose combining these estimates via data fusion, following which the most probable health state is selected and used in the remainder of the model. However, no explicit analysis of the benefit of such an approach is presented.

A broad range of approaches to the modelling of casualty health have been reviewed. While some have omitted any explicit representation of health in their models, others have included the capability to distinguish between casualties according to their injury levels. The health of casualties may be assumed to be constant and deterministic. Alternatively, methods for simulating and for forecasting how health evolves, and in particular how casualties go on to die, have been proposed. However, no analysis of the effect of error in the assessment of casualty health has been found.

Where health is assumed to be of a dynamically evolving nature, any decision support program which involves predicting the outcome of a proposed response operation must predict how such evolution will occur. Developing such a predictive model is an extremely challenging task, with any errors potentially leading to poor performance of the decision support program. A natural question to ask is whether such poor performance could be mitigated through allowing the decision support program to be continually updated with the latest observations of casualty health, thus ensuring the impact of any past errors in prediction mistakes is minimized.

3.3.2 Hospitals

A significant decision problem faced by emergency managers during an MCI is that of allocating casualties to available hospitals. An obvious consideration in such a decision is that of the distance between the casualty and the hospital, and the corresponding time it will take to transport them. However, other important factors may be taken into account. Specifically, the decision maker may consider both the *capabilities* of the available hospitals, in terms of specialist treatment units designed for specific injuries, and their *capacities*, in terms of how many beds, staff and other resources they currently have available to treat incoming casualties.

This observation is confirmed in the U.S. Department of Health and Human Services' Center for Disease Control and Prevention's report on 'preparedness and response to a mass casualty event resulting from terrorist use of explosives' [19]. The authors note that, when deciding how best to distribute casualties to hospitals, the *"medical command and control center should use updates of hospital capacities and capabilities and help emergency medical services (EMS) determine the optimal destination for each casualty."* This is echoed in a report on the distribution of casualties in an MCI [150], in which the authors conclude that optimized care in an MCI means *"matching patients with facilities that have the appropriate resources available in sufficient quantities."* The importance of recognizing the heterogeneity of hospital capacities and capabilities has been further established in retrospective analyses of past events such as [10] and [96].

In terms of hospital capacity, it is important to recognize variation not only between hospitals but also dynamic variation within them. One principal reason for this behavior arises from the enaction of major incident plans upon receiving notice of a major incident, which result in a steady increase in capacity as non-urgent patients are relocated and extra beds are freed up [64]. Another important factor is the tendency for many causalities to autonomously leave the disaster scene and transport themselves to a hospital of their choosing, resulting in an uncontrollable stream of casualties taking up capacity which may otherwise be perceived as available [5].

Past research has typically considered the problem of casualty allocation as part of a larger scheduling or routing problem, with varying degrees of detail incorporated. While [7] and [46] describe models which recognize the need for the transportation of casualties to hospitals in an MCI, the decision of where each casualty should be allocated is omitted. The work presented in [60], a model to assist in the allocation of ambulances to clusters of casualties in MCIs, assumes that casualties are always taken to the nearest hospital. This model is extended in [67] to allow for the transportation of casualties to any of the hospitals under consideration, using information of initial capacity levels, waiting times and distance to help make these decisions. However, the model considers only one allocation decision at a time, resulting in a 'greedy' approach rather than considering the whole problem at once.

A network flow model is employed in [143] and [144] to determine how casualties should be transported from disaster sites to hospitals, considering a fixed level of hospital capacity which can be redistributed amongst all hospitals in the model in an attempt to best meet demand. That is, the models assume that hospital resources are primarily represented by staff who can be transfered from one hospital to another if necessary. The models do not, however, consider the potential difference in hospital capability or the dynamic nature of capacity. A fixed initial hospital capacity is also used in [91], although the distribution of casualties is only considered as a constraint in the model. That is, a valid solution requires all casualties to be transported to some hospital, but the effectiveness of the resulting distribution is not measured.

3.3.3 Routing and the transport network

Transport networks are not always explicitly modeled within decision support programs designed for disaster response. For example, [136] present a scheduling model designed to assist in the allocation of response units to incidents, taking as input the travel times associated with each possible journey response units may make. Similarly, a travel time matrix describing the relation between points of interest is taken as problem input by [147] in the model of emergency responder allocation. This can be contrasted with work such as that of [143] and [61], where the transport network is represented as a graph, with each edge assigned a parameter describing the time needed to traverse it. Where such graphical representations of transport networks are included, it is common to assume their structure and parameters are deterministic and constant over time. This is true both of models designed to assist in commodity distribution over a large geographic area, such as those presented in [20, 113, 79, 124], decision support programs using a scheduling formulation to assign tasks to emergency responders [106, 138], and routing based formulation for the support of casualty transportation and evacuation [21, 144]. By assuming all necessary information regarding the transport network is readily available, routing decisions can be made with confidence using a standard shortest path algorithm.

It is common in past work to use a reduced simplification of the actual transport network when representing it as a graph, an approach which can help avoid excessive computational burden. In the problem scenarios considered by [143], for example, the most complex network considered contains 80 nodes connected by 1600 edges. Considering a geographic area large enough to encompass six cities in Turkey, the model presented by [92] represents the transport network using 12 nodes and 12 links, based upon motorway infrastructure. In contrast, a dense network comprised of 34,890 nodes and 43,445 links is used in the test problem considered by [67], with a hierarchical decomposition employed to assist route computation in a timely manner.

Uncertainty in the disruption of the transport network has been incorporated to a limited extent using stochastic programming formulations. Examples include [6, 85, 100], which consider a finite number of scenarios, each with assigned probability and associated network parametrization. Uncertainty is also acknowledged in the work of [67], which extends the ambulance allocation model presented by [60] by including a data fusion step to estimate the level of damage and disruption on each road link. A solution method for finding optimal paths in a disrupted network following a disaster is presented in [148]. The authors employ the network representation described by [145], where the travel time associated with each edge of the transport network is assumed to increase over time in a manner which reflects its proximity to the disaster. A dynamic transport network structure is also modeled in the work of [46], with nodes and edges being added or taken away to reflect the impact of both the disaster and the response operation.

In recent reviews of optimization models for emergency logistics [18, 34] it has been noted that there has been little research in the area employing stochastic models. Given the potential for an MCI to disrupt the transport network, directly or indirectly, and thus lead to uncertainty in routing and travel time prediction, this is clearly an area which merits further research. While some authors have acknowledged the possibility of disruption to the network and the subsequent uncertainty, it remains unclear whether or not this uncertainty will ultimately reduce the utility of a decision support program, and how any such effect depends on the choice of routing policy.

3.4 Modelling uncertainty

Related research into designing and implementing decision support systems for use in emergency response have taken a variety of approaches to addressing the presence of uncertainty in the environment they model. Three key themes can be identified: a *deterministic* approach, where all parameters are assumed to be known with certainty; a *stochastic* approach, where some parameters are assumed to follow a known probability distribution; and a *dynamic* approach, where models are designed to adapt to changes in the environment without explicitly predicting them. We proceed to review work utilizing each of these approaches to addressing uncertainty in emergency response modelling.

3.4.1 Deterministic

An early example of a deterministic approach is provided by [61], where exact supply and demand figures for various goods over time and space are required as input for the proposed commodity flow model. Considering the optimal assignment of tasks to engineer battalions, [14] assume that the resource requirements of the tasks are known and constant over time. More recently, [124] propose a model which assumes *"availability and accessibility of information"* and that any quantities which are subject to variation, such as demand for resources or the availability of a certain road, remain constant over the planning horizon they consider. [63] acknowledges that demand will vary with time, but assumes such variation is known in advance. Similarly, in [135] the authors note that all information required by the proposed resource assignment model is available at the time of planning.

In [91] and [93], only the current values of the required variables, such as the number of casualties and the demand for certain supplies, are included in the model.

That is, the models do not account for scenarios where these quantities are expected to vary over time. The aim of adopting this approach is to eliminate a set of integer decision variables related to the temporal nature of the problem, thereby enabling the resulting free space to be taken by other variable corresponding to the size of the network considered in the problem. Another example of a model which does not consider temporal variations in quantities such as casualty number and commodity demand is provided by [7], where the problem of determining the optimal use of helicopters for commodity transportation and evacuation during response is addressed.

3.4.2 Stochastic

Various models have included some degree of uncertainty in the parameters under consideration. Stochastic programming is employed in [20, 85, 110], and [6], where the problem is divided into two stages. In the first stage, some variables are random with a known probability distribution, while in the second stage these variables are realized. The decision process considers the likely outcome of various scenarios and specifies a two-stage plan, designed to be optimal in terms of the expected value of the objective function. These models assume that the stochastic information is revealed at a single point in time.

Another approach employed to incorporate uncertainty into the model is robust optimisation, as described in [12]. In this paper considering the distribution of resources to areas affected by a disaster, the authors acknowledge that the location and level of demand, the level of available supply, and the costs of procuring and transporting these supplies are all uncertain at the time of planning. Solutions are then found by considering a discrete set of scenarios designed to approximate the variability in these parameters, and optimizing for a weighted sum of known preparedness cost, expected response cost and the variance of the response cost.

3.4.3 Dynamic

One method for handling uncertainty in a model of emergency response is to enable the receiving of updated values of critical information in real time and modifying it's suggested solutions accordingly. As a result, any inaccuracies arising as a result of stochastic parameters can be corrected so that their impact is reduced, leading to a more robust model. This benefit of acknowledging changing information and updating the decision support model to reflect it has been acknowledged in the literature. In their review of operational research in emergency response [116], the authors state that any model of the immediate large scale emergency response problem must be capable of incorporating *"emergent aspects of any disruption that... defy prior anticipation and explicit modeling."*

A common approach to incorporate such capability is an iterative decision support framework where a sequence of static models are built and solved at regular intervals. For example, [46] presents an optimization model for the allocation of tasks to resources and the transportation of casualties to hospital, capable of adjusting the parameters of the transport network to reflect developments on the ground. In [77] a model for determining the allocation of relief supplies to victims is presented, designed to be solved at two hour intervals to allow for new information to be accounted for. A solution algorithm is presented in [106], designed to facilitate the solving of their proposed Mixed Integer Programming resource allocation model in near-real time. The authors argue this will allow decision makers to re-solve any particular response problem when conditions change, although this capability is not explicitly tested and evaluated. Specifically, although it is reported that the proposed algorithm can produce a solution to the problem in a matter of seconds (in a specific example scenario), no details are supplied regarding the time necessary to 're-build' the model to reflect the current problem environment before solving once again.

[92] [143] and [144] provide details of a vehicle routing model designed to be used in determining optimal commodity flow and evacuation plans over a discrete planning horizon. Plans are designed to be updated at regular, set intervals incorporating new information, where the current state of the system resulting from the implementation of one plan is considered when formulating the next. The length of the intervals between information updates varies between thirty minutes and an hour [113] also propose updating information, at set intervals of a third of a day in length.

In [60] the authors note that determining the appropriate length of this interval is crucial to performance, proposing that future work should look to develop models which operate in continuous time. This is echoed in [21], where the authors state that *"from a real-time implementation standpoint, a cyclic rolling horizon based updating and re-optimizing framework and scheme need to be developed to improve accuracy and robustness of the model under the highly unpredictable environment".* In the literature external to disaster response such an approach is often described as a rolling or receding control model [43, 88, 82].

In order to move towards such a real-time system, the work reported [47], [41], [42], and [67] link their proposed decision support models to simulations of the actual response environment, allowing for the testing of the models ability to cope with changes in information. In each case the whole decision problem is decomposed into a sequence of single decision points, where tasks are allocated to one responder at a time as and when the responder becomes available.

This literature demonstrates that the dynamic and uncertain nature of emergency response has been acknowledged to varying degrees in the literature. Many models require a substantial amount of information in order to be initialized, and for this information to be known with certainty. Stochastic programming and robust optimisation have been utilized to address some of the uncertainty in the environment, although the proposed models are designed to be used only once in a static manner. Examples of iterative decision support programs have been found, where a sequence of dependent static models are built and solved at regular intervals. However, there has yet to be any work examining the effect of stochastic parameters on the effectiveness of a scheduling model which requires accurate forecasts in order to facilitate better planning by considering the future of implications of current decisions.

3.5 Solution methods

Thus far, the focus of this review has been on the models proposed in the literature, the assumptions present in them and their capability to sufficiently represent some of the real problems faced in disaster response. These modelling considerations should always take precedence over the development of particular solution methodologies, as any solution which is of high quality with respect to an incorrect metric or model may be of very low quality in practice. In this section we proceed to briefly review the solution methodologies employed in order to find sufficiently high quality solutions to these models in a timely manner. Background regarding optimisation techniques can be found in Chapter 5.

Employing a graphical representation of the emergency environment, in this case including both the transport network and the layout within buildings, [74] employs established shortest path algorithms in order to find exactly optimal routes. The mixed integer programming formulation proposed in [61] as a model of commodity transportation is solved through the use of a linear relaxation of the full model, and so is not an exact method but an approximate one, using a heuristic to ensure computation times remain feasible. An exact solution method is used in [6], where the authors employ a commercial linear program solver to find solutions to their proposed stochastic programming formulation of another commodity distribution problem. It is noted, however, that stochastic programming problems can be computationally expensive in large instances and so custom built heuristic algorithms would possibly have to be developed to deal with particularly large instances. The multi-objective model proposed in [63] is formulated as a linear program through the use of a scalarizing method to produce a single weighted objective. This is then solved using the commercial optimisation software Dash Xpress 20.0.

The model of [92], for the inner city transportation of commodities, is decomposed into two multi-commodity network flow problems in order to be solved. The two models are of a different nature, with one being linear and the other an integer program. These two sub-models are combined through Lagrangian relaxation in the proposed solution method.

The design and implementation of appropriate solution algorithms to solve the presented model is mostly left to further work in [7], although the authors take the first steps towards an efficient solution method through the use of decomposition. Dividing the entire problem into two levels, operational and tactical, an iterative method for linking the two sub-models together during the solution process is presented.

Decomposition is also employed in [144], where both a mixed integer network flow problem, governing how commodities and vehicles should move across the environment, and the complementary sub-problem of specifying the itineraries of each vehicle are solved using the commercial optimisation software GAMS. Building upon this, [143] considers a similar problem and model as [144] but proposes a new solution method, again employing decomposition. The problem is decomposed into two separate models, which are linked together through an iterative procedure: first the ant colony optimisation method is used to construct routes for each of the vehicles, then, once this has settled, it models the problem of commodity/casualty transportation as a integer flow problem. The results of this are fed back into the route construction model, which re-evaluates and refines the solution. This methodology enables the complex problem faced to be solved in reasonable computation time without relaxation of key problem characteristics such as the integer nature of the casualty variables.

Two established metaheuristics are applied to the scheduling model formulated in [46], namely simulated annealing and tabu search, in order to ensure acceptable computation times. In contrast, the simulation-based model of [47] primarily uses agents, communicating with each other, to ensure coordination of plans. Different agents are assigned specific responsibilities in a way analogous to that of a real-life emergency operations center. Agents are equipped with a plan library, derived from analysis of emergency plans and standard operation procedures as well as expert surveys. Using these in conjunction with (unspecified) operational research models, near optimal allocations of resources are found. A genetic algorithm metaheuristic is used in [22] to solve their model of resource distribution.

The model proposed by [140], an agent-based simulation, is itself approximately re-modeled as an integer programming problem. It is argued that solving this approximate model is more efficient than using the simulation itself to determine optimal action, as such an approach (requiring many simulations to be run in succession) would require computation times which would be infeasible in the disaster response environment.

A heuristic approach is maintained in [4] as a means to find sufficiently high quality solutions to their problem of ambulance dispatch and relocation. In particular, a tree search algorithm is implemented.

A simple, custom heuristic is used in [60] when solving the problem of assigning ambulances to clusters of casualties. As this model is developed in [67], so too is the solution method. When deciding to which casualty cluster an ambulance should travel to, and subsequently to which hospital the casualties should then be taken, the model maximizes simple linear expressions containing details of locations and travel times, hospital capacities and the waiting times for each priority level of casualties at hospitals. Given a target destination for a vehicle, the model also delivers support on its routing. The authors note that solving shortest path problems on a full realistic road network can be computationally demanding, and so employ a relaxation of the problem. In particular, the model divides the road network into zones and uses information of road capacity and speed (e.g. classification of roads as motorways, A-roads, etc.) to find the shortest path to a high speed road which connects the start zone with the destination zone. It then details the route within the destination zone from leaving the high speed road to arriving at the destination. This approach is called Best Exit-Entry routing, and by employing it the overall complexity of the problem is reduced which leads to faster computation times.

Similarly, [14] also decomposes the entire problem under consideration (defining a work schedule for a number of engineering battalions) into tactical and operational sub-problems. Tactically, the model first assigns a set of tasks to each unit. Following this, the operational problem of how each unit should employ its resources to complete the work it has been assigned is then solved. The tactical problem is presented as a integer linear program, while the operational problem is a linear program, and a heuristic problem cascade approach is used to optimize. The authors note that, since the input data used by the model are rough estimates, the value of a truly 'optimal' solution is questionable and therefore the model accepts solutions with an objective value up to 25% worse than the optimal. To solve the model proposed by [137], two heuristic algorithms based on a linear relaxation of the problem are presented, which achieve fast run times (i.e. < 10 seconds) on problems based on real-world data.

A common theme emerging from this literature is the use of heuristic solution methodologies. The use of such methods over exact solution methodologies is generally justified by reasoning around computational performance or model fidelity. In the case of the former argument, the use of heuristic methods enables algorithms which find good solutions quickly to be designed and implemented, which may be preferable to exact methods which will typically take much longer to present a single 'best' solution. The latter argument notes that a solution which is 'optimal' in terms of the model may not in fact correspond to the best solution to actual problem due to model infidelity. These points will be borne in mind in the subsequent development of solutions methods for use within the proposed model, which will be discussed in Chapter 5.

3.6 Summary

A number of examples of previous research in the optimisation of disaster response operations have been reviewed over the course of this chapter. In comparing existing models, and the problems they are designed to solve, with the MCI response problem described in Chapter 2, we have noted (see Section 3.2.3) that there has yet to be a model of this specific problem proposed in the academic literature. More generally, the prevalence of the modelling assumptions described in Tables 3.1 and 3.2 has been confirmed. Looking in further detail at the modelling of casualties, hospitals and transport networks has provided guidance as to how to proceed, both in terms of identifying good practice upon which we may build and in identifying gaps in the literature which we may attempt to fill in the current work. Moreover, a similar review focused on how the inherent dynamics and uncertainties present in an MCI response operation may be accounted for in a model has further strengthened the base for going forward.

Specifically, we have shown that a model with the following capabilities has yet to be developed:

1. Addresses the decisions of allocating tasks regarding the rescue, treatment and transportation of casualties to available responders;

2. Addresses the decisions of allocating casualties to hospitals, accounting for transportation time, hospital capability, and hospital capacity;

3. Acknowledges that different casualties will be of different health states, and that this will influence the decision making process;

4. Acknowledges that the health of casualties may change over time, and that the measurement of casualty health may be subject to error;

5. Acknowledges that the times required to complete tasks and to travel around the MCI environment are not known and not deterministic, but estimated and subject to revision;

6. Acknowledges that the model will be used in real-time as the response operation progresses, and therefore should provide decision support in a timely manner and allow for changes in the environment to be reflected by the model;

It is hypothesized that an optimisation model with the aforementioned capabilities will have great potential to assist in the coordination of MCI response operations. Specifically, by avoiding making limiting assumptions, the model will be more applicable and more robust than previously published models designed for similar problems. Acknowledging the structured decision problems within MCI response, the model will be capable of using automated solution algorithms to locate and suggest response plans of a higher quality than those available through simpler heuristic generation.

In the next chapter, the central form of such a model will be developed, specifying in detail the exact form of the decision problem to which it is to be applied and the methods for evaluating the quality of a solution. For the initial treatment, attention will be devoted to the case of 'static' problems, that is, those which do not change over time. The extension of the model to more complex, and more realistic, dynamic problems will be described in Chapter 6. In the intervening chapter, solutions methods to be used with the model will be discussed.

Chapter 4

Model

A combinatorial multi-objective formulation

4.1 Introduction

In Chapters 2 and 3, a specific untreated decision problem faced in Mass Casualty Incident response, which we have called 'casualty processing', was identified and placed in the context of large scale emergency response in general. The broad decision making structure in place during such incidents has been reviewed, which will be given a formal treatment in this chapter. That is, the mathematical representation of the *decision space* of the problem will be set out. This will then be complemented by a description of the *objective space*, through definition of a number of objective functions based on the objectives held by practitioners in actual incidents. Both the decision space and objective space development will also be informed by similar problems found in the Operational Research literature, namely vehicle routing and job-shop scheduling, with an aim to make use of established methodologies in this novel application area. It is important to understand the empirical properties of any new model proposed, and so some computational analysis will be reported with this goal in mind.

4.2 Combinatorial optimisation frameworks

As discussed in Chapter 3, the casualty processing problem has yet to be mathematically modelled in the literature. In order to build such a model, we first consider some established archetypal combinatorial models outwith the disaster response application domain in order to identify any promising methodological techniques. In particular, we focus our attention on the Vehicle Routing Problem (VRP) and the Job-shop Scheduling Problem (JSP), both of which exhibit characteristics which mirror aspects of the casualty processing problem.

What follows is a brief discussion of both problems, with the goal of identifying key factors which may be incorporated into the design of a novel mathematical model for the casualty processing problem.

4.2.1 Vehicle routing

As described in [29], the standard VRP is a decision problem whereby a number of discrete points, typically arranged spatially, must be visited by one of a number of vehicles available to the decision maker. This involves both specifying the number of trips to be made (whether by different vehicles or the same one in succession), the locations to be visited on each route, and the order with which these locations should be visited. A simple example of a VRP and its solution is given in Figure 4.1, where three routes have been identified which cover all desired points.

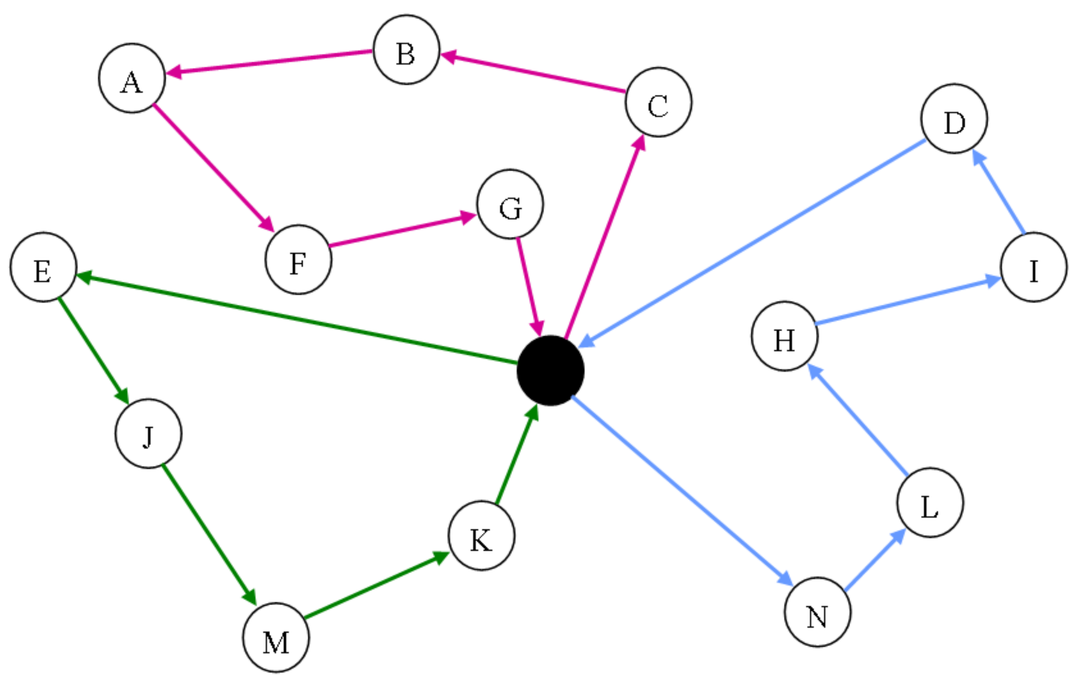

Figure 4.1: An example solution to a vehicle routing problem, with three routes (coloured lines) serving fourteen locations, beginning and ending at the central depot.

A number of variants of the standard VRP have been studied, including the following features noted in [29]:

- time windows, where locations must be visited within a pre-specified time frame,
- a heterogeneous fleet of vehicles, differing in characteristics such as capacity,
- the incorporation of collection in addition to delivery,
- some locations requiring several visits over time,
- and split deliveries, where goods may be delivered across several visits as opposed to all at once.

Considering its applicability to the casualty processing problem outlined in Chapter 3, it is clear that similarities exist with some variants of the VRP. In particular, the key requirement that casualties must be transported from their incident sites to hospitals by a fleet of Ambulance Service vehicles may be thought of as a VRP. As noted in [136], time windows could potentially be used to reflect the limited time a casualty will survive for before arriving at a hospital. The problem would naturally require the specification of collection and delivery, and furthermore would involve vehicles visiting locations several times across the duration of the response operation. Such similarities are discussed in [36]. Finally, work focusing on modelling VRPs as real-time problems, as in [59], may be applicable due to the potentially volatile nature of the casualty processing problem.

Solution methodologies employed in solving VRPs are surveyed in [75] and [39]. In the former, the authors describe exact optimisation methods, heuristic techniques and metaheuristics (see Section 5.2 for further discussion of these concepts as related to the work of this thesis). In the latter the focus is on metaheuristic techniques. The authors note the popularity of both local search and population-based metaheuristics. It is noted that early heuristic methods used to solve the VRP 'lack some of the necessary attributes to ensure their adoption by practitioners'. The authors go on to describe four such attributes: accuracy, speed, simplicity and flexibility. A further solution methodology is described in [84], where an algorithm known as the Active Guided Evolutionary Strategy in applied to the capacitated VRP. The algorithm employs heuristics such as choosing a number of 'customers' to be re-allocated a delivery slot.

In [89] a new variant of the VRP is proposed, the Cumulative Capacitated VRP. The key difference between it and the standard Capacitated VRP is in the objective measure employed to assess the quality of solutions. In this case, the focus is on minimizing the summed waiting times of all customers as opposed to minimizing the distance travelled or the cost of the journeys. Solutions to the problem are represented by permutations of the n customers, and the associated schedule and objective value is the best extractable from this encoding. To find this best schedule, the problem of splitting the permutation of customers into a number of distinct routes is solved using graph techniques and shortest path algorithms. Excessive computation in the evaluation of solutions is avoided through specifying ways in which to only partially re-evaluate a solution after a heuristic has been applied, as opposed to evaluating the entire solution again. The same problem is addressed in [105], where an Adaptive Large Neighbourhood Search metaheuristic is proposed as a solution method. The authors note that this solution method has performed well on other routing problems. The results presented demonstrate superior performance in comparison to the work of [89] in terms of objective value, although the computation time is larger in the case of small problems.

4.2.2 Job-shop scheduling

The JSP and its variants provides another archetypal combinatorial optimisation problem potentially applicable to the casualty processing problem. The standard

JSP involves a number of jobs, each of which is comprised of a number of operations, which must be completed by a number of identical machines. The operations which make up a job generally have to be completed in a specified sequence. The usual goal is to minimize the time at which the last job is completed, known as the *makespan*. An example solution to a simple JSP, involving four jobs of three operations each being assigned to four machines, is shown in Figure 4.2.

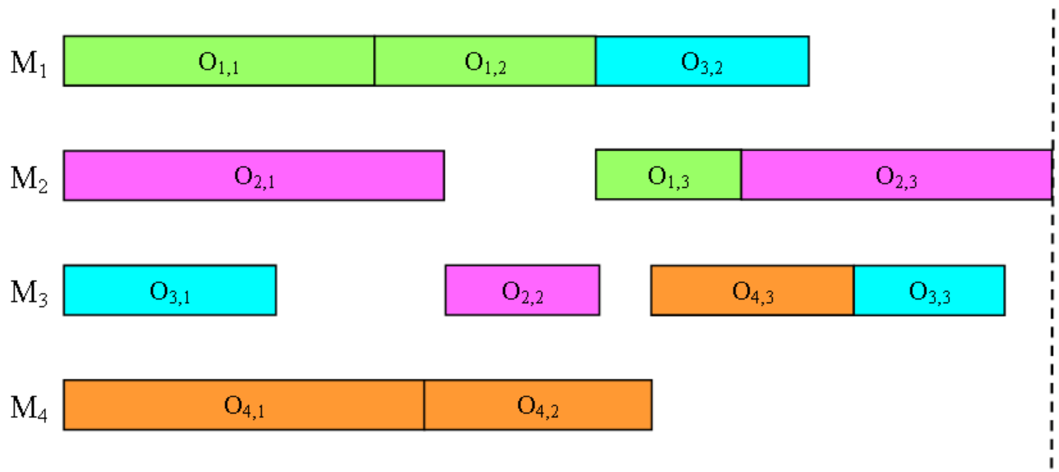

Figure 4.2: An example solution to a Job-shop Scheduling Problem, with four jobs (color coded) of four operations each scheduled across four machines.

As with the VRP, many variants of the JSP have been proposed over the years, increasing the general frameworks applicability. Characteristics which have been incorporated include:

- heterogeneous machines, where different types take different lengths of time to process operations,
- setup times, where each operation requires a period of time before work on it can be begin (with this time possibly dependent on the sequence of operations for the machine in question),
- other objective functions, such as the length of time spent by machines in an idle state,
- uncertain, as opposed to known and deterministic, processing times for operation,
- constraints regarding which machines can process which operations.

By considering the processing of a single casualty as a job, which itself may consist of a number of discrete tasks (operations in the language of the JSP), it is clear that the general framework could be applicable to some extent in the casualty processing problem. In this case, 'machines' would be the emergency response units available and able to work on the response operation. The spatial nature of the environment would require responder units to travel from the end location of one task to the start location of another, a feature which could potentially be encapsulated through using the sequence dependent setup times mentioned. The use of constraints to restrict the set of tasks which any response unit may complete is clearly applicable also, due to the well defined roles and responsibilities of different emergency response organizations as discussed in Chapter 2. However, as in the case of the VRP, a major difference will be in the objective function(s) used when evaluating the quality of a solution. In the case of casualty processing, it is unlikely that all available details will be encapsulated by the standard measures of efficiency employed in JSP scenarios, such as makespan.

As with VRPs, JSPs are typically solved using heuristic and metaheuristic techniques due to their combinatorial nature. Metaheuristics employed to find high quality solutions to the JSP include Ant Colony optimisation [142], Genetic Algorithms [130], Tabu Search [13] and Variable Neighbourhood Search [3]. Common low level heuristic operations included swapping the position of two operations, or re-assigning an operation to a different machine. The scale of problems considered varies considerably, but often involve tens of machines and hundreds of operations. For example, in [142] the authors test their proposed algorithm on problem instances where the number of jobs range from 50 to 200, the number of operations per job range from 10 to 30, and the number of machines range from 10 to 20. The resulting computation times are reported to range from 1.5 to 878 minutes.

As noted in [142], two general approaches are often employed when solving a JSP. A 'hierarchical' approach involves separating the two main decision tasks of assigning operations to machines and then determining the sequence of operations, whereas an 'integrated' approach considers the whole decision problem at once. Solutions are often represented using two vectors, one which denotes the assignment of operations to machines and another which defines their sequencing (for example, [78], [56]).

Multi-objective formulations of the problem have also been studied. In [78], a Pareto approach is employed, where a number of solutions which are not dominated by any other solutions (composing a Pareto front) are sought. In contrast, [56] present an approach whereby objectives are *a priori* ranked by their importance and then used in a lexicographic manner. That is, the second most important objective is only used when solutions are tied on the objective deemed most important. A further approach to the problem posed by multiple objectives is exhibited in [70]. In this work, the authors use fuzzy logic to compute the weights to be assigned to each objective, which are then used in a weighted-sum aggregated objective function. The Genetic Algorithm variant proposed in [69] is then employed. The authors note, however, that weights should be set by the decision maker when possible.

The two key components of an optimisation model are the decision space and the objective space. Following the preceding discussions, it is clear that there is little of direct relevance to the casualty processing problem in terms of the objective spaces employed in typical VRP or JSP formulations. While the objective measures commonly employed, such as makespan and cost, may be applicable to the casualty processing problem, it is unlikely that they will be capable of capturing the intricacies of the problem at hand. However, the decision spaces used in VRP and JSP formulations do have the potential to exhibit a structure similar enough to be used as a basis of a decision formulation for the casualty processing problem. We will return to these points as we develop the casualty processing model throughout this chapter.

4.3 The modelled environment

In this section we focus on the details of the problem environment as outlined in Section 2.3.5, and describe their representation within the proposed model. In turn we consider the transportation network, casualties, hospitals, types of responder units and finally the tasks which these responder units carry out.

4.3.1 Transport network and incident sites

The spatial environment is represented primarily through its transport network. This network is represented as an undirected graph, where nodes correspond to junctions and locations of interest, and edges represent the corresponding length of road. Each network link is characterized by a single parameter, representing its length in kilometres. This information is available for any region of the UK via Ordnance Survey GIS files, which provide a high degree of detail and thus ensure an accurate representation. An example transport network, corresponding to an area of central London, is provided in 7.2. This figure illustrates the dense, detailed nature of the available data.

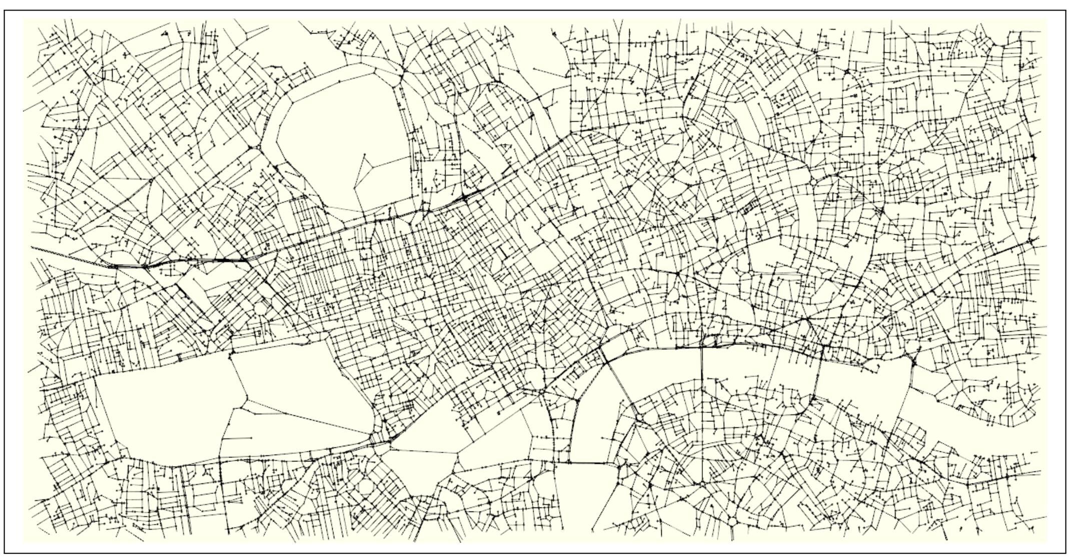

Figure 4.3: The transport network graph corresponding to a modelled area in central London.

Mass Casualty Incident sites are represented by a number of discrete ‘incident site’ nodes. Each incident site node is placed within the transport network graph by connecting it to the nearest road node using an edge of length zero. Given the dense network structures which are considered in this thesis, it is assumed that a road node will always exist sufficiently close to the specific (x, y) location of the incident site. Each incident site has a set of casualties, which, together with their characteristics, completely defines the nature of the incident. That is, the model does not discern between any two situations which result in the same number of casualties with identical profiles.

At each incident site a Casualty Clearing Station is assumed to be set up, as is standard practice, together with an associated ambulance loading point and ambulance parking point. It is assumed that for all incidents considered a suitable location or structure is available for this purpose.

4.3.2 Casualties

Casualties are defined as those individuals who have sustained injuries at an incident site and therefore require transportation to hospital. We denote the set of all casualties by $\mathcal{C}$ and an individual casualty by c_i, so

$$c_i \in \mathcal{C}, 1 \leq i \leq n_c, \tag{4.1}$$

where n_c is the total number of casualties. This set of casualties may grow as an incident evolves in time in the manner discussed in Section 3.4. However, in this chapter the focus is on static incidents, and so we do not consider such dynamic casualty sets until Chapter 6.

Casualties are defined by their location, their accessibility (i.e. to what extent they can be reached by responders, how long it will take for them to be extricated by the fire service) and their health. Specific details regarding how each of these factors is represented in the model follow.

Each casualty is associated with a specific incident site, which in turn defines their location within the environment. Thus, in terms of location, all casualties associated with a particular incident site are considered equivalent. This level of detail in spatial representation of casualties is common, as can be seen in the work of [45]. While it would be possible to further specify casualty location, perhaps to the level of individual buildings or to specific (x, y) coordinates, the benefits of including such further detail in a scheduling optimisation model would be limited. In particular, we argue that differences in casualties arising from their location may be subsumed into differences in the durations of the tasks associated with them. This approach is illustrated in Example 4.3.1.

Example 4.3.1. Consider a scenario where three casualties are located at an incident site. The node of the transport network which corresponds to this incident site is denoted by X, and each of the three casualties has a unique (x, y) coordinate near it.

As discussed, two possible approaches are available for the representation of the locations of each casualty. In the first case, an *explicit* representation may be employed, where the specific (x, y) coordinates of each casualty is held within the model.

In this case, any calculations regarding, for example, the time needed for a Search & Rescue response unit to extricate a specific casualty would be broken down in the following manner:

$$\begin{aligned} \text{Total time} &= \text{ time to travel to node } X + \text{ time to travel to (x,y) } + && (4.2)\\ &\text{extrication time} && (4.3)\\ &= 5.4 \text{ mins } + 0.5 \text{ mins } + 7.4 \text{ mins.} && (4.4) \end{aligned}$$

Alternatively, an *implicit* representation may be employed. In this case, the time needed for the extrication operation takes the form

$$\begin{aligned} \text{Total time} &= \text{time to node } X + \text{modified extrication time} && (4.5)\\ &= 5.4 \text{ mins} + 7.9 \text{ mins.} && (4.6) \end{aligned}$$

Here, the specific differences in locations of casualties associated with the same incident site are accounted for when setting times needed to complete tasks. It is this second representation which is employed in the model proposed in this thesis.

In addition to location, each casualty requires three further variables to be described fully. Specifically, we require:

- $c_i^s \in \{0, 1\}$, a binary stability variable indicating whether stabilizing treatment is required;

- $c_i^e \in \{0, 1\}$, a binary 'trapped' variable indicating whether or not the casualty requires extrication; and

- $c_i^t \in \{T1, T2, T3, Dead\}$, a discrete variable describing the triage level associated with their initial health.

As discussed in Section 2.3.5, triage is a procedure carried out in MCIs where the health of each casualty is briefly assessed in order to estimate the extent of their injuries. The result of this sieve procedure is a designated triage level, which can take one of the four categories shown in Table 4.1 [1] (as previously shown in Table 2.5 and repeated here for ease of reference). In the UK, following a major incident, it is standard policy to require a full triage operation be completed before any treatment is administered to any casualty. As such, we note that it is reasonable to assume that the proposed model can be initialized after this triage operation has been completed and will therefore have access to all relevant information including the number of casualties and their health, stability and need or otherwise of extrication.

While casualties are assumed to be mostly passive, and thus under the control of the response agencies, we do allow for the possibility of some casualties deciding to leave the incident site and travel to a hospital of their choosing. This is termed 'self-presentation'. Such an action is assumed to only occur for those casualties of triage state T3, the 'walking wounded'.

Table 4.1: Triage levels assigned to casualties [1].

Category	Description	Explanation
T1	Immediate	Require immediate life-saving procedure
T2	Urgent	Require surgical or medical intervention within 2-4 hours
T3	Delayed	Less serious cases whose treatment can safely be delayed beyond 4 hours
Dead		

4.3.3 Hospitals

The set of all hospitals considered in the model is denoted by $\mathcal{H}$, with individual hospitals represented by h_i. So,

$$h_i \in \mathcal{H}, 1 \leq i \leq n_h, \tag{4.7}$$

where $n_h = |\mathcal{H}|$ denotes the total number of hospitals. Each hospital is represented on the network as special nodes which are connected to the nearest road node in the same way as incident sites. That is, a hospital node is created and joined, via an edge of length zero, to the nearest node on the transport network graph. In addition to their locations, hospitals require further parameters to be set in order to describe their resource levels and how these are likely to change over time.

The 'resources' available at any given hospital are considered to be encapsulated by its capacity in terms of the number of patients who could be admitted, and subsequently treated, at that point in time. Thus, the term 'resources' is somewhat abstract, and for its full derivation one would require detailed information regarding quantities such as available staff, the number of free beds, and medical supplies. By defining resources at the aggregate level we ensure the resulting model is sufficiently flexible in its application to different scenarios. Resources are modelled to be dynamic, albeit in a simple, linear manner. Specifically, a hospital h's resource levels are defined through an initial level, cap_h, a maximum capacity max_h, and a linear rate at which capacity will increase upon enaction of a the hospital's major incident plan, $rate_h$. This plan will result in the freeing up of resources through the movement and re-allocation of those casualties currently in the hospital.

The rate at which hospital resources grow in response to the enaction of a major incident plan may not be as simple as a piecewise linear model. However, there is an absence of real data recording such growth in actual major incidents. Moreover, as discussed in Section 3.3.2, there has been limited research into MCI response optimisation models which has included the explicit modelling of hospitals. As such, we argue that the simple linear model is appropriate at this stage.

In addition to resource levels, we further allow for any hospital to have one or several specialist treatment facilities present. In our implementation and by means of example, we consider one such type of specialist fatalities, namely burns units. However, the model can be easily extended to any number of specialist facilities as required. The impact of such facilities is felt in the model when evaluating response plans, and so will be discussed in further detail in Section 4.6.3.

Where specialist treatment facilities do exist, it may be desirable to set resource level constraints specific to those facilities. Thus, a hospital holding a burns treatment unit could have a capacity of 30 casualties in total, 7 of whom may be within the burns unit. If such a configuration is desired, it may be obtained through the representation of the burns unit as a separate hospital at the same location, with resource level parameters set accordingly.

4.3.4 Responders

The model considers a number of specific emergency responders, coming from the Fire and Rescue Service, Ambulance Service and Health Service. In the case of the Fire and Rescue Service, we consider a single, homogeneous responder type, namely a Search And Rescue (SAR) unit. Each unit corresponds to a team of individuals assigned a generic fire appliance, containing the necessary equipment to carry out search and rescue operations. While, as discussed in Section 2.3.4, there is a large variety in fire appliances and resources available to the Fire and Rescue Service, the MCIs to be considered by our model will require only these most general appliances. All SAR units are initially located at local Fire stations, as specified within the modelled environment.

The responders made available by the Ambulance Service consist of two-person paramedic teams each with an allocated ambulance, in addition to Hazardous Area Response Teams (HART units). HART units [35] are specially trained and equipped paramedic teams whose primary purpose is to provide life-saving treatment to those casualties who are in hazardous environments. Both Ambulance and HART units are initially located at either local ambulance stations or hospitals.

Finally, we consider Medical Emergency Response Incident Teams (MERIT units) units. MERIT units consist of teams of clinicians capable of fast dispatch to an incident site in order to deliver stabilizing treatment to casualties within the Casualty Clearing Station.

All response units are suitably defined within the model through their specified initial location (i.e. their base station or hospital), parameters describing their maximum travelling speed, average rate of acceleration, and an optional delay time corresponding to, for example, the time needed from receiving notification of the incident to prepare for dispatch. The inclusion of this parameter helps to ensure the flexibility of the model. A further variable which could potentially have been included in the specification of a responder would be a time at which they may no longer work, corresponding to an end of a shift. However, given the scope of problems to be considered in this work and their relatively short time frames of a few hours, this detail was deemed unnecessary.

The total set of all response units is denoted $\mathcal{R}$, with individual response units denoted r_i. The total number of responder units is denoted by $n_r = |\mathcal{R}|$. A summary of the types of responders considered in the model is presented in Table 4.2.

Table 4.2: Responder units considered in the model.

Name	Description
Ambulance	An Ambulance unit includes a paramedic team, and can both administer treatment at a CCS and transport casualties to hospital.
HART	A Hazardous Area Response Team consists of paramedics equipped with the necessary equipment and training to allow them to administer stabilizing care to casualties in high risk environments, i.e. those who are trapped.
MERIT	A Medical Emergency Response Incident Team is a mobile team of clinicians who can travel to any mass casualty incident and administer treatment to the wounded at the CCS.
SAR	A Search And Rescue team can rescue trapped casualties from disaster sites and deliver them to the associated CCS.

4.3.5 Tasks

Having discussed the representation within the model of the environment, casualties, hospitals and responders, we now discuss the representation of the tasks which the responders are required to carry out in an MCI response operation. Firstly, the tasks corresponding to the Ambulance and Health Service will be described. Following this, tasks completed by the Fire Service will be considered.

As discussed in Section 2.3.4, the core duties of the Ambulance Service are the "three T's": Triage, Treatment, and Transportation. We model treatment and transport explicitly, with triage accounted for in an implicit manner as discussed below.

As previously noted, it is standard policy [128] for a full triage operation to be completed before any treatment work may be undertaken. As such, we assume that an initial triage operation has occurred by the time $\tau = 0$ in the model. At this stage, we consider this initial triage to be the only one to take place for each casualty. However, as discussed in Section 2.3.5, triage is in fact repeated throughout the response operation, allowing for any changes in the health of casualties to be monitored. The complexities introduced by such dynamic data collection will be considered in more detail in Chapter 6 of this thesis.

Treatment can take place in one of two environments at any incident site. For those casualties who are trapped within the hazardous inner cordon area and who are awaiting rescue, a HART unit may deliver pre-rescue stabilizing treatment. For all other casualties at the incident site, pre-transportation stabilizing treatment may be delivered by a HART, MERIT or Ambulance unit. We note that it is the stated goal of the UK Ambulance Service to ensure such treatment is minimal, providing only enough to ensure each casualty's safe progression in the next stage of their processing. Thus, we do not consider the health of a casualty to be improved by treatment - only that it will not deteriorate during the treatment operation. It should be noted that this approach to modelling treatment avoids the addition of many more decision variables corresponding to treatment time. That is, the preparation of any casualty for the next stage of their processing will take a specific, fixed amount of time. If the goal of treatment was instead to improve the health of the casualty, there would be no well-defined, natural length for the treatment operation. In this case, the decision space of the model would have to encompass the setting of the length of each treatment task.

