City Research Online

City St George's, University of London

Citation: Wilson, D.T., Hawe, G.I., Coates, G. & Crouch, R.S. (2016). Online optimization of casualty processing in major incident response: An experimental analysis. European Journal of Operational Research, 252(1), pp. 334-348. doi: 10.1016/j.ejor.2016.01.021

This is the published version of the paper.

This version of the publication may differ from the final published version. To cite this item please consult the publisher's version.

Permanent repository link: https://openaccess.city.ac.uk/id/eprint/14442/

Link to published version: https://doi.org/10.1016/j.ejor.2016.01.021

Copyright and Reuse: Copyright and Moral Rights remain with the author(s) and/or copyright holders. Copies of full items can be used for personal research or study, educational, or not-for-profit purposes without prior permission or charge, unless otherwise indicated, provided that the authors, title and full bibliographic details are credited, a hyperlink and/or URL is given for the original metadata page and the content is not changed in any way. For full details of reuse please refer to City Research Online policy.

City Research Online: http://openaccess.city.ac.uk/ publications@citystgeorges.ac.uk

European Journal of Operational Research 252 (2016) 334–348

Contents lists available at ScienceDirect

European Journal of Operational Research

journal homepage: www.elsevier.com/locate/ejor

Innovative Applications of O.R.

Online optimization of casualty processing in major incident response: An experimental analysis

Duncan T. Wilson*, Glenn I. Hawe, Graham Coates, Roger S. Crouch

School of Engineering and Computing Sciences, Durham University, South Road, Durham, DH1 3LE, UK

ARTICLE INFO

Article history:
Received 18 July 2013
Accepted 13 January 2016
Available online 22 January 2016

Keywords:
Scheduling
Combinatorial optimization
Emergency response
Disaster management
Dynamic

ABSTRACT

When designing an optimization model for use in mass casualty incident (MCI) response, the dynamic and uncertain nature of the problem environment poses a significant challenge. Many key problem parameters, such as the number of casualties to be processed, will typically change as the response operation progresses. Other parameters, such as the time required to complete key response tasks, must be estimated and are therefore prone to errors. In this work we extend a multi-objective combinatorial optimization model for MCI response to improve performance in dynamic and uncertain environments. The model is developed to allow for use in real time, with continuous communication between the optimization model and problem environment. A simulation of this problem environment is described, allowing for a series of computational experiments evaluating how model utility is influenced by a range of key dynamic or uncertain problem and model characteristics. It is demonstrated that the move to an online system mitigates against poor communication speed, while errors in the estimation of task duration parameters are shown to significantly reduce model utility.

© 2016 The Authors. Published by Elsevier B.V.
This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

1. Introduction

In the period immediately following a mass casualty incident (MCI), such as the London Bombings of July 7th 2005 (London Assembly, 2006), many decisions need to be made in a fast and effective manner within a high pressure environment (Paton & Flin, 1999). Within emergency response organizations such as the Ambulance Service and the Fire and Rescue Service, decision makers must decide how best to allocate their limited resources amongst the various sources of demand. This problem environment exhibits a large amount of structure, with well defined roles and responsibilities and a clear decision making system as defined through the command and control system (Wallace & de Balogh, 1985). In this respect, the problem represents a strong candidate for the application of mathematical modeling and optimization. However, significant challenges remain, particularly with respect to the volatile nature of the problem environment. That is, the nature of any decision problem is likely to change over time as the problem evolves, and the available information upon which a model can be built will typically be subject to a significant level of uncertainty (Galindo & Batta, 2013).

In the modeling of MCI response, as with the design of any optimization model, it is necessary to make certain assumptions in order to ensure the implementation remains feasible. In this paper we seek to gain a better understanding of several characteristics of the response problem, their associated assumptions, and the extent to which they affect the utility of a scheduling-based optimization model. In order to proceed we first discuss a number of assumptions common to optimization models for MCI response. We cover the modeling of casualty health, their allocation to hospitals for treatment, the transportation of casualties and responders around the response environment, and the representation of tasks which responders must carry out. We go on to focus on how others have considered the dynamic and uncertain nature of the response environment in their models. Based on our findings, we identify gaps that remain uncovered in the literature and we discuss how our research contributes to fill such gaps.

1.1. Common modeling assumptions

Some common assumptions made in the design of operational research models for disaster operations management are identified in Galindo and Batta (2013). Further common assumptions covering the more general area of disaster planning are listed in Auf der Heide (2006).