Transportation tasks consist of several distinct stages, namely loading, travel and unloading. We assume loading and unloading are both invariable and fast operations, and do not include their time demand in the model. The removal of this assumption, on recovering adequate data describing loading and unloading times, would be simple. Transportation tasks are unique within our model as their duration is a decision variable as opposed to a (possibly unknown) fixed and true quantity. That is, the choice of hospital allocation will define the length of the corresponding transportation task.

Rescue tasks are those undertaken by SAR units in order to remove casualties from a hazardous environment in which they are trapped. Just as treatment tasks were noted to have a well-defined length (the time needed to prepare the casualty for extrication or transportation), rescue tasks also have a set time associated with them. Thus, the decision space of the model does not need to include the setting of the length of time spent on each individual rescue operation.

While a transportation task is required for each casualty considered, the presence of all other tasks is dependant on the defining casualty characteristics as discussed in Section 4.3.2. Specifically, only those casualties whose 'trapped' variable c_i^e is set to `true` will have a rescue task associated with them, and only those whose stability variable c_i^s is set to `false` will have a pre-transport stabilization task associated with them. Where both these conditions hold, a pre-rescue stabilization task will also be required. For all tasks required for a given casualty c_i, a strict dependency exits in terms of the order in which they are undertaken. This ordering is illustrated in Figure 4.4. A summary of the tasks considered is provided in Table 4.3, and their relation to both responder types and casualty characteristics is illustrated in Figure 4.5.

Given that all tasks are associated with a specific casualty, and that an ordering of the tasks corresponding to each casualty exists (as illustrated in Figure 4.4, the jth task associated with the ith casualty will e denoted $t_{i,j}$. The set of all tasks will be denoted $\mathcal{T}$. If casualty i requires $n_{i,t}$ tasks to be completed, our notation is summarized as

$$t_{i,j} \in \mathcal{T}, \text{ where } 1 \leq i \leq n_c \text{ and } 1 \leq j \leq n_{i,t}. \tag{4.8}$$

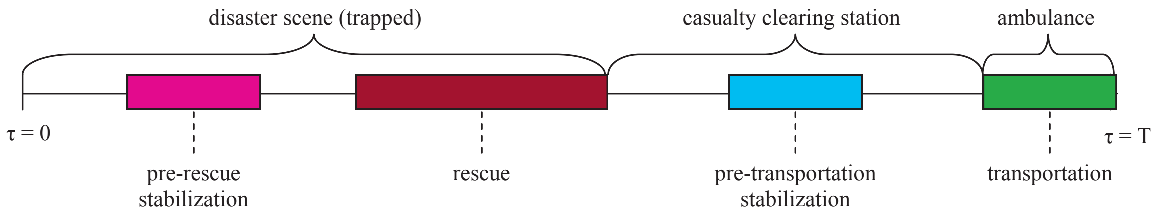

Figure 4.4: An example of the processing of a single casualty in an MCI

Table 4.3: Tasks considered in the model.

Name	Description
Transportation	All casualties require transportation to a hospital.
Pre-transportation stabilization	Those casualties whose condition is unstable require a period of treatment/stabilization to be carried out to ensure their safe transportation.
Rescue	Casualties may be trapped by debris at the disaster site, in which case a rescue task must be completed to ensure their extrication.
Pre-rescue stabilization	Of those casualties who are trapped, some may require a period of treatment/stabilization before the rescue task commences in order to ensure their safety.

In this section we have taken the qualitative description of the scope and characteristics of the decision problem we consider, as given in Chapters 2 and 3, and provided a concrete and quantitative definition. The dimensions along which a problem may vary have been described. The problem scope is sufficiently broad to allow for a large number of scenarios to be effectively modelled. In particular, the scope is not limited in terms of geographic area (any transport network may be used, at the fine detail provided by GIS, or at any other aggregated level which can be represented as a graph). Neither is there a limit upon the number of casualties or responders, other than that imposed by computational burden. The representation of a response operation in terms of a finite number of four task types helps to ensure flexibility of the decision space of the model, with the content of tasks open to interpretation and only the dependencies, both in terms of task sequences and responder capabilities, fixed.

4.4 Defining solutions

The previous section focused on how a problem and its environment may be represented within our model. We now proceed to focus on which aspects of the problem are controllable and amenable to decision support. That is, we seek to define the

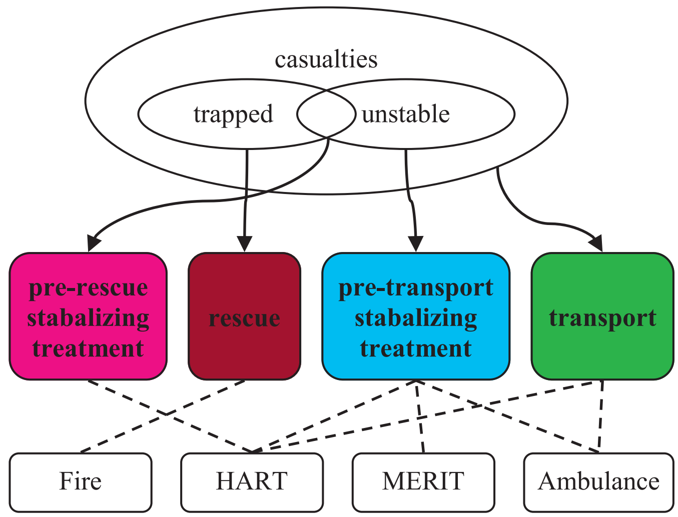

Figure 4.5: The relationships between casualties, tasks and resources.

solution space. We do so in an incremental manner, allowing for a progressively finer level of control for the decision maker. Challenges relating to infeasible solutions will then be outlined.

4.4.1 The solution space

Firstly, following the structure of general Flexible Job-shop Scheduling problems, we consider the allocation of tasks to the available response units. Formally, for the jth task associated with the ith casualty $t_{i,j} \in \mathcal{T}$ we associate a responder unit $t^r_{i,j} \in \mathcal{R}$. This decision making, in isolation, provides a limited amount of control of the solution to the decision problem. In such a situation, the decision maker could allocate all tasks to be completed to the available responder units, but would not have control of other aspects such as the order in which each responder completes their allocated tasks. The benefit of allocating this decision making stage to an optimisation process will be examined in Section 7.3. In terms of the size of the solution space defined by this single decision set, the solution space, denoted $\mathcal{S}_1$, will be of size

$$|\mathcal{S}_1| = |\mathcal{T}|^{|\mathcal{R}|}. \tag{4.9}$$

Given an allocation of tasks to responder units, a further dimension to the solution space may be imposed through allowing for the sequencing of tasks assigned to each responder unit. That is, for each task we define a priority level $t^p_{i,j} \in \mathbb{N}$ and these priority levels are used to order the tasks allocated to any given responder unit in the desired manner. The solution space is constrained such that the n tasks assigned to any given responder are assigned priorities from $1, \ldots, n$. This decision component confers significantly greater control over the solution to the response problem than that afforded by the allocation of tasks to responders alone. Now, the decision maker has full control over not only who does which tasks, but in what order they should

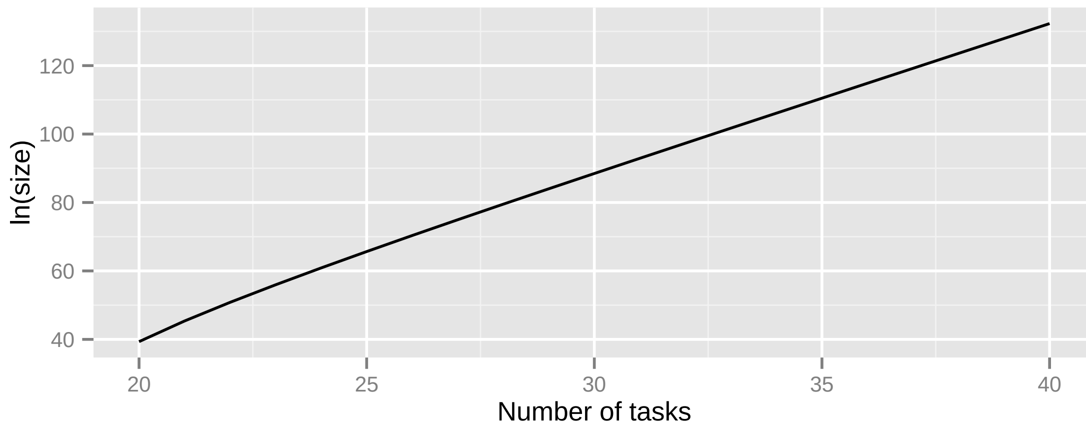

Figure 4.6: The exponential growth of the size of the solution space for a fixed number of responders and increasing number of tasks.

be completed. This increased level of control comes at the cost of an increase in complexity through the size of the solution space. Specifically, we now have a solution space $\mathcal{S}_2$ such that

$$|\mathcal{S}_2| = (|\mathcal{T}| - 1)! \times \frac{|\mathcal{T}|!}{(|\mathcal{T}| - |\mathcal{R}|)!}. \tag{4.10}$$

Note that this solution space may contain infeasible solutions. Such solutions could occur when tasks are allocated to responder units who are not capable of their completion, or when tasks are sequenced in a manner which violates dependency relations. Such infeasible solutions will be discussed in more detail shortly, in Section 4.4.2.

A final level of control is realized through the allocation of casualties to hospitals. That is, for each task we associate a hospital $t_{i,j}^h \in \mathcal{H} \bigcup \{0\}$, where $h = 0$ for all tasks other than transportation tasks. This decision dimension provides potential for an optimisation model to incorporate knowledge of hospitals as described in Section 4.3.3, namely their location, resource levels and availability of specialist treatment facilities. This final increase in the solution space leads to a solution space S_3, such that

$$|\mathcal{S}_3| = \left((|\mathcal{T}| - 1)! \times \frac{|\mathcal{T}|!}{(|\mathcal{T}| - |\mathcal{R}|)!} \right) \times \left(|\mathcal{C}|^{|\mathcal{H}|} \right). \tag{4.11}$$

We now have that a solution can be defined by a mapping $s : \mathcal{T} \to \mathcal{R} \times \mathbb{N} \times \mathcal{H} \bigcup \{0\}$. The size of this space, as is common with combinatorial optimisation problems, scales exponentially as we increase the number of tasks and/or responders, corresponding to the $\mathcal{T}$ and $\mathcal{R}$ dimensions. This growth with respect to the number of tasks is illustrated in Figure 4.6, where the size of the solutions space of a problem involving ten responders is given, on a log scale, against the number of tasks. A numerical example is provided in Example 4.4.1

Example 4.4.1. Consider a solution space arising from a problem involving 10 responder units, 25 tasks associated with 8 casualties, and 3 hospitals. Equation 4.11 gives

$$|\mathcal{S}_3| = \left((25-1)! \times \frac{25!}{(25-10)!}\right) \times \left(8^3\right) \tag{4.12}$$

$$= 3.28 \times 10^{50} \tag{4.13}$$

Increasing the size of the problem by 3 further casualties and an associated 9 further tasks leads to a solution space of size

$$|\mathcal{S}_3| = \left((34-1)! \times \frac{34!}{(34-10)!}\right) \times \left(11^3\right) \tag{4.14}$$

$$= 1.42 \times 10^{77} \tag{4.15}$$

An increase of only 3 casualties and 9 associated tasks in the problem to be solved has lead to an increase in the solution space of a factor of the order of 10^{27}. This illustrates the inflexible nature of combinatorial optimisation models.

More formally, we note that the solution space can be translated into that of a common Flexible JSP (FJSP) by setting parameters describing distances on the road network to zero, considering casualties as jobs and responder units as machines. Since the JSP has been shown to be NP-hard [95], we can deduce that the casualty processing problem is also NP-hard. The implication of this in terms of appropriate solution method will be discussed in Chapter 5.

4.4.2 Infeasible solutions

The solution space defined in Section 4.4.1 includes solutions which are infeasible. In addition to the aforementioned possibility of tasks being assigned to responder units incapable of their completion, it is possible that tasks may be allocated and ordered in such a fashion as to create a 'loop' in dependence relations. By dependency relations we include both normal sequencing dependency (that is, tasks must be completed in the order in which they were assigned, and so each task is dependent on all tasks preceding it in the sequence) and the dependence of task types as described in Section 4.3.5. A simple illustration is given in Figure 4.7, where a loop exists over a single responder unit's schedule due to task 3 being required to be completed before work on task 1 can begin, but task 3 being scheduled after task 1 in the same responder's schedule.

A more complex example involving two responder units is provided in Figure 4.8. In this case, dependencies between tasks assigned to different responders give rise to an infeasible solution, as can be seen by following the directed arrows from task 2 onwards, eventually leading back to task 2 again. Clearly, these simple examples can extend to far larger cases involving many responder schedules, and any method for evaluating the quality of a proposed solution must be capable of identifying any such dependency violations.

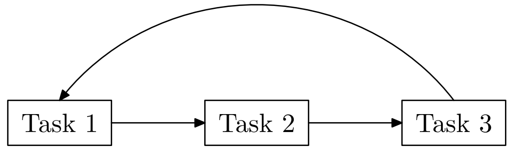

Figure 4.7: A simple example of a task dependency violation within a single responder unit's schedule.

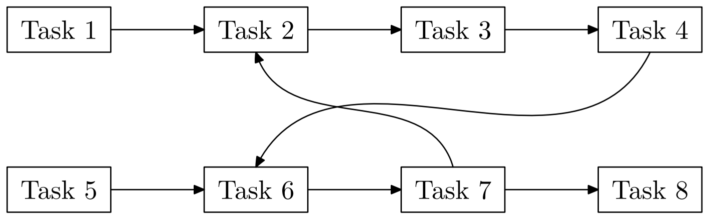

Figure 4.8: A task dependency violation involving two distinct responder units.

4.5 Creating schedules

In order to evaluate the quality of a solution as defined in Section 4.4, we must first combine the solution specification with knowledge of the problem environment to create a schedule predicting when each task will start and finish. To do so, we first consider the estimation of travel times before defining the scheduling algorithm.

4.5.1 Travel times

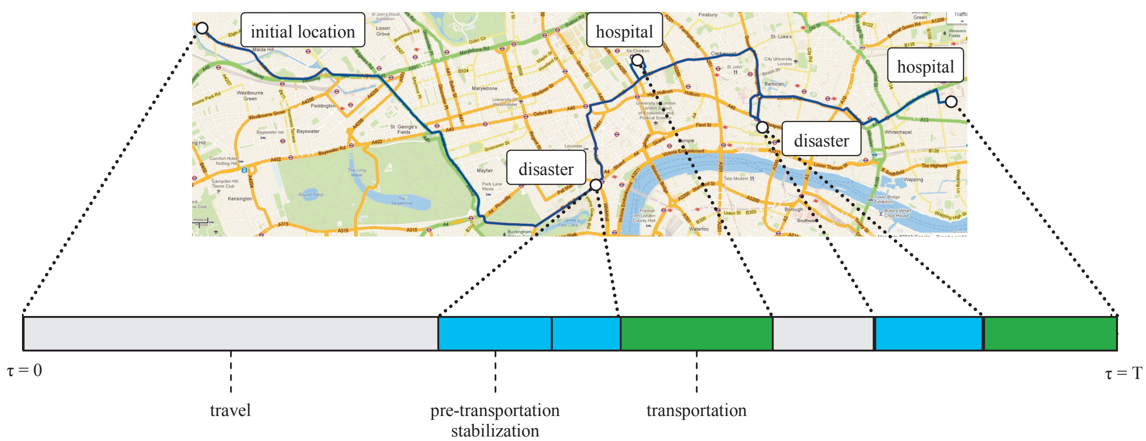

Figure 4.9: The work schedule and associated travel path of an ambulance.

As illustrated in Fig. 4.9 the tasks within our model are distributed across a geographical area and as such we must account for the time needed for responder units to travel from the end location associated with one task to the start location of the next task. The prediction of such travel times is accomplished through first estimating the route taken and then, using the resulting distance and parameters describing the speed and acceleration of the responder unit, estimating the corresponding travel time.

Considering the routing problem, we assume that, through personal knowledge and / or the availability of sat-nav routing technology, a responder unit will always take the shortest path available. Under this assumption, we may employ a shortest path algorithm to find this path. Djikstra's algorithm, as implemented in the Boost graph library [114], is employed in this manner. Denoting the distance (in kilometres) of the shortest path by d, we go on to use the model described by [72] (and more recently validated in [15]) to estimate the median travel time $\hat{m}(d)$:

$$\hat{m}(d) = \begin{cases} 2.42\sqrt{d}, d \leq 4.13 \text{ km} \\ 2.46 + 0.596d, d > 4.13 \text{ km} \end{cases} \tag{4.16}$$

where the time is given in minutes. The model was fitted to data describing the travel times of the ambulance service in Calgary, Canada [15]. In the absence of emergency service travel time data for the UK, this was taken to be a sufficient approximation. It should be noted that the data used referred to ambulances travelling to reach everyday emergencies, as opposed to travelling within an MCI environment. It may reasonably assumed that ambulance speeds and associated travel times will be affected by the disruptions caused by an MCI. This will be considered in further detail in Chapter 6, where a model for transport network disruption will be described.

Calculating the median travel time estimate in this manner for each journey, these times are then used when constructing the schedule. In terms of the Job Shop Scheduling methodology this corresponds to transforming a basic FJSP into a FJSP with *sequence dependent setup time* [109]. This same travel time estimation method is employed when calculating the duration of transportation tasks, which vary according to the location of the associated casualty and which hospital said casualty has been allocated to. The durations of all other tasks (that is, stabilizing treatment and rescue tasks) are considered to be available upon initialization of the model. This assumption will be relaxed in Chapter 6.

Implementation 4.5.1. Given the flexibility of the model in terms of the geographic area covered, it is possible that large, dense graphs representing transport networks will be encountered. In such cases, the calculations of shortest paths may incur a significant computational burden and as such should be kept to a minimum. This is achieved through the reduction of the full transport network to a travel time matrix in a pre-computation stage. Specifically, during the initialization of the model a list of all locations of interest is compiled, including responder bases, hospitals and incident sites. Repeated applications of shortest path calculations leads to a single matrix storing the distances of all routes between any two locations, which can then be translated into travel times using the formula 4.16. These values can then be looked up as and when they are required without the need for further, computationally expensive, shortest path calculations.

4.5.2 Scheduling algorithm

The creation of a schedule from a given solution, where the start and end times of all tasks are estimated, is described in Algorithm 1. In this algorithm, the set of tasks is iterated over, with the start and end times of each task being computed in turn. When considering any given task $t_{i,j}$, all tasks upon which $t_{i,j}$ depends are examined. These tasks may be located at an earlier position in the same responders schedule, or they may be associated with the same casualty. For example, when computing the start and end times of a rescue task, the algorithm will check if a pre-rescue stabilization tasks exists for the corresponding casualty. If such a task does exist, it is necessary to compute the start and end times of this before the times of task $t_{i,j}$ can be computed. This is because the start time of $t_{i,j}$ is dependent on the end time of this task. Once the start and end times of all tasks upon which task $t_{i,j}$ depends are known, the algorithm proceeds with the computation of the start and end times of task $t_{i,j}$.

Algorithm 1 Creating a schedule from a solution.

```
set VALID = true
for all tasks t ∈ 𝒯 do
    if t^r can NOT do t then
        set VALID = false
    end if
    if t is first task in t^r's schedule then
        x = getTravelTime(t^r's initial location, start location of t)
        set start time of t = x
        set end time = start time + duration
    else
        get task preceding t, t̂
        if t̂ times have not been updated then
            update times of t̂
        end if
        x = [end time of t̂] + getTravelTime(end location of t̂, start location of t)
        set y = 0
        if t depends on another task, t* then
            if t* times have not been updated then
                update times of t*
            end if
            y = end time of t*
        end if
        set start time of t, max(x, y)
        set end time = start time + duration
    end if
end for
```

Implementation 4.5.2. The scheduling algorithm allows for the recognition of infeasible solutions by adding a check to each task to denote if it is currently under evaluation. As described, it is possible that a specific task $t_{i,j}$ may be considered, but the algorithm proceeds to compute the timings of a task or tasks upon which $t_{i,j}$ depends. This process could continue, following a chain of dependency relations of the type illustrated in Figures 4.7 and 4.8. As in those illustrated cases, if following the chain of dependencies eventually leads back to the initial task being considered, a dependency loop exists and the solution is infeasible. Thus, by noting when each task is considered and therefore becomes part of a chain, the algorithm can recognize when such loops exist. At this point the scheduling process is terminated and a high number is assigned to all objectives of the current solution, allowing for the optimisation process to discount it.

4.6 Schedule evaluation

As discussed in Section 4.2.2, while similarities between the casualty processing problem and the more general class of JSPs can be seen in the similar creation of a schedule as part of the evaluation process, the problems are markedly different as we move beyond this point. While the objective functions typically used in FJSP's (such as makespan) may be applicable to the casualty processing problem, they are unlikely to fully capture its nature. Specifically, in order to properly evaluate a solution to the casualty processing problem, consideration must be given to the resulting health and well-being of casualties.

A multi-objective approach to the evaluation of solutions is proposed, considering the following five objective functions which can be applied to a solution s:

- $f_1(s)$, the expected number of fatalities,
- $f_2(s)$, measure of the time taken to deliver casualties to hospitals,
- $f_3(s)$, measure of how appropriate the hospital allocation choice is,
- $f_4(s)$, the total time spent idle by responders,
- $f_5(s)$, the latest time at which a casualty arrives at a hospital, i.e. the makespan.

We group the above objectives into three classes: *fatalities*, consisting of f_1 alone; *suffering*, consisting of f_2 and f_3; and finally *efficiency*, consisting of f_4 and f_5. In what follows we will describe each objective f_i individually, after which we will discuss how to combine them in a unified manner through multi-objective techniques.

4.6.1 Fatalities

In order to predict the number of fatalities resulting from a response operation s, we first note which casualties $c \in \mathcal{C}$ are in a dangerous environment at any point. As illustrated in Figure 4.4, a casualty can be in one of four environments during a response operation: trapped at the disaster site; at a Casualty Clearing Station; in an ambulance; or at a hospital. We assume that the latter three environments are of a relatively stable nature and casualties will not deteriorate in health over the course of the operation. For casualties trapped at the scene, however, we acknowledge the risk of further injury and the deterioration of health.

Given the discrete nature of triage classification we propose a discrete state Markov chain model of casualty health in a similar fashion as in [111], with a state space $L = \{T1, T2, T3, D\}$ denoting the four triage levels described in Table 4.1 (where D corresponds to *dead*). This approach allows the calculation of the probability that casualty c_i will be in state $T \in L$ at time τ under the proposed solution, which we shall denote by $p_i^T(\tau)$. The parameters used in the Markov chain are given in Fig. 4.10.

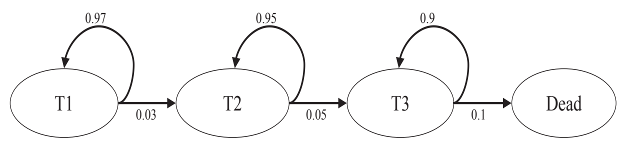

Figure 4.10: The Markov chain representing the stochastic process of the health of a trapped casualty, with transition probabilities for each state transition.

Two health states are linked if it is possible to move from one to the other in any given time step, where a time step represents one minute. As can be see from Figure 4.10, only negative health progression is possible when a casualty is trapped at the scene. The use of Markov chains in this manner is attractive as it facilitates the calculation of a number of quantities of interest in an efficient manner. In particular, in addition to providing an estimated probability that any trapped casualty will be deceased before they reach a Casualty Clearing Station, one may also calculate the probability of being in any other health state at any other time. This information can then be used when prioritizing casualties according to their current health and predictions of how it will change over time. These calculations are rendered feasible through the assumption of the Markovian property, which states that the stochastic process must be memoryless. In the context of our problem, this translates to assuming that the probability of a casualty's health deteriorating from one level to the next is dependent only on their current state, not on how long they have occupied it. Under this assumption, predicted health states can be calculated based on only the current health state of the casualty at that point in time. If a non-Markovian chain were employed, such calculations would require a full knowledge of the health history (that is, the specific time periods spent in each health state over the course of the response) of the casualty in question. This would represent a significant challenge in practice, in terms of collecting and storing the relevant data. A related favourable property of a Markov chain formulation is the compact nature of the model. Specifically, it requires only state-to-state transition probabilities to be specified. This is a more realistic requirement than that posed by a non-Markov chain, which would require the specification of transition probabilities to be dependant also on the health history of casualties. As the Markovian assumption cannot be explicitly tested in this case, due to a lack of data, it was deemed to be reasonable for the purposes of this work.

Denoting by τ_i^C the time at which casualty c_i arrives at a Casualty Clearing Station after being extricated, the Markov chain model is used to calculate $p_i^D(\tau_i^C)$ for each casualty. We can therefore define the fatality component of the objective function to be

$$f_1(s) = \sum_{i=1}^{n_c} p_i^D(\tau_i^C). \tag{4.17}$$

The parameters used in these probabilistic models should ideally be determined from an extensive analysis of data. However, in the field of MCI response such data is not freely available and as such the transition probabilities in Figure 4.10 have been estimated. In order to better understand the behaviour resulting from this parametrization it is useful to perform some descriptive analysis. Figure 4.11 illustrates the probability of a casualty dying, and how this probability varies according to the time taken to rescue the casualty. This analysis is conducted for casualties whose initial health state is T1, T2 and T3.

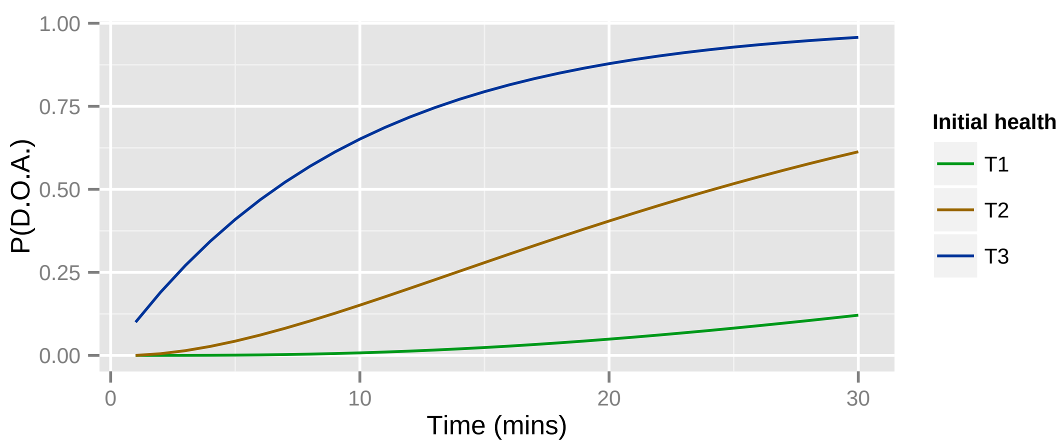

Figure 4.11: The probability of death on arrival at hospital over time spent trapped at incident site, dependent on initial health state.

The behaviour displayed in Figure 4.11 demonstrates that the transition probabilities given in Figure 4.10 do in fact lead to plausible behaviour. Specifically, the probability of a casualty of health state T3 (i.e. the 'walking wounded') dying whilst trapped at an incident site remains low over a period of 30 minutes, reaching a maximum of 0.125 within this time. In contrast, a casualty of the most severe health state, T1, will have a high probability of death in this same period. These behaviours correspond sufficiently to the qualitative descriptions of the triage health states as noted in Table 4.1.

To further illustrate the behaviour predicted by the Markov chain model described,

Figure 4.12 shows the probability vector of health for a single casualty over a 30 minute period as they receive pre-rescue stabilizing treatment, and then are rescued, with T3 as their initial state. The figure shows how the probability of being in state T3 decreases over time, from being completely certain (i.e. $P(D) = 1$) immediately following triage to around $P(D) = 0.6$ after 30 minutes. Corresponding to the decrease, the probability of the casualty being in either state T1 or T2 increases over time.

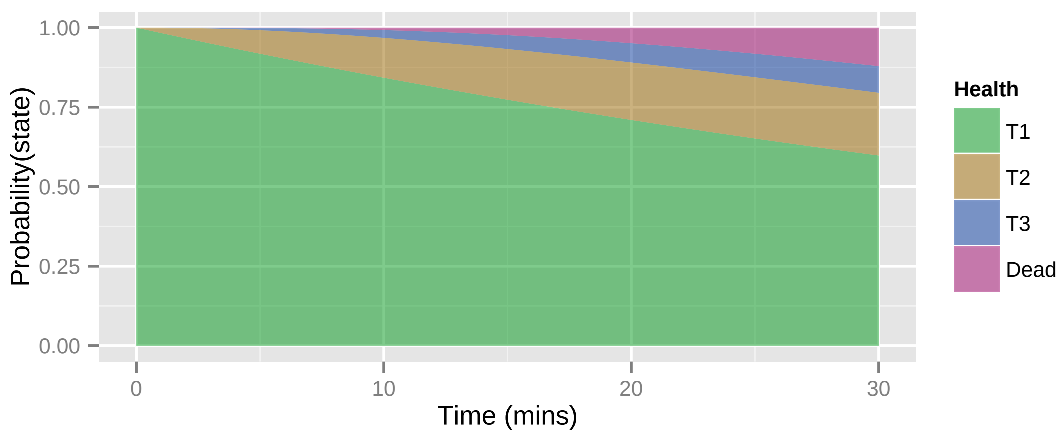

Figure 4.12: Probability of being in each health state whilst trapped at incident site.

4.6.2 Hospital arrival time

While the minimisation of fatalities is clearly a key objective held in any response operation, it will not sufficiently capture all relevant details of the problem. That is, it is possible for several response plans to lead to approximately the same predicted fatalities, but for these plans to be clearly separable on other measures. One such measure to be considered involves the speed at which the casualties involved in the incident are taken to hospital.

We note that the prioritization of casualties in a way which respects their triage level is essential to a high-quality response, by the very definition of triage. This can be achieved through the use of a weighted total flowtime measure. Here, we sum the completion times of each casualty's processing, i.e. the time at which they are delivered to a hospital (which we denote $t^c_{arrives}$), weighting each component by the parameter w_T, where $T \in L = \{T1, T2, T3, D\}$ denotes the triage levels. This defines the second objective function component,

$$f_2(s) = \sum_{T \in L} \left(\sum_{c \in \mathcal{C}_T} w_T t^c_{arrives} \right). \tag{4.18}$$

The weights used have been set in accordance with the description of triage levels given in Table 4.1, where we set a twenty four hour delay in the treatment of a T3 casualty to be equivalent to a four hour delay in the treatment of a T2 casualty. This in turn is set to be equivalent to a fifteen minute delay in the treatment of a T1 casualty. That is, taking $w_{T3} = 1$, we calculate $w_{T2} = \frac{24}{4} \times w_{T3} = 6$ and $w_{T1} = \frac{4}{0.25} \times w_{T2} = 96$. The weight corresponding to the dead is set at $w_{dead} = 0.1$ in order to ensure the model places only limited value on the prompt transportation of fatalities to hospital in comparison to the transportation of injured survivors.

4.6.3 Hospital allocation

In order to quantify how well casualties have been allocated to hospitals, we must consider two factors: the dynamic capacity of each hospital and the effect of over subscription; and the pairing of casualties' specific injuries to the corresponding specialist treatment facilities.

Hospital capacity

We consider two factors which will result in a dynamic variation of a hospital's available capacity. Firstly, the effect of a hospital enacting its major incident plan. Secondly, the effect of casualties autonomously leaving the disaster scene and transporting themselves to self-present at a hospital of their choosing.

The result of a hospital's major incident plan being enacted is modelled as a steady increase in its capacity. We characterize this process using the following parameters:

cap_h:	Initial free capacity of hospital h
max_h:	Maximum capacity of hospital h
$rate_h$:	Constant rate at which hospital h can increase capacity until max_h is reached

Given these values, the capacity of the hospital in question is modelled as increasing at the constant rate of $rate_h$ from time $\tau = 0$ to time $\tau = \frac{(max_h - cap_h)}{rate_h}$.

In order to forecast the effects of self-presentation we must estimate the number of casualties at each disaster site who will self-present, which hospitals they will choose, and how long they will wait before leaving the scene. We recognize that the severity of casualty injury plays an important role in determining whether or not self-presentation is an option. Accordingly, our model allows self-presentation only for casualties c such that $c^t = T3$. Under this assumption, the following parameters are required:

To determine a measure of the attractiveness of a given hospital h to a self-presenting casualty at site d, denoted $g_{d,h}$, we compute the measure

$$g_{d,h} = \exp(-sp^{attract}\delta(d,h)) \tag{4.19}$$

where $\delta(d,h)$ is the estimated travel time from site d to hospital h. This measure is computed for each h, after which all values are normalized to give the proportion of self presenting casualties at site d expected to travel to each hospital h, denoted $n_{d,h}$.

sp^{prob}: Probability of a priority 3 casualty self-presenting

$sp^{attract}$: Parameter describing how the attractiveness of a hospital varies with distance

sp^{wait}: Length of the interval over which self-presenters' waiting times will be uniformly distributed

These values are then used to create arrival distributions for each casualty-hospital pair, where casualties begin to arrive at time $\delta(d,h)$ and continue arriving at the constant rate $\frac{n_{d,h}}{sp^{wait}}$ until time $\delta(d,h) + sp^{wait}$.

For each hospital considered part of the model, we now have: a list of scheduled arrival times of casualties of each triage level; a list of anticipated arrival times of T3 casualties self-presenting; and an anticipated rise in capacity due to major incident plans. This information is combined to predict the total waiting time of casualties of each triage level at the hospital. In order to do so, we make the following assumptions:

(i) a casualty arriving at a hospital with free capacity is immediately admitted to a bed, thus consuming a capacity unit, regardless of their triage level;

(ii) once a casualty has been admitted, they will occupy a space in that hospital for the duration of the response operation;

(iii) when there is a queue of casualties at a hospital awaiting admittance, they will be allocated in an order which reflects their triage level irrespective of their time of arrival at the hospital.

While assumptions (i) and (iii) are not controversial, assumption (ii) may not be realistic for casualties with light injuries when the response operation continues for several hours. This assumption could be removed, given data regarding the length of stay of such lightly injured casualties. In the absence of such data, we note that our attention is restricted to MCIs where the response operation is anticipated to take 1-3 hours (see Section 2.2.2), which reduces the impact of this assumption.

An illustration of a hospital being over-subscribed is given in Fig. 4.13, where the cumulative casualty arrivals exceeds the available capacity over a period of time. The shaded areas denote the proportion of those waiting for treatment of triage level. The information we take from this are the areas Q_{T1}, Q_{T2} and Q_{T3}, representing the total untreated waiting time of casualties at hospital $h \in \mathcal{H}$ as grouped by triage level.

When calculating the quantities Q_{T1}, Q_{T2} and Q_{T3} associated with each hospital, the planned arrivals of all casualties of health state T3 are contrasted with the expected arrivals of all self-presenting casualties of health T3 arriving from that same disaster site. If the earliest expected leaving time of a self-presenting casualty is less than the latest processing time of a scheduled arrival, that latest scheduled arrival time is adjusted to correspond to the self-presentation arrival.

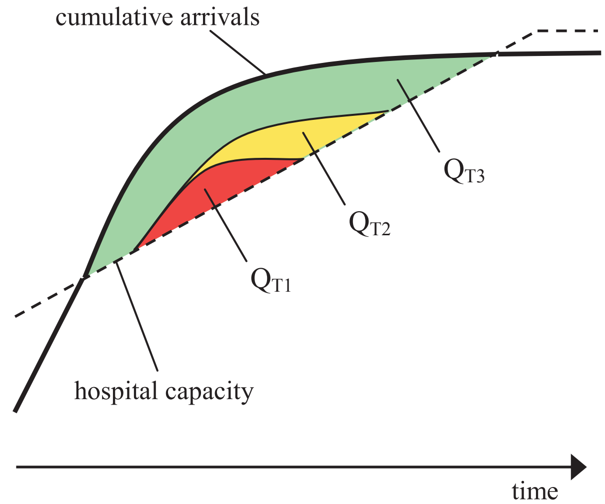

Figure 4.13: An illustration of casualty arrivals exceeding hospital capacity over a period of time. The shaded areas denote the proportion of 'excess' casualties of each triage level T1, T2 and T3.

Hospital capability

Given a specific injury type (e.g. severe burns or spinal injury) denoted ν, we wish to include in the model sufficient detail to ensure an allocation matching injury types to treatment facilities is preferred. In order to do so, we define a set of penalty terms $k_{\nu,T}$. Each of these quantities can be interpreted as 'the maximum delay in the treatment of a casualty with injury ν and priority p which could be tolerated in order to ensure they are treated at an appropriate specialist facility'. This interpretation helps to provide a clear understanding of the parameters involved to any user of the model. Given this set of terms, we calculate for each casualty c the value β_c, where $\beta_c = k_{\nu,T}$ if casualty c has injury of type ν and triage level T but is not taken to a hospital with the corresponding treatment facilities, and 0 otherwise.

Combining the above terms associated with both hospital capacity and capability gives a total measure of how well casualties have been allocated to hospitals, accounting for both dynamic capacity levels and heterogeneous treatment facilities:

$$f_3(s) = \left[\sum_{T \in \{T1,T2,T3,D\}} \sum_{h \in \mathcal{H}} w_T Q_T^h \right] + \left[\sum_{c \in \mathcal{C}} \beta_c \right]. \tag{4.20}$$

4.6.4 Responder idleness

We wish to include a measure of how much time is spent by responders in an idle state, neither completing a task nor travelling to their next location. In addition to an expected correlation between idleness and the other objectives (that is, a solution where all responders are constantly busy is likely to be of high quality in other respects), we also note that any perceived idleness can have a negative impact on the public's impression of the quality of the response operation. Total idleness can be calculated easily from any given schedule by summing all intervals between the end of one task and the time where the responder either leaves to travel to the site of their next task or if the next task is at the same location, begins work on this task.

4.6.5 Latest finishing time

As noted previously, the 'makespan' of a solution, i.e. the time at which the last task has been completed, is not an appropriate measure of solution quality when considered in isolation. However, when complemented by the objectives previously discussed, it is desirable to give some consideration to makespan since a low value corresponds to an early finish of the response operation. This is particularly desirable when in terms of objective (xii) as listed in Section 2.3.2, the 'time taken to return to normality'. While several other factors must be considered in defining what is necessary to return to normality after an MCI, an early completion of the response operation will clearly contribute towards this goal.

4.6.6 Multi-objective formulation

As previously mentioned, the five objectives considered in our model can be partitioned into three categories:

1. *fatalities* - $f_1(s)$,
2. *suffering* - $f_2(s), f_3(s)$,
3. *efficiency* - $f_4(s), f_5(s)$.

We view the optimization of these three categories in a lexicographic [49] sense, assuming that the minimization of fatalities is infinitely more important to the emergency response decision maker than the minimization of suffering, which in turn is infinitely more important than the minimization of efficiency. Within each category we employ a method of weighted metrics [86] to convert the multi-objective subproblem into a single objective one. In doing so we define three objective functions, namely

$$g_1(s) = (w_1|f_1(s) - z_1^*|^2)^{1/2} \tag{4.21}$$
$$g_2(s) = (w_2|f_2(s) - z_2^*|^2 + w_3|f_3(s) - z_3^*|^2)^{1/2} \tag{4.22}$$
$$g_3(s) = (w_4|f_4(s) - z_4^*|^2 + w_5|f_5(s) - z_5^*|^2)^{1/2} \tag{4.23}$$
$$\tag{4.24}$$

measuring repectively the fatalaties, suffering, and efficiency of solution s. When combining objective functions, the method of least squares was used[1]. In order to

[1]Other weighted metric methods were implemented for comparison, with no difference in performance observed.

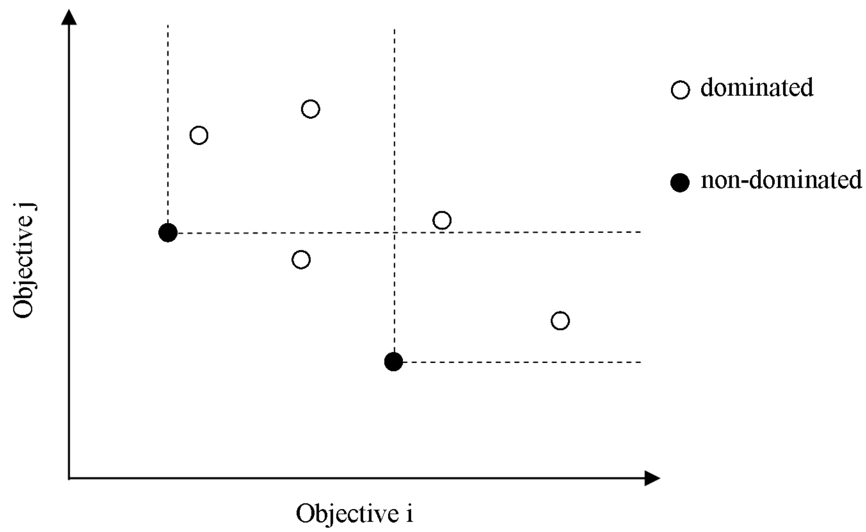

Figure 4.14: Illustration of Pareto optimality, highlighting non-dominated points in the solution space.

employ this method we must set the utopia point $\mathbf{z}^*$, an infeasible point in the objective space used to provide guidance to the search. For objectives f_1, f_3 and f_4 we simply set $z_i^* = 0$. For objective f_2, hospital arrival time, we obtain an infeasible lower bound by supposing each casualty arrives at hospital at the earliest possible time (i.e. all relevant tasks must be completed) and at the same triage state as at time $\tau = 0$. For objective f_5, the makespan, we use the latest of such idealized hospital arrival times. As described in [31] the use of a weighted metric method to aggregate separate objective measures will only be capable of finding points in all parts of the corresponding Pareto set when the Pareto curve is convex, which we cannot assume to always be the case in our model. However, due to the limited time available in which to search for high quality solutions in MCI response scenarios, we note that this shortcoming will rarely be felt.

In addition to defining the utopia point, we must also set the relevant weights w_i. In the case of suffering, we set $w_2 = 1$ and $w_3 = 0.5$, corresponding to a belief that a fixed time spent waiting at a hospital is twice as preferable as that same time spent waiting at a disaster scene. All other weights w_i have been set as 1.

The full model can now be defined as an unconstrained multi-objective optimisation problem

$$\min_{s \in \mathcal{S}} \; g_1(s), \; g_2(s), \; g_3(s), \tag{4.25}$$

where a solution is defined by

$$t^r_{i,j} \in \{1, 2, \ldots, n_r\} \tag{4.26}$$

$$t^p_{i,j} \in \mathbb{Z} \tag{4.27}$$

$$t^h_{i,j} \in \{1, 2, \ldots, n_h\}, \tag{4.28}$$

for $1 \leq i \leq n_c$ and $1 \leq j \leq n_{i,t}$.

Regarding model fitting, we note that the lexicographic approach employed helps to minimize the need for setting weights. What weights are required by our model have been estimated, in the absence of sufficient data from which they may be derived. Clearly there is a need to further consult with appropriate practitioners and seek relevant data in order to refine and validate the choice of parameters in order to maximise the applicability and acceptability of the model. However, the goal of this work is to take the first step in evaluating the potential of the proposed model, and so this is left to future research.

4.7 Summary

In this chapter a novel mathematical model for delivering decision support to the emergency services during mass casualty incidents has been introduced. The combinatorial, multi-objective model accounts for several key problem features including the uncertain health levels of casualties, the spatial nature of the problem and the importance of appropriate choice of hospital for any casualty. The model is of a temporal nature, using estimates of task durations and travel times to build a predicted schedule covering the course of the response operation, thus avoiding the myopic decision making which could result from the use of a sequential, heuristic decision making process.

Chapter 5

Solution

Heuristics and algorithms

5.1 Introduction

The model introduced in Chapter 4 is of a combinatorial nature, and as such is prone to exponential growth of decision space with increases in problem parameters such as the number of casualties in the problem. Accordingly, in this Chapter the solution methodologies described are of a heuristic nature, where the focus is on delivering high quality solutions in a timely manner as opposed to locating the best possible solution.

In order to develop appropriate solution algorithms, current heuristic and metaheuristic methodology will be briefly reviewed. Both constructive and iterative solution algorithms will be introduced, and their performance tuned through the experimental evaluation of parameter settings.

5.2 Combinatorial optimisation

5.2.1 Exact and heuristic methods

When addressing a problem with a large solution space, such as that of the casualty processing problem described in Chapter 4, it is necessary to implement intelligent search methods so that the high quality solutions may be found quickly. The proposed model is of a multi-objective nature, but for the purposes of this general introduction to combinatorial optimisation we assume a single objective representation, where a lower objective value is considered desirable. We refer to a problem of this sort as *minimization*. It is clear, then, that when we refer to seeking the 'best' solution we mean the solution with the lowest objective value. In the terminology of optimisation, such a solution is called a *global optimum*. Note that there may be several solutions with equal value, in which case several global optima exist.

When describing search methods, we may denote them as being either *exact* or *heuristic* methods. If an exact method is applied to a problem, it is guaranteed to locate a global optima. Finding a global optima is of value in two ways - not only have we located a solution with the best objective value, but our knowledge that there is no better solution means we can stop searching and apply our efforts to other problems. Exact methods, however, can take an impractical amount of time to arrive at their solution, if indeed this global optima can be found at all. Heuristic methods offer a compromise in this respect, performing their search in far less time at the expense of a guaranteed global optima.

It is worth noting, as Jungnickel did in [68], that “input data . . . always have a limited accuracy, so it may not even mean much to have a truly optimal solution”. This observation is particularly applicable to the problem domain in question, where we anticipate significant uncertainty in the dynamically collected data used to build the model. Furthermore, it is essential that the decision support framework delivered is capable of producing results in a timely manner as any delay in decision making will result in a delay in the response operation. For these reasons, it was decided to focus efforts on implementing and developing heuristic search methods, as opposed to exact algorithms. In doing so, significant challenges are encountered. In particular, we must consider the problems arising from the presence of *local optima*. Whereas a global optima denotes a solution with a better objective value than all others, a local optima denotes a solution which is better than the solutions nearby, where by nearby we mean in the local optima’s *neighbourhood*. The definition of a neighbourhood is an important aspect of modelling, and is discussed in more detail in section 5.4.1. We illustrate the properties of local and global optima in Figure 5.1.

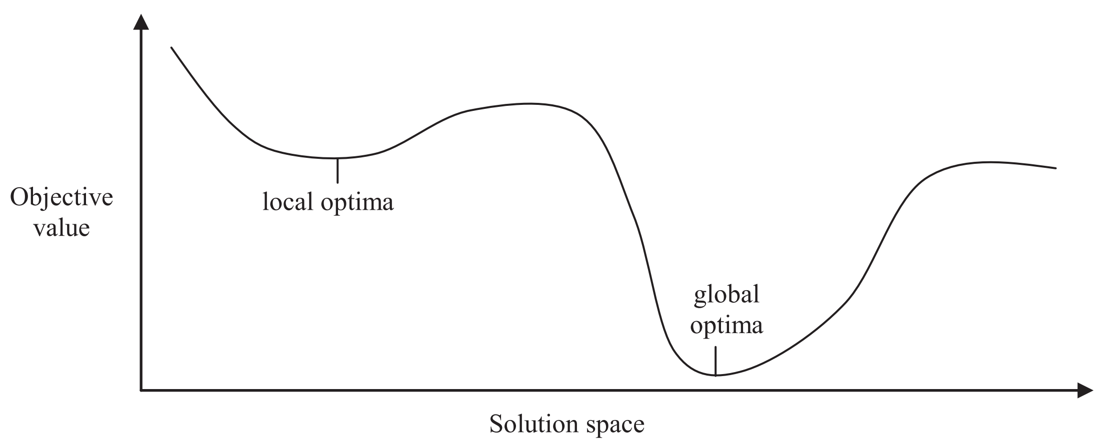

Figure 5.1: An example local optima and global optima within a simple solution space.

A potential problem is clear - we wish to design search methods which travel ‘downhill’ in our solution space, looking for solutions with lower objective values, but we do not want them to become ‘stuck’ in local optima. Metaheuristics provide potential solutions.

5.2.2 Metaheuristics

A heuristic is defined to be "A 'rule of thumb', based on domain knowledge from a particular application that gives guidance in the solution of a problem". As described in [16], heuristics can be broadly classified into two categories: *constructive* heuristics and *local search* heuristics. In the case of the casualty processing model from the preceding chapter, a constructive heuristic could specify how, given a set of unscheduled tasks and a set of responders, a specific task should be selected to put on the end of a specific responders schedule. Such a heuristic could be used repeatedly to 'construct' a full schedule. A local search heuristic, on the other hand, is one which takes a fully specified schedule and makes a small adjustment. For example, this could involve taking a specific task and moving it to another position in the schedule. An overview of heuristic methods can be found in [115]. Given one or some low-level heuristics which may be used with a given problem, *Metaheuristics* are heuristic methods which provide high-level strategies to use these low-level heuristics in a structured way. Due to their high-level nature, metaheuristics may be applied to a range of decision problems, each time requiring a set of problem-specific and generic design parameters to be set [102].

Consider, for example, a *random search* metaheuristic. As a local search method, this will examine all the neighbours of the current solution in turn. Once a neighbour with a lower objective value is found, it is accepted as the new current solution. This process continues until the search reaches a local optima, which by definition will have no neighbours with a lower objective value. When implementing a random search metaheuristic, the key design decision is the choice of neighbourhood structure.

A random search cannot be relied on to produce good results in practice since, although the process may be fast, it is incapable of 'escaping' local optima and so may miss much better solutions, as illustrated in Figure 5.1. Several metaheuristics have been developed in an effort to combat this issue, aiming to perform more comprehensive searches of the whole solution space while maintaining acceptable performance in terms of computation time, including simulated annealing, genetic algorithms and tabu search. Descriptions of these methods can be found in [102], [11], [16], [58].

In addition to detailed discussions of several popular metaheuristics, [11] also proposes a framework designed to create a unified view of metaheuristics. The authors argue the concepts of *intensification* (where a subset of the solution space is explored thoroughly) and *diversification* (where unexplored regions of the solution space are sought) are the important forces in metaheuristic performance. Noting that there may be a number of algorithm components to any given metaheuristic, and that each component will contribute in some way to intensification or diversification of the search, the question of how to strategically control the balance between these two aspects is discussed.

Cited in [11] as a very promising research issue is the *hybridization* of metaheuristics. Hybrid search methods arise when two or more heuristics or metaheuristics are combined with the aim of utilizing the benefits of all approaches while mitigating their faults. The recent rise of interest in applying hybrid metaheuristics to combinatorial problems, and their resulting success, is highlighted in [98] and [118]. Both papers present a classification framework to assist in the understanding and design of hybrid metaheuristics.

Another related area of research receiving much recent attention is that of *hyper* heuristics. They are described in [107] as "heuristics to choose heuristics", and so are heuristic methods similar to those discussed earlier, but which operate on a search space of other heuristics. Hyper heuristics potentially provide a methodology to address the problem of over parametrized metaheuristics. In the problem domain of large scale emergency response, it is anticipated that the mathematical models built to represent the real problems could significantly vary in characteristics as the environment changes. While we may be able to achieve good performance for a certain type of problem (for example, those involving significant fires), that same algorithm may not perform satisfactorily for other problem environments (for example, those involving chemical attacks) since the tasks and resources present in the model may differ. Methods which can intelligently adapt the optimisation algorithm in use to fit the problem observed could therefore be considerably useful in this research.

Recently, it has been argued [117] that a significant amount of metaheuristic research is of a low quality with regard to rigorous design and evaluation of novel solution methodologies. As such, caution should be exercised when considering the adoption of new, un-tested metaheuristics.

5.2.3 Fitness landscapes

In order to decide which algorithm should be used to search for optimal solutions, it is helpful to analyse the *fitness landscape*. This term is defined in [104] as consisting of three aspects, namely

1. a set X of solutions,
2. a notion of neighbourhood, nearness, distance or accessibility on X, and
3. an objective (or fitness) function $f : X \to \mathbf{R}$.

By analysing the fitness landscape of the model, information which can inform the design of appropriate algorithms may be gleaned. Furthermore, analysing the landscape may result in motivation to alter one of the three aspects listed. One idea presented in [103] revolves around the concept of a 'big valley' structure of a fitness landscape, where the high quality solutions to the problem tend to be clustered together. If such a structure exists, an algorithm which initially attempts to cover a large, diverse sample of the solution space but then focuses on intensification once the 'big valley' has been found could be very effective. In order to find evidence for such a structure, [103] suggests taking the best solution found and comparing the difference in objective value with the distance over the solution space to a set of other local optima and looking for a large correlation between the two measures.

Ultimately, a fitness landscape with only one optima would lead to very fast results when using almost any algorithm. While such a landscape is unlikely to be achievable when modelling a real problem, it illustrates another motivating aspect for the study of fitness landscapes. We can change the landscape by altering any of the three aspects listed, and by analysing the results we may be able to move closer to an ideal model in terms of algorithm performance.

5.3 Constructive solutions methods

Before describing the main solution algorithm which will be used to solve the model described in Chapter 4, which is of an iterative local search nature, two constructive algorithms are given. As constructive algorithms generate a single solution quickly, they are useful both as a solution method in their own right and as a way to generate initial solutions for an iterative solution method.

5.3.1 Simulating constructor

The proposed constructive heuristic Φ has been designed to give an approximation of decision making in casualty processing in practice. Whereas the model described in this thesis makes use of the temporal nature of the problem, forecasting over the whole of the response operation in order to better 'plan ahead', this is not achievable to any great extent under the current decision making structure. Rather, each decision is made in a 'greedy' fashion, selecting the option which gives the maximum benefit at that point in time.

Decision making occurs at two types of points, namely when a responder finishes a task and when a transportation task is issued. In the former case, we must decide which task the responder in question should next complete, whereas in the latter we decide to which hospital the casualty in question should be taken to.

Selecting a task

Given a set of n criteria which can be applied to any task t, $o_i(t), i = 1, \ldots, n$, we apply an evaluation process based on a lexicographic [49] approach and described in Algorithm 2. This approach allows for several criteria to be considered and requires only an ordering of these criteria, as opposed to a weighting. Lexicographic approaches to multi-criteria decision making are common [66], and can be considered appropriate in this situation due to their ease of interpretation.

In order to employ the general approach, we must first specify a set of criteria which can be applied. These are:

1. **P**riority - corresponding to the triage level of the casualty;

2. **T**ime - how soon the task can start;

3. **D**ependancy - the number of other tasks dependent on the completion of the task in question;

4. **L**ocation - the distance from the responders current location to the location of the task in question.

Algorithm 2 Constructive heuristic

```
set measure functions o_{1...n}
set initial task best
for all tasks t ∈ T do
  if r can do t & all tasks t depends on are issued then
    done ← false, i ← 1
    while done = false and i < n + 1 do
      if o_i(t) > o_i(best) then
        done ← true
      else if o_i(t) = o_i(best) then
        i++
      else
        best ← t
        done ← true
      end if
    end while
  end if
end for
issue task best to r
```

Criteria P and D take a naturally discrete form. In the case of criteria L, since all casualties are located at one of a finite number of shared incident site nodes of the transport network a discrete form emerges in its case also. We impose a discrete form on T by categorizing each time τ as one of $\{\tau = 0, 0 < \tau \leq 0.5, 0.5 < \tau \leq 1, 1 < \tau \leq 2, 2 < \tau \leq 5, 5 < \tau\}$, where the units are in minutes. This partitioning of time, with finer granularity as we approach 0, allows for the algorithm to find efficient schedules at this fine level of detail.

By assigning an order to each of the criteria, i.e. mapping each o_i to one of P, T, D or L, the algorithm is fully specified. Rather than pre-specifying the preference order of the proposed measures, in Section 5.5 we will empirically analyse the performance of each possible ordering in order to determine the optimal configuration.

Selecting a hospital

The selection of a hospital is carried out in a similar manner as the selection of a task. In this case, there are three measures used in making the choice: the distance of the hospital from the current location; the capacity level of the hospital under consideration; and the presence of treatment facilities appropriate to a casualty's injury. The decision is made by first restricting the choice of hospitals to those with the correct treatment facilities. Following this, if there are hospitals with free capacity then the closest of these is chosen. If not, the hospital least over capacity is chosen. This heuristic rests on an implicit assumption that the set of hospitals under consideration will be in the same geographical area, with no single hospital being particularly far from the incident site(s). While this will be true when incidents occur within large cities, in other incidents some hospitals could be a considerable distance from the incident site(s). There, it may be preferable to instead prioritise hospitals which are close and very over-subscribed, to one far away but only mildly over-subscribed.

5.3.2 Random constructor

In addition to the simulating constructor described in Section 5.3.1, a simpler constructive heuristic was implemented. The basic structure follows that of the simulating constructor, where a schedule is gradually constructed by identifying which responder is due to finish their current allocated task and then choosing from the available pool of unallocated tasks to find their next one. However, whereas in the simulating constructor a number of measures were computed to identify a particularly appropriate task, in the case of the random constructor tasks are selected at random and assessed only for the feasibility of their allocation (i.e. it is checked that the responder is capable of carrying out the tasks and that any other tasks which must be completed first have been issued by that point in time). Similarly, all hospital allocation decisions are made at random, with no regard to the the capacity or capability of the hospitals.

With no need to evaluate measures when allocating tasks or assigning casualties to hospitals, the random constructor will be less computationally demanding than the simulating constructor and as such will have a shorter runtime. However, for the scale of problems to be considered in this thesis the computational burden posed to the simulating constructor is not significant, and as such it is this method which will be employed throughout, with the exception being the random sampling of solution spaces as discussed in Section 7.3.1 where the random constructor is employed. We will, therefore, refer to the simulating constructor as simply “the constructor” in the remainder of this thesis.

5.4 Local search solution methods

As discussed in Section 4.4, a solution to our model can be defined by an association $s : \mathcal{T} \rightarrow R \times \mathbb{N} \times \mathcal{H} \bigcup \{0\}$, so that every task $t \in \mathcal{T}$ has an associated responder $r \in \mathcal{R}$, priority level $p \in \mathbb{N}$ and hospital $h \in H \bigcup \{0\}$, where $h = 0$ for all tasks other than transportation tasks. In order to implement any local search solution method, one or more neighbourhood structures must be designed in order to allow for the navigation around this set of solutions.

5.4.1 Local search neighborhoods

In specifying any neighbourhood structure, a number of characteristics should be considered. Firstly, does the neighbourhood allow for a fully connected solution space? That is, it is desirable that any two solutions within the solution space are connected by a finite number of neighbourhood moves. Secondly, does the neighbourhood reflect the similarity of solutions? Broadly speaking, neighbours of any given solutions should have relatively little difference in objective values to ensure the local search process can be guided by an implicit gradient in the objective space. Finally, it is important to consider the size of the neighbourhood and any computational issues which may arise from it.

Considering each of these features, the following four neighbourhood structures are proposed for the casualty processing problem:

1. Priority - adjusting the priority variable, t^p, of a single task;
2. Insert - removing and re-inserting a single task to another location in the schedule;
3. Hospital - adjusting the hospital allocation of a single casualty;
4. Swap - swapping the positions of two tasks.

Priority neighbourhood, P

Given a solution s, the operation $p_{t,b}(s)$ switches the priority value of task t, t^p (see Section 4.4.1), with the task preceding it (in the case $b = 0$) or succeeding it (in the case $b = 1$). The neighbourhood of s under P can therefore be defined by

$$P(s) = \{p_{t,b}(s) | t \in \mathcal{T}, b \in \{0, 1\}\}. \tag{5.1}$$

The neighbourhood P does not, on it's own, provide a fully connected solution space. In particular, it does not allow for tasks to be re-assigned to another responder or for casualties to be re-allocated to another hospital. In terms of ensuring a smooth objective space, the neighbourhood P performs well as it allows for only subtle changes within any one move. The neighbourhood is also relatively small, with a maximum size of

$$|P| = 2|\mathcal{T}| - 2. \tag{5.2}$$

Insert neighbourhood, IN

Given a solution s, the operation $in_{t,\hat{t}(s)}$ moves task t from its current position to before task $\hat{t}$, which may be located in a different responder's schedule. As a result of this operation, all tasks which were succeeding t in its original responder schedule have their priority parameter decreased, while all those with $t^p \geq \hat{t}^p$ in the new schedule will have their priority parameter increased. We therefore define the neighbourhood

$$IN(s) = \{in_{t,\hat{t}}(s) | t \in \mathcal{T}, \hat{t} \in \mathcal{T}\}. \tag{5.3}$$

The neighbourhood IN allows for greater connectivity than neighbourhood P, as it facilitates the re-allocation of tasks from one responder to another. However, it does not lead to a fully connected solution space as it does not allow for the re-assignment of casualties to hospitals. The neighbourhood is not as smooth as P, as it allows for more significant moves. The maximum size of the neighbourhood is relatively large at

$$|IN| = \binom{|\mathcal{T}|}{2}. \tag{5.4}$$

For example, for a solution space involving 40 tasks the size of the neighbourhood IN will be

$$|IN| = \binom{|\mathcal{T}|}{2} = 435. \tag{5.5}$$

Hospital neighbourhood, H

We define the operation $h_{t,h}(s)$ as one which changes the allocation of hospital of task t to h. Then

$$H(s) = \{h_{t,h}(s) | h \in \mathcal{H}\}, \tag{5.6}$$

where task t is a transportation task. Thus, neighbourhood H complements those previously discussed as it allows for the re-assignment of casualties to hospitals whilst providing no means with which to alter the scheduling of tasks. The neighbourhood will have a constant size of

$$|H| = |\mathcal{C}| \times (|\mathcal{H}| - 1). \tag{5.7}$$

Swap neighbourhood, SW

Finally, we define the operation $sw_{t_1,t_2}(s)$ as one which interchanges the responder allocation and priority values of tasks t_1 and t_2. Then

$$SW(s) = \{sw_{t_1,t_2}(s) | t_1, t_2 \in \mathcal{T}\}. \tag{5.8}$$

The neighbourhood SW could be thought of as consisting of paired operations from within the neighbourhood IN. The maximum size is also

$$|SW| = \binom{|\mathcal{T}|}{2}. \tag{5.9}$$

Implementation 5.4.1. Neighbourhood structures are implemented through both a 'neighbourhood' class and an associated 'neighbour' class. Each neighbourhood class contains a list of neighbour objects. This list is constructed each time the search process moves to a new point in the solution space. The construction of the list of neighbours avoids including any which would lead to infeasible solutions. For example, when identifying neighbours of the 'insert' nature for neighbourhood IN, any neighbour which would lead to a task being assigned to a responder which is incapable of carrying it out is discarded. This process will identify some infeasible moves, but not all. Where moves leading to infeasible solutions are allowed, their infeasible nature will be identified during the scheduling process of Algorithm 1.

Neighbour objects contain the necessary information to define the operation. For example, a Swap neighbour contains the unique ID's of the two tasks which are to be swapped. Each neighbour contains a function which returns a list of *directions*, low-level operations which act on solutions to make atomic changes. This implementation means that all possible local search moves, regardless of which neighbourhood structure they employ, are translated into a series of atomic operations which come from a shared library. As a result, it is possible to define a common, flexible metric on the space of solutions. The potential for this to allow further developments in terms of solution method performance are discussed in more detail in Chapter 8.

5.4.2 Variable Neighbourhood Descent

Each of the four neighbourhood structures listed in Section 5.4.1 can be generalized to capture a notion of size in an intuitive manner. Namely, a neighbourhood of size $i = 1$ consists of 100 random samples from the neighbourhood structure as defined. We limit ourselves to a finite sample due to the potentially large size of these combinatorial neighbourhoods. To generate a neighbourhood of size $i = 2$, we first generate a neighbourhood of size $i = 1$ before generating a further 100 neighbours of size 2 by selecting random pairs of size 1 neighbours and composing them. Composition, in this sense, is defined by performing each operation in turn, and is facilitated by the implementation of neighbourhood operations as described in Implementation 5.4.1. Similarly, for a neighbourhood of size $i = 3$ a further stage is carried out, where 100 random triples of size 1 neighbours are composed. This routine can be carried out for any desired neighbourhood size i.

The VND algorithm is given in Algorithm 3, and is an adapted version of the Variable Neighbourhood Search (VNS) algorithm presented in [3] where the VNS was shown to have competitive performance for the Flexible Job Shop Scheduling problem. The VNS has also been successfully applied to Vehicle Routing Problems (see, e.g., [51]). Specifically, the 'shaking' procedure of the VNS, which involves periodically performing a random walk for a number of iterations, is not carried out.

The algorithm continues to use a neighbourhood structure until that same structure fails to return an improving solution (i.e. we reach a local minimum) or $k = k_{max}$ successful iterations are performed, at which point the next neighbourhood structure in the list $\mathcal{N}$ is selected. The iteration limit k_{max} will be selected following an empirical analysis of performance in Section 5.5.2. In the case where a local minimum was found, the size of the neighbourhood structure used to find, say $\mathcal{N}[i]$, is increased. In this manner, when the algorithm completes a cycle of considering each neighbourhood structure and returns to consider $\mathcal{N}[i]$ again, the neighbourhood's larger size will increase the chance of finding an improving solution. The termination criteria used is that of a real-time threshold denoted τ^*. This is a practical measure for the problem domain in question, since emergency response decision makers will only wait a short time for decision support.

Algorithm 3 VND

```
generate initial solution s
define set of neighbourhoods 𝒩 ← {P, IN, H, SW}
let i ← 0
while time < τ* do
    let N ← 𝒩[i]
    set k ← 0
    while k < k_max do
        compute s* ← arg min N(s)
        if f(s*) < f(s) then
            s ← s*
            N_size ← 1
            ++k
        else
            N_size ++
            i ← (i + 1) mod 4
        end if
    end while
end while
```

5.5 Evaluation

Both the constructor and local search solution method require some parameters to be specified in order to complete their implementation. In this section results of experiments into how these parameters should be set are reported. In addition to this parameter setting, the effect of problem characteristics on these results is also examined. Throughout, performance will be measured in terms of objectives g_1, fatalities, and g_2, suffering, as defined in Section 4.6.6. Discussing results in terms of these two objectives reflects their priority over the third objective of efficiency, and will allow for a clearer graphical interpretation of results.

5.5.1 Constructive heuristic configuration

As described in Section 5.3, the principal constructive heuristic is fully specified upon being given an ordering of preference for four criteria used in its evaluation process. In order to identify which such ordering leads to best performance, a number of computational experiments were carried out. A total of twenty seven problem types were used, varying in three dimensions: the number of responders; the average number of tasks required for the processing of each casualty, which may be controlled through the number of tasks associated with each casualty and therefore with the amount of dependency that exists between tasks; and the number of incident sites over which the casualties are distributed. A full description and discussion of these problem scenarios can be found in Section 7.2. In terms of the configuration of the constructor, the twenty four possible orderings of measures are given in Table 5.1.

Table 5.1: Constructive heuristic configurations.

1.	**P - T - D - L**	7.	**T - P - D - L**	13.	**D - P - T - L**	19.	**L - P - T - D**
2.	**P - T - L - D**	8.	**T - P - L - D**	14.	**D - P - L - T**	20.	**L - P - D - T**
3.	**P - D - T - D**	9.	**T - D - P - L**	15.	**D - T - P - L**	21.	**L - T - P - D**
4.	**P - D - L - T**	10.	**T - D - L - P**	16.	**D - T - L - P**	22.	**L - T - D - P**
5.	**P - L - T - D**	11.	**T - L - P - D**	17.	**D - L - P - T**	23.	**L - D - P - T**
6.	**P - L - D - T**	12.	**T - L - D - P**	18.	**D - L - T - P**	24.	**L - D - T - P**

The complete experimental design is a factorial one, where each of the 3 levels of responders $\{R_{low}, R_{med}, R_{high}\}$, dependency $\{D_{low}, D_{med}, D_{high}\}$ and sites $\{S_{one}, S_{two}, S_{three}\}$ are crossed with the 24 constructor configurations and 50 runs of each experiment are performed, leading to a total data set of 32400 computational experiments.

Primary analysis

Taking the problem components of the design to be representative of the population (i.e. we assume these 27 problems are equally likely), we compare the performance in fatalities and suffering of each configuration on average. That is, we compute means, medians and standard deviations for each configuration and both objectives.

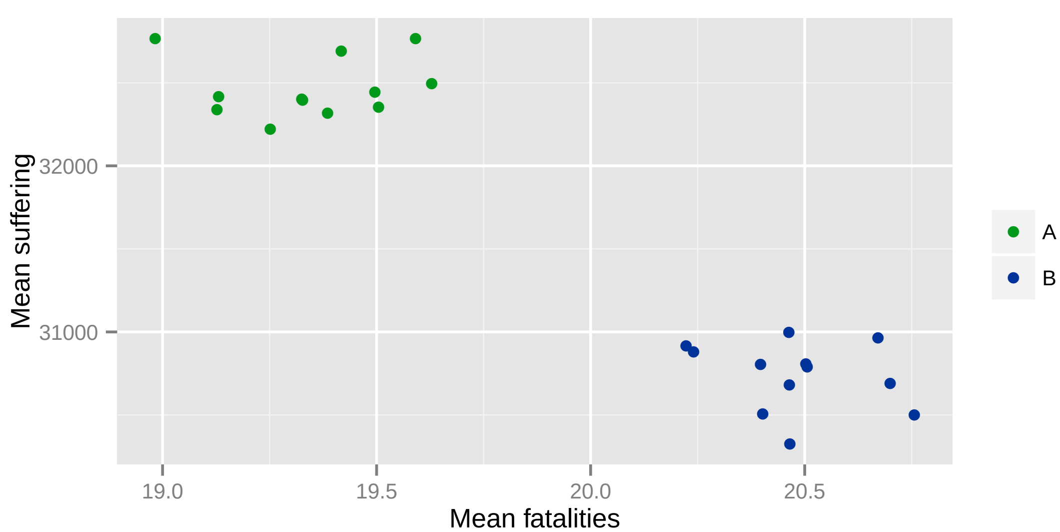

Figure 5.2: Scatter plot for fatalities and suffering means of the 24 constructor configurations.

A scatter plot of fatalities against suffering, Figure 5.2, shows two clear clusters. According to the indices assigned to the configurations in the experimental design given in Table 5.1, these two clusters are:

1. Group A, low fatalities and high suffering - $\{9, 10, 12, \ldots, 18, 22, \ldots, 24\}$.

2. Group B, high fatalities and low suffering - $\{1, \ldots, 8, 11, 19, \ldots, 21\}$.

In order to assess the statistical significance of the constructor configurations, a multivariate analysis of variance (MANOVA) test is carried out to compare the 24 configuration groups in terms of their mean vectors. The result is a rejection of the null hypothesis ($p \approx 0$) that there is no difference in the mean vectors associated with each configuration, as shown in Table 5.2. Given the lexicographic ordering of objectives introduced in Chapter 5, we restrict our focus to the cluster A, plotted in Figure 5.3.

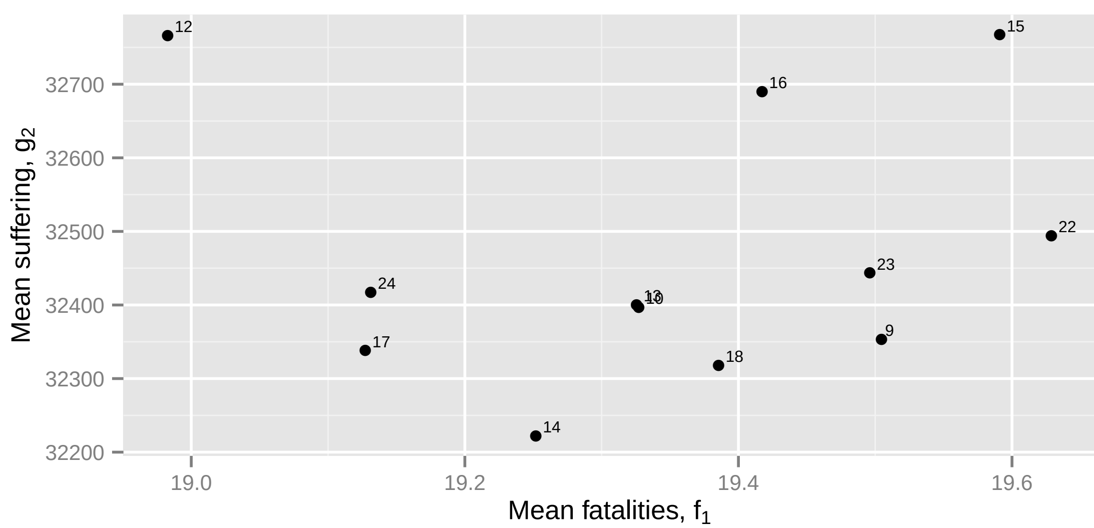

Figure 5.3: Scatter plot for fatalities and suffering means of the constructor configurations in group A.

Table 5.2: MANOVA testing for difference in constructive heuristic configurations.

		Response f_1			
	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Config.	23	791	34.41	0.0878	1
Residuals	6188	2425229	391.92		
		Response g_2			
	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Config.	23	1.0179e+10	442558907	2.2019	0.0007864
Residuals	6188	1.2437e+12	200990788		

Another MANOVA reports no statistically significant difference within the subset of group A (see Table 5.3), so we cannot conclude that any specific configuration is the best performer in the cluster. We choose the configuration with lowest fatalities, configuration 9, bearing this caveat in mind, to use in our subsequent experiments where the constructive heuristic in employed.

Table 5.3: MANOVA testing for difference in constructive heuristic configurations within group A.

		Response f_1			
	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Config.	11	106	9.67	0.0255	1
Residuals	3176	1206616	379.92		
		Response g_2			
	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Config.	11	4.3650e+07	3968161	0.0199	1
Residuals	3176	6.3309e+11	199335079		

Secondary analysis

Having evaluated the performance of each constructive heuristic configuration on average, it is now of interest to examine how performance varies according to problem characteristics. To do so, we examine the relative performance (i.e. the rankings) of configurations when restricted to specific levels of each problem characteristic (resource levels R, dependency levels D or number of incident sites S). In examining a specific characteristic, note that the results presented are averaged across the possible levels of the remaining two problems characteristics.

In Figure 5.4 the relative performance of constructor configurations, in terms of fatalities objective, is shown for varying levels of R, D and S. The figure shows that the *relative* performance of each constructor configuration, as represented by the shape of the connected lines, does not vary by any significant amount as the levels of R, D or S are adjusted. This in turn implies that a single choice of configuration could be made and that this would consistently provide best performance over a range of problems which varied in these characteristics. In contrast, when measuring performance in terms of suffering the interaction plots (Figure 5.5) highlight some interesting behaviour. In particular, the relative performance of configurations differ when the number of incident sites changes between one, two and three. This raises the possibility that average performance of this constructive heuristic method could be improved through adapting its configuration based on observations of the nature of the problem (in this case, the number of incident sites).

5.5.2 Local search VND configuration

As with the constructive heuristic, the local search-based Variable Neighbourhood Search solution method requires some parameters to be set in order for it to be fully specified and implemented. Specifically, the neighbourhoods to be used throughout must be chosen, and the iteration limit must be set. The iteration limit governs the

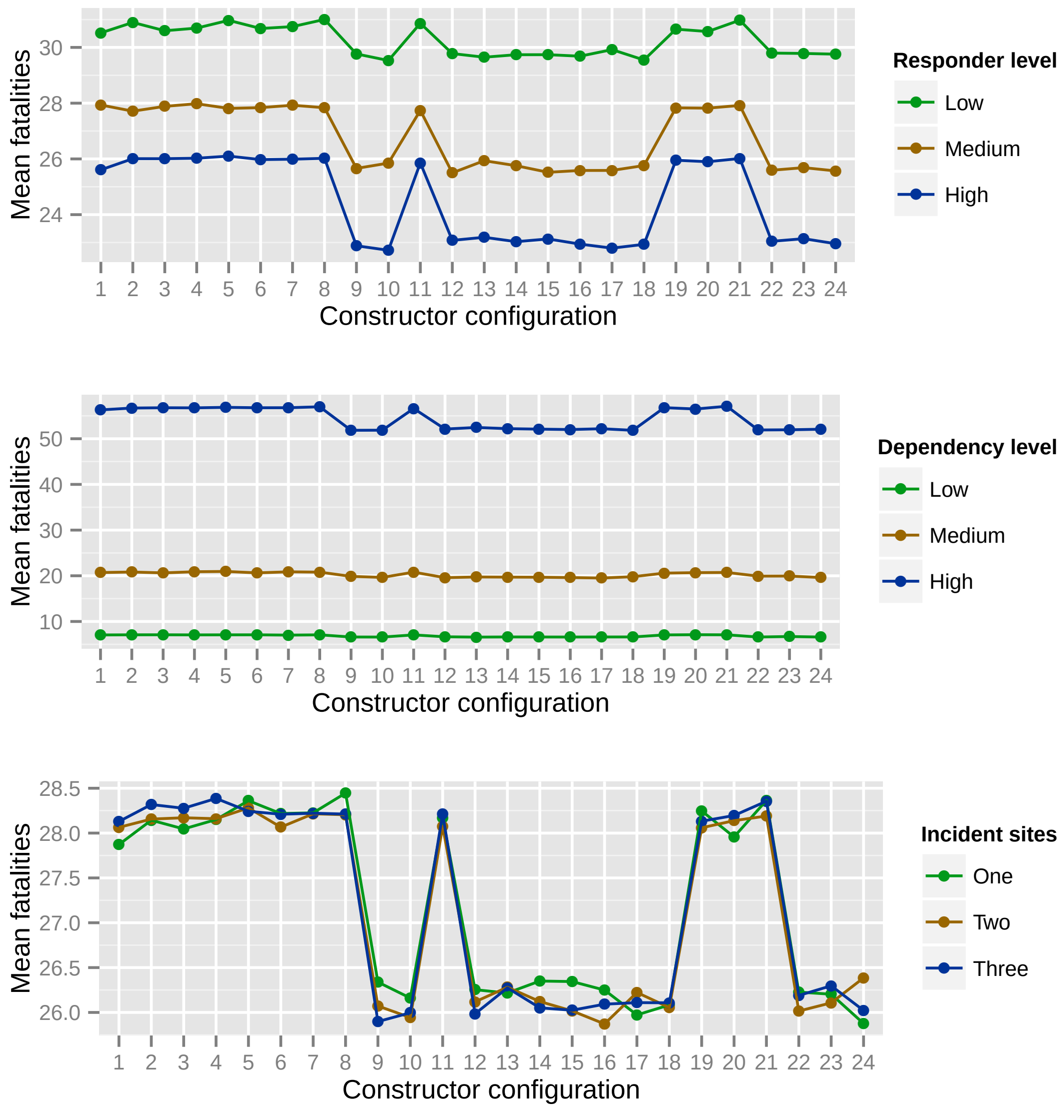

Figure 5.4: Interaction plot for constructive heuristic configuration against R/D/S, fatalities objective.

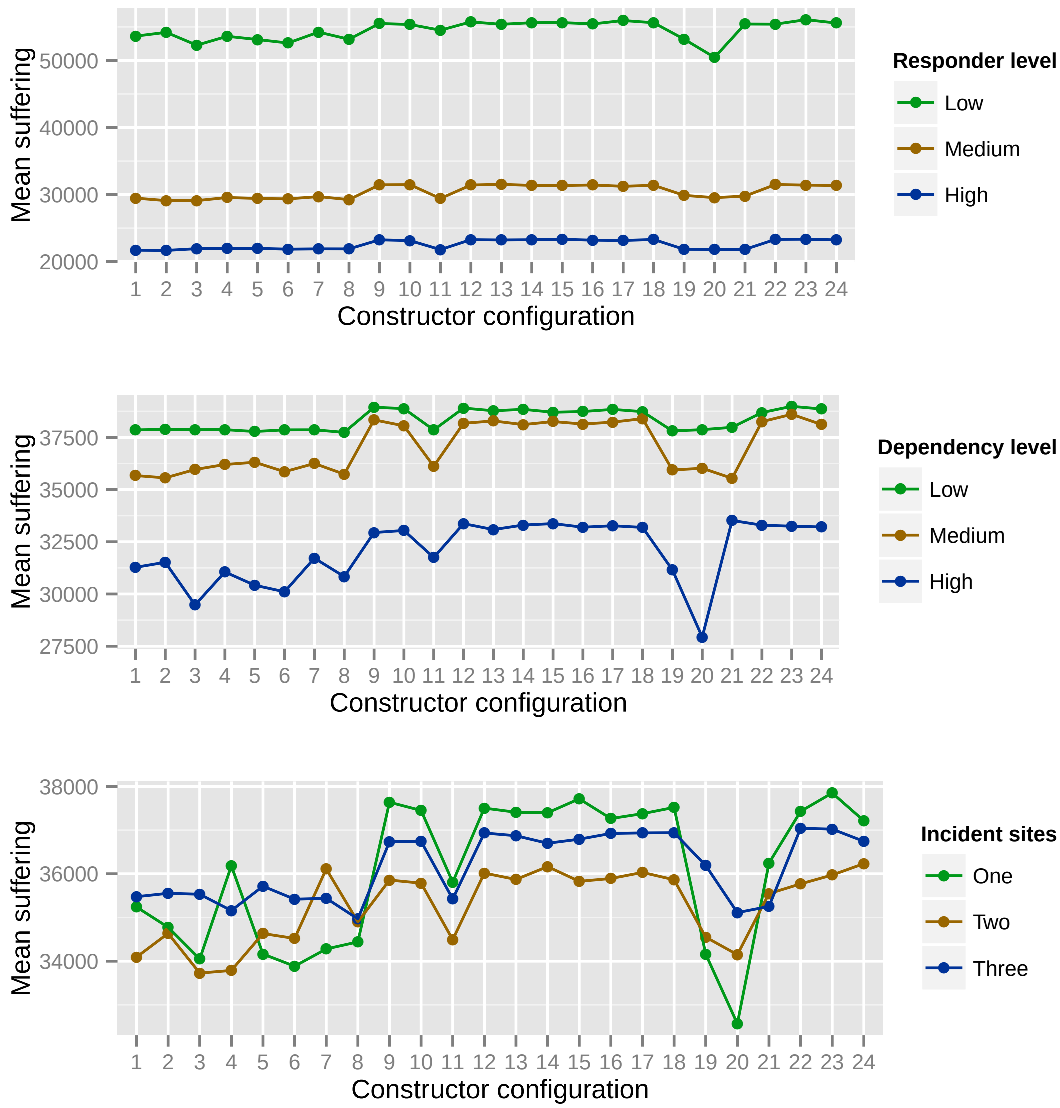

Figure 5.5: Interaction plot for constructive heuristic configuration against R/D/S, suffering objective.

maximum number of iterations which may be performed using one neighbourhood structure, without locating an improving neighbour, before the next neighbourhood structure is adopted. In terms of neighbourhoods included, neighbourhood P is used in every configuration. Crossing all possible selections of the remaining three neighbourhoods with three values of the iteration limit k_{max}, (10, 20 and 40) leads to a total of 24 distinct local search VND configurations.

Table 5.4: Local search VND configurations, showing which neighbourhoods are included (black) and iteration limit.

1.	P, IN, H, SW, 10	9.	P, IN, H, SW, 20	17.	P, IN, H, SW, 40
2.	P, IN, H, SW, 10	10.	P, IN, H, SW, 20	18.	P, IN, H, SW, 40
3.	P, IN, H, SW, 10	11.	P, IN, H, SW, 20	19.	P, IN, H, SW, 40
4.	P, IN, H, SW, 10	12.	P, IN, H, SW, 20	20.	P, IN, H, SW, 40
5.	P, IN, H, SW, 10	13.	P, IN, H, SW, 20	21.	P, IN, H, SW, 40
6.	P, IN, H, SW, 10	14.	P, IN, H, SW, 20	22.	P, IN, H, SW, 40
7.	P, IN, H, SW, 10	15.	P, IN, H, SW, 20	23.	P, IN, H, SW, 40
8.	P, IN, H, SW, 10	16.	P, IN, H, SW, 20	24.	P, IN, H, SW, 40

As in the case of the constructive heuristic, the complete experimental design is a factorial one, where each of the 3 levels of responders $\{R_{low}, R_{med}, R_{high}\}$, dependency $\{D_{low}, D_{med}, D_{high}\}$ and sites $\{S_{one}, S_{two}, S_{three}\}$ are crossed with the 24 constructor configurations.

Primary analysis

Figure 5.6 plots average performance, in terms of percentage improvement on the initial solution, of each configuration. In order to assess the statistical significance of any effect of local search VND parameters on performance, a linear regression model was fitted to the data. The model included categorical variables denoting the inclusion of each neighbourhood structure, together with a continuous variable representing the iteration limit. The analysis found that the effect of iteration limit was not statistically significant. As such, we proceed to focus on the effect of neighbourhood choice.

As configurations 6 and 14 (which involved identical neighbourhood choice but varying iteration limits) exhibit best performance in terms of fatalities, we can conclude that in these problem scenarios the addition of the hospital neighbourhood structure is not necessary in order to attain best performance, as measured by the lexicographic ordering of objectives.

Secondary analysis

As in the case of the constructive heuristic, it is of interest to examine the effect of problem characteristics on solution method performance.

As can be seen in Figures 5.7 and 5.8, there is little interaction between problem characteristics and the configuration of the local search solution method. One point of

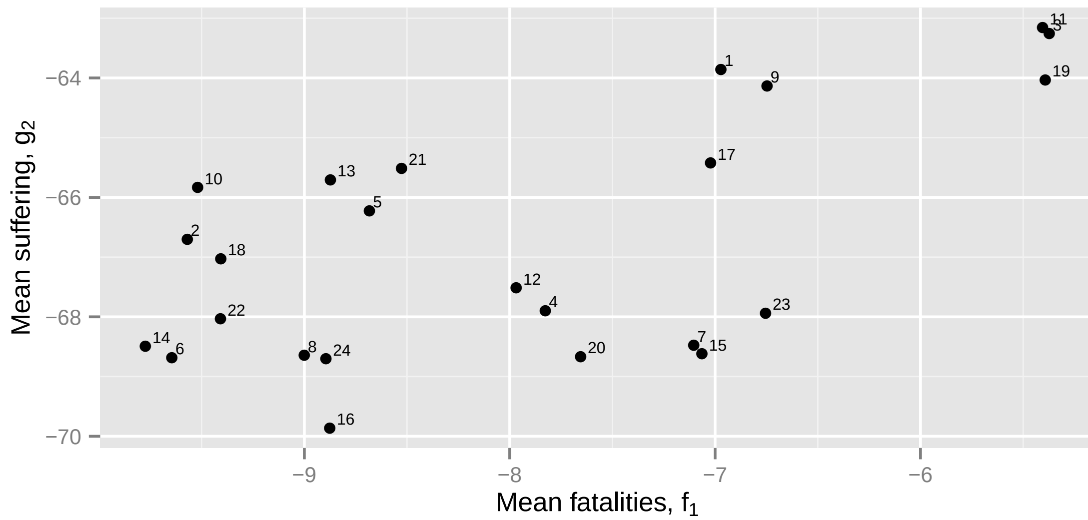

Figure 5.6: Average performance across all problem instances of each search configuration.

note is that when dependency levels are low, performance is relatively homogeneous. As dependency levels increase, differences in performance between the configurations begin to emerge.

5.6 Summary

In this Chapter two key solution methodologies have been introduced. Firstly, a constructive heuristic method allows for the fast generation of response schedules in a manner designed to reflect the decision making processes in place during an MCI response operation (as discussed in Section 2.3). The algorithm can be parametrized, and these parameters have been tuned for optimal performance through a series of computational experiments. Secondly, an iterative solution method based on the Variable Neighbourhood Descent metahueristic has been introduced. Making use of several neighbourhood structures with associated parameters, this method has also been tuned for optimal performance. Together, these solution methodologies provide a sufficient means for all solution generation required in the remainder of this thesis.

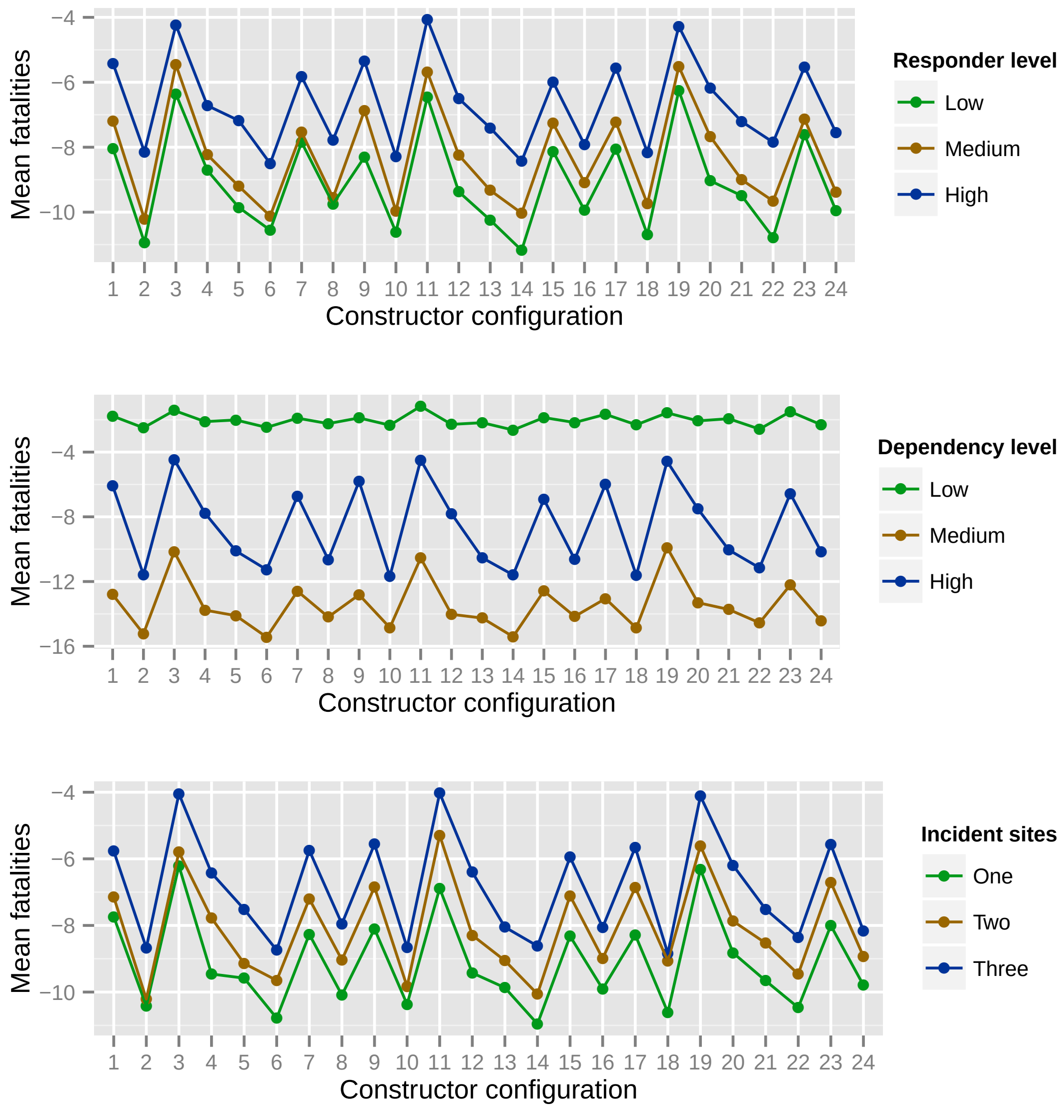

Figure 5.7: Interaction plot for local search configuration against R/D/S, fatalities objective.