Depending on the general form of the model, the parameters needed to specify its form can include variables such as

* Corresponding author. Tel.: +44 1904 659005.
E-mail address: d.t.wilson@dur.ac.uk, d.t.wilson@leeds.ac.uk, duncantwilson@gmail.com (D.T. Wilson).

http://dx.doi.org/10.1016/j.ejor.2016.01.021
0377-2217/© 2016 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

commodity supply and demand levels, resource requirements for specific tasks, and the number and nature of casualties. As noted in Galindo and Batta (2013), it is common for models to assume that

1. the information needed to deduce these parameters is available and accurate upon initialization of the model, and
2. the parameters are not required to change over time.

The extent to which these assumptions are justified depends on the specific problem under consideration, but will often be limited by the intrinsic uncertainty and volatility common to all emergency response problems. Some specific examples follow.

1.1.1. Casualty health

Some authors assume there are no meaningful differences between the health levels of casualties (Barbarosoglu & Arda, 2004; Barbarosoglu, Ozdamar, & Cevik, 2002; Mete & Zabinsky, 2010; Rolland, Patterson, Ward, & Dodin, 2010; Wex, Schryen, & Neumann, 2011; 2012). Where differences are acknowledged, it is common to assume all casualties have been partitioned into discrete categories reflecting the urgency of their treatment (Galindo & Batta, 2013), as in the work of Chiu and Zheng (2007); Gong and Batta (2007); Yi and Ozdamar (2007). This is reasonable, as it is normal for an assessment of the health of each casualty (known as triage) to be completed before the remainder of the response is enacted (Group, 2011). It is often assumed that individual casualty health will not change over time, and that assessments of health are always accurate. The attraction of the former assumption is understandable, as the task of accurately forecasting the changing health of casualties in these environments is challenging. Some attempts are described in Cotta (2011); Fiedrich, Gehbauer, and Rickers (2000); Tatomir and Rothkrantz (2006). These models, however, do not provide any way to correct errors in prediction, an occurrence which we can assume to be likely due to the complexity of the underlying process.

1.1.2. Hospitals

Many models assume that the allocation of casualties to hospitals will be done automatically and appropriately. Limited examples of including hospital allocation into a wider decision problem can be found in Jotshi, Gong, and Batta (2009); Mysore et al. (2005); Wilson, Hawe, Coates, and Crouch (2013a). In Wilson et al. (2013a) an often ignored aspect of casualty management, *self presentation*, is discussed. It is often assumed that all casualties are transported to hospital by the Ambulance Service only (Auf der Heide, 2006), with the casualty undergoing triage and treatment operations prior to this. In reality, it is common for some casualties to remove themselves from the incident site and transport themselves to a hospital of their choosing. In Wilson et al. (2013a) it is assumed that this process could be predicted accurately. In scenarios where this is not possible, a dynamic approach, updating the model regarding the number of casualties who have left the incident scene and who have arrived at each hospital, may be effective.

1.1.3. Transportation

The transport network within the problem environment is often assumed to be known, both in terms of topology and the travel times between locations (Yi & Kumar, 2007; Zhang, Li, & Liu, 2012). As noted in Galindo and Batta (2013), the former assumption is more justified than the latter. Examples of removing the latter assumption include (Wilson, Hawe, Coates, & Crouch, 2013b). In this work it is demonstrated that disruption to the network resulting in uncertainty in travel times can have a significant effect on the performance of an optimization model. As such, this problem characteristic should not be ignored.

Uncertainty in the disruption of the transport network has been incorporated to a limited extent using stochastic programming formulations. Examples include (Barbarosoglu & Arda, 2004; Mete & Zabinsky, 2010; Rawls & Turnquist, 2010), which consider a finite number of scenarios, each with assigned probability and associated network parametrization. Uncertainty is also acknowledged in the work of Jotshi et al. (2009), which extends the ambulance allocation model presented in Gong and Batta (2007) by including a data fusion step to estimate the level of damage and disruption on each road link. A solution methodology for finding optimal paths in a disrupted network following a disaster is presented in (Zhang, Zhang, Zhang, Wei, & Deng, 2013). The authors employ the network representation described by Yuan and Wang (2009), 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 Fiedrich et al. (2000), with nodes and edges being added or removed to reflect the impact of both the disaster and the response operation.