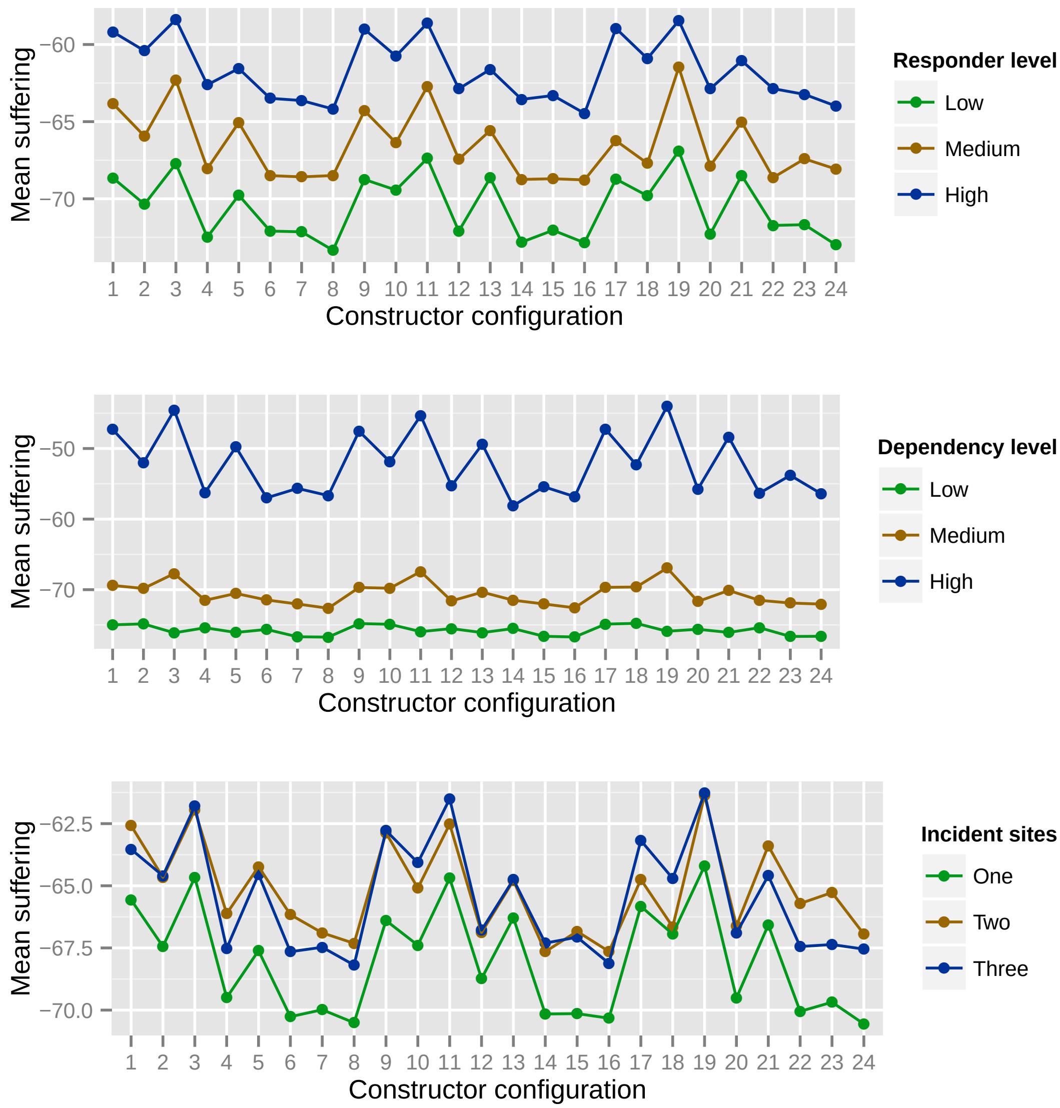

Figure 5.8: Interaction plot for local search configuration against R/D/S, suffering objective.

Chapter 6

Simulation

Uncertain and dynamic incidents

6.1 Introduction

The model described in Chapter 4 is designed to be used in what we shall refer to to as an *offline* manner. By this we mean that the communication between the problem and the model is of a simple, non-iterative nature. In this case, all necessary information (such as the number of casualties and the duration of tasks) is assumed to have been collected at the outset of the incident, when it is passed on to the model. The model is initialized using this information, and a solution is generated using the methods described in Chapter 5. This solution is then communicated back to the problem environment, so that the response operation can follow the specified schedule. We will refer to this model from this point onwards as M_{off}. In this chapter we seek to develop and extend the model M_{off} to better cope with the dynamic and uncertain nature of MCI response. The resulting model will be of an *online* nature, and will be denoted by M_{on}.

The extension from M_{off} to M_{on} is motivated by the fact that many of the parameters of the model are not known, deterministic and stable. Rather, they may be estimated, subject to error, and liable to change as the response operation progresses. As an example, consider the duration of a particular rescue task. This duration is required in order to specify the model M_{off}, but in reality it will not be known with certainty until the task has been completed. Thus, it must first be estimated, with this estimate subject to error. Moreover, as the response progresses and the task is completed, we wish to update the model to reflect the fact that this parameter is now known.

Many parameters, in addition to task durations, will be of a similar nature. In order to evaluate the ability of an improved online model M_{on} to address the challenges posed by such parameters, it is necessary to develop a simulation model. Such a model will be described in Section 6.2. Following this, in Section 6.3 details will be given of the general interface between the optimisation model and the simulation of the problem environment. Finally, modifications to the optimisation model allowing for greater ability to address the challenges brought about through the simulation will be described in Section 6.4.

6.2 Simulation

In the following discussion we shall partition all simulated parameters into two sets. By *solution space* parameters, we refer to those which affect the nature of the solution space, as described in Section 4.4. That is, a change in a solution space parameter will alter the set of possible solutions. In contrast, *objective space* parameters are those which, when altered, result in a change in the objective value(s) of one or more solutions.

6.2.1 Solution space parameters

Consider first the solution space parameters. As described in Section 4.4.1, the decision problem modelled consists of assigning an ordered list of tasks to a number of responder units, and allocating casualties to appropriate hospitals. Since the set of tasks $\mathcal{T}$ is determined by the set of casualties $\mathcal{C}$, we can reduce the parameters associated with solution space change to be:

- $\mathcal{C}$, the set of all casualties,
- $\mathcal{R}$, the set of all responder units,
- $\mathcal{H}$, the set of hospitals.

Considering the set of all hospitals, it is clear that the hospitals available for use in the response operation are unlikely to alter over time. Accordingly, no change in the hospital set is simulated.

Regarding the set of available responder units, we note that this can both increase and decrease as the response operation progresses. As discussed in [5], it is common for responders from areas neighbouring the affected district to self-dispatch, thus arriving with little or no notice and increasing the set of responders. Although a reduction can occur due to injury sustained when working in a hazardous environment, given the short timescale of problem scenarios considered in this thesis we do not account for this possibility. Thus, we wish to simulate increases in the set of responder units.

In terms of the set of casualties, an increase can occur in both a gradual manner, as more casualties are discovered during search and rescue operations, or in a sudden manner, if a secondary incident were to occur nearby. In terms of the latter, this is simulated by setting a time parameter to be associated with each incident. This parameter determines the length of time which will pass in the simulation before the incident is created. In the case of the offline model discussed in Chapter 4, these time parameters would all be set to zero, corresponding to a situation where all incidents occur at the same time. By modifying the time parameter of any incident, the simulation will delay the creation of the set of casualties associated with that incident until that time arrives. Thus, setting a time parameter of 00:05:30 will lead to the corresponding set of casualties being created and added to the model at exactly five and a half minutes following the start of the response operation.

In terms of gradual increases in casualty numbers, this is achieved in a manner similar to that described above but at the level of individual casualties. That is, each casualty may be assigned a time parameter denoting the time which will pass before their addition to the simulation. Unlike incident sites, which will be few in number, many casualties will exist in any MCI problem scenario. As such, it would be infeasible to manually specify the time parameter of each casualty in the same manner as incident sites. Rather, the time parameter for each casualty c_i, denoted τ_i^*, is generated randomly in the simulation. An exponential distribution is used,

$$\tau_i^* \sim exp(\lambda^{cas}), \tag{6.1}$$

where λ^{cas} is the rate of casualty 'discovery'. The use of an exponential distribution in generating these times ensures an intuitive pattern, where times are most likely immediately following the time of the incident and become less likely as time passes.

The set of casualties may also decrease as the operation progresses. In particular, the possibility of casualties with health state T3 leaving the incident site and self-presenting at a hospital of their choice should be allowed for in the simulation. The process of self-presentation has already been described in Section 4.6.3, in the context of *predicting* how it will happen. There, it was described how the number of casualties at incident site d who will self-present at hospital h was calculated. This quantity was denoted $n_{d,h}$. Whereas when predicting self-presentation we assumed that these casualties will arrive at the hospital at a constant rate $\frac{n_{d,h}}{sp^{wait}}$ until time $\delta(d,h) + sp^{wait}$, when simulating we instead generate a random variate $X_i^{sp} \sim exp(\lambda^{sp})$ for each casualty c_i, defining the maximum time they will wait without receiving attention before they self-present. Having generated these 'threshold' times for all casualties upon their initialization, the simulation will progress, in real time, monitoring these threshold times. When the current time elapsed in the simulation τ equals X_i^{sp} for some i, the status of casualty c_i is examined. If they have not received any attention (that is, if no tasks associated with them have commenced) then the casualty is marked as leaving to self-present.

The nature of the solution space, in terms of the number of solutions which exist for a given problem formulation, has been discussed in Section 4.4.1. It is of interest to revisit these considerations in light of the discussion presented in this section thus far. In particular, we wish to illustrate how the size of the solution space might vary as a simulation progresses and changes are made to the sets of responders and casualties. Such an illustration is provided in Figure 6.1, which charts the solutions space size over the first fifteen minutes of an example problem.

As noted when considering solution methods in Chapter 5, the size of the solution space is a key factor in determining the performance of solution algorithms. It is worth noting, then, that a dynamic solution space of the type illustrated in Figure 6.1 could present interesting challenges to the design and implementation of such algorithms, as a method known to be effective for 'medium' sized problems may be ineffective when faced with 'large' problems, or indeed vice versa. These challenges will not be

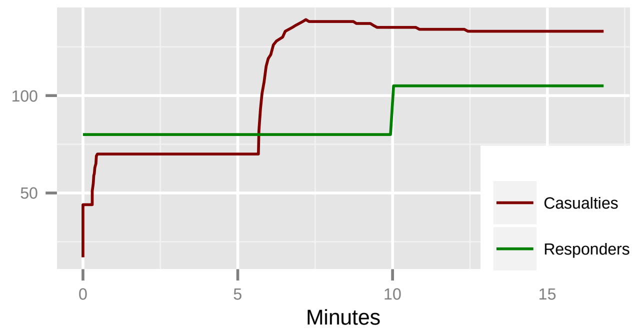

Figure 6.1: Changes over time in numbers of casualties and responder units in an example problem.

taken up in this thesis, however, as the present focus is on making a first evaluation of the proposed optimisation model in these dynamic environments.

Having discussed the simulation of solution space parameters, we now turn to objective space parameters. Specifically, the following groups of parameters are of interest in this respect:

- Casualty health;
- Task duration.
- Travel times;

We will consider each type of parameter in detail in the remainder of this section, describing in each case how to simulate their dynamic and/or uncertain nature.

6.2.2 Casualty health

As discussed in Section 4.6, the health of a casualty is described through the discrete triage classification system. The simulation of this parameter is accomplished by both introducing an element of error and by allowing it to evolve in a stochastic manner over time.

We allow for *error* in triage assessment to better reflect the realities observed in actual response environments [52], which we parametrized by $e^{tri} \in (0, 1)$, in the following manner. Denoting by $A[Ti]$ the event that a triage assessment has led to a casualty being classified in state Ti, the probability of these events conditional on the true health state of the casualty is given in Table 6.1.

For example, if the true underlying health state of a casualty is T1, the simulated outcome of a triage operation will be T1 (i.e. correct) with probability $1 - e^{tri}/2$ and T2 (i.e. under triaged) with probability $e^{tri}/2$. This system implies that a casualty could only be misclassified into a state adjacent to their true state, and that a casualty who is alive will never be mistakenly classified as dead. As a numerical example, if the error parameter $e^{tri} = 0.4$ a casualty whose true state is T1 will

Table 6.1: The probability of triage assessment outcomes with error level e.

True health state	$P(A[T1])$	$P(A[T2])$	$P(A[T3])$
$T1$	$1 - e^{tri}/2$	$e^{tri}/2$	0
$T2$	$e^{tri}/3$	$1 - 2e^{tri}/3$	$e^{tri}/3$
$T3$	0	$e^{tri}/2$	$1 - e^{tri}/2$

have a probability of $1 - 0.4/2 = 0.8$ of being correctly classified as T1 in a triage operation, with a probability of 0.2 of being classified as T2. An advantage of using these probabilities is that it allows for the general 'accuracy' of triage operations to be controlled through a single parameter, e^{tri}. A limitation, however, is its symmetric nature. It has been shown that, in practice, a casualty is more likely to be *over*triaged than *under*triaged, and so it could be argued that the probabilities used should better reflect this fact. However, our goal is to provide an adequate level of uncertainty and dynamic behaviour in order that its effect on the utility of the proposed optimisation model may be studied. By simulating only *error* and not simulating *bias*, we can properly assess the implications of the former without fear of it's effects becoming entangled with the latter.

In addition to simulating errors in the health parameters of casualties, we also consider how the changing of health over time may be modelled. This dynamic behaviour of casualty health is simulated using the same Markov chain methodology which was introduced in Section 4.6 as a means of *predicting* the future health of casualties. We note that our model assumes the parameters of this chain are known. In practice, this may not be possible due to the inherent low frequency of MCIs and the lack of data collection which occurs during them. The specific choice of transition probabilities and the rationale for their their choice were given in Section 4.6.1.

6.2.3 Task durations

In Section 6.2.2 it was described how the health of casualties could be represented in a manner which reflects its uncertain and dynamic nature. In this subsection we seek to achieve the same with task durations. That is, in order to be able to evaluate the ability of the proposed optimisation model to be of use in realistic scenarios where the durations of tasks are not all known, deterministic and unchanging, we must describe how tasks can be represented in a more realistic manner.

We propose to encapsulate the uncertain nature of task durations through a hierarchical probability model. Such a model will allow for the tasks durations to be simulated in a manner which allows for the durations of tasks associated with the same incident site to be correlated, as opposed to being independent. We discuss the case of rescue tasks, noting that the model for treatment tasks (pre-rescue and pre-transportation) is identical.

In the simulation model the true duration of a rescue task relating to casualty i

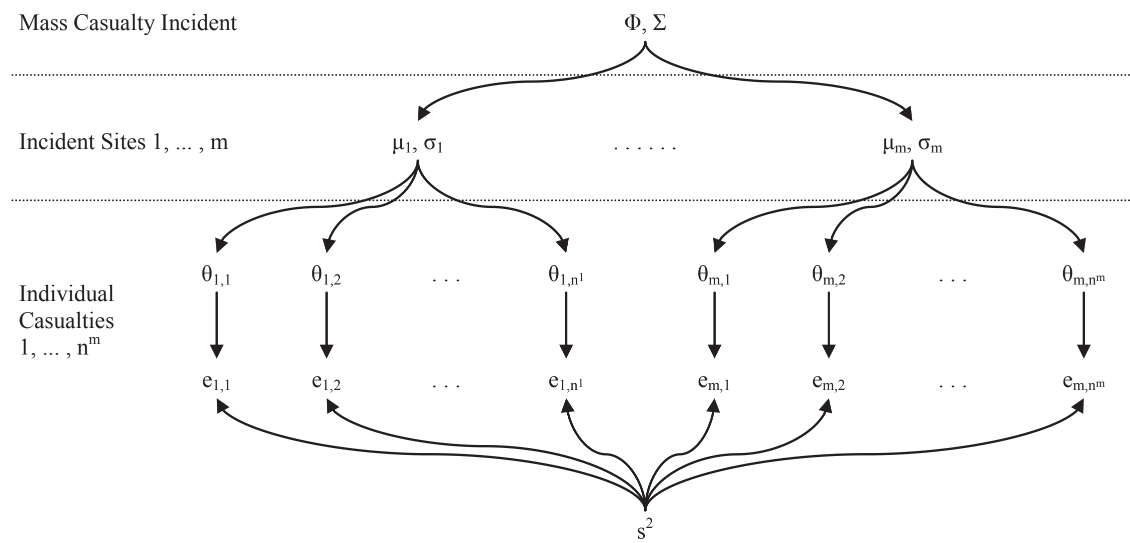

Figure 6.2: A hierarchical model of task durations.

at site j, $\theta_{j,i}$, is normally distributed around μ_j with variance σ_j^2,

$$\theta_{j,i} \sim N(\mu_j, \sigma_j^2). \tag{6.2}$$

The parameters μ_j and σ_j^2 are specific to the incident site j. These site-specific parameters are themselves governed by a bi-variate normal distribution with mean Φ and covariance Σ,

$$(\mu_j, \sigma_j^2)^T \sim N(\Phi, \Sigma). \tag{6.3}$$

Thus, given values of Φ and Σ as problem input the site-specific parameters are first sampled, after which individual task durations are sampled using the resulting values μ_j and σ_j^2. The true durations of all tasks in the model are simulated in this manner.

Uncertainty in task durations is introduced by generating (unbiased) estimates of each true duration $\theta_{j,i}$, denoted by $e_{j,i}$, by sampling from normal distributions with mean $\theta_{j,i}$ and variance s^2,

$$e_{j,i} \sim N(\theta_{j,i}, s^2). \tag{6.4}$$

These estimates correspond to the judgements of the responder units as they survey the MCI, estimating how long each required task will take. The variance s^2, which determines the accuracy of these task duration estimates, is specified as problem input. This will facilitate experiments examining the effect of varying accuracy of estimates upon the utility of the optimisation model. For example, the true duration of task $t_{j,i}$ may be simulated as $\theta_{j,i} = 5$ minutes, with an initial estimate of the duration given as $e_{j,i} = 7$ minutes. The hierarchical system of generating task durations is illustrated in Figure 6.2.

6.2.4 Travel times

Having discussed the simulation of task durations, the case of travel times is now considered. As discussed in Chapter 4, the spatial nature of the casualty processing problem ensures that routing and travel time prediction is of paramount importance. Over the course of a schedule many responder units will be required to move back and forth between several locations. This is illustrated in Figure 6.3, where the paths of two responder units (yellow and purple lines) are shown as they move through their assigned schedule. In this example, the responders make 4 and 2 journeys in order to complete 3 and 2 tasks, respectively. As such, it is clear that the simulation of travel times will have a similar impact as the simulation of task durations, with inaccurate estimates of travel times significantly impacting the utility of the optimisation model.

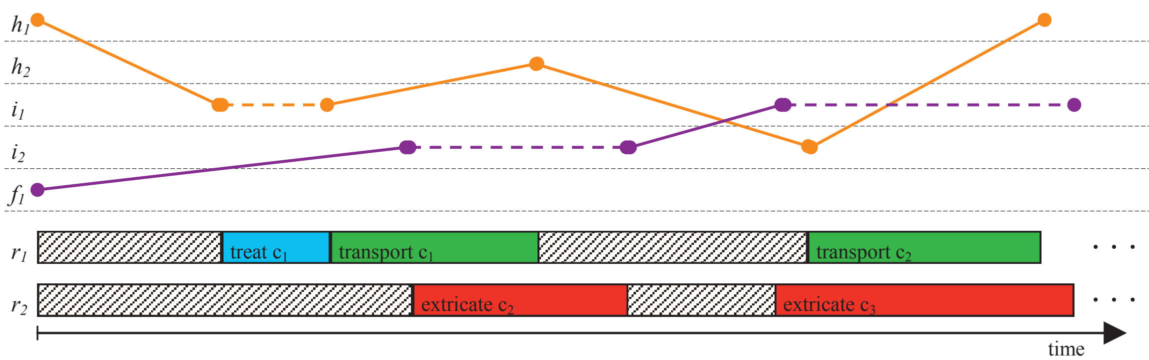

Figure 6.3: An extract from a simple schedule involving two responder (r_1, an ambulance, and r_2, a fire appliance) processing two casualties (c_1 and c_2), involving travel between a fire station (f_1), two hospitals (h_1 and h_2) and two incident sites (i_1 and i_2).

In Section 4.5.1 it was shown how the travel time of a given journey was estimated. In particular, a shortest path algorithm was applied to the road network which covers the problem environment. This procedure generates a route, following which a distance travelled can be calculated using the known distance of each edge on that route. An estimate for the travel time using this distance is provided by the model of [72], as recently validated by [15]. The function, denoted $KWH(d)$, gives an estimate of the median travel time along the route with distance d, where d was found using Dijkstra's algorithm The median travel time is then estimated as

$$\hat{m} = KWH(d) = \begin{cases} 2.42\sqrt{d}, d \leq 4.13 \text{ km} \\ 2.46 + 0.596d, d > 4.13 \text{ km} \end{cases} \tag{6.5}$$

where $4.13 = v_c^2/2a$ denotes the distance required to travel in order to reach 'cruise speed' v_c and a is the average acceleration of the vehicle as it increases speed to v_c. Again, the values of these parameters are taken from the analysis of ambulance travel times in Calgary, Canada, presented in [15].

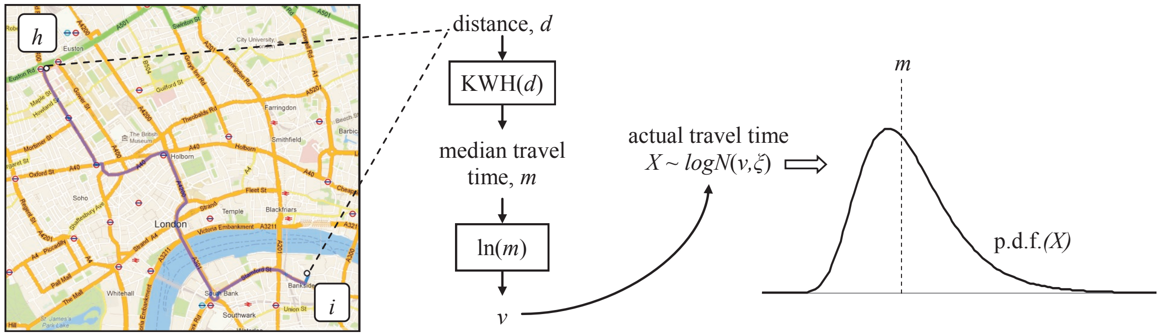

Figure 6.4: The simulation of travel times associated with a certain route, using a log-normal distribution parametrized indirectly through the length of the route.

To begin our simulation of travel times, we follow the above methodology in order to generate an estimated median travel time for the route in question. Variation around this median time is then generated by sampling from the corresponding log-normal distribution. That is, denoting the simulated travel time by X, we have

$$X \sim logN(\nu, \xi), \tag{6.6}$$

where $\nu = \ln(m)$ and ξ denotes the variance parameter, which controls the possible deviation of the simulated travel time around the estimated median travel time. A log-normal distribution of travel times is employed following [134], where it was shown that this form fitted well to empirical travel time data. This whole process, simulating a travel time for a specified shortest path between two locations h and i, is illustrated in Figure 6.4.

In the remainder of this subsection we go on to consider further details regarding the simulation of travel times. First, the possibility of the road network experiencing some disruption as a result of the MCI is considered. Here, the aim will be to simulate the effect of such a disruption in terms of the travel time associated with any given route. Secondly, the routing of responder units will be considered in more detail. In particular, we introduce alternative approaches to simulating the selection of route for a given journey, beyond the use of a shortest path algorithm.

Disruption

In order to simulate the effects of disruption to the transport network, the original network G, is transformed into one which has been disrupted, denoted G^*. This transformation preserves the structure of the graph, but alters the distance parameter associated to each edge. Specifically, a random variable $Y \sim exp(\lambda^{dist})$ is sampled for each link, the distance of which is then multiplied by the factor $(1 + Y)$. As such, we can interpret λ^{dist} as a parameter representing the level of disruption to the transport network. For example, setting $\lambda^{dist} = 0.5$ will lead to the distance parameter of each link on the road network being increased on average by a factor of

$(1 + E[Y]) = (1 + 2) = 3$, noting that the expectation of Y is $1/\lambda^{dist} = 2$.

The purpose of this disruption process is to generate uncertainty regarding travel times and optimal routing within the transport network, as opposed to realistically simulating the effects of a MCI on the network. Whilst such a realistic simulation model would be desirable, we are only required to show that the approach described does indeed generate uncertainty in both travel time estimation and optimal route choice.

In order to achieve this, we consider an example of a specific journey from a hospital to an incident site. Under normal conditions the median travel time for the journey can be calculated as 2.92 minutes using equation 6.5, where the shortest path was calculated using Dijkstra's algorithm on the standard network parametrization. This shortest path was stored, denoted by p. Following this, we simulated 500 instances of disruption with $\lambda^{dist} = 0.5$. In each simulated case, the following three steps were performed:

1. The median travel time of path p was calculated under the disrupted network G^* using equation 6.5,
$$\hat{m} = KWH(D_{G^*}(p)). \tag{6.7}$$

2. The actual shortest path under the disrupted network G^* was found, denoted p^*.

3. The median travel time of path p^* was calculated under the disrupted network G^* using equation 6.5,
$$\hat{m}^* = KWH(D_{G^*}(p^*)). \tag{6.8}$$

This analysis allows for two things to be examined. Firstly, we contrast the estimate of median travel time of the path p made under the undisrupted network G (which was shown to be 2.92 minutes) with the estimate found using the disrupted network G^*, $\hat{m}$. The distribution of the simulated values of $\hat{m}$ is given in the histogram at the top of Figure 6.5. A range of values over the interval (6.5, 9.5) minutes is observed, centred around 8 minutes. This illustrates the effect of applying disruption with the parameter $\lambda^{dist} = 0.5$: a route which would have normally had a median travel time of 2.92 minutes now has one of 8 minutes (on average). We can therefore conclude that the method of simulating road disruption does indeed achieve its first goal, namely introducing further uncertainty into travel times.

The second goal of the disruption simulation was to introduce uncertainty into the optimal route choice. That is, we wish for the shortest path for a given journey under the disrupted network to be different to the shortest path under the standard network. This would not be achieved through simpler disruption simulations, such as increasing the distance parameter of each edge by the same factor. However, the stochastic approach described here does achieve this goal. To see this, observe the central scatter plot of Figure 6.5. This plots the median travel times of route p^* against that of route p for each simulated instance. It is clear that these values are significantly different, with the value of $\hat{m}^*$ being 1.19 minutes lower than $\hat{m}$ on average. This allows us to conclude that the paths are indeed different. Specifically, the results show that if the disruption of the network is completely known, a better path for a specific journey will exist than that which was found using the standard undisrupted network.

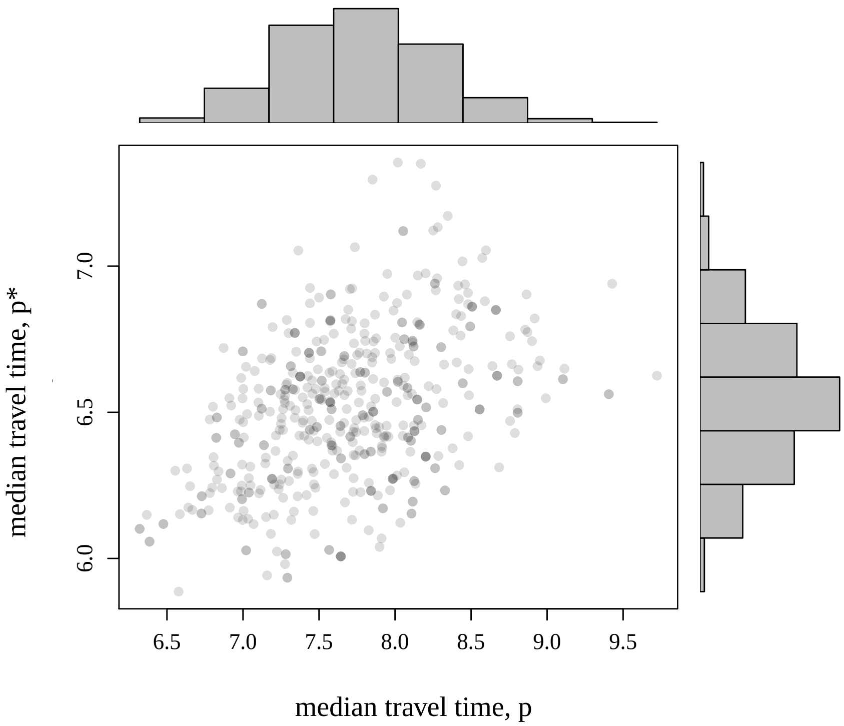

Figure 6.5: The joint and marginal empirical distributions of travel times on a disrupted network using both pre-calculated (p) and actual (p^*) shortest path routes.

Considering the MCI scenario, we assume that detailed information of the nature of any disruption to the transport network is not available to the optimisation model at the start of the response operation. As such, any instructions given to responders regarding the routes they should take could only be made using the standard network representation G. But, we have shown that these routes will not necessarily be optimal in the disrupted network. The question of how decisions regarding the routing of responders should be made emerges. In attempting to answer this question, we introduce four routing *policies*. Two policies (Static Routing and Centralized Adaptive Routing) are of a centralized nature, in that the routing choice of each journey is made by the central optimisation model. The remaining two policies (Autonomous Individual/Collective Adaptive Routing) are of an autonomous nature, where routing decisions are made by individual responders without recourse to the central optimization model. The remainder of this subsection will focus on describing how these strategies can be simulated, with issues arising to the optimisation model considered after this in Section 6.4.

Static routing (SR)

The first routing policy is termed *static routing* and will be denoted throughout by SR. Here, for any given journey a single route is specified *centrally* at the outset of the response operation and is used for the duration by all responders. By centrally, we mean that the optimisation model is used and that the result is communicated out to responder units. This policy employs a network reduction approach, reducing the full transport network to a simplified network connecting each point of interest to be used throughout the response operation. This policy will allow for accurate travel time prediction but will likely involve relatively poor routing decisions. This fact was demonstrated in Figure 6.5 of Section 6.2.4, where our analysis showed an average difference of 1.19 minutes between median travel times of paths p and p^*.

In terms of the simulation of SR, note that the specific routing choices do not change as the response operation progresses. That is, for a specific journey from A to B, whenever that journey is made the exact same route will be followed. Moreover, that route is the shortest path as calculated using the standard network G, which in our notation above is denoted route p. Given that the route is always known, the simulation of the specific travel time for a single journey is carried out using the procedure outlined at the start of this subsection and summarized in Figure 6.4.

Central adaptive routing (CAR)

The second routing policy is termed *central adaptive routing.* As with SR, this policy is of a centralized nature in that the choice of route for any given journey is determined by the optimisation model. Unlike SR, however, the choice of route will vary as the response operation progresses. The rationale for this is that as more journeys are completed, it will be possible to learn about the extent of any disruption to the network and to use this knowledge to suggest alternative routes to responders.

As in the case of SR, the simulation of a travel time can be achieved using the procedure of Figure 6.4 as the specific route to be taken is known. The generation of these routes is accomplished by the following procedure. At the start of each journey a route is calculated centrally using the current perception of the transport network, denoted G_i. This perception is modified each time a journey is completed and the travel time information is communicated to the optimisation model, giving a new network parametrization G_{i+1}. The process used to update the current perception of the transport network, G_i, upon receiving information regarding a travel time for a particular route p_i employs a simple heuristic which adjusts the weights of all links in the route p_i by a factor determined by comparing the realized travel time m with the predicted travel time $\hat{m}$. Specifically, the distance parameter of each link is multiplied by the factor $m/\hat{m}$. This results in the updated network parametrization G_{i+1}, which is used in subsequent routing decisions in conjunction with Dijkstra's shortest path algorithm.

Example 6.2.1. Consider the simple transport network representation illustrated in Figure 6.6. The network on the left of the figure represents the current perception of the optimisation model at iteration i, where the labels of each edge denote the perceived distance parameter. Using this network, a shortest path from A to B is found (shown in red). The estimated travel time, as calculated using equation 6.5, is found to be $\hat{m} = 4$ minutes.

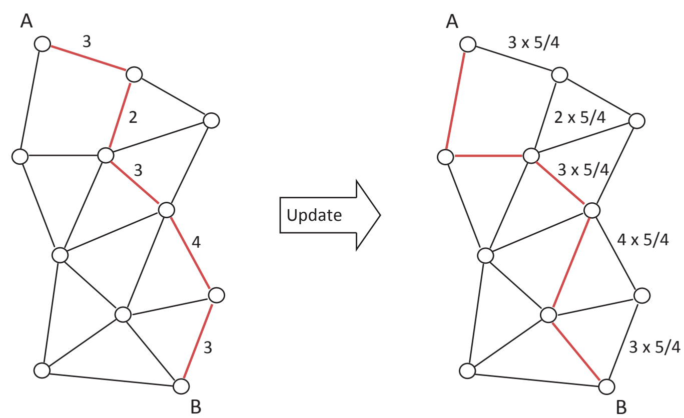

Figure 6.6: An example of central adaptive routing.

Upon completion of the journey, the true travel time is noted to be $m = 5$ minutes. In order to incorporate this new information into the network representation, each edge which formed part of the path travelled has its distance parameter multiplied by the factor $m/\hat{m} = 5/4$. For example, the edges with a distance parameter of 3 in network G_i are updated to 15/4 in representation G_{i+1}. Following this update procedure, the route selected for the next journey from A to B (shown in red) is modified in response to the changes in the network parameters.

This policy may improve upon the routing decisions of SR as it allows for routes alternative to the initial route p to be explored. However, the introduction of more variation into route choice may impact the ability of the optimisation model to predict travel times.

Autonomous individual/collective adaptive routing ($AIAR$ and $ACAR$)

In *Autonomous Individual Adaptive Routing* ($AIAR$) each responder makes their own routing decisions upon making a journey. They do so in an isolated manner with no communication with either the central optimisation model or other responders, and learn from their experiences during the response operation such that their routing choices improve, on average, as more journeys are completed. By introducing uncertainty in route choice, over and above that arising from disruption to transport network parameters and standard travel time variance, this policy will lead to larger errors in travel time prediction than in SR and, possibly, CAR. *Autonomous Collective Adaptive Routing* ($ACAR$) is identical to $AIAR$ but with information being freely shared among responders, allowing the learning process to be done in a collective manner, resulting in faster convergence towards the optimal routing decisions.

In order to simulate the natural improvement in routing choices which would be made by a responder, or set of responders under the $ACAR$ policy, we use a Markov process to generate a sequence of true median travel time values which can then be used, as illustrated in Figure 6.4, to define log-normal distributions from which travel times can be sampled.

We assume that the first route chosen by a responder will be the shortest route under normal conditions, which has been denoted as p in our discussion thus far. We can use the disrupted road network to find the true median travel time associated with this initial route, which we denote m_1. The travel time of the first trip is then simulated as $X_1 \sim logN(\ln(m_1), \varsigma)$. In order to generate the next median travel time, m_2, we sample from a normal distribution with mean αm_1 and variance β^2. This process then leads to an improvement of a factor of α each time a journey is taken, reflecting the assumption that responders will be capable of improving their routing decisions over time. By allowing for random variation around this improvement, controlled by the variance parameter β, we allow for situations where routing choices are not necessary monotonically improving with each journey. That is, it may be possible that exploring a new route leads to a worse travel time than the last journey. The process will continue until a median travel time less than or equal to the best possible route, under the disrupted network, has been reached, and is illustrated in Figure 6.7.

The setting of α and β^2 will determine the average improvement per journey and the variance in improvement per journey respectively. Without a comprehensive source of data relating to travel times in a MCI environment it is not possible to derive empirical estimates for these parameters. Accordingly, in the experiments described in Chapter 7 they will be adjusted to explore sensitivity.

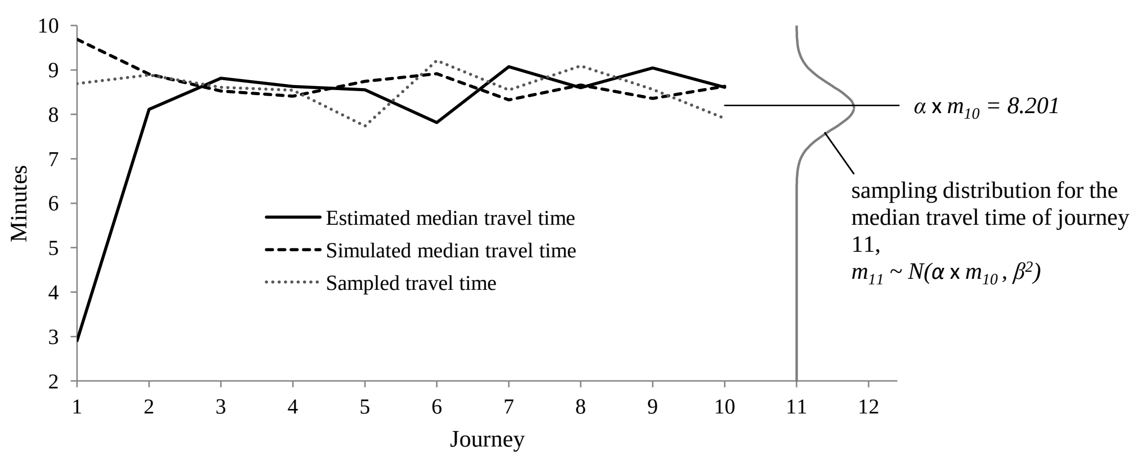

Figure 6.7: An instance of route choice progression under the $ACAR$ policy.

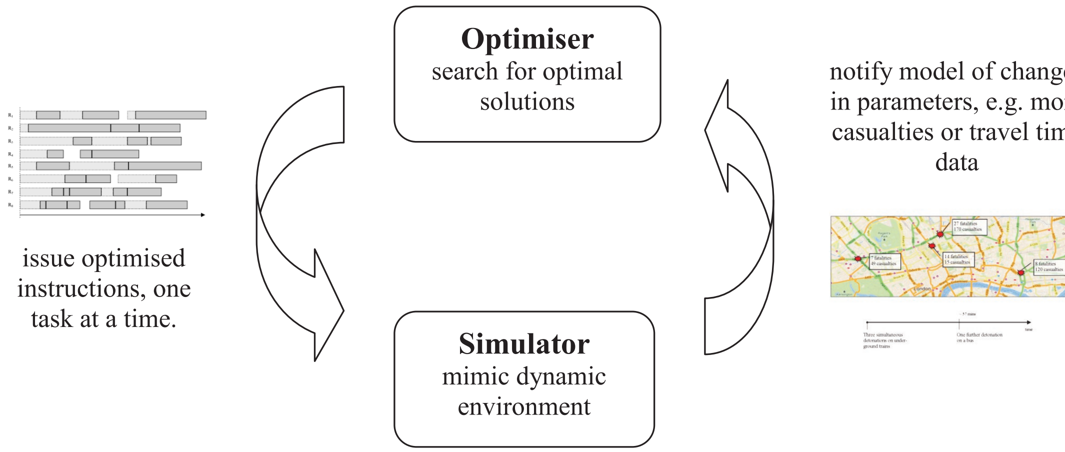

Figure 6.8: Two-way communication between optimisation and simulation models

6.3 Interface

As illustrated in Figure 6.8, the interface between optimisation and simulation functions in both directions.

Communication from the optimisation model to the simulation of the problem environment occurs whenever instructions are issued to responder units. This may be carried out in the manner discussed in Chapters 4 and 5, where an entire schedule is handed over at a single point in time and is assumed to provide adequate instruction for the duration of the response operation. This strategy will be explored further in Section 7.3, along with alternatives based on issuing instructions one task at a time throughout the response operation.

Communication from the simulation to the optimisation model occurs when something pertaining to the problem changes, and this change could influence the decisions being made in the optimisation process. Such developments could range from the fundamental (e.g. another attack during a terrorist incident, resulting in another incident site and associated set of casualties) to the subtle (e.g. the time taken to complete a task being more or less than originally estimated). The nature of the developments simulated in this work were discussed in Section 6.2, while in Section 6.3.2 we focus on how the optimisation model is notified of them.

6.3.1 Issuing instructions in real-time

Recall that the offline model, as described in Chapter 4, involves the passing of information from the optimizer to the problem environment at a single point in time. This communication occurs once a sufficient termination criteria of the search method, described in Chapter 5, has been reached, and will consist of a fully specified schedule covering the (expected) duration of the response operation. This schedule includes full details regarding which responder units should complete which tasks, at what times these tasks should begin and end, and to which hospital each casualty should be sent to.

In adapting the offline model for use in an online manner, we must allow for the model to pass instruction to the problem environment in a gradual manner as opposed to at one single point in time. Prior to this development, however, we first make the more fundamental acknowledgement of the real-time nature of the casualty processing problem. Specifically, we note that there is a natural conflict between the desire to enact the response operation as soon as possible and the potential benefits which could be gained through spending more time using the proposed search method to locate schedules of high(er) quality. Three potential approaches for tackling this problem are identified:

1. Find a (near) optimal search time *a priori*, using either an analytical or experimental approach;

2. Find a (near) optimal search time at runtime, through monitoring the effect of delays in enacting the response operation and incorporating this information into the evaluation process;

3. Issue schedules gradually, one task at a time, thereby removing the requirement that an optimal search time must be found at all.

Considering the first approach, we note that an analytical approach is not feasible due to the complex nature of the objective functions described in Section 4.6. An experimental approach would involve setting a range of potential search times and, through computational experiments, identifying that which leads to best performance on average. This approach will be explored during the evaluation process detailed in Chapter 7.

The second option is also explored in Chapter 7. Here, the delay in the start of the response operation is explicitly acknowledged in the evaluation of solutions during the search process. This is achieved by monitoring the length of time elapsed and setting this as the earliest start time of all tasks in the model. The effect of this approach is illustrated in Figure 6.9. The figure shows a hypothetical objective value trajectory (solid black line) as would be perceived during an optimisation process under the normal procedure. The red line shows the penalty to the objective value of the solution incurred due to the delay in the response operation being enacted. Accordingly, the delay penalty increases steadily as time passes. The sum of these two functions is illustrated by the dashed line. This represents the true value of the solutions over time, where the delay penalty has been added to the perceived objective value.

The point 'C' in Figure 6.9 corresponds to the perceived objective value at the point where the search process has terminated. The point 'B' shows the true value of this solution. When not taking the delay penalty into account, the best value is achieved at 'C', but when the delay penalty is incorporated we can see that the best value was in fact obtained at point 'A'. Note that the penalty of delaying the start of the response operation is a simplified illustration; in practice, due to complexities

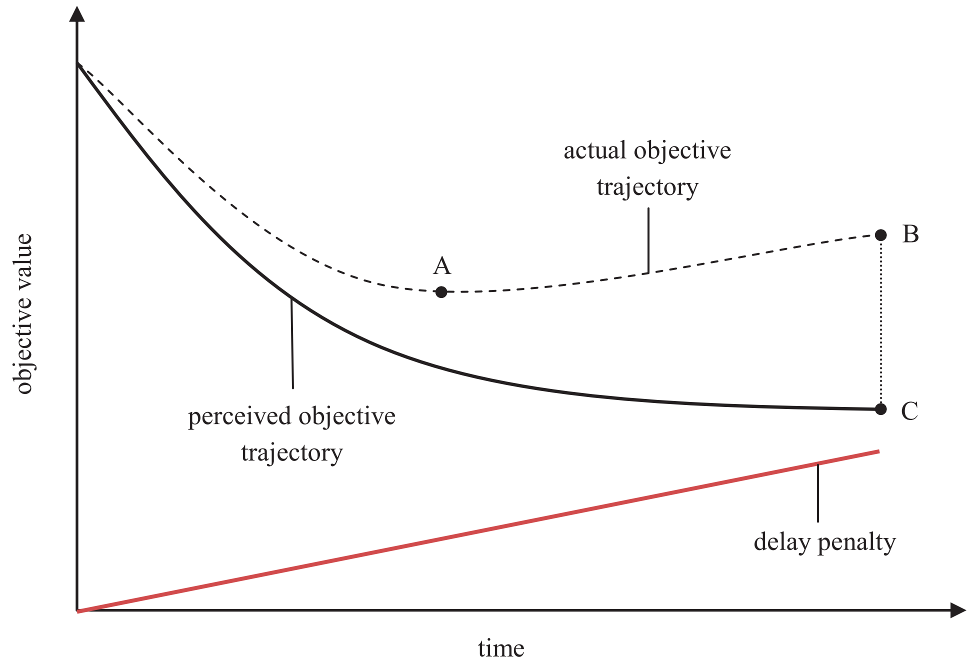

Figure 6.9: Illustration of the perceived objective trajectory of an optimisation process being affected by the associated delay incurred.

such as the dependencies which exist between tasks, the penalty may not evolve in a simple linear manner.

Figure 6.9 also illustrates the potential for the second approach to outperform the first. The second approach could be implemented by having the optimisation algorithm use the *actual* as opposed to *perceived* objective values of solutions when making comparisons. Nevertheless, this approach is not ideal. In particular, some delay in the enaction of the response operation will still be incurred (see point 'A' in Figure 6.9). As such, we now go on to consider the third strategy described: issuing instructions to responder units in a gradual manner. Here, rather than constraining the model to hand over an entire schedule at a single point in time, it is allowed for single tasks to be issued as instructions to individual responder units as and when the unit becomes free. This process is illustrated in Figure 6.10.

This is accomplished through the partitioning of all tasks within the model at a specific moment of time into two complementary sets: *fixed*, denoting tasks which have been given as instructions and which a responder unit has begun; and *floating*, denoting tasks yet to be issued to a responder unit. Employing the optimisation procedure in real time involves issuing a single task to a responder at the point where they become idle, whilst continuing to search the remaining parts of the solution space corresponding to the known tasks which are yet to be fixed. Thus, a responder's

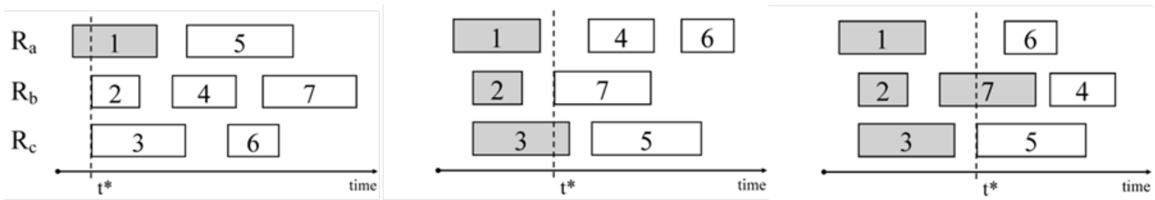

Figure 6.10: A simple example illustrating the gradual issuing of tasks to responders, resulting in them taking on 'fixed' status (shaded) whilst the remaining 'floating' tasks (not shaded) continue to be manipulated.

schedule is not fixed at time τ=0, but rather continuously built as the response operation progresses. Communication of information from the optimisation model now involves the passing of single instructions to single responders, each one specifying the task that responder is to begin working on immediately. These communications will occur continuously throughout the response operation.

6.3.2 Updating parameters

As has been discussed in detail throughout this Chapter, a variety of information will be gathered over the course of the response operation. For example, when a responder completes a task, the duration of this task is now known and this information will be passed to the optimisation model. This information may be used by the optimisation model to revise parameters to ensure a higher level of fidelity to the real problem. It is assumed that the passing of much of this information to the optimisation model is instantaneous. For example, when a casualty undergoes a further triage assessment the outcome of this assessment (which, as was discussed in further detail in Section 6.2, may be subject to error) is immediately available to the optimisation model. However, in order to explore how delays in information transfer may affect the utility of the optimisation model, we will allow for a delay in the communication of information relating to the start and end times of tasks and of periods of travel. The length of such delays will be set to varying levels in the experimental analyses to be presented in Chapter 7.