1.1.4. Task durations

Where the modeling methodology involves the allocation of discrete tasks to available responder units, the times needed to complete these tasks are necessary problem parameters. Examples include the scheduling models presented in Rolland et al. (2010) and Wex et al. (2011). In the former, the authors propose a specific solution algorithm which, through its fast execution, is designed to facilitate the solving of their proposed 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. In Wex et al. (2011) a similar modeling methodology is proposed, where all necessary parameters are assumed to be fixed and known upon model initialization. This model is extended in Wex, Schryen, and Neumann (2012), allowing for task durations to be represented by fuzzy values in an effort to acknowledge the uncertainty inherent in available information. The authors suggest the model should be regularly rebuilt and solved when the problem environment has evolved by some significant degree.

1.2. Modeling uncertainty and dynamicity

All the assumptions mentioned relate to model parameters which change over time, either because they are estimates of unknown real values and can therefore be revised as new information comes to light, or because the real values themselves are of a dynamic nature, or both. In the worst cases these assumptions will render a model unusable in many realistic scenarios. General strategies to their removal tend to take either a stochastic yet static approach, applying stochastic (Barbarosoglu & Arda, 2004; Chang, Tseng, & Chen, 2007; Mete & Zabinsky, 2010) or robust (Bozorgi-Amiri, Jabalameli, Alinaghian, & Heydari, 2012) programming to find solutions which will remain valid as the problem evolves over time, or a dynamic approach, allowing for the model to be updated at a number of set length intervals to help ensure it remains applicable (see, for example, Lee, Ghosh, & Ettl, 2009; Ozdamar, Ekinci, & Kucukyazici, 2004; Yi & Kumar, 2007). Only limited steps have been taken with the latter approach. In the context of manufacturer or retailer response to hurricanes, the supply chain models proposed in Lodree and Taskin (2009); Taskin and Lodree (2011) employ a Bayesian approach to allow for dynamic information to be incorporated into future decisions. In Gong and Batta (2007) the authors note that determining the appropriate length of update interval is crucial to performance, proposing that future work should

Table 1
Task types and their dependency relations.

Task:	Pre-rescue stabilization	→	Rescue	→	Pre-transportation stabilization	→	Transportation
Responder:	HART		SAR		Ambulance / MERIT / HART		Ambulance / HART
Condition:	Trapped and unstable		Trapped		Unstable		All

look to develop models which operate in continuous time. This is echoed in Chiu and Zheng (2007), 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 order to move towards such a real-time system, the work reported in Engelmann and Fiedrich (2007); Englemann and Fiedrich (2009); Fiedrich (2006); Jotshi et al. (2009) link their proposed decision support models to simulations of the actual response environment, allowing for the testing of the ability of each model to cope with changes in information. However, in these cases the whole decision problem is decomposed into a sequence of single decision points, where tasks are allocated to responders one at a time as and when the responder becomes available. This structure does not allow for the potential benefits of forward planning, as would be available in a scheduling model, to be explored. There has been no detailed investigation of the potential for real-time decision support considering the entire planning horizon. The need for such research is further highlighted in the review of Jiang, Yuan, Huang, and Zhao (2012).

1.3. Contribution

In this paper we describe such a real-time system, building upon the static model presented in Wilson et al. (2013a). The model, coupled with addressing many of the limiting assumptions discussed above, allowing for information to be updated in a realistic manner and for this information to be used to improve future predictions as well as correct past errors, forms the principal contribution of this paper. In addition to this contribution, the paper presents a detailed computational analysis of model performance, identifying a number of potential explanatory parameters and exploring to what extent they impact upon the utility of the optimization model.

The remainder of this paper is structured as follows. In Section 2 we briefly describe a previously published decision support model for casualty processing, given in Wilson et al. (2013a). Following this, Section 3 details how this model has been extended to allow for its use in real-time during an MCI characterized by uncertainty and volatility. The results of extensive computational experiments are then reported and discussed in Section 4. Finally, we draw conclusions and identify promising avenues for future research.

2. A static model of casualty processing

In this work we build upon the multi-objective combinatorial optimization model described in Wilson et al. (2013a). Originally, the model was designed for use in a static manner, being initialized at a point where all relevant information was available and running for the desired length of time before delivering the solution output, which took the form of a work schedule detailing the allocation and ordering of response tasks to available responders. While the model did incorporate a probabilistic approach when describing the evolution of casualty health, no other parameters were of a stochastic nature.

As this model is designed to perform a period of pre-computation before delivering a single solution, we denote it as model M_{pc}. In this section we will describe the key components of this model, with the aim of conveying its nature while minimizing the technical detail which can instead be found in Wilson et al. (2013a). In the following section we will discuss its extension for use in dynamic, evolving problems where many more parameters are subject to uncertainty.

2.1. Solution space

A solution to the casualty processing problem faced in MCI response consists of:

- an allocation of casualties $c \in \mathcal{C}$ to hospitals $h \in \mathcal{H}$,
- an allocation of tasks $t \in \mathcal{T}$ to responders $r \in \mathcal{R}$,
- an ordering of the tasks assigned to each responder r.

The types of tasks which can be found in $\mathcal{T}$ are summarized in Table 1. Each casualty requires the completion of a transportation task, to be carried out by an Ambulance responder unit, in order for them to be taken from the incident site to their allocated hospital. In addition, if the casualty's health is unstable they will also require a pre-transportation stabilizing treatment task to ensure their safe transportation. Such tasks may be carried out by ambulance responder units, a Medical Emergency Response Incident Team (MERIT), or a Hazardous Area Response Team (HART). MERIT units are medical teams who attend incident sites to assist the triage and treatment of casualties (London Emergency Services Liaison Panel, 2015). HART teams are specially trained and equipped for working within the hazardous inner cordon area. In some cases, casualties may require extrication from the incident site by a Search And Rescue (SAR) responder unit, which we shall refer to as a 'rescue' task. Should this be the case, it is possible that a pre-rescue stabilizing treatment task be required in order to reduce the likelihood of the health of the casualty deteriorating during the extrication operation. These tasks may only be completed by HART units.

2.2. Objective functions

Given a solution as defined in Section 2.1, a schedule can be constructed detailing the work plan for each responder, identifying the time at which the responder (a) begins traveling to the location of their next task, (b) begins work on this task, and (c) finishes work on this task. In constructing a schedule from a solution, the spatial nature of the problem is taken into account in estimating the travel times of responders as they move between sites and/or hospitals. These estimates are combined with estimates of task duration when constructing the schedule.

For a given (estimated) schedule, a number of measures are calculated and used to evaluate and compare solutions during the optimization process, together measuring fatalities and suffering. We will briefly describe these functions here and refer the reader to Wilson et al. (2013a) for further details and discussion.

2.2.1. Fatalities

In many countries it is standard practice in MCI response for a full triage of casualties to be carried out before any subsequent tasks may begin. The result is an assessment of the health of each casualty, which is classified according to the four possible categories listed in Table 2.

Our model uses a Markov chain consisting of a state space {*T*1, *T*2, *T*3, *dead*} to predict how the health of a casualty will evolve

Table 2
Triage levels assigned to casualties.

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		

over the course of the response operation. We assume that the health of casualties will only ever decrease when in an unstable environment, that is, before they have been extricated and taken to a safe designated area. For each casualty, the model is used to calculate the probability that they will have died before they reach hospital. These probabilities are summed together to produce an objective of the casualty processing problem,

$$\min f_1(s) = \sum_{c \in \mathcal{C}} \mathrm{P}(c \text{ dead on arrival at hospital}), \tag{1}$$

where s is a solution to the model.

We note that our model assumes the transition probabilities 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. However, we have attempted to set transition probabilities which reflect the qualitative descriptions of triage states, as given in Table 2. For example, we have ensured that the probability of death of a $T1$ casualty who is left untreated in an unstable environment for thirty minutes approaches 1, while for a $T3$ casualty it reaches only 0.1. In selecting these transition probabilities we aim to consider the most generic MCI scenarios. Were the model to be applied to more unique and idiosyncratic scenarios, these parameter values should be adjusted accordingly. It is noted that due to the inherent low frequency of MCIs and the lack of data collection which occurs during them, estimating these probabilities presents a significant challenge. However, one suggested approach to estimate transition probabilities would be to analyze patient data from non-MCI emergency situations in which there are fewer casualties and their health states are monitored more closely. While acknowledging that such data would originate from non-MCIs, it would provide a more realistic basis for their estimation.

2.2.2. Suffering

A second objective of MCI response, f_2, is to minimize suffering. We consider suffering to be quantified through two components. Firstly, for each casualty the time taken from moment of injury to their arrival at hospital is noted. These times are summed together with each individual contribution weighted by the severity of that casualty's health. Secondly, the standard of treatment available at the hospitals to which casualties have been assigned is measured. This is done through forecasting the arrival times of casualties at each hospital and contrasting with predicted resource levels in order to estimate the amount of time casualties will collectively wait at a hospital before treatment is administered. To this we add a penalty term for every casualty who has been assigned to a hospital which does not provide any specialist treatment their injuries require (e.g., those suffering from severe burns should be encouraged to be sent to a hospital with a specialist burns unit). These two measures are combined to form the single suffering objective, f_2, using the weighted metric method of least squares.