For each task within the model, information regarding its processing is passed from the simulation model to the optimizer at three points. Firstly, when a responder unit begins to travel to the appropriate location to undertake the task in question, this time is communicated. Note that in some cases the responder will already be at the location in question, and so this travel time would be zero. Secondly, upon the responder arriving at the location of the task and commencing its processing, again this time is communicated. Finally, the time at which the task is completed is communicated to the optimizer. Each of these communications may be subject to a delay in its reaching the optimisation model. This delay is simulated by drawing a random variate $X \sim exp(\lambda^{del})$, where λ^{del} is given as an input parameter to the simulation model. In cases where a responder reaches one of these points (beginning to travel, arriving at location to commence a task, or completing a task) earlier than expected, this information is incorporated into the schedule of the optimisation model as soon as it arrives. In cases where a communication has yet to be received by the expected time, the appropriate parameters of the task are continuously adjusted to track the elapsed time until the information has been received, at which point the parameter is set to its true value.

A simple example of an evolving schedule is given in Figure 6.11. The illustration shows the schedule of a single responder, as viewed from the perspective of the optimisation model, and how this schedule changes with time. These changes are illustrated on the vertical axis. Note that, in this case, the tasks assigned to the responder do not change in their ordering, only in the parameters describing their timings. As time progresses, we see tasks moving from a *floating* state (dark green or blue) to a *fixed* state (light green or light blue). We also observe the points at which information regarding the timings of tasks are sent, and the delay in these messages reaching the optimizer, at which point the schedule is updated to reflect the new information. For example, the initial estimated completion time of task t_1 is shown to be 7 minutes. However, the true duration is in fact 5 minutes. Thus, at the 5 minute mark, a message is sent from the simulation to the optimisation model notifying it that the true duration of task t_1 is 5 minutes. However, there is a delay of 1 minute in this message reaching the optimisation model. It is therefore not until the 6 minute mark that the optimisation model is updated, with the duration parameter of task t_1 changed from the original estimate of 7 to the true value of 5. During the simulation and optimisation of a full problem instance, such evolution of model parameters will clearly occur on a much larger scale.

The communication of information from the problem environment to the optimisation model consists of task timing information of the type illustrated in Figure 6.11 together with casualty health states as encoded by triage labels. Information is sent as single items, in real time, as and when it comes to light. For example, a communication could be of the form 'casualty 25 was noted to be in health state T1 at time 14:47', and many such communications will occur as the response operation progresses. In the language of the rolling horizon control literature [82], this system can be described as employing a continuous scheduling strategy [43] as the model being optimised is changed immediately upon receiving any information, at which point optimisation continues on the revised model. An alternative approach would be to consider a periodic and event-driven rescheduling strategy [23]. This would involve collecting information over a period of time and sending it as a 'packet' to the optimisation model, which would then be updated. This process would continue unless an event designated as particularly important occurs, in which case information would be passed on and the optimisation model would be revised immediately. For example, the discovery of a new casualty could be considered an important event to be communicated immediately, while the recording of task timings could be communicated periodically. The continuous scheduling strategy is most appropriate in the present case since the optimisation algorithms employed, as described in Chapter 5, are of an 'anytime' nature. By this we mean that the algorithm can return the best solution found at any point. As a result, the algorithm can be interrupted at any stage to

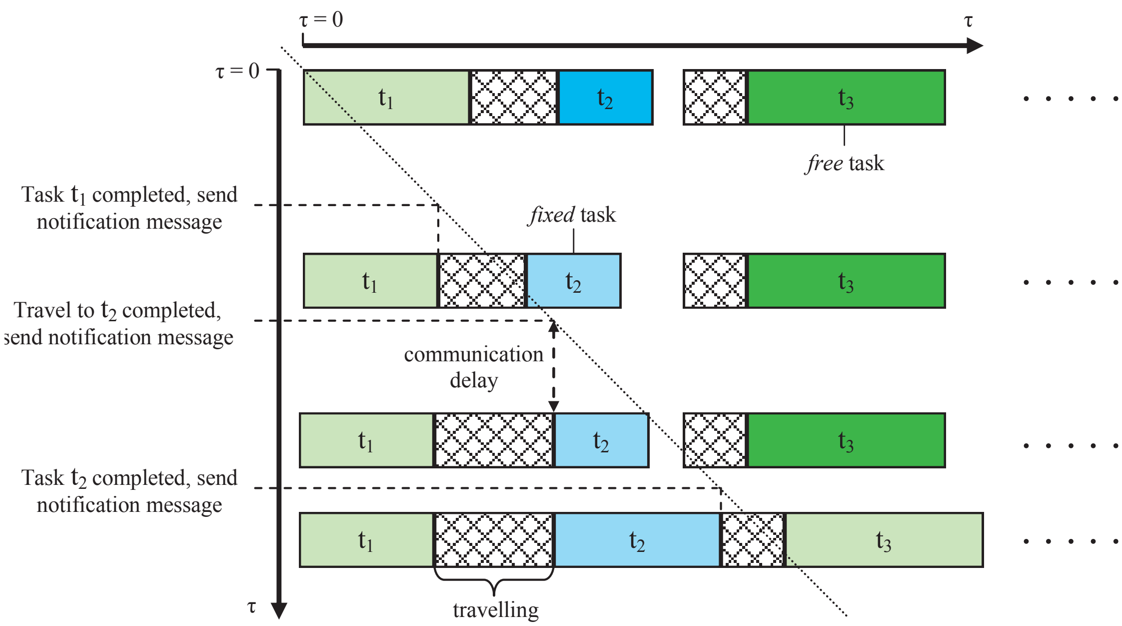

Figure 6.11: An illustration of the dynamics inherent to the temporal parameters within the model, examining the schedule of a single responder unit. Tasks which are issued are of the lighter shades. Travel times and task durations are revised to correct initial estimates, with delays in these communications from the simulator to the optimizer.

allow for information to be received and the model updated. This then implies that there is no benefit to collecting information in packets to be sent on at less frequent intervals.

The model can be extended to allow for the changes discussed. This can be achieved through the use of a heuristic procedure to govern the assignment of tasks associated with a newly discovered casualty. Note that the arrival of more responder units does not strictly require any modification of the solution method described thus far, as the iterative VND algorithm described in Algorithm 3 will automatically begin to allocate existing tasks to them as such allocations will lead to higher quality solutions. The constructive heuristic described in Section 5.3.1 can be employed when a group of new tasks are to be incorporated into the schedule, specifically deciding to which responder an un-allocated task should be assigned, and, in the case of transportation tasks, to which hospital the casualty should be taken to.

6.4 Improving predictions

The utility of the proposed optimisation model rests, in part, on its ability to make accurate predictions. Indeed, the ability of an optimisation model to *look ahead* when making decisions, as opposed to looking only at the immediate consequences of them, was a key motivation for the work described in this thesis. The objective functions used to evaluate and compare candidate solutions all require the prediction of quantities such as task durations, travel times, and casualty health. Accordingly, it is of interest to consider how the predictions made by the model will be affected by the uncertain and dynamic environment introduced through the simulation described in Section 6.2, and how prediction could be improved in this situation.

One key component of this approach has already been described in Section 6.3, where it was shown how the optimisation model can accept information relating to *past* events (e.g. the duration of a completed task). By ensuring the optimization model's view of past events is accurate, the accuracy of *future* predictions is improved. However, more can be done. For example, consider the uncertainty introduced through a disruption of the transport network, as described in Section 6.2.4. An unknown disruption will lead to inaccurate predictions of travel times in the optimisation model. The notification of the travel times of completed journeys allows the optimisation model to correct its view of the past, but can this same information be used to improve predictions of future travel times? The application of Bayesian statistics allows for this to be done. The methodology needed for such improved prediction of task durations, self-presentation, and travel times, will be detailed in this section.

6.4.1 Self-presentation

Recall that the simulation of self-presentation of a casualty with health state T3 required a parameter λ^{sp} which governed the rate at which casualties will leave their incident site to self-present. The parameter was used in an exponential distribution, such that

$$X_i^{sp} \sim exp(\lambda^{sp}) \tag{6.9}$$

where X_i^{sp} is the maximum time which casualty c_i will wait without receiving attention before self-presenting. The parameter λ^{sp} is unknown at the start of the response operation. However, as self-presentations are recorded at hospitals, data regarding realizations of the random variables X_i^{sp}, denoted x_i^{sp}, will become available, and these data can be used to improve any estimate of λ^{sp}. Specifically, for each individual datum x_i^{sp} we have

$$P(\lambda^{sp}|x_i^{sp}) \propto P(x_i^{sp}|\lambda^{sp}) \times P(\lambda^{sp}). \tag{6.10}$$

We use a gamma distribution with parameters a and b as the prior distribution for $P(\lambda^{sp})$, in which case the posterior distribution is also gamma with parameters $a+1$ and $b + x_i^{sp}$. The mean of this posterior distribution is then taken as the subsequent estimate of the self-presentation parameter:

$$\lambda_*^{sp} = \frac{a+1}{b + x_i^{sp}}. \tag{6.11}$$

In this manner, each time a data point is observed the estimate of λ^{sp} is revised, allowing for more accurate predictions of self-presentation in the remainder of the response operation.

Example 6.4.1. Consider a point mid-way through a response operation, where the current values of the prior distribution are $a = 5$ and $b = 50$, giving an estimate of the self-presentation rate as

$$\lambda^{sp} = a/b = 0.1. \tag{6.12}$$

This value corresponds to an average threshold waiting time of

$$E(X_i^{sp}) = 1/\lambda^{sp} = 10. \tag{6.13}$$

At this point, a self-presentation is registered at hospital h, where the casualty came from incident site d. As we know the time the incident at site d occurred, and also an estimate of the time needed to travel from site d to hospital h, the value x_i^{sp} can be deduced. In this example, $x_i^{sp} = 15$. Accordingly, the self-presentation rate is revised to be

$$\lambda_*^{sp} = \frac{a+1}{b+x_i^{sp}} \tag{6.14}$$

$$= \frac{5+1}{50+15} \tag{6.15}$$

$$= 0.0923. \tag{6.16}$$

This value corresponds to an average threshold waiting time of

$$E(X_i^{sp}) = 1/\lambda_*^{sp} = 10.834, \tag{6.17}$$

Thus, we see that the observation of a larger waiting time than expected has led to an increase in the expected waiting times of all future self-presenting casualties.

6.4.2 Task duration

Having described how the durations of tasks are simulated, we turn our attention to how the optimisation model should predict task durations when given only the initial estimated values.

Given the hierarchical structure described in Figure 6.2 and the estimates of tasks durations $e_{j,i}$, the true duration of each task is known to follow the distribution $N(e_{j,i}, s^2)$. An estimate of the duration can then be made according to a desired level of confidence. That is, for a given confidence level $\Psi \in [0, 1]$ we estimate the duration of tasks to be

$$\hat{\theta_{j,i}} = F_{j,i}^{-1}(\Psi) \tag{6.18}$$

where $F_{j,i}$ denotes the relevant cumulative distribution function.

The hierarchical probabilistic model illustrated in Figure 6.2 shows that the true duration of a task $\theta_{i,j}$ influences the associated estimate $e_{i,j}$ and is itself influenced by the site-level parameters μ_j, σ_j. When considering how to predict the true values of these task durations, we consider three levels of assumed knowledge which could be modelled in this situation:

(a) Perfect knowledge - the error in the estimate, $s^2 \to 0$ so that $e_{i,j} \approx \theta_{i,j}$;

(b) Basic knowledge - only the estimated durations $e_{i,j}$ are known, with a significant variance s^2;

(c) Intermediate knowledge - as in (b) but with known site parameters μ_j, σ_j^2.

Case (a) is clearly unrealistic, but provides a useful comparison for (b) and (c). In the case of (a), the initial estimate of task duration $e_{i,j}$ is approximately equal to the true duration $\theta_{i,j}$ and so the optimisation model will never need to revise this initial estimate. This is the scenario which was assumed throughout our development in Chapter 4, where it was noted that all task durations were known with certainty at the outset of the response operation.

Case (b) is more realistic, assuming a minimal amount of information is available, and enables the scheduling model to represent duration $d_{i,j}$ as a normally distributed random variable with mean $e_{i,j}$ and variance s^2. Finally, case (c) can be thought of representing the best possible result of an *online learning* process. That is, given some (possibly non-informative) prior distributions over the parameters $\mu_j, \sigma_j^2, \Phi, \Sigma$ we can employ Bayesian inference to calculate the posterior distributions of these parameters given information on the actual durations $\theta_{i,j}$ which will be obtained as the response operation progresses and further information becomes available through the online process described. An idealized result of such a process would be a perfect knowledge of site parameters μ_j, σ_j^2. Given these, we can set a prior distribution for all unknown $\theta_{i,j} \sim N(\mu_j, \sigma_j^2)$ and, given evidence in the form of an estimate $e_{i,j} \sim N(\theta_{i,j}, s^2)$ the posterior distribution of $\theta_{i,j}$ can be calculated from standard Bayesian formulae. Specifically, $\theta_{i,j} \sim N(x, y^2)$ where

$$x = \frac{\frac{\mu_j}{\sigma_j^2} + \frac{e_{i,j}}{s^2}}{\frac{1}{\sigma_j^2} + \frac{1}{s^2}}, \text{ and } y = \frac{1}{\frac{1}{\sigma_j^2} + \frac{1}{s^2}}. \tag{6.19}$$

6.4.3 Travel times

Static routing (SR)

Under the static routing policy, the route taken by responders does not change over the course of the response operation for any specific journey. As such, our goal is to generate and update the probability distribution of the travel time associated with each journey. Recall that any simulated disruption to the transport network (see Section 6.2.4) will introduce a significant amount of uncertainty in travel time estimation.

As has been previously discussed, we assume travel times follow a log-normal distribution with an assumed, constant precision ξ,

$$X \sim logN(\nu, \xi), \tag{6.20}$$

as discussed in [134]. In a manner similar to that described in subsections 6.4.1 and 6.4.2 we employ a Bayesian approach in revising the estimate of the unknown parameter ν as more travel time data becomes available. Specifically, using the conjugate prior of the log-normal distribution for ν,

$$\nu \sim N(\nu_0, \xi_0), \tag{6.21}$$

we can calculate the posterior distribution following the observation of n data x_i,

$$\nu \sim N(\nu_n, \xi_n) \tag{6.22}$$

where

$$\nu_n = \frac{\xi_0 \nu_0 + \xi \sum_{i=1}^{n} \ln(x_i)}{\xi_0 + n\xi} \tag{6.23}$$

and

$$\xi_n = \xi_0 + n\xi. \tag{6.24}$$

The expectation of this posterior distribution, $\hat{\nu} = E(\nu)$, is then used as an estimate of ν, giving $X \sim logN(\hat{\nu}, \xi)$. As noted previously, the median travel time for the route in question can then be estimated as $m = e^{\hat{\nu}}$.

CAR

Recall that the Centralized Adaptive Routing policy involved updating the parameters of edges on the transport network when those edges were part of a travelled route, with the aim of obtaining a more accurate view of the disruption on those edges. Following this updating step, the next route calculated for the same journey may be different, as other edges could have a lower perceived disruption level.

Given the transport network representation G_j used to find the route in question, p_j, the travel time is estimated to be the median time according to equation 6.5, $m = KWH(D_{G_j}(p_j))$.

AIAR and ACAR

Whereas under policies SR and CAR the routing choice was known to the optimisation model, in the case of autonomous routing this is not the case. Accordingly, making accurate predictions of travel times is more challenging as there is now uncertainty in the route choice over and above uncertainty in the level of disruption of the network. However, predictions for future journeys may be informed by the observed travel times of those journeys up to that point in the response operation.

We propose using an exponential smoothing method when predicting the travel time of the ith journey based on past observed travel times. Specifically, denoting the travel time of journey i as τ_i,

$$\tau_{i+1} = \gamma \times \tau_i + (1 - \gamma) \times \tau_{i-1}. \tag{6.25}$$

This simple model is easy to interpret and computationally lightweight (only requiring the last value to be stored in memory). Exponential smoothing has also been noted to be successful in short term forecasting [57].

In the experimental analysis presented in this thesis we will use a smoothing factor of $\gamma = 0.5$. This corresponds to an equal weight being placed on the most recent observation and on the observation preceding that. The optimal value of γ will depend on the nature of the responder units' routing. Specifically, if routing choices improve travel times in a smooth manner, without a great deal of random variation, a high value of γ may lead to more accurate predictions. However, in the case where routing choices are leading to erratic travel times, a lower value of γ could improve predictive performance.

6.5 Simulating decision making

Having specified how a range of parameters relating to the solution and objective spaces of the model are simulated, it is left to consider the simulation of the decision making of responder units during the response operation. We consider two cases, corresponding to which of the two solution methods (constructive heuristic and local search) described in Chapter 5 is being employed when identifying schedules.

6.5.1 Using the search method

When a responder unit completes a task the best schedule found up to that point is consulted to determine which task they are to complete next. By employing this method we are effectively assuming that responder units will always 'obey' the instructions of the optimisation model. While such an assumption does not pose any problem when considering only static problems, issues are raised when it is applied in the type of dynamic and uncertain incidents to be simulated using the methods described in this chapter. In particular, note that it is possible that the world-view of the optimizer may be inaccurate, whereas the local world-view of the responder may be more reflective of the actual problem environment. This could lead to situations where instructions are given to a responder unit which are based on inaccurate information, with the responder unit keen to override these instructions as they believe they would be better equipped to make a decision themselves using the information available to them.

Situations like these could occur with respect to information regarding the health of casualties, or the timings of tasks. For example:

1. A casualty is incorrectly classified during as of health state T1 during a triage operation, but an Ambulance responder unit instructed to deliver pre-transportation stabilizing treatment identifies the error and notes they are in fact of health state T2. Other casualties of health state T1 are present and also require treatment.

2. A SAR responder unit is instructed to begin an extrication task corresponding to a T3 casualty as opposed to a nearby T1 casualty, as the optimisation model believes a pre-rescue stabilizing treatment task for the T1 casualty is underway. In fact, this task took less time than estimated and has been completed, although this information has yet to be received by the optimisation model.

In scenario 1, it may be prudent for the Ambulance responder to ignore their instruction and instead administer treatment to a casualty of higher priority. Similarly, in scenario 2 the SAR unit may prefer to extricate the T1 casualty who has received treatment sooner than expected. Under the simulation and interface described in this chapter, these plausible outcomes will not be possible, and this represents a limitation of the methodology in terms of realism. Nevertheless, the proposed simulation and corresponding interface allow for a level of realism not yet considered in this context, and so we will leave the modelling of scenarios such as those described to future work.

6.5.2 Using the constructor

When employing the constructive heuristic described in Section 5.3 to determine the schedule being issued, we note that the solution method was originally designed to mimic the decision making of responders during an actual incident, to enable meaningful comparisons with the proposed local search solution method. As such, the constructive method can be applied when faced with more realistic problems simulated using the methods described in this chapter in order to give a full simulation of a response operation enacted without the use of the optimisation model. Recall that the constructive heuristic required some parametrization in order to be fully specified, and that these parameters were set through an experimental analysis which identified the configuration which performed best, on average, when applied to a range of problem scenarios.

Implementation 6.5.1. When introducing the constructive heuristic solution method in Section 5.3, it was shown that a full schedule is generated in any one application. However, note that the algorithm has a basic structure whereby the responder who is due to finish their assigned jobs is identified, and a task to assign to the end of their schedule is then chosen. It is clear, then, that the algorithm could be run in a staggered manner in real-time, where a task is selected and assigned to a responder unit when they complete their current task (see Algorithm 2). Another possible implementation of the constructive heuristic in real time would be to repeatedly construct whole schedules, respecting the position of any tasks which have been fixed. These schedules would be consulted in the same manner as those located during a local search procedure, with tasks being transferred from a floating state to a fixed state when responder units become free. This second implementation route was employed in this work, as the computational burden imposed is insignificant due to the real-time nature of the problem.

6.6 Summary

The dynamic and uncertain nature of the MCI response environment is clear, yet, as discussed in Section 3.4, this has only been acknowledged to a limited extent in related work. In this Chapter, an approach to accounting for several sources of uncertainty and/or dynamic behaviour has been described. Firstly, the use of a simulation methodology has been employed in order to replicate the stochastic nature of casualty health, task durations and travel times. Following this development, the necessary details of the interface between this simulation and the optimisation model described in Chapter 4 were given. Finally, the possibility of improving the predictions made in the optimisation model in response to the observation of simulated data was considered. As a result, the offline optimisation model M_{off} has been extended to an online optimisation model M_{on}, which should be of far greater utility in an actual MCI environment. This last hypothesis, among others, will be tested through a series of computational experiments to be described in Chapter 7.

Chapter 7

Evaluation

Assessing model utility in realistic scenarios

7.1 Introduction

In this thesis a multi-objective combinatorial optimisation model for use in MCI response has been described. The initial model development of Chapter 4 took place under a number of assumptions relating to the nature of information in an MCI. Specifically, it was assumed that

1. All information necessary to specify the model is *available* at the outset of the response operation;

2. All such information was known with *certainty* and not subject to error;

3. All such information was *constant* over time as the response operation progressed.

We denoted this model as the *offline* model, M_{off}. By offline, we mean that the model did not require regular interaction with the problem environment. Rather, all necessary information would be gathered at a single point in time (i.e. upon initialization), and all instructions to responder units would be issued at a single point in time (i.e. when the solution method described in Chapter 5 terminated). We note that the models could use either the constructive heuristic or local search solution methods described in Chapter 5. Throughout this chapter, when employing either model it should be assumed that the local search method is being used, unless explicitly stated otherwise.

The model M_{off} was further developed in Chapter 6 to produce an *online* model, M_{on}. By online, we mean that the model is capable of regular interaction with the problem environment. Specifically, the model M_{on} could receive information (such as the recorded duration of a certain task) from the problem environment at any time, and could issue instructions to responder units in a gradual manner as opposed to handing over a full schedule. The model M_{on}, therefore, does not require the assumptions made by the model M_{off} and listed previously.

In this chapter, we seek to evaluate the utility of both models M_{off} and M_{on} over a range of possible problem scenarios. Our goal is to assess applicability in general, and also to explore in detail to what extent performance is dependent on specific components of the models, or on specific characteristics of problems. Given the resources available and the large number of parameters which could potentially be varied, it is neither feasible nor desirable to exhaustively analyse each one in this thesis. Rather, we proceed to identify a number of key parameters to be studied.

7.1.1 Goals

In terms of problem characteristics, the following will be examined:

- The number of incident sites comprising the MCI;
- The number of responder units available;
- The level of dependency which exists between tasks;
- The level of dynamic behaviour present.

Given the potential multi-site nature of some of MCIs, we clearly wish to consider some multi-site incidents in our experiments. Moreover, we are interested in examining the effect of the incident being spread over multiple sites. As such, the number of sites comprising an incident is one characteristic we will vary.

In conjunction with keeping the number of casualties in each problem constant, varying the number of responder units will vary the ratio of available resources to demand for resources. This has clear potential to impact upon the utility of the optimisation model. In particular, we might expect that the model would be particularly useful in situations where resources are sparse, and of less benefit when they are plentiful. Varying the number of responders will allow for insight into this matter to be gained.

Given the scheduling nature of the optimisation model, it is reasonable to expect that the level of dependency that exists between the set of tasks could significantly impact its utility. It is not clear, however, what the nature of such impact would be. An increase in the number of inter-dependent tasks will certainly pose an additional challenge to the optimisation model, and the question is whether the model is better or worse equipped to tackle this challenge in comparison to the alternative approach (i.e. the constructive heuristic method).

Finally, the extent to which the problem exhibits dynamic behaviour is clearly of interest, as such behaviour provided the motivation for the development of Chapter 6. However, the 'level of dynamic behaviour' is not as clear a term as those discussed so far, and will not be easily represented by a single parameter which can be varied. Rather, a number of parameters, both quantitative and qualitative, will need to be considered when exploring this key aspect of the MCI problem.

Of the parameters which will *not* be varied in the course of experimentation, some do not require any significant justification in this respect. For example, it would be possible to examine the effect of changes in the average speed of response units;

however, what would the results tell us about the future conduct of either MCI response operations or the application of optimisation models to them? In contrast, some parameters would have been desirable to include in our experimental design. In particular, the Markov chain transition probabilities (as given in Figure 4.10) which govern the simulation of the health of casualties are together important parameters which could have significant influence upon experimental results. However, running computational experiments of the simulation and optimisation models is a resource intensive task. In particular, the simulation of a response operation may take a number of hours, as it must be carried out in real time in order to provide an accurate reflection of how the optimisation model would be employed in reality. Given the need to ensure any insights gained from the experimental data are not spurious and arising through chance alone, a large number of such experimental runs (corresponding to a significant amount of CPU time) will be required when investigating any one parameter or characteristic. This lead to the necessary omission of some potentially interesting investigations, in order that those which are carried out are done with the appropriate degree of rigour.

7.2 Problem scenarios

In this section we describe the characteristics of the set of problem scenarios which will be used in the computational experiments evaluating the model developed in this thesis, as set out in 7.1.1. Further details regarding the rationale for the choice of problems is provided, as is a discussion of potential limitations. We first consider the fundamental characteristics which will not be varied, these being the transport network topology, the set of casualties, and the set of hospitals. Those characteristics which will be varied over the course of experiments will then be detailed. Following this, Section 7.2.7 will describe problem scenarios in terms of their dynamic characteristics.

7.2.1 Transport network

All problems considered in this thesis take place within a single geographic area, namely central London, UK. This location was chosen for several reasons:

1. A dense inner city transportation network ensures the model is tested in its ability to use graphs and routing algorithms;

2. Information regarding hospitals and emergency services is available in the public domain;

3. The area chosen includes that affected by the London terrorist bombings, a major motivating incident for this study.

The primary limitation of performing all analysis in this environment is in terms of the dense transport network. In less built-up or rural areas the topology of the transport network would be significantly different. Although a sparser network would not pose challenges with regard to the computational issues of routing, response performance may be more sensitive to disruption in individual roads.

The area considered is illustrated in Figures 7.1 and 7.2, where the latter shows only the graphical representation of the transport network.

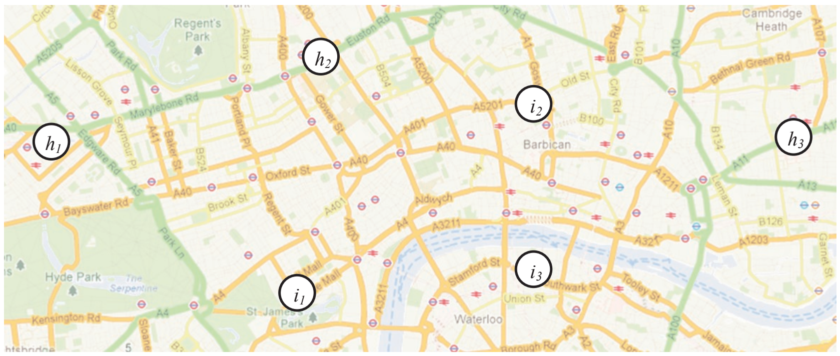

Figure 7.1: The geographic area considered in all problem scenarios, where h's denote the location of hospitals and i's the locations of incident sites.

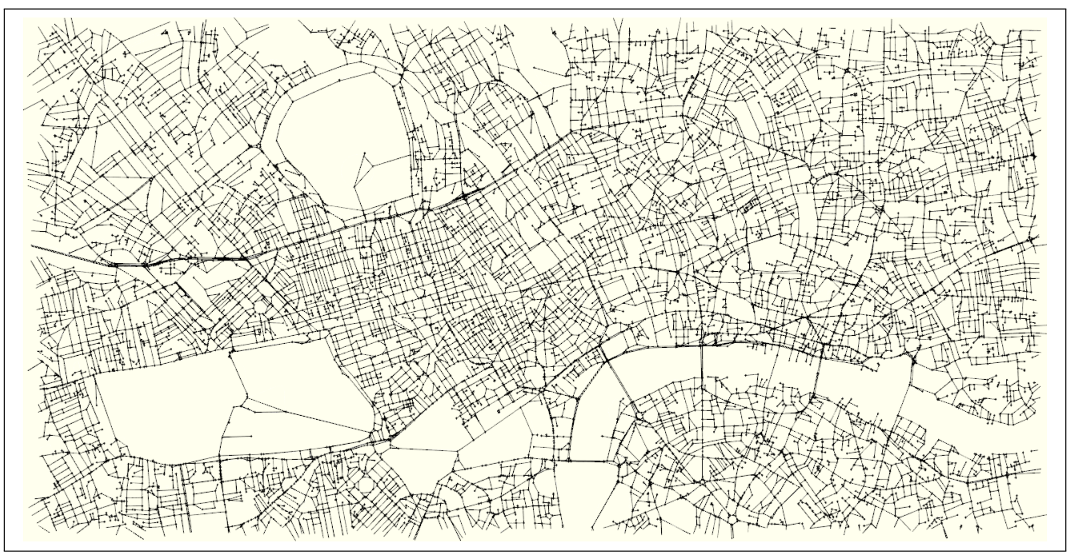

Figure 7.2: The graphical representation of the transport network considered in all problem scenarios.

7.2.2 Casualties

In all problems the total number of casualties does not change. Moreover, we do not alter their initial health states or their specialist injury status. The number of casualties designated as trapped will be changed in the context of varying the level of dependency between tasks (to be described in Section 7.2.5), as will their locations in the context of varying the number of incident sites (to be described in Section 7.2.6).

The total number of casualties in each health state (initially) is:

- T1 - 42;
- T2 - 60;
- T3 - 108.

A total of 210 casualties was considered to represent an adequate challenge to the optimisation model, providing scope for an interesting examination of how its utility is affected by other parameters. Moreover, this figure complies with the definition of "large scale" as set out in Chapter 2.

7.2.3 Hospitals

As discussed in Section 4.3.3, each hospital considered within the model must be specified by their location, an initial resource level, a maximum capacity, a resource increase rate (where such increases are due to the enaction of major incident plans) and an indication of any specialist treatment facilities present. In the experiments of this chapter the set of hospitals and their specification will be held constant. This allows for the experimental analysis to focus on identifying effects on model utility other than that which would arise from changes to hospitals and their parameters, as was discussed in Section 7.1.

Three hospitals will be considered in the problem scenarios, with their locations shown on Figure 7.1. By restricting the number of hospitals to three, the goal is to aid interpretation of both the problem and the experimental results, which may otherwise be unnecessary confounded by excessive information. The parameter values for the three hospitals are given in Table 7.1.

Table 7.1: Parameter values for the three hospitals included in all problem scenarios.

Hospital	cap_h	max_h	$rate_h$	A & E	Burns unit
h_1	50	90	2/3	✓	×
h_2	30	80	5/3	✓	×
h_3	25	40	3/7	✓	✓

The values chosen lead to an initial total casualty capacity of 105 and a maximum capacity of 210. Given that the total number of casualties in the problems considered will always be 210, this scenario will allow for issues regarding hospital capacity to be felt initially during the response operation but for an adequate capacity to be available by its end.

7.2.4 Responders

While the transport network, casualties and hospitals described in Sections 7.2.1, 7.2.2 and 7.2.3 are to be held constant throughout all computational experiments, the number of responder units will be varied throughout in order to ascertain the associated effects upon the utility of the optimisation model. In order to explore this whilst maintaining a feasible number of experiments, the number of responders takes one of three levels which will be denoted R_{low}, R_{med} and R_{high}.

In the first scenario, R_{low}, the responder units available are 12 ambulance, 3 MERIT units, 3 HART units and 13 SAR teams. This is increased to by a further 12 ambulances, 3 MERITs, 3 HARTs and 7 SAR teams for the scenario R_{med}, with these same amounts added once again for the final scenario R_{high}. These specifications are summarized in Table 7.2.

Table 7.2: Resource levels and associated responder units.

Scenario	Ambulance	MERIT	HART	SAR
R_{low}	12	3	3	13
R_{med}	24	6	6	20
R_{high}	36	9	9	27

The numbers of responder units present in each scenario allow for a reasonable range of responder/casualty ratios to be considered. For example, the lowest ration of ambulances to casualties considered in $12/210 = 0.057$ while the highest is $36/210 = 0.171$.

In all problem scenarios we assume that all responder units listed in Table 7.2 are available at time $\tau = 0$, i.e. upon the initialization of the model. We note that it would be possible to change these setting so that some or all responders are unavailable for a period from $\tau = 0$. Due to the fact that the number of casualties is to be held constant throughout, variations in the number of responders will correspond to variations in the ratio of responders to casualties. Recall that in Section 2.2.3 the scale of a MCI was considered to be *relative* to the resources available, and so variations in responder level will reflect this notion of scale.

7.2.5 Tasks

As mentioned in Section 7.2.2, the number of casualties designated as trapped will be varied in order to vary the level of dependency that exists between the set of tasks in the optimisation model. That is, an increase in the number of casualties who are trapped corresponds to an increase in the number of rescue tasks, which will increase the number of dependency relations in the set of all tasks. As in the case of variations in the level of responders, three discrete levels of dependency denoted D_{low}, D_{med} and D_{high} are defined. These levels are summarized in Table 7.3, which describes the total number of casualties in each health state who are designated as trapped for each dependency level.

Also given in Table 7.3 is the probability that any given casualty will require pre-transportation stabilisation treatment. As with rescue tasks, an increase in the number of casualties requiring a pre-transportation treatment task lead to an increased level of dependency within the set of tasks. In the case of pre-*rescue* stabilizing treatment, a similar probabilistic method is used to determine which casualties require

Table 7.3: Dependency levels and associated casualty characteristics.

Scenario	T1	T2	T3	P(*treat*)
D_{low}	6	6	6	0.3
D_{med}	12	18	18	0.6
D_{high}	30	36	42	0.9

these tasks. Here, though, the probability does not change over the levels D_{low}, D_{med} and D_{high}, and is instead set at 0.5 throughout.

Random sampling is also employed when determining task parameters. Specifically, the true duration of rescue and treatment tasks are sampled using the hierarchical model described in Section 6.2.3, which is specified by the parameters Φ and Σ, the mean and covariance at the highest level of the hierarchy. A consequence of employing these random methods when generating problem instances is the introduction of significant levels of heterogeneity, which will allow for conclusions to be drawn without significant concern regarding their robustness. That is, if the problems considered had little variation in terms of these task parameters, any results would be difficult to generalize as it would not be clear to what extent they were dependent upon the specific values chosen.

7.2.6 Incident sites

The scenarios considered may consist of one, two or three incident sites. As illustrated in Figure 7.1, these sites - denoted i_1, i_2, i_3 - are located at St James' Park, Queen Margaret University, and the Tate Modern Gallery respectively. The scenarios corresponding to each potential number of incident sites will be denoted S_{one}, S_{two} and S_{three}.

The 210 casualties considered will be evenly distributed across the number of sites. For example, a three site incident will involve $210/3 = 70$ casualties at each incident site. Moreover, the pre-defined characteristics of casualties (i.e. their initial health state and their being trapped or otherwise - see Sections 7.2.2 and 7.2.5) will also be distributed in an even manner. For example, each incident site will have the same number of casualties whose initial health state is T2.

The use of up to three incident sites will allow for the examination of the effects of spatial distribution of the MCI on the utility of the optimisation model. Moreover, by ensuring the casualty profile at each incident site is similar, effects due to the number of incident sites will be easier to attribute.

7.2.7 Dynamics

As discussed in Chapter 6, dynamic and uncertain behaviour within the problem environment may be thought of as affecting the objective space of the optimisation model, its solution space, or both. In order to fully evaluate the proposed optimisation model we propose using three stages of increasingly complex, and realistic, problem types: *static*, *partially dynamic* and *fully dynamic*.

Static problems

In static problem scenarios we assume that all necessary information required in specifying the model is available, accurate and constant over time. As mentioned in Section 7.1, this is exactly the type of scenario which was considered when developing the initial *offline* version of the optimisation model in Chapter 4. As such, the offline model M_{off} is clearly applicable to problems of a static type.

The online model M_{on} may also be applied to this type of problem, although there is not a great deal of motivation in this respect. That is, model M_{on} was designed specifically to address cases where the assumptions mentioned do *not* hold, and so we may question the usefulness of its application to those problems where they do. However, when we consider the nature of communication between the model M_{on} and the (simulation of the) problem environment, we can see that there is the potential of benefit. Although the ability to accept information from the problem environment will be redundant (since all information in the static problem case is known, accurate and unchanging), the ability to issue instructions in a gradual manner, as opposed to at a single point in time, could lead to improved performance.

Partially dynamic problems

In problems which shall be termed *partially dynamic*, we allow for changes over time in parameters relating to the objective space of the model. The nature of these parameters and how they may change over time has been described in Section 6.2. In this case, it is clear that any application of the offline model M_{off} can be significantly affected, in terms of its utility, by such changes. This is because the initial view of the problem which was passed to the model M_{off} will become increasingly inaccurate as the response operation progresses, because parameters (such as an estimated task duration) will change while the model M_{off} is not capable of recognizing such changes. By applying the offline model M_{off} to partially dynamic problems and comparing performance with that obtained through the constructive heuristic method, we will be able to estimate the effect of such an inappropriate application. This result is important, particularly when considering the number of optimisation models surveyed in Chapter 3 which had limited or no capability to interact with the problem environment in an online manner, but were nonetheless developed for problem types which would be expected to be of a dynamic nature.

The application of the online model M_{on} to partially dynamic problems will allow for the model to be evaluated in a more complete manner than when applied to static problems. The ability of the model to respond to incoming information regarding model parameters (as described in Sections 6.3 and 6.4), and the associated effect on the utility of the model, will be analysed.

Fully dynamic problems

Finally, further dynamic behaviour of the problem environment can be introduced through allowing for changes in parameters relating to the solution space. Problems which allow for such changes, over and above the changes relating to the objective space described for partially dynamic problems, will be termed *fully dynamic*.

In the case of fully dynamic problems, the application of the offline model M_{off} would lead to non-sensical situations. Recalling that changes parameters relating to the solution space include increases and decreases to the set of casualties (see Section 6.2.1), it is clear that the model M_{off} could find itself in a situation where it is providing instructions on the processing of casualties which are no longer in the scope of the problem (as they have left to self-present), whilst also ignoring the processing of casualties which were not known on initialization. We therefore note that an experimental analysis of the utility of model M_{off} in fully dynamic problems would not be of sufficient interest.

As in the case of partially dynamic problems, the online model M_{on} is again naturally applicable to fully dynamic problems. Experimental analysis of the utility of M_{on} in these types of problem scenarios will provide the most realistic test of the applicability of the work developed in this thesis. Again, the comparison will be made between the optimisation model M_{on} and the constructive heuristic method in order to estimate the potential benefits of the optimisation approach.

7.2.8 Summary of problem scenarios

Four dimensions of variation have been introduced. Firstly, we allow for three responder levels, R_{low}, R_{med} and R_{high}, governing the number of responder units available. The level of dependency within the tasks of the problem may also take one of three levels, D_{low}, D_{med} and D_{high}. The total disaster can be distributed over one of three sites, scenarios denoted by S_{one}, S_{two} and S_{three}. Finally, each problem may also vary in the level with which it exhibits dynamic and uncertain behaviour, from static problems, through partially dynamic, to fully dynamic. These variations in problem scenario characteristics are summarized in Table 7.4, where we also note which characteristics are to be held constant.

Table 7.4: Summary of the problem scenarios considered.

Characteristic	Levels
Transport network	Constant
Casualties	Constant (210 in total)
Hospitals	Constant
Responders	$R_{low}, R_{med}, R_{high}$
Dependency	$D_{low}, D_{med}, D_{high}$
Incident sites	$S_{one}, S_{two}, S_{three}$
Dynamicity	Static, partially dynamic, fully dynamic

In the remainder of this chapter we evaluate the performance of both offline and online search, in comparison to the standard constructive heuristic method, in these increasingly complex problems. The structure of this evaluation process is outlined in Table 7.5.

Table 7.5: Summary of the problem scenario and model type pairs evaluated.

	Problem		
Model	Static	Partially dynamic	Fully dynamic
Offline, M_{off}	7.3.1	7.4.2	n/a
Online, M_{on}	7.3.3	7.4.1	7.5

7.3 Static Problems

In this section we focus on evaluating model utility when applied to static problems, as defined in Section 7.2.7. First, the offline model M_{off} will be evaluated in Section 7.3.1. In doing so, we will make comparisons between the solution produced by model M_{off} with the solution produced by the constructive heuristic for the same problem instance. In fact, the constructive heuristic will be used to generate the initial solution used by M_{off} in the iterative search process described in Chapter 5.

Initially, results will be presented in an aggregated manner considering all problem scenarios described in Section 7.2. This will allow for an initial impression of the utility of model M_{off} in static problem scenarios, and will introduce the experimental methodology which will be used throughout. Following this, variations of the model M_{off} will be introduced, where specific components of the model have been removed. Comparing the performance of these model variants will allow for insight regarding the importance of these model components. The effect of the problem characteristics, as set out in Section 7.2, will then be evaluated.

Before turning our attention to the online model M_{on}, we examine the utility of M_{off} in what will be termed a *real-time* application. By this we mean an application of M_{off} to a problem which retains the characteristics deriving from its static nature, but where we acknowledge the fact that any time spent using the model M_{off} to search for high quality solutions will, by definition of the offline model and its methodology, be felt as a delay in enacting the response operation. The analysis of model performance in this application will provide empirical motivation for the use of the online model M_{on} to static problem scenarios.

Accordingly, the utility of model M_{on} will then be examined in Section 7.3.3. The rationale for this analysis is to enable a comparison of offline and online models in the case of static problems, and accordingly this section will less intensive than Section 7.3.1. Rather, a more detailed examination of the online model M_{on} will be presented in the context of partially and fully dynamic problems.

7.3.1 Offline model M_{off}

The stochastic nature of both the problem instances and the solution method employed by model M_{off} necessitate a statistical approach to any computational experimentation. That is, any experiment will involve the repeated application of the model to the same problem scenario, where with each run the specific problem instance will be different (due to the random sampling of several parameters; see Section 7.2). The overall problem scenario, as specified by, e.g., the resource level, will remain the same across these repeated runs. The output of all runs will then be considered together as a data set, and will be examined appropriately using summary statistics, graphical illustrations, and statistical models (if appropriate).

Initial evaluation

Here, we consider all possible static problem scenarios. That is, given the three variable problem characteristics described in Section 7.2 (responders, dependency and number of sites), each of which had three possible levels, we consider each of the $3 \times 3 \times 3 = 27$ possible problems scenarios. As previously stated, the initial solution generated by the constructive heuristic method provides the point of comparison with the solution presented by model M_{off} after application of the search process of Chapter 5. Accordingly, each experiment run returns a single vector, constituted of the difference between the initial and final objective values for each of the five objective measures defined in Section 4.6:

- $f_1(s)$, the expected number of fatalities,
- $f_2(s)$, measure of the time taken to deliver casualties to hospitals,
- $f_3(s)$, measure of how appropriate the hospital allocation choice is,
- $f_4(s)$, the total time spent idle by responders,
- $f_5(s)$, the latest time at which a casualty arrives at a hospital, i.e. the makespan.

These differences are presented in the form of *percentage improvement*, calculated as

$$[(\text{initial value - final value}) \ / \ \text{initial value}] \times 100$$

Experiments consisted of 50 runs for each of the 27 problem scenarios, giving a total of 50 results. Descriptive statistics of this data, as broken down by objective measures, are provided in Table 7.6. Note that a negative value corresponds to an improvement.

From the results shown in Table 7.6 we can see that, averaging across all static problem scenarios, the model M_{off} leads to better performance (on average) in terms of the objectives f_1 (fatalities), f_2 (hospital arrival time), f_4 (idleness) and g_2 (suffering) than the constructive solution method. For example, the median change in the expected number of fatalities is -7.5115%.

It should be noted that while while the *average* (i.e. mean) change in objective f_3 (hospital allocation) is 3.97%, the median change is -9.388%. This suggests the

Table 7.6: Average % change in objective values, across all problem instances, under model M_{off}.

	f_1	f_2	f_3	f_4	f_5	g_2
Min.	-25.40	-81.68	-92.03	-98.77	-31.84	-81.79
1st Qu.	-12.77	-71.06	-64.12	-50.75	-3.30	-71.05
Median	-7.51	-57.92	-9.39	-26.46	5.12	-57.91
Mean	-7.26	-47.73	3.97	-18.78	8.28	-47.56
3rd Qu.	-0.38	-21.24	45.13	0.00	19.03	-20.60
Max.	0.00	72.17	1084.42	443.19	211.45	90.27

distribution of these changes is being skewed in the positive direction, which is confirmed to be the case when noting the 3rd quartile and the maximum values of the distribution: 45.134% and 1084.417% respectively. The median puts less weight upon such extreme values than the mean, and provides a better estimate of the center of the distribution in such instances. Thus, when considering the median improvement rather than the mean, we observe improvements in objectives f_1 to f_4.

From these results we can conclude that the search method typically leads to a significant improvement in the estimate of fatalities, and typically an even stronger improvement in terms of suffering. However, due to the multi-objective lexicographic formulation employed, only objective f_1 is guaranteed to never worsen. We observe that, on average, the makespan of the schedules produced (objective f_5) will usually increase as a result of the search method. As such, if the only objective used to guide the model in selecting a solution were that of makespan, the resulting solution would likely be of poor quality in terms of the other objectives considered. This underscores the point made in Section 4.6, that the bespoke, detailed objective functions defined in this thesis are essential when considering the optimisation of casualty processing operations.

In order to gain further understanding of this data, we consider a subset of all problems, namely those of type $R_{high} - D_{high} - S_{three}$. The results are plotted as pairwise scatter plots in Figure 7.3. The figure also provides marginal histograms for each objective along with correlation coefficients for each pair of objectives.

The charts of Figure 7.3 are pairwise scatter plots, where each pair of objectives are taken together and the values of all solutions in terms of these two objectives are plotted. These plots are of interest in that they provide insight into the degree of correlation between each of the objective functions. This is important to understand, as if any two objective functions were highly correlated in their outcomes it would suggest that one of the objective functions is superfluous; effectively, not providing any valuable information that was not already known from the other objective function. However, both the scatter plots and the correlation coefficients provided in Figure 7.3 suggest that there is no pair of objective functions which exhibit such behaviour. The strongest correlation observed is between objectives f_4 (idleness) and f_5 (makespan), with a correlation coefficient of 0.674.

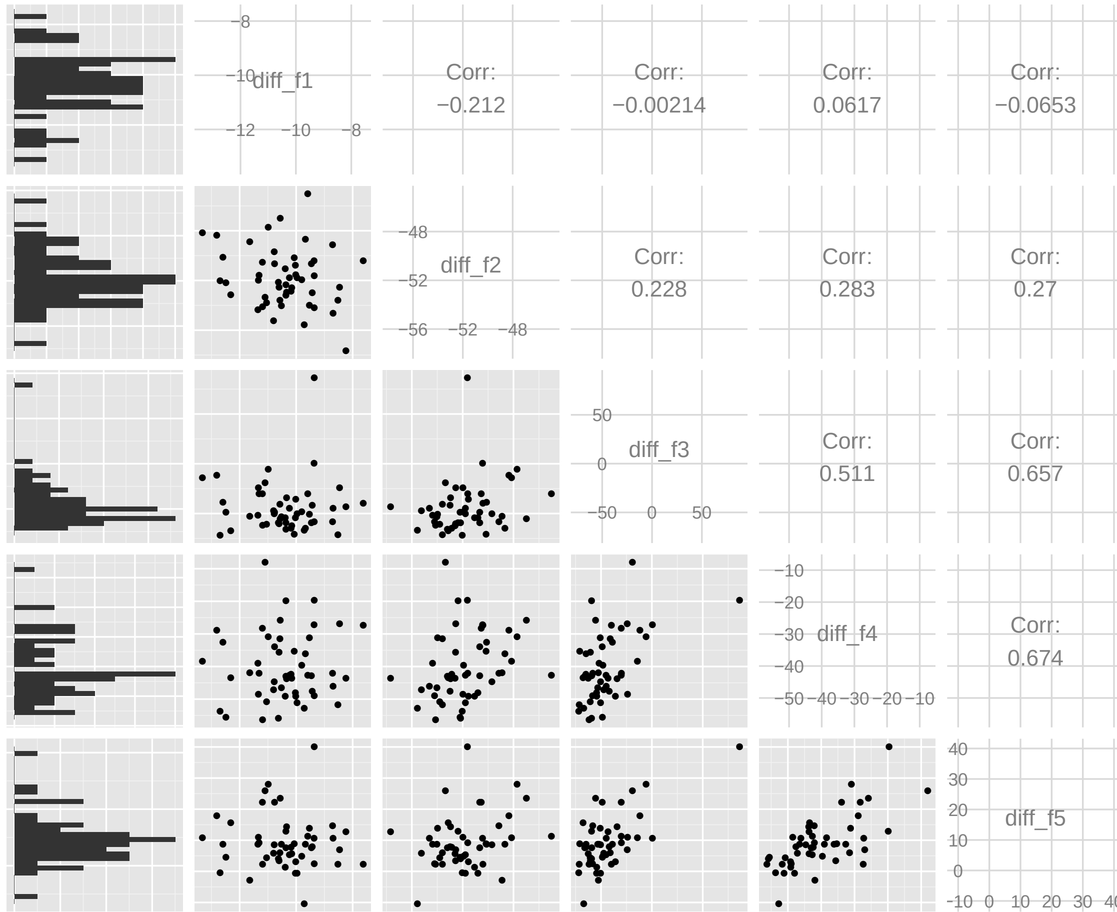

Figure 7.3: Pairwise scatter plots of % change in objective values for $R_{high} - D_{high} - S_{three}$ problem instances.

The effect of model components

In addition to evaluating the improvement resulting from our model when compared to the constructive heuristic, we also wish to determine to what extent the components of our model are required in order to achieve such improvements. In particular, we are interested in answering the following questions -

1. Is it beneficial to include hospital allocation decisions within the model?
2. Is it beneficial to include the hospital allocation term, f_3, in the objective function?
3. Is it beneficial to include task sequencing decisions within the model?

In order to answer these questions, a series of modified models were constructed. Letting M^4_{off} denote the full model described in Chapter 4, in M^3_{off} we remove the hospital allocation neighborhood from the Variable Neighbourhood Descent solution method. As a result, the iterative search procedure will not have any control over hospital allocation decisions. Rather, these decisions are made using the relevant component of the constructive heuristic. As the constructive heuristic is designed to mimic the decision making procedure which would be used in reality, model M^3_{off} corresponds to one which suggests an allocation of tasks to responders, and a sequencing for those tasks, but does not suggest to which hospital casualties should be taken.

Model M^2_{off} is identical to M^3_{off} except in the objective function, where the term associated with hospital allocation, f_3, is omitted. This modification implies that not only is model M^2_{off} incapable of offering suggested decisions regarding the hospital allocations of casualties, it is also incapable of evaluating the impact of decisions made in this respect externally (that is, decisions made through the constructive heuristic).

Finally, in model M^1_{off} we remove one further element, namely the neighborhood used to alter tasks sequencing. As in the case of removal of the hospital allocation neighborhood, the decisions are instead made using the relevant components of the constructive heuristic. The model M^1_{off} is therefore considered basic, in the sense that it is only capable of suggesting to which responder each task should be assigned, not the order in which they should be undertaken. This model configuration and the others described are summarized in Table 7.7.

Table 7.7: Variants of model M_{off}.

Model	Task assignment	Task sequencing	Hospital allocation	f_3 included
M^1_{off}	search	constructor	constructor	no
M^2_{off}	search	search	constructor	no
M^3_{off}	search	search	constructor	yes
M^4_{off}	search	search	search	yes

The results of this analysis are given in Table 7.8, which gives the average percentage change in each objective function across all problem scenarios along with the associated standard deviation. To illustrate the relation between the key objective functions f_1 and g_2, Figure 7.4 provides a scatter plot of these values as grouped by model variant.

Table 7.8: Average (standard deviation) % change in objective values, across all problem instances, for each offline model variant.

	f_1	f_2	f_3	f_4	f_5	g_2
M^1_{off}	-0.8 (1.4)	-8.9 (9.4)	69.0 (78.5)	15.8 (58.2)	13.8 (16.7)	-8.7 (9.4)
M^2_{off}	-9.7 (5.8)	-64.4 (10.5)	27.4 (77.5)	-36.3 (27.6)	7.6 (15.7)	-64.4 (10.4)
M^3_{off}	-9.8 (5.9)	-52.4 (19.2)	-13.5 (74.9)	1.5 (43.2)	21.7 (20.1)	-51.7 (20.2)
M^4_{off}	-9.3 (5.7)	-68.9 (9.1)	-72.5 (21.1)	-59.4 (23.8)	-10.4 (14.3)	-69.1 (9.2)

The results described in Table 7.8 and illustrated in Figure 7.4 show some interesting behaviour of the offline model variants. When looking at the objective of

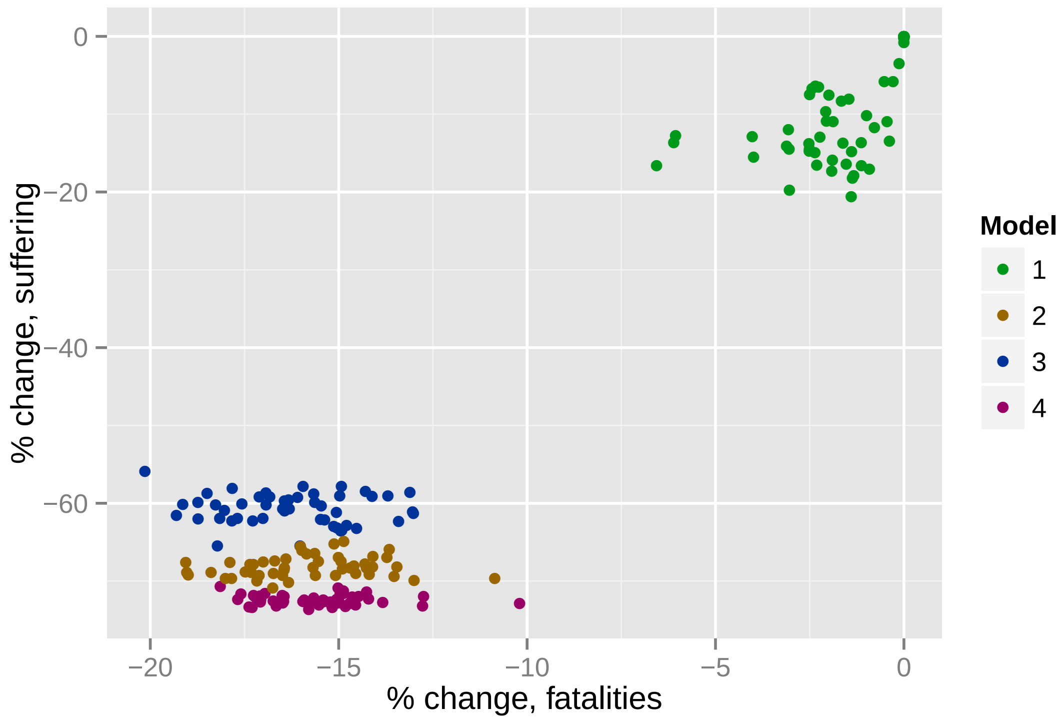

Figure 7.4: Scatter plot illustrating performance of each model variant in terms of % change in objectives f_1 and g_2, for problem scenario $R_{med}, D_{med}, S_{two}$.

fatalities, f_1, we observe a large difference between model M^1_{off} and the other three models, where M^1_{off} is considerably worse. Within these other three model variants, however, there is little to distinguish their performance in this objective. When considering the objective of suffering, g_2, a clearer distinction emerges. Again, there is a large difference in performance between model M^1_{off} and all others, with M^1_{off} obtaining an average improvement of -8.7% while the best performing model (with respect to this objective), M^4_{off}, obtains -69.1%. The next best model variant is M^2_{off} with -64.4%, followed by M^3_{off} with -51.7%.

The observation that the variant M^2_{off} outperforms M^3_{off} is of interest. From the definitions of these model variants, this result suggests that allowing the model to use objective f_3 during its search process will actually lead to *worse* performance in terms of objective g_2 *unless* the model also has control over the hospital allocation decisions, as in the case of model M^4_{off}.

These results also reflect the experiments conducted into the performance of different configurations of the local search algorithm described in Section 5.5. There, it was shown that the inclusion of a neighbourhood allowing for alterations to the allocation of casualties to hospitals lead to improved performance in terms of fatalities, but worse performance in terms of suffering.

The effect of problem characteristics

The effect of problem characteristics on the performance of the offline model is of interest. It is feasible that changes in the numbers of responders, the level of dependency in tasks or the number of sites across which the MCI is distributed may affect the utility of the model. To assess to what extent this is true, a linear model with categorical coefficients representing the levels of each problem characteristic was fitted to the data for each objective function. These models are summarized in Table 7.9.

Table 7.9: Fitted linear regression models for each objective value for model M_{off} in static problems.

	Dependent variable:					
	f_1	f_2	f_3	f_4	f_5	g_2
	(1)	(2)	(3)	(4)	(5)	(6)
R_{med}	0.716	4.294	3.219	2.715	−0.657	4.245
	(0.139)	(0.148)	(1.125)	(1.163)	(0.607)	(0.159)
R_{high}	2.363	8.717	5.001	11.826	1.064	8.677
	(0.139)	(0.148)	(1.125)	(1.163)	(0.607)	(0.159)
D_{med}	−12.270	4.149	4.820	5.085	3.971	4.111
	(0.139)	(0.148)	(1.125)	(1.163)	(0.607)	(0.159)
D_{high}	−8.530	18.755	26.595	34.285	23.637	18.760
	(0.139)	(0.148)	(1.125)	(1.163)	(0.607)	(0.159)
S_{two}	0.847	2.601	11.310	7.763	8.845	2.738
	(0.139)	(0.148)	(1.125)	(1.163)	(0.607)	(0.159)
S_{three}	1.826	2.253	3.569	3.464	6.367	2.262
	(0.139)	(0.148)	(1.125)	(1.163)	(0.607)	(0.159)
Constant	−4.279	−82.518	−90.710	−81.076	−24.840	−82.649
	(0.150)	(0.160)	(1.215)	(1.256)	(0.655)	(0.172)
R^2	0.866	0.941	0.362	0.466	0.595	0.933
Adjusted R^2	0.866	0.941	0.359	0.463	0.594	0.933

Studying the coefficient estimates of the linear models summarized in Table 7.9 provides some clear insight into how variations in problem characteristics affect the utility of the offline model. Consider first the case of variations in the number of responders, as controlled by the levels R_{low}, R_{med} and R_{high}. The models have taken R_{low} as the baseline value, and so the coefficients for R_{med} and R_{high} should be interpreted relative to that. For example, the model relating to objective f_1 estimates the change in f_1 to be 2.363 percentage points worse when the level of responders is set to R_{high} than when it is set to R_{low}. Note that this is not implying that the offline model leads to an increase in fatalities, as the coefficient of the constant is estimated to be -4.279.

In general, we see that in all objective functions an increase in the level of responders leads to a positive increase in the change in each objective value. The only exception to this is in terms of the makespan objective f_5, where the level R_{med} leads to slightly better performance than R_{low}, both of which lead to significantly better performance than R_{high}. However, comparing the estimated coefficient associated with R_{med}, which takes the value -0.657, with the associated standard error, 0.607, implies that the difference between R_{med} and R_{low} is not statistically significant.

The fact that increasing the responder level generally leads to worse performance of the optimisation model may seem counter-intuitive, as we might expect *better* performance as responders are added. However, it is important to remember that the coefficients are associated with *relative change*, not absolute values. From this perspective, these results are not surprising. They suggest that an increase in the number of responders leads to a problem which is easier to 'solve', with the implication that the initial solution generated by the constructive heuristic is in fact of quite high quality and is therefore not as easy to improve upon through application of the iterative search procedure.

Turning our attention to the levels of dependency existing between tasks, a similar relationship is found. Similar to the case of increases in the level of responders, increases in the level of dependency also leads to a positive increase in the change of objectives values, with one exception: the fatalities objective, f_1. In this case, the dependency level which results in the best change in objective value is D_{med}. This is followed by D_{high}, with D_{low} resulting in the worst change in objective value. This irregular behaviour demonstrates the sensitivity of the utility of the optimisation model to the level of dependency which exists between tasks.

Finally, we consider the effect on offline model utility of changes in the number of incident sites which comprise the MCI. We emphasize here that, as explained in Section 7.2, an increase in the number of sites *does not* imply an increase in the number of casualties; rather, the same number of casualties are used throughout but will be distributed across more incident sites. Considering the fatalities objective, f_1, the coefficients of the corresponding linear model show that as the number of sites increases from S_{one} to S_{three} the utility of the offline model decreases. In terms of suffering, g_2, model utility decreases as we move from S_{one} to S_{two}, but no further (statistically significant) decrease is observed as we move from S_{two} to S_{three}.

In general, the presented analyses demonstrate that the model M_{off} does provide improved performance in static problems, in comparison with the alternative constructive solution method. The dependency of this performance on components of the model has been investigated, as has the effect of different problem characteristics.

7.3.2 Real-time model M_{rt}

Having demonstrated the effect of the offline model M_{off} in the case of static problems, we will go on to apply the online model M_{on} in Section 7.3.3. Prior to carrying out this analysis, though, it is of interest to briefly consider the real-time nature of the problem and the impact of this on the utility of the offline model. Specifically, we wish to acknowledge the fact that any time spent running the optimisation model results in a delay to the response operation beginning.

To evaluate the impact of this aspect of the problem, a number of computational experiments were performed. The model M_{off} was employed, with the search time, denoted by x, taking the values of 1, 2, 3, 4 or 5 minutes. Following the termination of the search process, the schedule was shifted by x minutes to the right and re-evaluated in order to encapsulate the effect of delaying action for x minutes in order to obtain an optimized schedule.

The resulting distributions of objectives f_1 and g_2 are shown in the box plots of Figure 7.5 and Figure 7.6 respectively. It can be seen that in the case of f_1, any benefit brought through the optimization process is not enough to counteract the penalty of delaying action for any value of x considered. While this is not the case for g_2, which shows moderate improvement for all values of x, the lexicographic ordering of these two objectives (as discussed in Section 4.6) enable us to conclude that, in an offline system, optimization will in fact lead to worse performance than would be obtained through using the constructive heuristic.

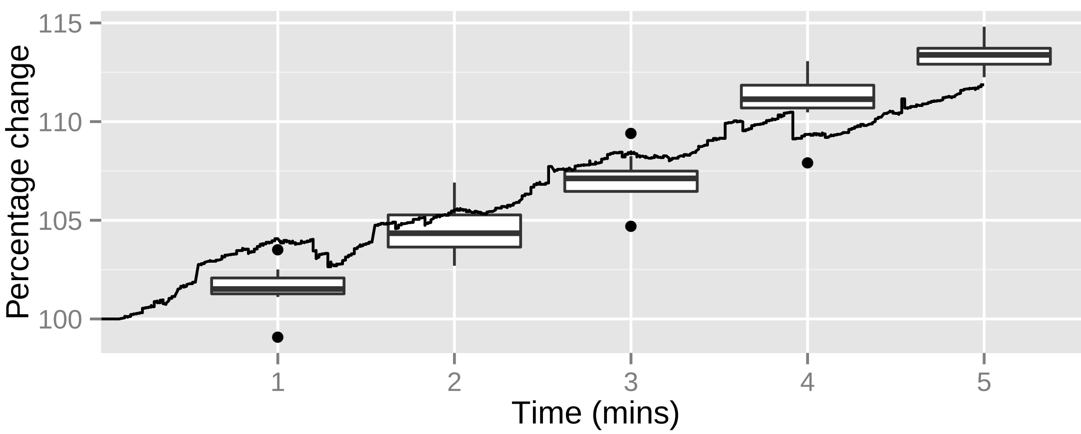

Figure 7.5: Percentage change in f_1 as search time increases (box plots). Solid line shows the objective trajectory of a single run using the real-time model M_{rt}.

In order to explore this behaviour further, a model variant which lies between the offline and online models, which we refer to as the *real-time* model M_{rt}, was defined. Whereas the online model M_{on} was developed to be capable of issuing tasks in a schedule in a gradual manner, ensuring responders are busy while continuing to search the remaining regions of the solution space, the real-time model M_{rt} issues the entire schedule at a single time point, as in the offline case. However, M_{rt} differs from M_{off} in that it continuously updates its parameters to ensure that the earliest point at which a task can start is set to the amount of time that has elapsed up to that point. As such, the optimization process is aware of the increasing penalty being accrued due to the delay in issuing instruction, as it is built into the evaluation process.

The resulting progress of objective values over the course of a single five minute run of this system are given as lines in Figure 7.5 and Figure 7.6. It can be seen that

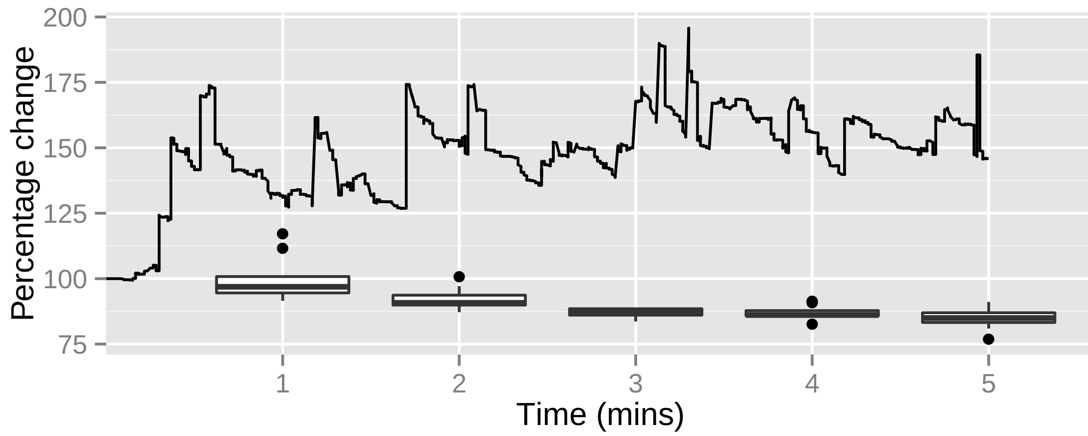

Figure 7.6: Percentage change in g_2 as search time increases (box plots). Solid line shows the objective trajectory of a single run using the real-time model M_{rt}.

the changes in f_1 map to the values observed previously, i.e. increasing at a steady rate over time. The progress of g_2, however, is erratic with no discernible trend. We conclude that optimizing in real-time, given the static problem and solution method described, will not deliver any improvement over the constructive heuristic.

7.3.3 Online model M_{on}

As discussed in Section 7.2.7, static problems are not the primary motivation for the development of the online model M_{on}. Rather, the motivation stems from an inability of the offline model to be used in the more realistic dynamic problems. The principle analysis of the utility of the online model will therefore be reported in Sections 7.4.1 and 7.5, where these dynamic problems will be considered. It is nonetheless of interest to report a brief analysis of performance of M_{on} in the case of static problems, as it is possible that even in these situations the online model could out-perform the offline model. We report the results of such an analysis in this subsection.

We restrict our attention to a subset of the problems outlined in Section 7.2, in order to focus on new aspects of the problem as opposed to those already explored. Specifically, we consider problems of the type $R_{high} - D_{high} - S_{three}$, and therefore do not further explore the effect of changing responder or dependency levels, or the number of sites, on the utility of the model. As a result, any resulting conclusions will be restricted to this specific problem type. This specific problem was selected from the set of 27 potential problems as it is, in a computational sense, the most challenging problem, with the high responder and dependency levels leading to the largest solution space. The results, in terms of the percentage change observed in each objective function, are given in Table 7.10. A scatter plot displaying the results of the objectives f_1 and g_2 is also provided in Figure 7.7.

The results show that, in this specific problem type, the online model outper-

Table 7.10: Average (standard deviation) % change in objective values, in problems of type $R_{high} - D_{high} - S_{three}$, for offline and online models.

	f_1	f_2	f_3	f_4	f_5	g_2
M_{off}	-10.4 (1.2)	-51.8 (2.4)	-44.3 (25.7)	-40.3 (10.5)	9.2 (8.6)	-51.8 (2.5)
M_{on}	-7.2 (1.0)	-54.1 (1.6)	-77.2 (8.7)	-63.8 (4.3)	-23.6 (6.8)	-54.5 (1.7)

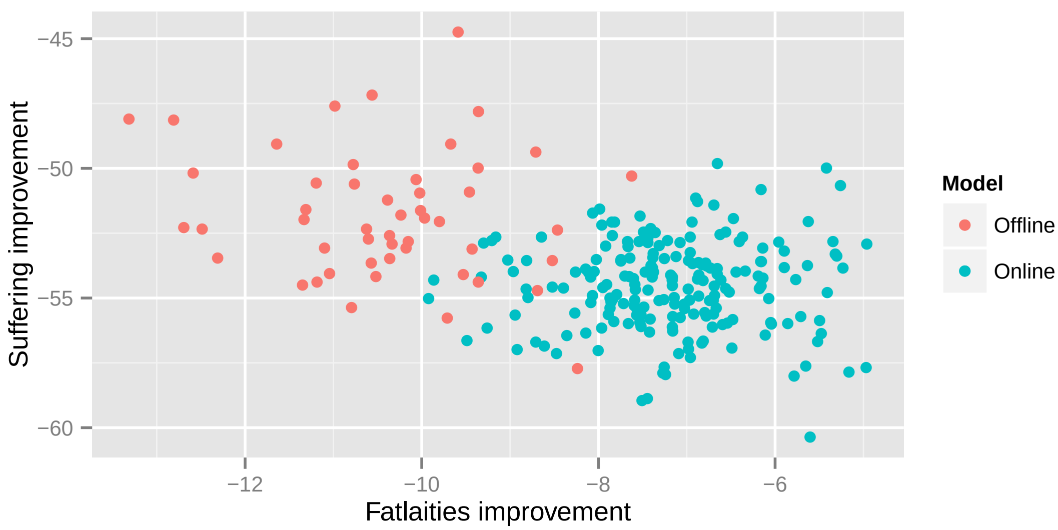

Figure 7.7: Scatter plot of % improvement in f_1 and g_2 for offline and online models in problems of type $R_{high} - D_{high} - S_{three}$.

forms the offline model in all objectives *except* objective f_1, fatalities. Given the lexicographic ordering of objective functions introduced in Section 4.6.6, which stated that preference should always be given to reductions in fatalities over other measures, these results suggest that the offline model should be preferred in the case of static problems. However, an important point should be emphasized in this regard. The offline model operates under a number of assumptions, one of which is that the time taken to search for solutions does not impact the starting time of the response operation. Effectively, the offline model assumes it has an 5 minute *head start*. As discussed in Section 7.3.2, this is clearly not the case in reality. Moreover, it was shown that removing this assumption has a significant negative impact upon the utility of the model. As the online model M_{on} does not make any such assumption, the comparison reported in Table 7.10 is not, strictly speaking, fair. However, the results remain of interest. They show that *even in static problems* (where the benefit of the online model is expected to be least felt), and *even when the offline model is given an advantage* (specifically, 5 minutes 'free' optimisation time), the online model still outperforms on all but one objectives whilst providing competitive performance in the remaining objective.

7.3.4 Discussion

In concluding this section on static problems, we note a number of important points that have emerged. First and foremost, it was shown in Section 7.3.1 that the offline model M_{off}, employing the iterative search solution method of Chapter 5, did indeed lead to improved solutions in comparison to those obtained using the constructive heuristic method. This represented the first key point in assessing the utility of the work proposed in this thesis.

A critical view of the design of model M_{off} was then taken, with experiments undertaken to examine to what extent the features of the model are essential to ensuring this utility. The results confirmed that the inclusion of hospital allocation within the decision model, with associated objective measures to assess the quality of such decisions, did indeed contribute to the observed performance. This result should be placed in the context of the discussion presented in Section 3.3.2, where it was shown that related research has often omitted such features in developing optimisation models. The analysis of model variants also shed light on an interesting characteristic regarding hospital allocation decisions, namely that if the model is not given control over such decisions it should not, in terms of maximizing performance, be allowed to measure the quality of them.

The problems used throughout the experimental analysis, as set out in Section 7.2, involved varying levels of a number of key characteristics. The effect of these characteristics on the utility of the offline model was considered in Section 7.3.1, in order that we may understand to what types of problems the proposed model would be particularly well-suited (and, indeed, to which problems the model would not be expected to perform well on). The results of this analysis provided a valuable insight into the sensitivity of the model to the type of problem, with significant effects observed when varying the number of responders, the level of dependency within tasks and the number of sites which the MCI is spread over.

The results of Section 7.3.2 are important, as they arise from a first investigation into the robustness of a key assumption in the offline model: that the time needed by the model to locate a solution using the search procedure is somehow insignificant, with no impact on the response operation. The results demonstrate the danger of making this assumption, as it was shown that if the response operation is held back to reflect the time spent searching for solutions, the model will lead to worse performance than would be obtained using the constructive heuristic and benefiting from its immediate nature. Moreover, it was shown that this behaviour is not dependent upon the length of time allowed for optimisation.

Finally, the application of the online model M_{on} to static problems led to some valuable insights. Given the loss of performance observed when applying the model in real-time, it was not clear if the benefits afforded by online optimisation would be enough to overcome this penalty. The results demonstrate that this is, in fact, the case, and that it is highly likely that the online model will lead to improvements over the offline model in these problem types.

Given this last result, we now turn our attention to the more realistic problems, those which were classified as partially and fully dynamic in Section 7.2.7.

7.4 Partially dynamic problems

Before looking in detail at the utility of the optimisation model in partially dynamic problems, a general point regarding the measurement of performance should be made. In the preceding analysis of Section 7.3, performance was measured in terms of the percentage change in objective functions when comparing the initial solution, as generated using the constructive heuristic, with the final solution generated by the optimisation model. This enabled us to understand the utility of the model in terms of the improvement it offers over the standard solution methodology in each single experimental run. That is, each experiment produced a single data point describing performance.

In the case of dynamic problems, this is no longer a sensible measure to compute. Due to the fact that the many parameters of the model are initially available only as estimates, the value of the solution generated by the constructive heuristic is the *perceived* value. In order to measure the *actual* value of that proposed solution, one would have to run the simulation model for the duration of the response operation while following the proposed schedule. The optimisation methodology of the online model M_{on} is designed to run over the course of this simulation, and so the objective values reported on completion of the response operation are the actual values. Accordingly, instead of comparing the constructive and optimisation methodologies on each single generated problem instance, we will instead compare the performance of each on average, over a number of generated problem instances.

7.4.1 Online model M_{on}

As discussed in Section 7.2.7, partially dynamic problems exhibit dynamic and uncertain behaviour in the objective space. This behaviour arises due to a number of factors, including casualty health, task durations and transportation times. In Section 7.4.2 we will consider all these factors at once. Here, we focus only on the challenges brought about by uncertainty in transportation times, and how these may be addressed through the use of routing policies introduced in Section 6.2.4 (namely static routing (SR), centralised adaptive routing (CAR), and autonomous individual or collective routing (AIAR/ACAR)). By focusing on this issue, we aim to fully evaluate the routing policies introduced in Section 6.2.4 and explore how their performance depends on relevant parameter levels.

Experimental design

A number of factors exist which may affect the performance of the various routing policies. In particular, we have:

1. The parameter describing the disruption to the transport network, λ (see Section 6.2.4);

2. The parameters describing the ability of responders to autonomously search for high quality routes within the disrupted transport network, α (the average

improvement rate) and β^2 (the variation around this average improvement, see Section 6.2.4);

3. The number of sites which comprise the MCI, S_{one}, S_{two} and S_{three} (see Section 7.2.6).

Our principal goal in the analysis that follows is to compare the performance of autonomous routing policies $AIAR$ and $ACAR$ with a baseline policy SR and an alternative CAR. In particular, we aim to determine which values α, β^2 will lead to autonomous policies outperforming centralized policies. That is, how well must responders be able to route themselves to justify removing the routing decision from the optimisation model? Answering this question will inform the design of future optimisation models for MCI response. The variation of both disruption level and the number of sites is important to explore how this relationship varies with underlying problem characteristics.

Given the several sources of uncertainty within the model, a Monte Carlo approach is employed. That is, for any given point in the experiment space n instances of the problem are generated and solved in order to estimate the distribution of the final objective values. The points within the experiment space are defined through a standard factorial design based on the following factors:

- Routing policy - {SR, CAR, AIAR, ACAR}
- Road disruption, λ - {2, 1, 0.5, 0.25}
- Number of incident sites - {1, 2, 3}

This gives a total of $4 \times 4 \times 3 = 48$ experimental design points, where each point corresponds to a unique combination of the three factors. For the purposes of the initial evaluation, the autonomous improvement parameter pertaining to the routing policies was set as $\alpha = 0.95$ throughout, corresponding to an average improvement in routing of 5% with each journey.

Following this initial set of experiments, a second set was designed to focus on the effect of altering the α parameter governing the rate of improvement in the autonomous routing policies. The points within the experiment space are defined through another standard factorial design based on the following factors:

- Routing policy - {AIAR, ACAR}
- Road disruption - {2, 1, 0.5, 0.25}
- Number of incident sites, λ - {1, 2, 3}
- Autonomous improvement, α - {0.9, 0.8}

This gives a total of $2 \times 4 \times 3 \times 2 = 48$ experimental design points. These results were combined with those of the previous set of experiments, considering only the routing policies AIAR and ACAR. This provided a total of three levels of α in total - 0.95, 0.9 and 0.8.

Results

The results of the first set of experiments are summarized in Tables 7.11 and 7.12, which focus on the objectives of fatalities (f_1) and suffering (g_2) respectively. Each table provides the average objective value obtained across all experiments for a given pair of road network disruption levels (denoted by λ) and routing strategy. In order to indicate the level of variability around these average values, the standard deviation is also provided.

Table 7.11: Average (standard deviation) values of f_1 of model M_{on} under each routing strategy, for varying levels of disruption.

	$\lambda = 0.25$	$\lambda = 0.5$	$\lambda = 1$	$\lambda = 2$
SR	67.3 (2.9)	56.1 (1.8)	49.4 (1.7)	45.6 (1.8)
CAR	67.8 (2.4)	56.2 (2.1)	49.1 (1.8)	45.6 (1.9)
$AIAR$	80.1 (2.7)	65.6 (1.9)	55.6 (1.8)	49.7 (1.6)
$ACAR$	68.1 (2.3)	56.0 (1.6)	49.1 (1.8)	45.5 (1.8)

As can be seen in Table 7.11, the routing strategies SR (static routing) and CAR (central adaptive routing) lead to similar performance in terms of the fatalities objective, f_1. This is consistent across each of the four levels of road network disruption considered. Moreover, with the rate of improvement set to $\alpha = 0.95$, similar performance is also observed when using the routing strategy $ACAR$ (autonomous collective adaptive routing). As would be expected, the strategy $AIAR$ leads to worse performance in comparison to $ACAR$. This is due to the fact that in $ACAR$ the improvements in routing are designed to reflect a sharing of information amongst responders, whereas in $AIAR$ each responder works independently to improve their routing choice.

Table 7.12: Average (standard deviation) values of g_2 of model M_{on} under each routing strategy, for varying levels of disruption.

	$\lambda = 0.25$	$\lambda = 0.5$	$\lambda = 1$	$\lambda = 2$
SR	19750 (1251)	14840 (889)	12048 (630)	10670 (581)
CAR	19589 (1393)	14904 (983)	12035 (673)	10608 (521)
$AIAR$	23263 (1679)	17050 (962)	13305 (646)	11153 (603)
$ACAR$	19164 (1243)	14517 (803)	11832 (636)	10584 (591)

Table 7.12 presents corresponding summary statistics in relation to the suffering objective, g_2. It is clear that the relation between routing strategies broadly mirrors that of the f_1 case, with routing strategies SR and CAR leading to similar performance, much better than that obtained through strategy $AIAR$. However, in this case the routing strategy $ACAR$ appears to offer some benefits over strategies SR and CAR. This benefit diminishes as the extent of disruption to the road network decreases.

The results presented in Tables 7.11 and 7.12 are broken down by routing strategy and level of network disruption, but averaged over the number of sites comprising the MCI. In order to assess the extent to which this factor influences performance, a linear regression model was fitted to the data. Specifically, two models were constructed, one for each of the objectives of interest. Included as predictor variables were the number of sites, the level of disruption and the routing strategy. Each variable was taken to have a qualitative, or categorical nature. This is the natural representation for the routing strategy, but the quantitative nature of both the number of sites (which can take values 1, 2 or 3) and the level of disruption (which can take values 0.25, 0.5, 1 or 2) suggests that a numerical representation may be the natural choice for these variables. However, such a representation would correspond to assuming a linear relationship between the predictor and the dependent variable. A qualitative representation does not require this assumption, and therefore provides greater flexibility.

The resulting regression models are described in Table 7.13, where estimates of the effects of each predictor variable are given along with their standard error. The reference values are taken as S_{one}, $\lambda_{0.25}$ and $ACAR$ for the number of sites, road network disruption and routing strategy respectively. Asterisks denote to what extent the effects are judged to be statistically significant (i.e., the extent to which the observed effect is unlikely to be due to chance alone).

As shown by the high value of the adjusted R^2's, both models fit the data well and as such no further terms (such as quadratic or interaction terms) were added. The models confirm what was suggested in the summaries of the data given in Tables 7.11 and 7.12, as they show that the choice of routing strategy has no significant effect upon the fatalities objective (discounting the choice of $AIAR$), but does have a significant effect upon the suffering objective.

This analysis used a value of $\alpha = 0.95$ for the rate of improvements in routing choice in the autonomous routing strategies, $AIAR$ and $ACAR$. As described in Section 6.2.4, this means that every time a responder makes a specific journey, the length of the route chosen will be (on average) 0.95 times the length of the last route chosen for that same journey. Using this parameter value, the routing strategy $ACAR$ was shown to lead to comparable performance to the centralized routing strategies, SR and CAR. It remains to be seen to what extent lower values of α will lead to the strategy $ACAR$ out-performing SR and CAR.

In order to investigate this, the second set of experiments described at the beginning of this sub-section were run. The resulting data set was combined with the previous data, together allowing for comparisons in the performance of strategy $ACAR$ for three values of α, 0.95, 0.9 and 0.8. The results are summarized in Tables 7.14 and 7.15, which show the average (standard deviation) values of objectives f_1 and g_2, respectively, factored by the level of road network disruption and value of α.

Table 7.14 shows that, in terms of objective f_1, altering the parameter α in the manner described does not lead to any significant change in the performance of the optimisation model when using the routing strategy $ACAR$. However, some difference is observed in objective g_2. The benefit afforded through the routing strategy $ACAR$ observed when $\alpha = 0.95$ (see Table 7.12) is further enhanced as the learning rate is

Table 7.13: Fitted linear regression models for objectives f_1 and g_2 for model M_{on} in partially dynamic problems, examining routing strategies.

	Dependent variable:	
	f_1	g_2
S_{two}	0.154	795.390***
	(0.128)	(48.273)
S_{three}	1.137***	1,457.455***
	(0.128)	(48.280)
$\lambda_{0.5}$	−12.389***	−5,141.224***
	(0.148)	(55.524)
λ_1	−20.058***	−8,155.442***
	(0.149)	(56.181)
λ_2	−24.261***	−9,692.291***
	(0.148)	(55.729)
$AIAR$	8.109***	2,180.700***
	(0.148)	(55.748)
CAR	−0.025	249.071***
	(0.148)	(55.775)
SR	−0.052	321.971***
	(0.149)	(55.939)
Constant	68.399***	19,013.970***
	(0.157)	(59.016)
R^2	0.945	0.948
Adjusted R^2	0.944	0.948
Note:	*p<0.1; **p<0.05; ***p<0.01	

Table 7.14: Average (standard deviation) values of f_1 of model M_{on} under autonomous routing strategies, for varying levels of disruption.

	$\lambda = 0.25$	$\lambda = 0.5$	$\lambda = 1$	$\lambda = 2$
$\alpha = 0.95$	68.1 (2.3)	56.0 (1.6)	49.1 (1.8)	45.5 (1.8)
$\alpha = 0.9$	68.1 (2.7)	56.7 (2.1)	49.0 (1.9)	45.5 (1.6)
$\alpha = 0.8$	67.3 (2.1)	56.2 (1.9)	49.2 (1.8)	45.1 (1.8)
SR	67.3 (2.9)	56.1 (1.8)	49.4 (1.7)	45.6 (1.8)

Table 7.15: Average (standard deviation) values of g_2 of model M_{on} under autonomous routing strategies, for varying levels of disruption.

	$\lambda = 0.25$	$\lambda = 0.5$	$\lambda = 1$	$\lambda = 2$
$\alpha = 0.95$	19163 (1243)	14517 (803)	11833 (636)	10584 (591)
$\alpha = 0.9$	18949 (1194)	14396 (686)	11779 (627)	10639 (559)
$\alpha = 0.8$	18818 (1120)	14383 (809)	11758 (725)	10443 (596)
SR	19750 (1251)	14840 (889)	12048 (630)	10670 (581)

improved. In the most extreme case considered, where $\alpha = 0.8$ and $\lambda = 0.25$, the strategy $ACAR$ leads to 4.7% better performance in terms of g_2 in comparison to strategy SR.

Similarly to the previous analysis conducted on the first set of experiments, linear regression models were also fit to the data allowing for different values of α in order to further quantify its effect. Considering only the routing strategy $ACAR$, models were fitted to both fatalities and suffering objective values using the number of sites, the level of disruption and the rate of autonomous improvement, α. As in the models described in Table 7.13, number of sites and level of disruption were coded as qualitative variables. The rate of autonomous improvement was coded as a quantitative variable. The resulting models are described in Table 7.16.

Again, the models are judged to fit the data well (see the high values of adjusted R^2's), and so no further terms were considered. Focusing on the effects of the rate of autonomous improvement, α, we see that no statistically significant effect on the fatalities objective is observed. However, the effect on the suffering objective is judged to be significant ($p < 0.01$). While *statistically* significant, it should be emphasized that the effect size itself is modest. Estimated to be 1169, this corresponds to a change in suffering of 117 units for any 0.1 change in the rate of autonomous improvement.

Discussion

The routing strategies described in Section 6.2.4 were noted to have different effects on both the length of the resulting routes and the predictability of the associated travel times. The strategy of static routing (SR) was noted to favour the ability to accurately predict travel times at the expense of finding shorter routes. This is

Table 7.16: Fitted linear regression models for objectives f_1 and g_2 for model M_{on} in partially dynamic problems, examining autonomous improvement rate.

	Dependent variable:	
	f_1	g_2
S_{two}	-0.058	763.799***
	(0.167)	(55.259)
S_{three}	1.042***	1,403.707***
	(0.167)	(55.044)
$\lambda_{0.5}$	-11.832***	$-4{,}613.958$***
	(0.192)	(63.389)
λ_1	-18.827***	$-7{,}262.311$***
	(0.193)	(63.815)
λ_2	-22.507***	$-8{,}497.965$***
	(0.192)	(63.471)
α	1.163	1,169.000***
	(1.256)	(414.328)
Constant	66.548***	17,279.780***
	(1.166)	(384.609)
R^2	0.955	0.967
Adjusted R^2	0.955	0.966
Note:	*p<0.1; **p<0.05; ***p<0.01	

achieved through consistently using a single route for each journey, namely that which would be expected to be the shortest using baseline data describing the transport network (i.e. not considering the effects of disruption). By using the same route for each journey, the optimisation model is capable of revising its estimate of the travel time on that route as the relevant data is collected. The strategies of CAR, $AIAR$ and $ACAR$, on the other hand, sacrifice this ability to 'learn' the travel times of a specific route, as the routes taken for each journey are allowed to change each time in the hope of finding shorter routes.

The analysis presented in this subsection demonstrate that the choice of routing strategy does not have any significant impact upon the utility of the model, when assessing performance through the fatalities objective. By this measure, the strategies of SR, CAR and $ACAR$ all lead to similar performance. In terms of the objective of suffering, however, some differences can be noted. In particular, autonomous routing can lead to improved performance in certain situations. Specifically, in scenarios where the level of disruption to the network is large, where responders are assumed able to share knowledge regarding routing (that is, where the routing strategy of $ACAR$ as opposed to $AIAR$ is adopted), and where the rate of autonomous improvement is set to $\alpha = 0.8$, autonomous routing can lead to improved performance. This result is of interest, particularly when viewed in the context of related research into decision support for large scale emergency response involving routing decisions. As was discussed in the review of such related work given in Section 3.5, it is common for decision support to include the specification of which route responders should take when enacting response operations.

7.4.2 Offline model M_{off}

In the subsequent section, problems of a fully dynamic nature will be examined. In particular, the utility of the online model M_{on} will be assessed. It would be natural to also assess the performance of the offline model M_{off} in these problem types, in order that a comparison between the two models could be made. However, as discussed in Section 7.2, it does not make sense to apply the offline model to fully dynamic problems due to the changes in the solution space observed during such problems. The offline model can be applied to partially dynamic problems, however, as will be demonstrated in this sub-section.

Experimental design

In this analysis, problems of the type R_{high}, D_{high} and S_{three} will continue to be employed. By keeping these factors at constant levels, we are free to focus in detail on the effect of other problem characteristics specific to the partially dynamic case. Individual problem instances may vary in the seven dimensions described in Table 7.17. The table describes each factor which may vary, gives the notation which will be used to denote the factor throughout, and describes the range of values which will be considered. For some parameters, namely λ_1 and λ_5, the choice of range is a natural one which follows directly from their definitions. For other parameters, a judgement has been made regarding feasible levels; for example, we consider it unlikely that the communication delay described in Section 6.3 will be larger than five minutes.

Table 7.17: Parameters to be varied in designed experiments for partially dynamic problems.

Parameter	Notation	Range	Description
Triage assessment error	λ_1	(0, 1]	Error in the triage classification process.
Triage frequency	λ_2	[1, 20]	Time between each triage assessment of any given casualty (mins).
Task duration variance	λ_3	(0, 3]	Inherent variation in the durations of all tasks.
Task duration assessment error	λ_4	(0, 2]	Error in the estimation of durations of all tasks.
Task duration confidence	λ_5	[0.1, 0.9]	Level of confidence required that task duration estimates will not be short.
Communication delay	λ_6	[0, 5]	Average wait between a temporal event being recorded and the optimization model being notified (mins).
Road network disruption	λ_7	[0.5, 2]	Extent to which the road transport network is disrupted.

The set of experiments is determined in a manner which fills the space of possible experiment points, with each point being defined by a unique setting of each parameter. The experiments follow a Sobol sequence of 500 points in the 7 dimension experimental design space, as constructed using the R package randtoolbox [38]. Such a design ensures that non-linear relationships between the objective function values and the listed parameters may be detected. This would not be possible if a standard factorial design was used.

Results

The applicability of the static scheduling model to dynamic environments may now be evaluated through employing the simulation routine described in Chapter 6. As it is the offline model M_{off} under study, each experiment involves spending five minutes searching the solution space. At the end of this time the best solution found is issued and the response operation proceeds to follow the corresponding schedule, with the dynamic and uncertain nature of all objective space parameters being simulated as the response operation progresses. By means of comparison, the same problem setup was addressed using the constructive heuristic defined in Chapter 5 which, recall, was designed to replicate how decisions would be made in reality when faced with an evolving problem. As such, the use of the constructive heuristic will have an advantage over the use of model M_{off}, due to its ability to adapt to the observed changes in the environment.

Descriptive statistics of these experiments are provided in Table 7.18. Figure 7.8 shows the joint distribution of objective values as contour plots for both cases, where we label the constructive heuristic method as 'Heur'.

Table 7.18: Descriptive statistics of final objective values across all partially dynamic problem instances, for both offline search and constructor solution methods.

	f_1				g_2			
	Mean	Min	Median	Max	Mean	Min	Median	Max
M_{off} search	20.82	9	21	33	31124	18265	30768	51365
M_{off} constructor	17.63	7	18	29	49808	33172	49295	76954

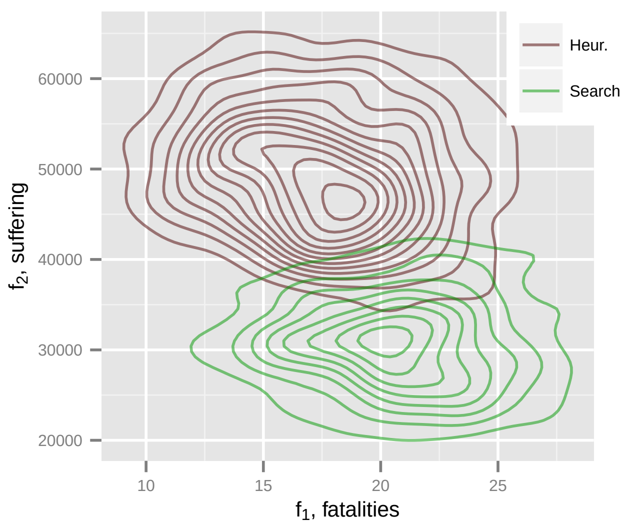

Figure 7.8: Density plots of final objective values across all partially dynamic problem instances, for both offline search and constructor solution methods.

The results show that in applying the model M_{off} to partially dynamic problems, the resulting outcome in terms of fatalities will be high (on average) than would be observed had the constructive solution method been applied to that same problem. The situation is reversed when considering the suffering objective, in which case the model M_{off} outperforms the constructive method. This suggests that the uncertainty and dynamic behaviour contained within these problems has a particularly negative impact on the fatalities objective which, given the lexicographic ordering of objectives, implies that it is inappropriate to apply model M_{off} to these problem types. This result should be contrasted with the analysis of model M_{off} in the case of static problems, where significant improvements in performance were found. These results clearly demonstrate the danger of developing an optimisation model under a set of assumptions which do not hold in reality.

The application of the model M_{off} to partially dynamic problems involves the search for a solution based on perceived parameter values at the start of the response operation. This suggested solution then has perceived objective values. The simulation of the response operation, enacting the solution suggested by the model, then leads to the realisation of the actual objective values of that solution. Contrasting this initial perceived evaluation with the final actual evaluation provides insight into the extent to which the initial value tends to be incorrect. This is demonstrated in Figure 7.9. The scatter plot shows the difference in the initial and final values of fatality and suffering objectives for each of the experimental runs conducted. It is clear that the number of fatalities is consistently *over*estimated, while the level of suffering is consistently *under*estimated.

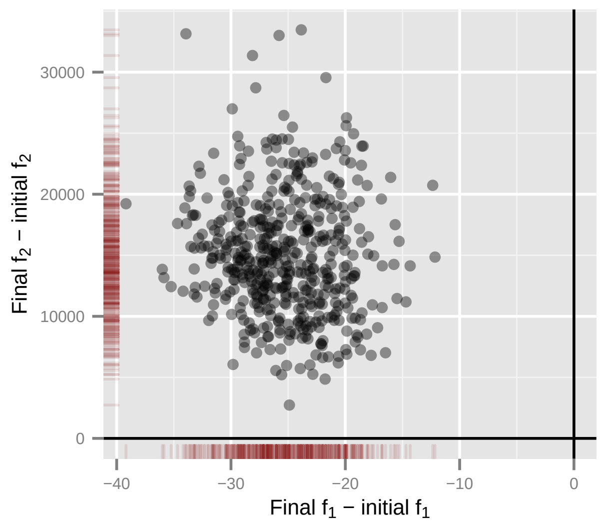

Figure 7.9: Final simulated values minus initial estimated values of the offline search procedure in partially dynamic problems.

It is of interest to consider the performance of the offline model in more detail, and in particular to identify any particular factors (in terms of the parameters listed in Table 7.17) which demonstrate significant relationships with performance. To do so, linear regression models relating each objective measure to the varied problem parameters were fitted by first considering a 'full' model, including potential interaction and higher order terms, and performing a backwards stepwise variable selection procedure. The resulting estimates of coefficients remaining in the model following a backwards stepwise elimination procedure are given, with associated standard errors in Table 7.19.

In order to better visualize the relationships suggested by the regression models described in Table 7.19, a number of Component And Residual (CAR) plots are provided. Each plot gives the residuals corresponding to a certain parameter, adjusted

Table 7.19: Fitted linear regression models for objectives f_1 and g_2 for model M_{off} in partially dynamic problems.

	Dependent variable:	
	f_1	g_2
	(1)	(2)
initial value	0.218***	1.220***
	(0.050)	(0.108)
λ_1		−1,101.085*
		(605.431)
λ_6		663.474***
		(134.435)
λ_7	−6.660***	−26,229.010***
	(2.282)	(2,623.779)
λ_7^2	1.528*	8,141.205***
	(0.903)	(1,038.463)
Constant	16.428***	28,742.410***
	(2.674)	(2,392.497)
R^2	0.151	0.474
Adjusted R^2	0.146	0.468
Note:	*p<0.1; **p<0.05; ***p<0.01	

by the fitted regression relationship. As such, each plot both provides a better intuition regarding the strength of the relationship identified, but also places this in the context of the remaining variation due to both randomness and variation in other parameters. In Figure 7.10, component and residual plots describing the relationships between the final fatalities objective and both the initial estimate of that objective and the level of disruption in the road network are given. These plots are generated by modifying a normal plot of residuals around a given predictor variable through adding the corresponding component of the fitted model. This allows both the fitted linear relationship to be visualised, and also for it to be placed in the context of all other sources of variation. As shown in Table 7.19, these parameters were the only ones deemed to have a statistically significant effect on the fatalities objective.

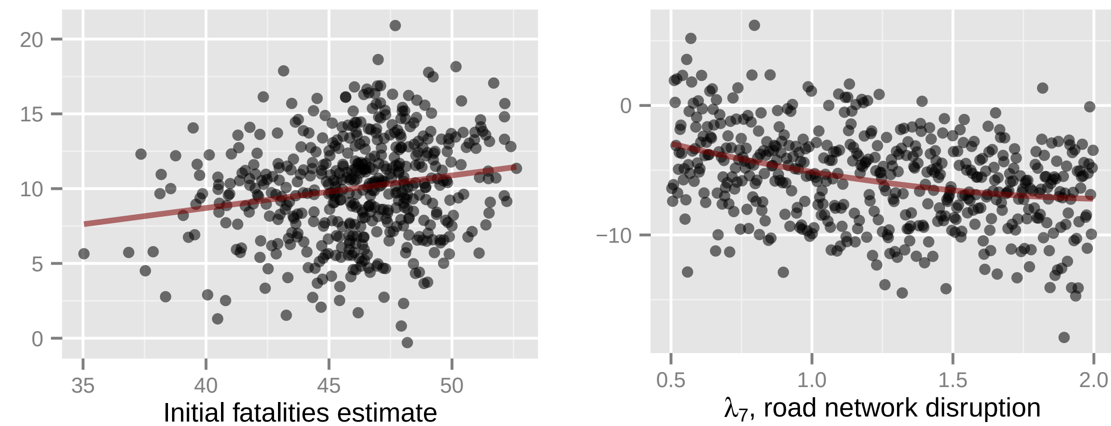

Figure 7.10: Component and residual plots of statistically significant relationships, considering the fatalities objective of model M_{off} applied to partially dynamic problems.

There is a positive relationship between the value of the initial solution generated at the start of model M_{off}'s search procedure and the % improvement attained. This suggests that it is in a sense 'easier' for the model to improve upon a solution of lower quality than a higher quality solution. This is intuitive in the sense that we know there is a lower limit to the objective values attainable, and the closer we are to this limit in the first instance the harder it will be to locate an improving solution.

Similar plots are given in Figure 7.11, in this case describing all parameters which were found to have a statistically significant relationship with the objective of suffering. In addition to the initial objective estimate and the level of road network disruption, the amount of error in triage classification and the length of delays in communication were found to have a significant influence on the final level of suffering. The former effect demonstrates the importance of having accurate information with regards to the health of casualties, and shows that an assumption that all health data is known with complete accuracy could produce misleading conclusions regarding the utility of the model.

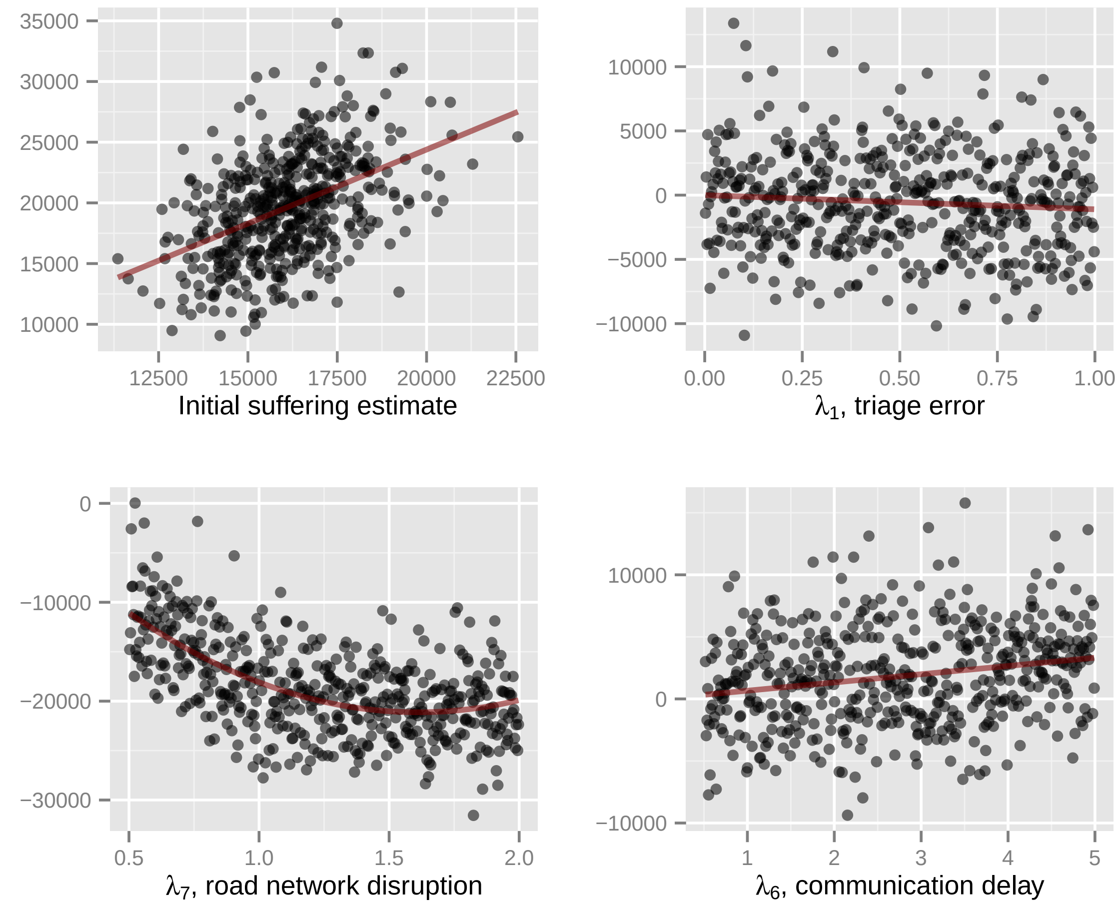

Figure 7.11: Component and residual plots of statistically significant relationships, considering the suffering objective of model M_{off} applied to partially dynamic problems.

7.5 Fully dynamic problems

Having evaluated the performance of both offline and online models in the case of static and partially dynamic problems, we now go on to consider fully dynamic problems. As has been discussed in Section 7.2.7, only the online model M_{on} may be applied to fully dynamic problems. In this section, a detailed analysis of how the utility of this model is affected by various parameters will be conducted with the aim of identifying key features of both the model and the problem which must be carefully considered with designing optimisation models for use in MCI response.

7.5.1 Experimental design

The following experiments employ the online model M_{on} and, by means of comparison, the constructive heuristic solution method. The experimental design builds upon that described in Table 7.17, in that each of the parameters listed there are subjected to variation using the same ranges. In addition to those parameters, which related exclusively to the objective space of the model, the parameters given in Table 7.20 are also subject to variation. These parameters affect the rate at which casualties both enter (through being discovered) and leave (through self-presentation) the modelled environment. As in the case of Section 7.4.2, the experimental design was determined through a Sobol sequence in order to ensure an even coverage of the design space.

Table 7.20: Parameters to be varied in designed experiments for fully dynamic problems.

Parameter	Notation	Range	Description
Casualty discovery rate	λ_8	[0.1, 10]	Average time taken to locate a casualty following an incident (mins).
Casualty self-presentation rate	λ_9	[5, 20]	Average time an eligible casualty will wait at scene before leaving to self-present (mins).

7.5.2 Results

Before considering any relationships which exist between the parameters listed in Tables 7.17 and 7.20, we fist examine the results averaged across all problem instances considered. These are summarised in Table 7.21, which gives descriptive statistics for both the fatalities and suffering objectives arising from application of both the online model M_{on} and the constructive heuristic. Comparing the two methods, we see that M_{on} outperforms the constructive heuristic in terms of both measures. These results allow us to conclude that in the most realistic problems considered, and across a wide range of problem and model parameters, application of the online model M_{on} leads to improved performance in comparison to the simulation of current decision making as represented by the constructive heuristic.

Table 7.21: Descriptive statistics of final objective values across all fully dynamic problem instances, for both M_{on} and constructor solution methods.

	f_1				g_2			
	Mean	Min	Median	Max	Mean	Min	Median	Max
M_{on}	19.49	11	20	28	29008	18920	28784	44789
Constructor	21.50	10	21	31	31663	21793	31329	43168

These results are also given graphically in the form of a contour plot in Figure 7.12. This chart can be contrasted with Figure 7.8, where it was shown that the offline model M_{off} resulted in worse performance in terms of the number of fatalities than when the constructive heuristic was employed.

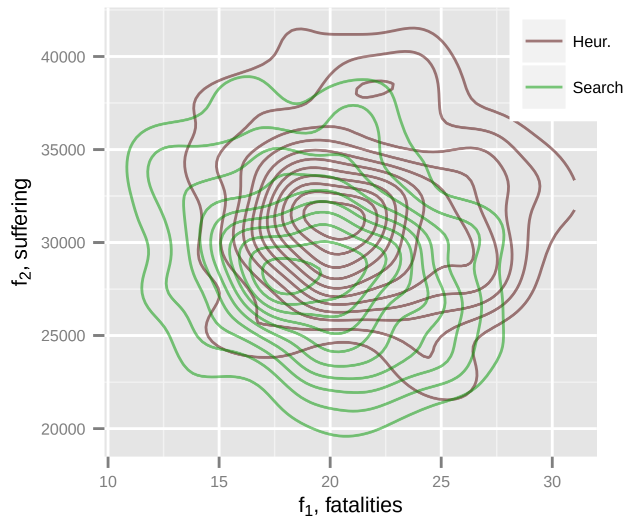

Figure 7.12: Density plots of final objective values across all fully dynamic problem instances, for both search and constructor solution methods.

Constructive heuristic

Before reporting a detailed analysis of the performance of the online model, it is of interest to first examine the case of the constructive heuristic. As this solution method was designed to mimic decision making as it stands, without the use of an optimisation model, this represents a baseline case. Understanding the extent to which performance is affected by the parameters listed in Tables 7.17 and 7.20 will allow for a more informed evaluation of the online model.

As in the case of the evaluation of the offline model in partially dynamic problems, it is informative to construct a regression model of performance, in terms of f_1 and g_2, against the various model and problem parameters which are varied in the experimental design. The results of these models for the case of the heuristic solution method are given in Table 7.22, with component and residual plots given in Figures 7.13 and 7.14.

Similar to the case of applying the offline model in partially dynamic problems, a clear relationship between the level of road network disruption and both fatalities and suffering objectives can be seen when considering the application of the constructive method to fully dynamic problems. In addition to this, a linear relationship between the rate at which casualties are discovered and the number of fatalities has been detected. This suggests that the performance of the model, in terms of f_1 is negatively impacted as the average length of time taken for casualties to be added to the model increases. One possible explanation for this is that a low level of parameter λ_8 will lead to a large number of casualties whose tasks are in a 'floating' state early on in the response operation, providing the constructive method more scope in arranging their tasks in an optimal manner. In contrast, as the average discovery time is increased the number of casualties to consider at any one time decreases, and so the scheduling of

Table 7.22: Fitted linear regression models for objectives f_1 and g_2 for the constructor method in fully dynamic problems.

	Dependent variable:	
	f_1	g_2
	(1)	(2)
λ_3		422.135*
		(231.500)
λ_4		932.003***
		(351.675)
λ_5		3,030.828***
		(874.526)
λ_6		301.396*
		(155.671)
λ_7	−3.140***	−14,621.610***
	(0.507)	(2,875.982)
λ_7^2		4,150.673***
		(1,138.955)
λ_8	0.160**	
	(0.078)	
Constant	24.579***	38,636.590***
	(0.776)	(1,868.724)
R^2	0.140	0.340
Adjusted R^2	0.134	0.325
Note:	*p<0.1; **p<0.05; ***p<0.01	

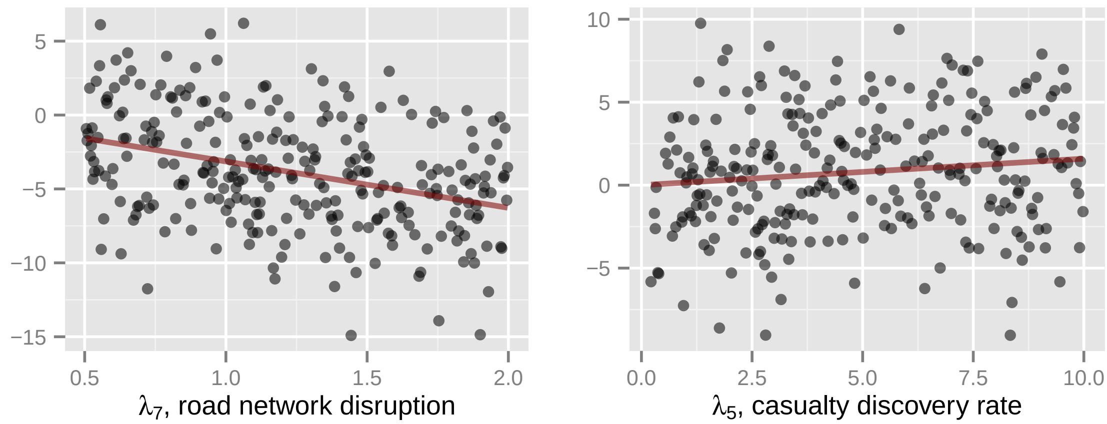

Figure 7.13: Component and residual plots of statistically significant relationships, considering the fatalities objective of the constructor method applied to fully dynamic problems.

tasks will be dictated by the discovery times of the casualties rather than the decision making of the constructor.

Online model M_{on}

Following the same format as in the case of the constructive solution method, the results of applying the model M_{on} to fully dynamic problems are given in Table 7.23. Component and residual plots are given in Figures 7.15 and 7.16.

In terms of the fatalities objective, the level of road network disruption was again found to have a significant effect. In addition to this, a number of parameters relating to the durations of tasks have been found to have significantly predictive relationships with the number of fatalities. Increases in the standard deviation parameter used when simulating task durations (λ_3) is found to predict slightly lower fatalities, as is an increase in the error of task duration estimates (λ_4). While the effect of the latter is quite intuitive, it is not as clear why an increased variation of task duration parameters should have a negative impact on objective f_1. One possible explanation is that the increased variability is felt particularly through those durations which are particularly large, in that the effect of these larger durations is not 'cancelled out' by an opposite effect arising from those tasks which take on shorter durations.

The parameter λ_5 was also shown to be a predictor of performance as measured by fatalities. In particular, lower values of this parameter (corresponding to attempting to consistently underestimate task durations) corresponds to best performance.

In the case of the suffering objective, g_2, both road network disruption and the error in task duration assessment were found to have relationships of the same type as was found in the case of fatalities. In addition to these, increases in the parameter corresponding to communication delay (λ_6) was shown to correspond to increases in

Table 7.23: Fitted linear regression models for objectives f_1 and g_2 for model M_{on} in fully dynamic problems.

	Dependent variable:	
	f_1	g_2
λ_3	-0.303^{*}	295.859
	(0.180)	(193.149)
λ_4	-0.979^{***}	$1{,}208.967^{***}$
	(0.268)	(287.114)
λ_5	2.477^{***}	
	(0.679)	
λ_6	0.184	286.583^{**}
	(0.121)	(129.505)
λ_7	-6.597^{***}	$-8{,}167.673^{***}$
	(2.352)	(2,518.210)
λ_7^2	1.764^{*}	$1{,}697.023^{*}$
	(0.930)	(995.911)
λ_8		106.037^{*}
		(58.314)
Constant	24.344^{***}	$33{,}284.630^{***}$
	(1.495)	(1,595.636)
R^2	0.131	0.225
Adjusted R^2	0.120	0.215
Note:	*p<0.1; **p<0.05; ***p<0.01	

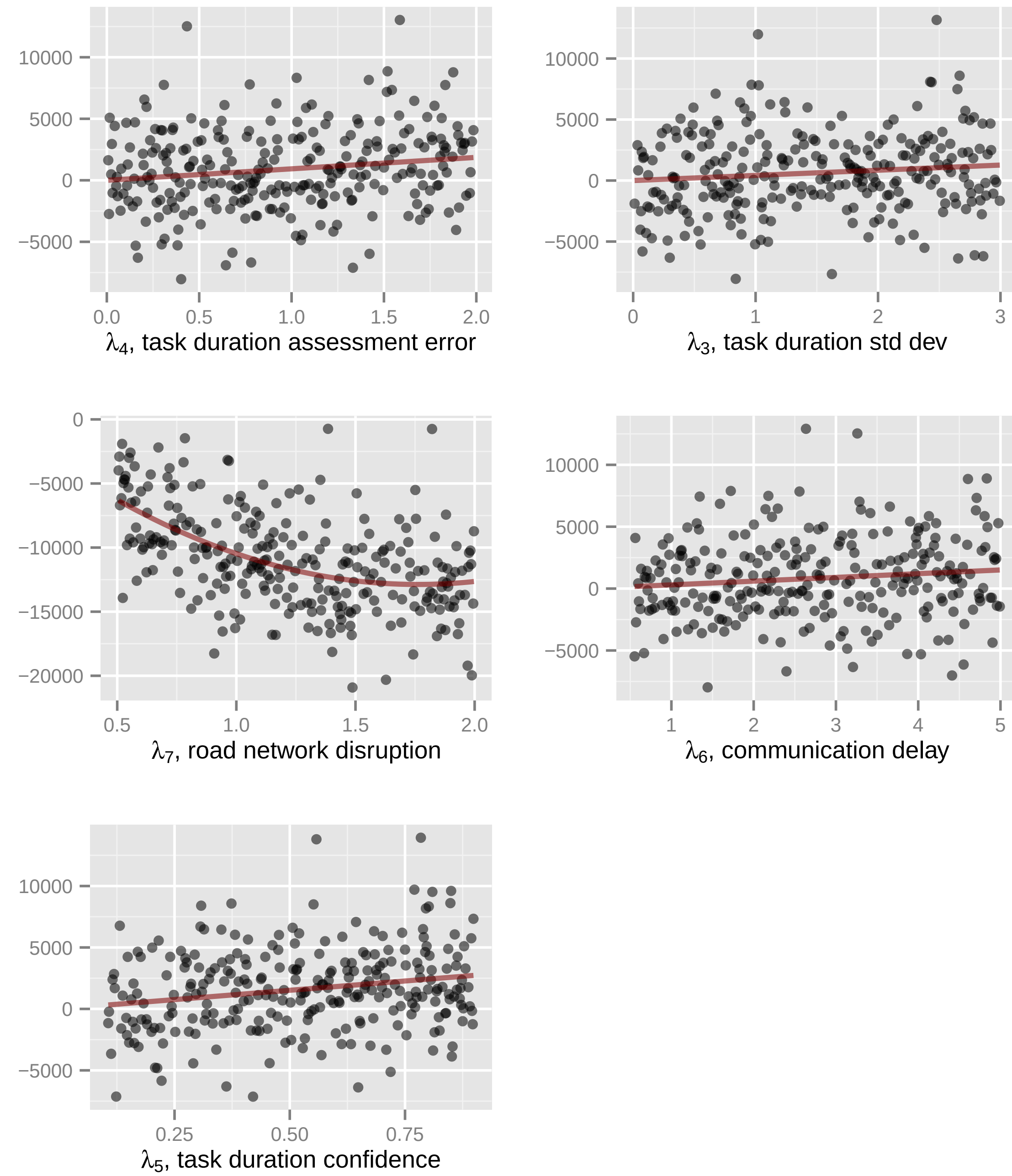

Figure 7.14: Component and residual plots of statistically significant relationships, considering the suffering objective of the constructor method applied to fully dynamic problems.

suffering. This was also the case of parameter λ_8, denoting the rate at which casualties are discovered. The latter relationship may be best explained through the same argument as was applied in the context of applying the constructive solution

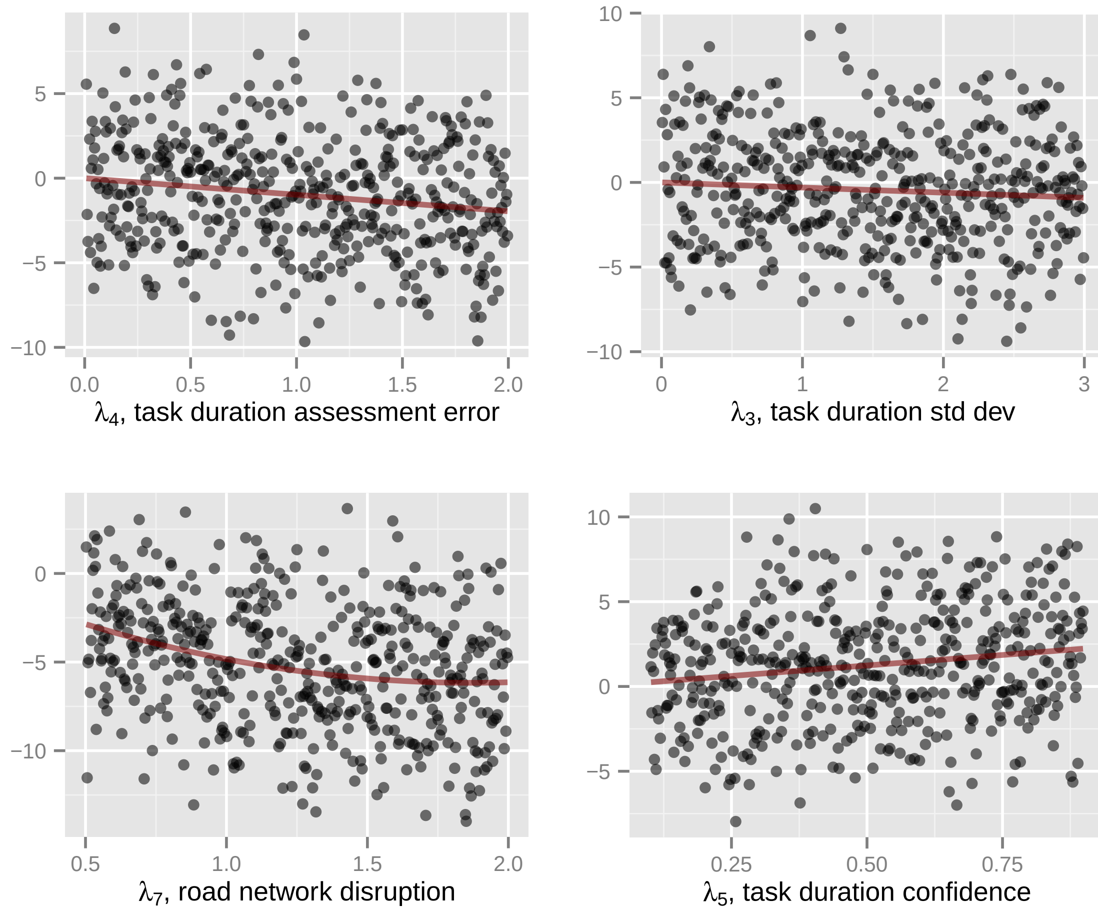

Figure 7.15: Component and residual plots of statistically significant relationships, considering the fatalities objective of model M_{on} applied to fully dynamic problems.

method; by reducing the number of casualties, and therefore tasks, available to the model for optimisation, the scheduling of tasks is determined to a greater degree by the 'discovery' times of casualties as opposed to the decision making of the model. It is also of interest to contrast these results with those obtained when applying model M_{off} to partially dynamic problems. In that case, as was shown in Table 7.19 and Figure 7.11, the accuracy of triage operations was shown to have a significant relationship with the suffering objective. This is not the case when applying model M_{on} to fully dynamic problems. This suggests that the model M_{on} is capable of receiving updates of information regarding the health of casualties, which could mitigate the effect of any error in such estimates.

7.5.3 Discussion

The results described in Section 7.5.2 provide valuable insight into the behaviour and utility of the proposed optimisation model in the context of realistic problems

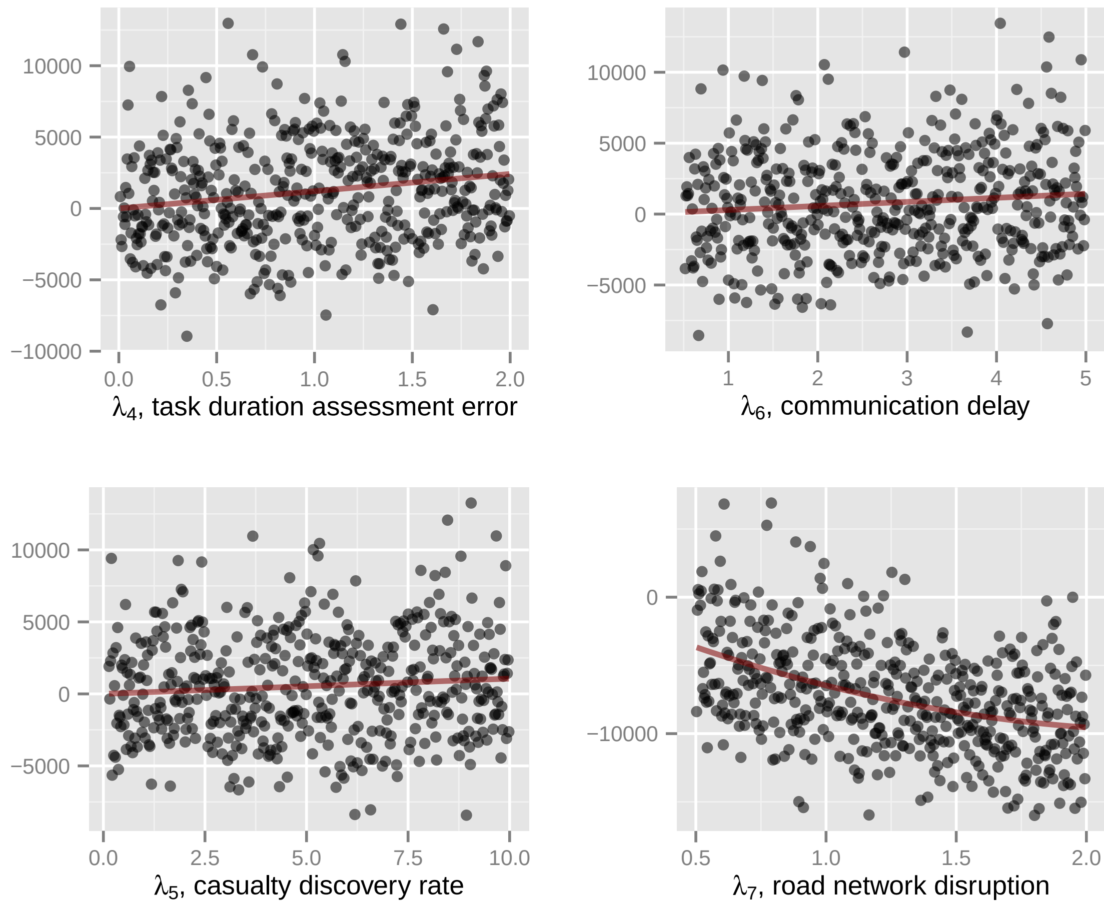

Figure 7.16: Component and residual plots of statistically significant relationships, considering the suffering objective of model M_{on} applied to fully dynamic problems.

involving uncertain and dynamic behaviour in both the solution space and objective space. We may explore how the results generated can illuminate a number of key questions.

To what extent are offline static schedules applicable in dynamic problems?

Considering the results of the offline model as applied to partially dynamic problems (Table 7.18 & Figure 7.8), we note that model M_{off} outperforms the constructive approach in terms of suffering. However, in the case of fatalities it is the constructive method which results in best average performance. In contrast, the results presented in Section 7.3 suggested that the offline model could lead to improvements in both objectives. These results highlight a shortcoming of the method which was not evident from previous studies. The dangers of evaluating optimization models for MCI response using unrealistically static and predictable problems scenarios is clear.

We note that the initial estimates of fatalities and suffering resulting from pre-computed solutions are systematically different from those realized upon completing the simulation, as shown in Figure 7.9. Specifically, the number of fatalities is typically overestimated whereas the level of suffering is typically underestimated. These discrepancies illustrate the difficulty in ensuring accurate forecasts within the dynamic and uncertain MCI response environment.

Considering the linear regression models fitted in Section 7.4.2 (Table 7.19) we note that the level of error in the estimation of task duration was not identified as a significant predictor in terms of either fatalities or suffering. In contrast, errors in the triage assessment of casualties do have a significant and negative association with suffering, as does a delay in communication. As would be expected, the initial expected value of the pre-computed solution influences the final objective values, confirming that some value is retained throughout the simulation.

Finally, we note that the problem scenarios considered in the pre-computed case only allowed for dynamic and uncertain behaviour in the objective space. Were such behaviour to exist in the solution space, the pre-compacted approach would lead to schedules which would quickly become irrelevant as the response operation progressed.

Can the search-based solution method cope with solution space dynamics as well as non-predictive methods can?

The online model M_{on} results in improved performance in both fatalities and suffering objectives, on average (Table 7.21 & Figure 7.12). We can therefore conclude that the model M_{on} can deliver improved performance over the alternative heuristic approach.

How sensitive is the model to underlying variation and uncertainty in objective space parameters?

Considering the regression models fitted (Tables 7.19 & 7.23) we see that, in comparison to the offline model, a larger number of relationships between parameters and objective values were identified in the case of the online model. Indeed, all parameters other than λ_2 (frequency of triage) and λ_9 (rate of casualty self-presentation) were identified as potential predictors of at least one objective outcome. The omission of triage frequency may be explained by the relatively short period of time in which casualty health may change, i.e. while they remain within the hazardous area.

Road network disruption is important, as we would expect since it leads to longer travel times. The delay in communication between the problem environment and the model also has a significant effect on performance, as does the accuracy of initial task duration estimates. However, while statistically significant linear relationships with these parameters were identified, a large amount of otherwise unaccounted for variance in performance remains.

Both fatalities and suffering are associated with a delay in communication in the case of the online model, although the associated parameter estimate is lower than in the offline case. This suggests that the online methodology successfully reduces the impact of poor communication by allowing for flexible adaptation of the schedule.

We note that the parameter describing the 'task duration confidence', i.e. the confidence the decision maker requires that their estimates of task durations will be greater than or equal to the realized value, is identified as a significant predictor of the fatalities objective for the online model. The relationship is positive and linear, suggesting that by estimating task duration in a manner which will result in under-estimates on average can lead to improved performance.

7.6 Summary

The development of any optimisation model for MCI response must be carefully evaluated before its further development or implementation can be considered. In this Chapter we have presented such an evaluation, focussing not only on specific details of the models described in this thesis but also on general problem and model characteristics, in an effort to gain valuable insight into the applicability of optimisation models in general to the MCI response environment.

Considering problems with increasing levels of realism in turn, both the offline model M_{off} (as developed through Chapter 4) and the online model M_{on} (as developed through Chapter 6) were evaluated through extensive simulation experiments. Generating a large number of data for each individual evaluation allowed for a statistical approach to the analysis of the results, thus helping to avoid spurious conclusions regarding patterns or relationships which were primarily due to chance variation.

While it was demonstrated that the model M_{off} led to improved performance in comparison to the constructive heuristic method when applied to static problems, this was shown to break down when partially dynamic problems were considered. This clearly demonstrated the danger of making inappropriate assumptions when evaluating the utility of an optimisation model in MCI response, as it implies that a model thought to be of considerable potential (evaluated in unrealistic static problem scenarios) could lead to *worse* performance than the current standard when applied to more realistic, partially dynamic problems.

However, in evaluating the utility of the online model M_{on} in the case of fully dynamic problems, performance was demonstrated to improve upon the alternative heuristic method. This provides valuable motivation for the continued study and development of optimisation models for MCI response, providing that such models are capable of online use, i.e. receiving regular updates of relevant information from the problem environment while delivering decision support in a gradual manner throughout the entire response operation. These conclusions contribute significantly to calls observed in the literature (as discussed in Chapter 3) for future work in this field to pursue the development of online optimisation models for use in real-time.

In addition to the evaluation of the proposed models, on average, results have also been broken down to allow for the analysis of individual problem and model characteristics. This has allowed for several important insights to be generated, providing further understanding to exactly why an online modelling methodology leads to improved performance, and to which types of problems we can expect this improved performance to be observed in.

Given the large number of characteristics which could be varied, both in terms of the problems considered and of the models themselves, only a subset of potential experiments were conducted and analysed in this Chapter. While many more could be suggested, the real-time nature of the simulation experiments coupled with limited computational resources meant a focussed approach was essential. The results generated through the experiments of this Chapter will be valuable in terms of deciding which factors should be the focus of subsequent evaluations.

Chapter 8

Conclusion

The application of optimisation methods to the area of Mass Casualty Incident response poses a significant challenge. In this thesis, such an application has been developed. Building upon the base of academic literature in the field, this work has focussed on developing an optimisation model which both addresses a problem not yet considered by other researchers (casualty processing), and on ensuring that this model improves upon similar ones before it in terms of the number and nature of assumptions which underpin it.

In this final chapter, a summary of the work described in this thesis is first given in Section 8.1. Here, the original objectives of the project, as described in Chapter 1, will be revisited. A focus will be maintained on the elements of the work which can be considered as both significant and original contributions of knowledge to the academic literature. Following this, in Section 8.2 the work will be critically assessed, with weaknesses identified and promising avenues for further research discussed.

8.1 Summary

Recall that the objectives of this thesis, as set out in Section 1.2, were to:

1. identify a particular decision problem faced in large scale emergency response which has the potential to be amenable to optimisation based decision support;

2. develop and implement a mathematical model of this decision problem;

3. implement an appropriate solution methodology such that high quality solutions to the decision problem can be found in a timely manner;

4. develop and implement a stochastic simulation of the large scale emergency response environment to be run in conjunction with the optimisation model;

5. evaluate the proposed model and solution algorithm through a Monte Carlo analysis using the simulation.

In what follows we summarise the work contained in this thesis and the extent to which the above objectives were achieved.

8.1.1 Background and foundations

The task of identifying a specific decision problem was detailed in Chapters 2 and 3. In the former, the area of large-scale emergency response was reviewed from a practical standpoint, with particular attention given to UK emergency planning and emergency response practitioner documentation. Throughout Chapter 2, the focus of the work was narrowed to a specific class of large scale emergency, namely the Mass Casualty Incident (MCI). Specific details regarding the decision making which takes place during MCI response were set out, providing the foundations for their subsequent mathematical formulation.

Following this establishment of the problem area, previous research into the design and implementation of optimisation models for decision support in emergency response was reviewed in Chapter 3. Given the large number of potential decision problems within MCI response from which a single decision problem was to be identified, the research reviewed was of an accordingly broad nature in terms of application. Reviewing models focussed on logistics, resource allocation and facility location led to a gap in the literature being identified. The problem of scheduling the extrication, treatment and transportation to hospital of all casualties, within time scales of a few hours following the MCI, and including in the decision formulation the choice of hospital to which each casualty should be delivered, had yet to be modelled and solved. This problem was denoted the *Casualty Processing Problem*, and was the focus of the remainder of the thesis. Objective 1 was therefore successfully achieved.

8.1.2 Model development and solution

The initial formulation of the Casualty Processing Problem was described in Chapter 4. Here, the specific structure of the decision space was set out, moving forward from the qualitative details provided in Chapter 2. A combinatorial structure was found to be the most natural representation of the decision space. Specifically, a task scheduling approach was employed, using the established Job-shop Scheduling Problem and its variants as a starting point. The task of evaluating any given solution lying in this decision space was separated into two parts. Firstly, the procedure used to compute a forecast of a schedule of the response operation was described. This routine made use of estimates of task durations, knowledge of dependency relations between tasks and a graphical representation of the spatial environment to arrive at estimated start and end times of all tasks within the model. The second component of solution evaluation made use of these estimated times when computing five separate measures of solution quality. These consisted of an estimate of the number of fatalities, a weighted sum of the times taken to deliver casualties to hospital, a measure of the suitability of hospital choice for each casualty, together with two standard metrics for schedule evaluation: the time spent idle by all resources, and the makespan. These five objective functions were combined through a mixture of lexicographic ordering and weighted scalarizing methods in order to specify the full multi-objective model with minimal reliance of the setting of weights. Together, the formal representation of the decision problem and the multi-objective formulation constitute the model of objective 2.

Having developed the initial model, Chapter 5 addressed the problem of automatically finding high quality solutions. Basic heuristics were identified, both of a constructive and a iterative nature. The former were employed in a constructive heuristic solution methodology. This algorithm was capable of quickly producing candidate schedules in a manner which aimed to reflect the decision making process which would be carried out as routine. In addition to this solution methodology, a local search routing was also described. The algorithm, an implementation of the Variable Neighbourhood Descent metaheuristic, used several neighbourhood structures of variable size in order to avoid becoming trapped in the local optima typical of combinatorial optimisation problems. Both the constructive and local search solution methodologies were evaluated in order to identify the best performing parameterisations. The development and evaluation of these solution methodogies corresponds to objective 3.

The final stage of model development was described in Chapter 6, where the important issue of the inherently volatile nature of MCI response was tackled. With the aim of removing common assumptions regarding the completeness, accuracy and stability of information used to build the decision support model, an interface allowing for the continual passing of information from the problem environment to the model was introduced. This was complemented by the ability to communicate in the opposite direction, from model to problem environment, in real time also. In preparation for the evaluation of the model in a dynamic environment subject to uncertainty, a simulation of the problem environment was described. In this simulation, corresponding to objective 4, parameters previously assumed to be fixed and known were allowed to be subject to error in their estimation. Parameters including the duration of tasks, the time needed to make a certain journey, and the health of a casualty, were all allowed to be estimated with a degree of error and subsequently corrected once the real values had been realised. The specific case of uncertainty in travel times, arising from uncertainty regarding the level of disruption to the road network, was considered in detail. A number of routing strategies for use in such situations were provided.

8.1.3 Experimental analysis

The development of the simulation and interface methods of Chapter 6 allowed for a comprehensive evaluation of the proposed model, as detailed in Chapter 7. Variations of both the model and the problem simulation were compared, in the process allowing for the validity of model assumptions to be examined. Features of the model introduced in Chapter 4 were evaluated in order to demonstrate their benefit. Furthermore, the extent to which features of the problem environment, such as the average time taken to update the model of a realised parameter (e.g. the end time of a task), affected the utility of the model was evaluated. Computational experiments of the Monte Carlo nature were employed in this process, with statistical models fitted to the results in order to identify statistically significant predictors of performance. This represented the achievement of objective 5.

8.2 Critical appraisal and further work

8.2.1 Decision problem

The casualty processing problem identified in Chapters 2 and 3 was represented as, primarily, a task scheduling formulation. In many respects, this is a natural representation of the problem. In particular, the key tasks of extrication, treatment and transportation all have well-defined durations - that is, the length of time spent on a task is not a decision in itself. The natural ordering of these tasks was also effectively captured through dependencies in the scheduling process. However, with up to four tasks to be scheduled for each casualty, this representation incurred a significant computational burden when used in problems involving large numbers of casualties. Furthermore, there is a degree of redundancy in this representation. Given two casualties with matching initial characteristics (i.e. health level, need of extrication, need of treatment) at the same incident site, their equivalence propagates to become an equivalence of tasks within the model. That is, throughout the optimisation process the corresponding tasks of these two casualties could be interchanged with no effect on the resulting schedule. A solution representation which recognised this equivalence would result in a reduced solution space, which in turn could be more comprehensively searched by a given optimisation methodology in a given amount of time.

The general form of the decision problem could be extended to other specific MCI response problems. In particular, other task types and responder types could be included within the general task scheduling framework. As with all tasks included in the proposed model, any additions would require to be of a discrete nature and to be completed by a single responder unit in a single sitting. Stochastic task durations could be incorporated in the same manner as for extrication tasks. Extensions to consider more fundamental difference, such as tasks which could be completed by multiple responders, or tasks which could be completed over multiple sessions, would require more extensive development.

8.2.2 Model

The process of constructing a schedule from a given solution took into account the expected durations of all tasks, their dependency relations, and any time needed by responders to travel between the final location of one task to the initial location of the next. Additional information regarding any ‘setup’ or ‘turnaround’ time required at the start or end of tasks could also be built into the scheduling process. For example, upon reaching the destination hospital while transporting a casualty, the transfer of said casualty from ambulance to hospital will likely require some time. This time could be relatively consistent or highly variable - data would have to be collected in order to properly infer such characteristics. In terms of the mathematical description of the model, setup or turnaround times could be included as properties of tasks in the same manner as duration. In terms of implementation, limited adjustment to the scheduling algorithm and the representation of tasks would be required.

During the evaluation of schedules, a Markov chain is employed in the representation of casualty health as a discrete state stochastic process. This reflects well the discrete triage system used in practice in classifying casualty health. However, modelling health in this manner imposes the *Markovian assumption*. This states that the stochastic process is 'memoryless', with the probabilities of transition into another state determined only by the current state. In our case this can be interpreted as saying that the probability of a casualty's health deteriorating, given that they are in state Ti, does not change whilst they stay in state Ti. Thus, a casualty who has been in state T1 for two minutes has the same probability of dying in the next minute as a casualty who has been in state T1 for an hour. The Markovian assumption is difficult to examine in the absence of data recording the health progression of a sufficiently large sample of casualties in MCI situations, and such data are not available.

Regarding the objective measure of hospital choice, we note that the process whereby hospitals increase their available resources through enacting a major incident plan is assumed to have a simple, linear form. This could be improved in terms of realism, if appropriate data were gathered. More generally, improvements in this measure could be found through introducing a more specific definition of hospital resources, with explicit representations of staff and of beds.

The combination of these and other objectives was achieved through a lexicographic approach. As demonstrated in Chapter 7, this can lead to large losses in suffering being accepted in exchange for small gains in fatalities. In some cases, this may not be viewed as desirable by the decision makers involved. The problem could be addressed through relaxing the priority levels employed when separating the objective classes, allowing for some degree of trade-off between the two objectives to be recognised as acceptable. Another approach to the formulation of the multi-objective problem would be to consider a Pareto view. Here, rather than focussing the search on a single solution which should involve a good balance between the individual objective functions, the model would instead attempt to find a set of solutions, each of which is not dominated by any other known solution (in the Pareto sense). It would then be left to the decision maker to select, from this set of solutions, that which best reflected their current priorities with regard to each objective function. Such a representation would, however, require significant modifications to the presented model, particularly in terms of the solution methodology employed and the interface between the online model and the simulation of the problem environment.

8.2.3 Parameterisation

Whilst the computational experiments described in Chapter 7 included analysis of the sensitivity of model performance to changes in several types of parameters, there remain some which did not undergo this process and so should be highlighted here.

Firstly, the transition probabilities used in the parameterisations of the Markov chain governing health states were estimated. As discussed previously, there is a sparsity of data documenting the progression of casualty health in MCI environments, and so it was not possible for these transition probabilities to be derived from empirical evidence. Instead, parameters were estimated and the resulting behaviour analysed to ensure that they were sensible choices. It would be of interest in future work to examine the sensitivity of the results reported in this thesis to changes in these parameters, as this would provide further insight in terms of to which types of problems we may expect the model to be successfully applicable to.

Weights were also assigned to casualty health states for the calculation of objective f_2, evaluating the time taken to deliver casualties to hospital. As described in Chapter 4, these weights were set in a manner which reflected the rough guidelines of how soon casualties should be treated for each triage state. Again, the sensitivity of performance to these parameters should be assessed.

Finally, we note that calculations estimating the time needed for responders to travel across the disaster environment required parameters of 'cruise speed' and acceleration to be set. In this work the values of these parameters were taken from a previously published analysis of ambulance travel times in Calgary, Canada [15]. While these values should provide a solid foundation for the specific model environment discussed in this thesis, it would clearly be desirable for them to be based on local ambulance travel time data.

8.2.4 Optimisation

The constructive heuristic introduced in Chapter 5 was designed to generate good quality solutions quickly, in a manner which would reflect how decisions would be made by responders on the ground in a real incident. Other constructive heuristic methodologies could be designed and implemented without such constraints. In particular, a simple 'greedy' constructor could, potentially, deliver solutions of higher quality with minimal increase in computation time.

In addition to development of constructive heuristic solution methodologies, further work could focus on the improvement of iterative search procedures. The Variable Neighbourhood Descent metaheuristic described in Chapter 5 was of a local search nature, using several neighbourhood structures. These same neighbourhood structures could be used in other metaheuristic search strategies of a local search nature, such as Simulated Annealing or Tabu Search. In addition, population based metaheuristics such as Genetic Algorithms could be implemented.

The neighbourhood structures detailed in Chapter 5 are implemented in a highly coherent manner, where all neighbourhood moves are translated into a series of atomic operations using a shared library. As noted previously, this in turn allows for the definition of a metric on the space of solutions independent of the neighbourhood structures employed. This feature was not exploited in this thesis, but represents a promising avenue for future research. In particular, imposing a common underlying structure on the solution space could facilitate its automatic characterisation, in terms of its fitness landscape features, using techniques from the field of machine learning. Solution algorithms which could adapt their heuristics to best suit the type of fitness landscape in which it finds itself could lead to improvements in terms of robustness.

Finally, it should be noted that the dynamic, evolving nature of the casualty processing problem presents specific challenges to optimisation. Given the variation in performance of both the constructive heuristic Φ and the VND algorithm over the range of problems considered in Section 5.5, it is suggested that hyper-heuristics [107] could provide more robust and adaptive performance over the wide range of potential scenarios. In particular, hyper-heuristics could be combined with the type of machine learning analysis described to self-learn how to proceed with searching a particular solution space.

8.2.5 Simulation

The simulation of volatility presented in this thesis is restricted to that resulting from factors external to the response operation itself. That is, we assume that all responders will follow the instruction issued by the optimization model regardless of their own personal view of events. This assumption should be examined in further detail. In particular, it would be of interest to consider situations where an individual responder has access to information which is significantly more accurate and up-to-date than that which was used by the optimization model in formulating its instruction. Such an analysis would require a detailed model of individual decision making; an agent-based simulation, such as that described in [62], would be well suited to this task. Allowing for the responder to override the model in such a situation could improve overall performance, although the impact of introducing further uncertainty and volatility into the model should be examined.

8.2.6 Information processing

As discussed in [55], a Bayesian approach to processing information in the dynamic environment of disaster response may be applicable. While the model presented in this work employs such an approach when considering parameters which govern travel times in the road network and the rate at which casualties self-present, other parameters may benefit from similar treatment. In particular, we note that the full hierarchial model representing the duration of response tasks could be estimated and adjusted as information is accrued during the response. This would, however, require further computational resources. If such a learning routine were to be implemented, it would be of value to analyze to what extent performance is robust to changes in the underlying model.

Considering the potential progression from a simulation of the problem environment to use in a real scenario, the results of Chapter 7 underscore the importance of effective and fast communication of information. While communication from the optimisation model to the problem environment should not pose a significant challenge, the transfer of information in the other direction may be difficult [97]. Further research into the availability and suitability of technology allowing for such gathering and transferring of information would be an essential next step in the application of optimisation models to real MCI response operations.

8.2.7 Computational experiments

Throughout Chapter 7 a number of experiments were designed with the aim of evaluating the performance of the model in a range of problem types. The design of these experiments sought to include the most relevant and interesting parameters of the model and the problem simulation as variables, in order to allow for their influence on performance to be measured. However, given further resources, yet more experiments could be run in order to evaluate a number of variables held fixed in Chapter 7. In particular, it is noted that the effect of the numbers of responders, the dependency level of tasks and the number of sites remained the same in all experiments from Section 7.4 onwards. It would be of interest to measure the effect changes in these parameters have on performance in the partially and fully dynamic cases.

Another parameter which was held constant was the number of casualties. This was done so as to make changes in the number of responders more meaningful, leading to a change in responders relative to the demand for their time. Repeating some of the analyses presented in this thesis with varying numbers of casualties would shed light on the stability of the conclusions derived. In particular, due to the expected 'combinatorial explosion' which would accompany an increase in casualty numbers, it would be of value to ascertain a threshold above which the application of the local search methodology is not worthwhile in comparison to the constructive methodology.

In some cases, the regression models fitted to data in Chapter 7 only captured a limited proportion of the variability in results. For example, the model describing the relation of the suffering objective to model and problem characteristics (see Table 7.19) had an adjusted R^2 values of 0.474. This demonstrates that the covariates included in the model, while explaining some of the variation in results, leave 52.6% of it unexplained. In one sense, this is a reassuring assessment of the complexity of the system. If a regression model using these covariates accounted for a high proportion of variation, this would indicate the system is highly predictable, a characteristic we would not associate with an MCI response operation. However, it would nevertheless be desirable to further categorise sources of variation in order to better understand the simulation model. For example, it would be of interest to estimate how much variability is due to the stochastic nature of the optimisation process, or how much is due to the stochastic nature of casualty health progression.

Appendix A

Description of raw data files

This appendix provides descriptions of the data files used in the analyses throughout this thesis. The data files described are available in electronic format, both on a CD attached to this thesis and online. For each data file, a table is given which describes each of the fields used in the analysis. The online version of the files can be found at 'https://www.dropbox.com/sh/a1gnxn6zkanae6x/AAAQhyoJZzuO2muQZotEGeB6a'.

A.1 Chapter 5

The data used in the analysis of various configurations of the constructive heuristic, presented in Section 5.5.1, are given in the file 'data_constructor_config.txt'. The format of that file is given in Table A.1. The data used in the analysis of various configurations of the local search procedure, presented in Section 5.5.2, are given in the file 'data_search_config.txt'. The format of that file is given in Table A.2.

Table A.1: Raw data: constructor configurations. Used in the analysis of Section 5.5.1.

Column	Name	Description
1	R	Responder levels, $12 = R_{low}, 24 = R_{med}, 36 = R_{high}$
2	D	Dependency levels, $18 = D_{low}, 48 = D_{med}, 102 = D_{high}$
3	S	Number of sites, $5 = S_{one}, 6 = S_{two}, 7 = S_{three}$
4	C	Constructor configuration
5	end_f1	Final value of objective f_1
6	end_f2	Final value of objective f_2
7	end_f3	Final value of objective f_3
8	end_f4	Final value of objective f_4
9	end_f5	Final value of objective f_5
10	end_g2	Final value of objective g_2
11	end_g3	Final value of objective g_3

Table A.2: Raw data: local search configurations. Used in the analysis of Section 5.5.2.

Column	Name	Description
1	R	Responder levels, $12 = R_{low}, 24 = R_{med}, 36 = R_{high}$
2	D	Dependency levels, $18 = D_{low}, 48 = D_{med}, 102 = D_{high}$
3	S	Number of sites, $5 = S_{one}, 6 = S_{two}, 7 = S_{three}$
4	P	Local search configuration
5	end_f1	Final value of objective f_1
6	end_f2	Final value of objective f_2
7	end_f3	Final value of objective f_3
8	end_f4	Final value of objective f_4
9	end_f5	Final value of objective f_5
10	end_g2	Final value of objective g_2
11	end_g3	Final value of objective g_3

A.2 Chapter 7

The data used in the analysis of routing strategies, presented in Section 7.4.1, are given in the file 'data_routing.txt'. The format of that file is given in Table A.3.

Table A.3: Raw data: routing strategies. Used in the analysis of Section 7.4.1.

Column	Name	Description
1	R	Responder levels, $12 = R_{low}, 24 = R_{med}, 36 = R_{high}$
2	D	Dependency levels, $18 = D_{low}, 48 = D_{med}, 102 = D_{high}$
3	S	Number of sites, $5 = S_{one}, 6 = S_{two}, 7 = S_{three}$
5	end_f1	Final value of objective f_1
6	end_f2	Final value of objective f_2
7	end_f3	Final value of objective f_3
8	end_f4	Final value of objective f_4
9	end_f5	Final value of objective f_5
10	end_g2	Final value of objective g_2
11	end_g3	Final value of objective g_3
12	strat	Routing strategy
13	dist	Network disruption parameter
14	rate	Autonomous routing improvement rate

The data used in the analysis of the model M_{off} in static problems, presented in Section 7.4.1, are given in the file 'data_static_offline.txt'. The format of that file is given in Table A.4.

The data used in the analysis of the model M_{rt} in static problems, presented in Section 7.3.2, are given in the files 'data_static_real_time1.txt', 'data_static_real_time2.txt',

Table A.4: Raw data: static, offline. Used in the analysis of Section 7.3.1.

Column	Name	Description
1	R	Responder levels, $12 = R_{low}, 24 = R_{med}, 36 = R_{high}$
2	D	Dependency levels, $18 = D_{low}, 48 = D_{med}, 102 = D_{high}$
3	S	Number of sites, $5 = S_{one}, 6 = S_{two}, 7 = S_{three}$
4	Exp	Offline model variant index, $\{1, 2, 3, 4\}$
5	end_f1	Final value of objective f_1
6	end_f2	Final value of objective f_2
7	end_f3	Final value of objective f_3
8	end_f4	Final value of objective f_4
9	end_f5	Final value of objective f_5
10	end_g2	Final value of objective g_2
11	end_g3	Final value of objective g_3
12	start_f1	Initial value of objective f_1
13	start_f2	Initial value of objective f_2
14	start_f3	Initial value of objective f_3
15	start_f4	Initial value of objective f_4
16	start_f5	Initial value of objective f_5
17	start_g2	Initial value of objective g_2
18	start_g3	Initial value of objective g_3

and 'data_static_real_time3.txt'. The format of that file is given in Tables A.5, A.6 and A.7 respectively.

Table A.5: Raw data: static, real-time, fixed optimisation time. Used in the analysis of Section 7.3.2 to generate the box plots in Figures 7.5 and 7.6.

Column	Name	Description
1	time	Time spent optimising before issuing the schedule (mins)
2	g2	Value of objective g_2 after correcting for delay
3	f1	Value of objective f_1 after correcting for delay

Table A.6: Raw data: static, real-time, continuous adjustment, objective g_2. Used in the analysis of Section 7.3.2 to generate the lines in Figures 7.5 and 7.6.

Column	Name	Description
1	cont_t	Time spent optimising
2	cont_g2	Value of objective g_2 at that time, accounting for delay

The data used in the analysis of the local search and constructive solution methods in partially dynamic problems, presented in Section 7.4.2, are given in the files

Table A.7: Raw data: static, real-time, continuous adjustment, objective f_1. Used in the analysis of Section 7.3.2 to generate the lines in Figures 7.5 and 7.6.

Column	Name	Description
1	cont_t	Time spent optimising
2	cont_f1	Value of objective f_1 at that time, accounting for delay

'data_PD_search.txt' and 'data_PD_const.txt'. The data used in the analysis of the local search and constructive solution methods in fully dynamic problems, presented in Section 7.5, are given in the files 'data_FD_search.txt' and 'data_FD_const.txt'. The format of all four files is identical, and given in Table A.8.

Table A.8: Raw data: partially and fully dynamic problems. Used in the analysis of Sections 7.4 and 7.5.

Column	Name	Description
2	end_f1	Final value of objective f_1
3	end_f2	Final value of objective f_2
4	end_f3	Final value of objective f_3
5	end_f4	Final value of objective f_4
6	end_f5	Final value of objective f_5
7	end_g2	Final value of objective g_2
8	end_g3	Final value of objective g_3
9	task_err	Task duration assessment error λ_4
10	tri_err	Triage assessment error λ_1
11	tri_freq	Triage frequency λ_2
12	task_sd	Task duration variance λ_3
14	sp_rate	Casualty self-presentation rate λ_9
15	road	Road network disruption λ_7
17	comm_del	Communication delay λ_6
19	cas_rate	Casualty discovery rate λ_8
20	conf	Task duration confidence λ_5

Appendix B

Publications arising from this thesis

B.1 Journal papers

B.1.1 Published:

(i) **Wilson, D.T.**, Hawe, G.I., Coates, G. & Crouch, R.S., "Evaluation of centralized and autonomous routing strategies in major incident response". *Safety Science*, Accepted.

(ii) **Wilson, D.T.**, Hawe, G.I., Coates, G. & Crouch, R.S. (2013), "Modeling uncertain and dynamic casualty health in optimization-based decision support for mass casualty incident response", *International Journal of Information Systems for Crisis Response and Management*, 5 (2), 32-44.

(iii) **Wilson, D.T.**, Hawe, G.I., Coates, G. & Crouch, R.S. (2013), "A multi-objective combinatorial model of casualty processing in major incident response", *European Journal of Operational Research*, 230, 643-655.

B.1.2 Under review:

(iv) **Wilson, D.T.**, Hawe, G.I., Coates, G. & Crouch, R.S., "Online optimisation for casualty processing in major incident response: an experimental analysis". *European Journal of Operational Research*, under 2nd review following requested revision.

B.2 Refereed conference papers

(v) **Wilson, D.T.**, Hawe, G.I., Coates, G. & Crouch, R.S. (2013), "Scheduling response operations under transport network disruptions", *Proceedings of the 10th International Conference on Information Systems for Crisis Response and Management (ISCRAM 2013)*.

(vi) **Wilson, D.T.**, Hawe, G.I., Coates, G. & Crouch, R.S. (2012), "Stochastic task durations in the scheduling of emergency response operations", *Proceedings of the 30th Workshop of the UK Planning And Scheduling Special Interest Group (PlanSIG 2012).*

(vii) **Wilson, D.T.**, Hawe, G.I., Coates, G. & Crouch, R.S. (2012), "A hyper-heuristic approach to optimizing emergency response", *Proceedings of the 4th International Conference on Metaheuristics and Nature Inspired Computing (META 2012).*

(viii) **Wilson, D.T.**, Hawe, G.I., Coates, G. & Crouch, R.S. (2012), "Effective allocation of casualties to hospitals in mass casualty incidents", *Proceedings of the 3rd IEEE International Conference on Emergency Management and Management Sciences (ICEMMS 2012).*

(ix) **Wilson, D.T.**, Hawe, G.I., Coates, G. & Crouch, R.S. (2012), "Estimating the value of casualty health information to optimization-based decision support in response to major incidents", *Proceedings of the 9th International Conference on Information Systems for Crisis Response and Management (ISCRAM 2012).*

(x) **Wilson, D.T.**, Hawe, G.I., Coates, G. & Crouch, R.S. (2011), "A decision support framework for large scale emergency response", *Proceedings of Disaster Management 2011.*

B.3 Timeline and description

An initial formulation of the casualty processing model and solution methodology described in this thesis was published in (x). In this paper, some key features of the final model were introduced. The task allocation and scheduling form of the decision was set out, with an initial identification of the associated tasks and response units. The objective function described in this paper employed the Markov chain technique described in Chapter 4, although in this case the health sub-model was used to describe fluctuations in casualty health in all environment the casualty found themselves in - not just the hazardous environment of within the inner cordon. The estimate of fatalities obtained through employing this method was the sole objective function used in the optimisation process.

Development of the model continued in (viii), where the importance of hospital allocation was explored in more detail. This involved improving the model to allow for more detail to be used when deciding to which hospital each casualty should be take, in particular making allowance for heterogeneous and dynamic capacities and capabilities of hospitals. The fact that casualties often autonomously self-present at a hospital of their choice was also included in this development.

Progressing to a multi-objective formulation, the full offline casualty processing model was described in (iii). In comparison to the initial formulation given in (x), the model took on a multi-objective nature, using a number of measures in addition to the expected number of fatalities. These measures included one corresponding to the quality of hospital allocation, as described above. Solution methodologies were also described in more detail in this paper, including both the constructive heuristic and local search routines described in Chapter 5. The key point of further work discussed in the conclusion of this paper was the need to recognise the dynamic nature of the disaster response environment and its potential impact upon the utility of the proposed scheduling model or other similar approaches.

This question directed the remainder of the research on this project. In (ix), an initial evaluation of dynamics was given by focussing on casualty health. The initial results presented in this conference paper were expanded and elaborated upon in the journal manuscript (ii). The challenge of a dynamic problem environment to effective optimisation was explored in (vii), where a hyper-heursitic approach to developing an adaptive local search procedure was described. Following this point, all subsequent research focussed on the evaluation of the model in dynamic environments, through the development of the simulation and interface methods described in Chapter 6. In particular, temporal parameters of the model were developed to allow for their dynamic and uncertain nature. In (vi), the impact of uncertain task durations upon the utility of the model was evaluated, allowing for the model to continuously update parameters as and when information regarding the true duration of tasks was gathered from the simulation. Similar ideas were detailed in (v), in this case relating to the uncertainty inherent in travel time predictions when there is an unknown level of disruption in the transport network. This work was subsequently elaborated in (i). A final journal paper examining, in a comprehensive manner, a host of dynamic and uncertain problem characteristics and their impact upon performance was also submitted for journal publication (iv).

Bibliography

[1] Advanced Life Support Group. *Major Incident Medical Management and Support: The Practical Approach at the Scene (3rd Edition)*. Wiley-Blackwell, 2011.

[2] Nezih Altaya and Walter G. Green. OR/MS research in disaster operations management. *European Journal of Operational Research*, 175 (1):475–493, 2006.

[3] M. Amiri, M. Zandieh, M. Yazdani, and A. Bagheri. A variable neighbourhood search algorithm for the flexible job-shop scheduling problem. *International Journal of Production Research*, 48:5671–5689, 2010.

[4] T. Andersson and P. Varbrand. Decision support tools for ambulance dispatch and relocation. *Journal of the Operational Research Society*, 58 (2):195–201, 2007.

[5] Erik Auf der Heide. The importance of evidence-based disaster planning. *Annals of Emergency Medicine*, 47 (1):34–49, 2006.

[6] G. Barbarosoglu and Y. Arda. A two-stage stochastic programming framework for transportation planning in disaster response. *Journal of the Operational Research Society*, 55 (1):43–53, 2004.

[7] Glay Barbarosoglu, Linet Ozdamar, and Ahmet Cevik. An interactive approach for hierarchical analysis of helicopter logistics in disaster relief operations. *European Journal of Operational Research*, 140:118–133, 2002.

[8] S. Belardo, J. Harrald, W. A. Wallace, and J. Ward. A partial covering approach to siting response resources for major maritime oil spills. *Management Science*, 30 (10):1184 – 1196, 1984.

[9] Michael R. Berren, Allan Beigel, and Stuart Ghertner. A typology for the classification of disasters. *Community Mental Health Journal*, 16:103–111, 1980.

[10] Yuval H. Bloch, Dagan Schwartz, Moshe Pinkert, Amir Blumenfeld, Shkolnick Avinoam, Giora Hevion, Meir Oren, Avishay Goldberg, Yehezkel Levi, and Yaron Bar-Dayan. Distribution of casualties in a mass-casualty incident with three local hospitals in the periphery of a densely populated area: Lessons learned from the medical management of a terrorist attack. *Prehospital and Disaster Medicine*, 22:186–192, 2007.

[11] Christian Blum and Andrea Roli. Metaheuristics in combinatorial optimization: Overview and conceptual comparison. *ACM Computing Surveys*, 35(3):268–308, 2003.

[12] Ali Bozorgi-Amiri, Mohammad Jabalameli, Mehdi Alinaghian, and Mahdi Heydari. A modified particle swarm optimization for disaster relief logistics under uncertain environment. *International Journal of Advanced Maufacturing Technology*, 60:357–371, 2012.

[13] Paolo Brandimarte. Routing and scheduling in a flexible job shop by tabu search. *Annals of Operations Research*, 41:157–183, 1993.

[14] G.G. Brown and A.L. Vassiliou. Optimizing disaster relief: Real-time operational and tactical decision support. *Navel Research Logistics*, 40:1–23, 1993.

[15] Susan Budge, Armann Ingolfsson, and Dawit Zerom. Empirical analysis of ambulance travel times: The case of Calgary emergency medical services. *Management Science*, 56:716–723, 2010.

[16] Edmund K. Burke and Graham Kendall. *Search Methodologies: Intordurctory Tutorials in Optimization and Decision Support Techniques*. Springer, 2005.

[17] Peter A. Cameron, Biswadev Mitra, Mark Fitzgerald, Carlos D. Scheinkestel, Andrew Stripp, Chris Batey, Louise Niggemeyer, Melinda Truesdale, Paul Holman, Rishi Mehra, Jason Wasiak, and Heather Cleland. Black saturday: the immediate impact of the february 2009 bushfires in victoria, australia. *The Medical Journal of Australia*, 191(1):11–16, 2009.

[18] Aakil M. Caunhye, Xiaofeng Nie, and Shaligram Pokharel. Optimization models in emergency logistics: A literature review. *Socio-Economic Planning Sciences*, 46:4–13, 2012.

[19] Centers for Disease Control and Prevention. *Interim planning guidance for preparedness and response to a mass casualty event resulting from terrorist use of explosives*. National Center for Injury Prevention and Control, Atlanta, GA, 2010.

[20] Mei-Shiang Chang, Ya-Ling Tseng, and Jing-Wen Chen. A scenario planning approach for the flood emergency logistics preparation problem under uncertainty. *Transportation Research Part E: Logistics and Transportation Review*, 43 (6):737–754, 2007.

[21] Yi-Chang Chiu and Hong Zheng. Real-time mobilization decisions for multi-priority emergency response resources and evacuation groups: Model formulation and solution. *Transportation Research Part E: Logistics and Transportation Review*, 43 (6):710 – 736, 2007.

[22] Chang Chunguang, Ma Xiang, Song Xiaoyu, and Gao Bo. Emergency goods scheduling model and algorithm during initial stage of disaster relief. In *2010 International Conference on Logistics Systems and Intelligent Management*, 2010.

[23] Laura K. Church and Reha Uzsoy. Analysis of periodic and event-driven rescheduling policies in dynamic shops. *International Journal of Computer Integrated Manufacturing*, 5(3):153–163, 1992.

[24] Graham Coates, Glenn I. Hawe, Duncan T. Wilson, and Roger S. Crouch. Adaptive co-ordinated emergency response to rapidly evolving large-scale unprecedented events (rescue). In *Proceedings of the 8th International ISCRAM Conference*, 2011.

[25] Louise K. Comfort and Naim Kapucu. Inter-organizational coordination in extreme events: The World Trade Center attacks, September 11, 2001. *Natural Hazards*, 39(2):309–327, 2006.

[26] Louise K. Comfort, Kilkon Ko, and Adam Zagorecki. Coordination in rapidly evolving disaster response systems. *American Behavioural Scientist*, 48(3):295–313, 2004.

[27] Clive Cookson. Sciences 10 hottest fields, 2011. Financial Times.

[28] Damon P. Coppola. Introduction to International Disaster Management. Elsevier Science & Technology, 2011. Retrieved 16 July 2012 from http://lib.myilibrary.com?ID=297664.

[29] J-F. Cordeau, M. Gendreau, G. Laporte, J-Y. Potvin, and F. Semet. A guide to vehicle routing heuristics. *Journal of the Operational Research Society*, 53:512–522, 2002.

[30] Carlos Cotta. Effective patient prioritization in mass casualty incidents using hyperheuristics and the pilot method. *OR Spectrum*, 33:699–720, 2011.

[31] I. Das and J. E. Dennis. A closer look at drawbacks of minimizing weighted sums of objectives for pareto set generation in multicriteria optimization problems. *Structural Optimization*, 14:63–69, 1997.

[32] Jan de Boer. Definition and classification of disasters: Introduction of a disaster severity scale. *The Journal of Emergency Medicine*, 8:591–595, 1990.

[33] J. Peral G. de Ceballos, F. Turgano-Fuentes, D. Perez-Diaz, M. Sanz-Sanchez, C. Martin-Llorente, and J.E. Guerrero-Sanz. 11 march 2004: The terrorist bomb explosions in madrid, spain an analysis of the logistics, injuries sustained and clinical management of casualties treated at the closest hospital. *Critical Care*, 9:104–111, 2005.

[34] Luis E. de la Torre, Irina S. Dolinskaya, and Karen R. Smilowitz. Disaster relief routing: Integrating research and practice. *Socio-Economic Planning Sciences*, 46:88–97, 2012.

[35] Department of Health. Ambulance staff in the inner cordon: Hazardous area response teams (hart). *http://www.ambulancehart.org.uk/site-content-files/files/hart-dl-13.05.08-final.pdf, accessed 26th August 2011*, 2008.

[36] Karl F. Doerner and Richard F. Hartl. Health care logistics, emergency preparedness, and disaster relief: New challenges for routing problems with a focus on the austrian situation. In Bruce Golden, S. Raghavan, and Edward Wasil, editors, *The Vehicle Routing Problem: Latest Advances and New Challenges*. Springer, 2007.

[37] Wolf Dombrowsky. Again and again: Is a disaster what we call disaster? some conceptual notes on conceptualizing the object of disaster sociology. . *International Journal of Mass Emergencies and Disasters*, 13(3):241 – 254, 1995.

[38] Christophe Dutang and Petr Savicky. *randtoolbox: Generating and Testing Random Numbers*, 2013. R package version 1.13.

[39] Burak Eksioglu, Arif Volkan Vural, and Arnold Reisman. The vehicle routing problem: A taxonomic review. *Computers & Industrial Engineering*, 57:14721483, 2009.

[40] Hagen Engelmann. DMT-EOC - an integrated approach for training and decision support of EOC members. In *International Symposium on Strong Vrancea Earthquakes and Risk Mitigation*, 2007.

[41] Hagen Engelmann and Frank Fiedrich. Decision support for the memebers of an emergency operation centre after an earthquake. In *Proceedings of the 4th International ISCRAM Conference*, 2007.

[42] Hagen Englemann and Frank Fiedrich. DMT-EOC a combined system for the decision support and training of EOC members. In *Proceedings of the 6th International ISCRAM Conference*, 2009.

[43] Jian Fang and Yugeng Xi. A rolling horizon job shop rescheduling strategy in the dynamic environment. *The International Journal of Advanced Manufacturing Technology*, 13(3):227–232, 1997.

[44] Samer Faraj and Yan Xiao. Coordination in fast-response organizations. *Management Science*, 52:1155–1169, 2006.

[45] FEMA. National preparedness guidelines, September 2007. U.S. Department of Homeland Security.

[46] F. Fiedrich, F. Gehbauer, and U. Rickers. Optimized resource allocation for emergency response after earthquake disasters. *Safety Science*, 35:41–57, 2000.

[47] Frank Fiedrich. An hla-based mulitagent system for optimized resource allocation after strong earthquakes. In *Proceedings of the 2006 Winter Simulation Conference*, 2006.

[48] Frank Fiedrich and Paul Burghardt. Agent-based systems for disaster management. In *Communications of the ACM*, volume 50, March 2007.

[49] Peter C. Fishburn. Axioms for lexicographic preferences. *The Review of Economic Studies*, 42:415–419, 1975.

[50] G. Fitzgerald and N.L. Russo. The turnaround of the London Ambulance Service Computer-Aided Despatch system (LASCAD). *European Journal of Information Systems*, 14 (3):244–257, 2005.

[51] Krzysztof Fleszar, Ibrahim H. Osman, and Khalil S. Hindi. A variable neighbourhood search algorithm for the open vehicle routing problem. *European Journal of Operational Research*, 195:803–809, 2009.

[52] Eric R. Frykberg. Medical management of disasters and mass casualties from terrorist bombings: How can we cope? *Journal of Trauma-Injury Infection & Critical Care*, 53:201–212, 2002.

[53] Eric R. Frykberg. Terrorist bombings in madrid. *Critcal Care*, 9:20–22, 2005.

[54] M. Gad-El-Hak. The art and science of large-scale disasters. *Bulletin of the Polish Academy of Sciences*, 57:3–34, 2009.

[55] Gina Galindo and Rajan Batta. Review of recent developments in OR/MS research in disaster operations management. *European Journal of Operational Research*, 230:201–211, 2013.

[56] Jie Gao, Linyan Sun, and Mitsuo Gen. A hybrid genetic and variable neighborhood descent algorithm for flexible job shop scheduling problems. *Computers & Operations Research*, 35:2892–2907, 2008.

[57] Everette S. Gardner. Exponential smoothing: The state of the art. *Journal of Forecasting*, 4(1):1–28, 1985.

[58] Michel Gendreau and Jean-Yves Potvin. Metaheuristics in combinatorial optimization. *Annals of Operations Research*, 140:189–213, 2005.

[59] Gianpaolo Ghiani, Francesca Guerriero, Gilbert Laporte, and Roberto Musmanno. Real-time vehicle routing: Solution concepts, algorithms and parallel computing strategies. *European Journal of Operational Research*, 151(1):1–11, 2003.

[60] Qiang Gong and Rajan Batta. Allocation and reallocation of ambulances to casualty clusters in a disaster relief operation. *IIE Transactions*, 39 (1):27–39, 2007.

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

[62] Glenn I. Hawe, Graham Coates, Duncan T. Wilson, and Roger S. Crouch. Investigating the effect of overtriage on hospital arrival times of critically injured casualties during a major incident using agent-based simulation. In *Proceedings of the 6th International Conference on Soft Computing and Intelligent Systems and the 13th International Symposium on Advanced Intelligent Systems (SCIS-ISIS)*, 2012.

[63] Zhi hua Hu. Multi-objective model for emergency logistics distribution with multiple supply points and multiple resource catagories. In *2nd Internation Conference on Industrial and Information Systems*, 2010.

[64] Kenneth V. Iserson and John C. Moskop. Triage in medicine, part I: Concept, history, and types. *Annals of Emergency Medicine*, 49:275–281, 2007.

[65] Yiping Jiang, Yufei Yuan, Kai Huang, and Lindu Zhao. Logistics for large-scale disaster response: Achievements and challenges. In *45th Hawaii International Conference on System Sciences*, 2012.

[66] Dylan F. Jones and Mehrdad Tamiz. Goal programming in the period 1990 - 2000. In Matthias Ehrgott and Xavier Gandibleux, editors, *Multiple Criteria Optimization: State of the Art Annotated Bibliographic Surveys*, volume 52 of *International Series in Operations Research and Management Science*, pages 129–170. Springer US, 2003.

[67] Arun Jotshi, Qiang Gong, and Rajan Batta. Dispatching and routing of emergency vehicles in disaster mitigation using data fusion. *Socio-Economic Planning Sciences*, 43:1–24, 2009.

[68] Dieter Jungnickel. *Graphs, Networks and Algorithms*. Springer, 2003.

[69] Imed Kacem, Slim Hammadi, and Pierre Borne. Approach by localization and multiobjective evolutionary optimization for flexible job-shop scheduling problems. *IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews*, 32:1–13, 2002.

[70] Imed Kacem, Slim Hammadi, and Pierre Borne. Pareto-optimality approach for flexible job-shop scheduling problems: hybridization of evolutionary algorithms and fuzzy logic. *Mathematics and Computers in Simulation*, 60:245–276, 2002.

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

[72] Peter Kolesar, Warren Walker, and Jack Hausner. Determining the relation between fire engine travel times and travel distances in New York City. *Operations Research*, 23:614–627, 1975.

[73] Kathleen M. Kowalski-Trakofler, Charles Vaught, and Ted Scharf. Judgment and decision making under stress: an overview for emergency managers. *International Journal of Emergency Management*, 1:278–289, 2003.

[74] Mei-Po Kwan and Jiyeong Lee. Emergency response after 9/11: the potential of real-time 3d gis for quick emergency response in micro-spatial environments. *Computers, Environment and Urban Systems*, 29:93113, 2005.

[75] Gilbert Laporte. Fifty years of vehicle routing. *Transportation Science November*, 43:408–416, 2009.

[76] Richard C. Larson, Michael D. Metzger, and Michael F. Cahn. Responding to emergencies: Lessons learned and the need for analysis. *Interfaces*, 36(6):486–501, 2006.

[77] Y. M. Lee, S. Ghosh, and M. Ettl. Simulating distribution of emergency relief supplies for disaster response operations. In *Proceedings of the 2009 Winter Simulation Conference (WSC)*, 2009.

[78] Junqing Li and Quanke Pan. A hybrid pareto-based local search for multi-objective flexible job shop scheduling problem. In *IEEE Congress on Evolutionary Computation (CEC)*, 2010.

[79] Yen-Hung Lin, Rajan Batta, Peter A. Rogerson, Alan Blatt, and Marie Flanigan. A logistics model for emergency supply of critical items in the aftermath of a disaster. *Socio-Economic Planning Sciences*, 45:132–145, 2011.

[80] D.J. Lockey, R. MacKenzie, J. Redhead, D. Wise, T. Harris, A. Weaver, K. Hines, and G.E. Davies. London bombings July 2005: The immediate pre-hospital medical response. *Resuscitation*, 66(2):ix – xii, 2005.

[81] London Assembly. Report of the 7 July review committee., 2006.

[82] J. Mattingley, Y. Wang, and S. Boyd. Receding horizon control. *Control Systems, IEEE*, 31(3):52–65, June 2011.

[83] M.T. Melo, S. Nickel, and F. Saldanha da Gama. Facility location and supply chain management a review. *European Journal of Operational Research*, 196(2):401 – 412, 2009.

[84] David Mester and Olli Brysy. Active-guided evolution strategies for large-scale capacitated vehicle routing problems. *Computers & Operations Research*, 34:29642975, 2007.

[85] Huseyin Onur Mete and Zelda B Zabinsky. Stochastic optimization of medical supply location and distribution in disaster management. *International Journal of Production Economics*, 126:76–84, 2010.

[86] Kaisa Miettinen. *Nonlinear Mulitiobjective Optimization*. Kluwer Academic Publishers, 1999.

[87] Ministry Of Defence. Operations in the uk: The defence contribution to resilience, 2nd edition. *http://www.mod.uk/NR/rdonlyres/80095393-A7B0-4B3A-B69A-D13742C29AAF/0/20071218-JDP02-2nd-Edition-U-DCDCIMAPPS.pdf, accessed 26th August 2011*, 2007.

[88] G.F Newell. The rolling horizon scheme of traffic signal control. *Transportation Research Part A: Policy and Practice*, 32(1):39 – 44, 1998.

[89] Sandra Ulrich Ngueveu, Christian Prins, and Roberto Wolfler Calvo. An effective memetic algorithm for the cumulative capacitated vehicle routing problem. *Computers & Operations Research*, 37:18771885, 2010.

[90] L. Ozdamar and G. Ulusoy. A survey on the resource-constrained project scheduling problem. *IIE Transactions*, 27(5):574–586, 1995.

[91] Linet Ozdamar. Planning helicopter logistics in disaster relief. *OR Spectrum*, 33:655–672, 2011.

[92] Linet Ozdamar, Ediz Ekinci, and Beste Kucukyazici. Emergency logistics planning in natural disasters. *Annals of Operations Research*, 129:217–245, 2004.

[93] Linet Ozdamar and Chandra Sekhar Pedamallu. A comparison of two mathematical models for earthquake relief logistics. *International Journal of Logistics Systems and Management*, 10:361–373, 2011.

[94] Douglas Paton and Rhona Flin. Disaster stress: an emergency management perspective. *Disaster Prevention and Management*, 8:261–267, 1999.

[95] F. Pezzella, G. Morganti, and G. Ciaschetti. A genetic algorithm for the flexible job-shop scheduling problem. *Computers & Operations Research*, 35:3202–3212, 2008.

[96] Moshe Pinkert, Ofer Lehavi, Odeda Benin Goren, Yaron Raiter, Ari Shamis, Zvi Priel, Dagan Schwartz, Avishay Goldberg, Yehezkel Levi, and Yaron Bar-Dayan. Primary triage, evacuation priorities, and rapid primary distribution between adjacent hospitals - lessons learned from a suicide bomber attack in downtown tel-aviv. *Prehospital and Disaster Medicine*, 23:337–341, 2008.

[97] Gary Preece, Duncan Shaw, and Haruo Hayashi. Using the viable system model (VSM) to structure information processing complexity in disaster response. *European Journal of Operational Research*, 224:209–218, 2013.

[98] Gunther R. Raidl. A unified view on hybrid metaheuristics. *Lecture Notes in Computer Science*, 4030:1–12, 2006.

[99] M. Rauner, M. Schaffhauser-Linzatti, and H. Niessner. Resource planning for ambulance services in mass casualty incidents: a des-based policy model. *Health Care Management Science*, 15:254–269, 2012.

[100] Carmen G. Rawls and Mark A. Turnquist. Pre-positioning of emergency supplies for disaster response. *Transportation Research Part B: Methodological*, 44:521–534, 2010.

[101] Carmen G. Rawls and Mark A. Turnquist. Pre-positioning and dynamic delivery planning for short-term response following a natural disaster. *Socio-Economic Planning Sciences*, 46:46–54, 2012.

[102] Colin R Reeves, editor. *Modern Heuristic Techniques for Combinatorial Problems*. McGraw-Hill, 1995.

[103] Colin R. Reeves. Fitness landscapes. In *Search Methodologies*. Springer Science+business Media, inc., 2005.

[104] Christian M. Reidys and Peter F. Stadler. Combinatorial landscapes. *SIAM Review*, 44:3–54, 2002.

[105] Glaydston Mattos Ribeiro and Gilbert Laporte. An adaptive large neighborhood search heuristic for the cumulative capacitated vehicle routing problem. *Computers & Operations Research Volume 39, Issue 3, March 2012, Pages 728735*, 39:728735, 2012.

[106] Erik Rolland, Raymond A. Patterson, Keith Ward, and Bajis Dodin. Decision support for disaster management. *Operations Management Research*, 3:68–79, 2010.

[107] Peter Ross. Hyper-heuristics. In *Search Methodologies*. Springer Science and business Media, inc., 2005.

[108] William J. Sacco, D. Michael Navin, Katherine E. Fiedler, Robert K. Waddell II, William B. Long, and Robert F. Buckman Jr. Precise formulation and evidence-based application of resource-constrained triage. *Academic Emergency Medicine*, 12:759770, 2005.

[109] Mohammad Saidi-Mehrabad and Parviz Fattahi. Flexible job shop scheduling with tabu search algorithms. *The International Journal of Advanced Manufacturing Technology*, 32:563–570, 2007.

[110] Javier Salmeron and Aruna Apte. Stochastic optimisation for natural disaster asset prepositioning. *Production and Operations Management*, 19:561–574, 2010.

[111] Narjes Bellamine-Ben Saoud, Tarek Ben Mena, Julie Dugdale, Bernard Pavard, and Mohamed Ben Ahmed. Assessing large-scale emergency rescue plans: an agent-based approach. *Int. Jour. of Int. Cont. & Syst.*, 11:260–271, 2006.

[112] Bhaba R. Sarker, Jr. Lawrence Mann, Evangelos Triantaphyllou, and Srinivas Mahankali. Power restoration in emergency situations. *Computers ind. Engng*, 31:367 – 370, 1996.

[113] Jiuh-Biing Sheu. An emergency logistics distribution approach for quick response to urgent relief demand in disasters. *Transportation Research Part E: Logistics and Transportation Review*, 43:687 – 709, 2007.

[114] Jeremy G. Siek, Lee-Quan Lee, and Andrew Lumsdaine. *The Boost Graph Library: User Guide and Reference Manual.* Addison-Wesley, 2002.

[115] E.A. Silver. An overview of heuristic solution methods. *Journal of the Operational Research Society*, 55:936–956, 2004.

[116] N.C. Simpson and P.G. Hancock. Fifty years of operational research and emergency response. *Journal of the Operational Research Society*, 60:SI26–SI39, 2009.

[117] Kenneth Srensen. Metaheuristics - the metaphor exposed. *International Transactions In Operational Research*, 00:1–16, 2013.

[118] E. G. Talbi. A taxonomy of hybrid metaheuristics. *Journal of Heuristics*, 8:541–564, 2002.

[119] Bogdan Tatomir and Leon Rothkrantz. Ant based mechanism for crisis response coordination. *Lecture Notes in Computer Science*, 4150:380–387, 2006.

[120] Steven Thompson, Nezih Altay, Walter G. Green, and Joanne Lapetina. Improving disaster response efforts with decision support systems. *International Journal of Emergency Management*, 3(4):250–263, 2006.

[121] Suleyman Tufekci and William A. Wallace. The emerging area of emergency management and engineering. *IEEE Transactions on Engineering Management*, 45 (2):103–105, 1998.

[122] Fernando Turgano-Fuentes, P. Caba-Doussoux, J. M. Jover-Navaln, E. Martn-Prez, D. Fernndez-Luengas, L. Dez-Valladares, D. Prez-Daz, P. Yuste-Garca, H. Guadalajara Labajo, and et al. R. Ros-Blanco. Injury patterns from major urban terrorist bombings in trains: The madrid experience. *World Journal of Surgery*, 32:1168–1175, 2008.

[123] Fernando Turgano-Fuentes, Dolores Prez-Daz, Mercedes Sanz-Snchez, and Javier Ortiz Alonso. Overall asessment of the response to terrorist bombings in trains, Madrid, 11 March 2004. *European Journal of Trauma and Emergency Surgery*, 34:433–441, 2008.

[124] Gwo-Hshiung Tzeng, Hsin-Jung Cheng, and Tsung Dow Huang. Multi-objective optimal planning for designing relief delivery systems. *Transportation Research Part E: Logistics and Transportation Review*, 43:673 – 686, 2007.

[125] UK Cabinet Office. Civil contingencies act 2004: a short guide (revised). *http://www.cabinetoffice.gov.uk/media/132428/15mayshortguide.pdf, accessed 26th August 2011*, 2004.

[126] UK Cabinet Office. Logistics operations for emergency supplies, April 2009.

[127] UK Cabinet Office. Emergency response and recovery - non statutory guidance accompanying the civil contingencies act 2004, April 2010.

[128] UK Cabinet Office. Responding to emergencies: the UK central government response, Concept of Operations, 2010. Retrieved 11th February 2013 from http://www.cabinetoffice.gov.uk/sites/default/files/resources/conops-2010.pdf.

[129] UK Cabinet Office. National risk register for civil emergencies, 2013.

[130] Prakarn Unachak and Erik Goodman. Solving multiobjective flexible job-shop scheduling using an adaptive representation. In *Proceedings of the 12th annual conference on Genetic and evolutionary computation*, 2010.

[131] Vector Command. Command support system™, September 2013.

[132] William A. Wallace and Frank de Balogh. Decision support systems for disaster management. *Public Administration Review*, 45:134–146, 1985.

[133] Y. Wang, K. L. Luangkesorn, and L. Shuman. Modeling emergency medical response to a mass casualty incident using agent based simulation. *Socio-Economic Planning Sciences*, 46:281–290, 2012.

[134] Bradford S. Westgate, Dawn B. Woodward, David S. Matteson, and Shane G. Henderson. Travel time estimation for ambulances using bayesian data augmentation. Technical report, Cornell University, 2011.

[135] Felix Wex, Guido Schryen, and Dirk Neumann. Intelligent decision support for centralized coordination during emergency response. In *Proceedings of the 8th International ISCRAM Conference*, 2011.

[136] Felix Wex, Guido Schryen, and Dirk Neumann. Operational emergency response under informational uncertainty: A fuzzy optimization model for scheduling and allocating rescue units. In *Proceedings of the 9th International ISCRAM Conference*, 2012.

[137] W.E. Wilhelm and A.V. Srinivasa. Prescribing tactical response for oil spill clean up operations. *Management Science*, 43 (3):386–402, 1997.

[138] Duncan T. Wilson, Glenn I. Hawe, Graham Coates, and Roger S. Crouch. Stochastic task durations in the scheduling of mass casualty incident response operations. In *Proceedings of the 30th Workshop of the UK Planning And Scheduling Special Interest Group (PlanSIG)*, 2012.

[139] Shengnan Wu, Larry Shuman, and Bopaya Bidanda. Disaster policy optimization: A simulation based approach. In *Proceedings of the 2007 Industrial Engineering Research Conference*, 2007.

[140] Shengnan Wu, Larry Shuman, Bopaya Bidanda, Oleg Prokopyev, Matthew Kelley, Ken Sochats, and Carey Balaban. Simulation-based decision support system for real-time disaster response management. In *Proceedings of the 2008 Industrial Engineering Research Conference*, 2008.

[141] Jean-Luc Wybo. Fmis: A decision support system of forest fire prevention and fighting. *IEEE TRANSACTIONS ON ENGINEERING MANAGEMENT*, 45(2):127–131, 1998.

[142] Li-Ning Xing, Ying-Wu Chen, Peng Wang, Qing-Song Zhao, and Jian Xion. A knowledge-based ant colony optimization for flexible job shop scheduling problem. *Applied Soft Computin*, 10:888896, 2010.

[143] Wei Yi and Arun Kumar. Ant colony optimization for disaster relief operations. *Transportation Research Part E: Logistics and Transportation Review*, 43:660–672, 2007.

[144] Wei Yi and Linet Ozdamar. A dynamic logistics coordination model for evacuation and support in disaster response activities. *European Journal of Operational Research*, 179:1177–1193, 2007.

[145] Yuan Yuan and Dingwei Wang. Path selection model and algorithm for emergency logistics management. *Computers & Industrial Engineering*, 56:1081–1094, 2009.

[146] Dongsong Zhang, Lina Zhou, and Jay F. Nunamaker Jr. A knowledge management framework for the support of decision making in humanitarian assistance/disaster relief. *Knowledge and Information Systems*, 4:370 – 385, 2002.

[147] Jiang-Hua Zhang, Jin Li, and Zhi-Ping Liu. Multiple-resource and multiple-depot emergency response problem considering secondary disasters. *Expert Systems with Applications*, 39:11066–11071, 2012.

[148] Xiaoge Zhang, Zili Zhang, Yajuan Zhang, Daijun Wei, and Yong Deng. Route selection for emergency logistics management: A bio-inspired algorithm. *Safety Science*, 54:87–91, 2013.

[149] Kostas G. Zografos, Christos Douligeris, and Panagiotis Tsoumpas. An integrated framework for managing emergency-response logistics: The case of the electric utility companies. *IEEE Transactions on Engineering Management*, 45 (2):115 – 126, 1998.

[150] Richard M. Zoraster, Cathy Chidester, and William Koenig. Field triage and patient maldistribution in a mass-casualty incident. *Prehospital and Disaster Medicine*, 22:224–229, 2007.