2.2.3. Lexicographic ordering

The objectives f_1 and f_2 are combined in a lexicographic manner to reflect the fact that the saving of lives is always of higher priority than the reduction of suffering. The full multi-objective model can now be defined as

$$\min_{s \in \mathcal{S}} \; f_1(s), \; f_2(s). \tag{2}$$

2.3. Solution methodology

2.3.1. Local search

A Variable Neighborhood Descent metaheuristic is employed in order to find high quality solutions to the scheduling problem described above. Four neighborhood structures are employed, each with variable size, which facilitates the local search process escaping local optima through consideration of larger neighborhoods. A similar approach has been shown to perform well on a flexible job shop problem (Amiri, Zandieh, Yazdani, & Bagheri, 2010), which is of a similar structure to the model described in Section 2.1. As described in Wilson et al. (2013a), the algorithm employs four different neighborhood structures, cycling between them at each iteration. When a certain neighborhood structure results in no neighboring solutions which improve upon the current solution, the size of that neighborhood is increased. For example, one neighborhood structure allows for any two tasks to be swapped, in terms of their responder allocation and their position in that responder's schedule. Increasing the size of this neighborhood allows for two of these 'swap' operations to be carried out in a single step. Accordingly, increasing the size of the neighborhood increases the likelihood of finding an improving solution. This strategy enables the search process to escape any local optima it finds itself in.

2.3.2. Constructive heuristic

In addition to a local search solution methodology, Wilson et al. (2013a) also provides details of a heuristic routine which can be applied in a constructive manner. Specifically, the constructor builds a solution by allocating tasks to the end of responders' schedules until all tasks have been allocated. At each decision point, the responder chosen is the one which is due to finish all their tasks first. A task to be allocated to the end of their schedule is chosen by considering a number of criteria, such as the time at which the task could begin and the health of the associated casualty, in a lexicographic manner. The constructor is designed to approximate how decisions would be made on the ground of an MCI, focussing on the immediate situation as opposed to planning ahead.

3. An online model of casualty processing

Having described the pre-computation model M_{pc} in Section 2, we now consider its extension to more realistic problems subject to high volatility and associated uncertainty in model parameters. We denote this online model by M_o. In the following discussion we shall partition all such parameters into two sets. By *solution space* parameters, we refer to those which affect the nature of the solution space, as described in Section 2.1. 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.

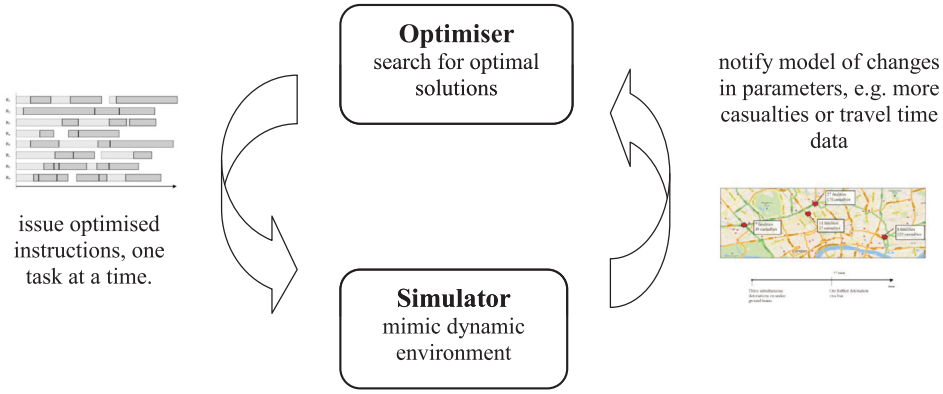

Fig. 1. Two-way communication between optimization and response environment.

We describe problems which do not include any dynamic characteristics as *static* problems. Those which include dynamic objective spaces are described as *partially dynamic* problems. Finally, those problems which exhibit dynamic behavior in both the solution and objective spaces are denoted *fully dynamic* problems. In this section we will first describe how the model can be used in real-time by allowing instructions to be issued to responders gradually, one task at a time, as opposed to issuing a full schedule at a single time point. We then go on to describe dynamic features of the problem which result in changes to solutions space parameters, before finally considering features leading to changes to the objective space.

3.1. Real-time and online optimization

A necessary first step in adapting the model is allowing for the model to pass instruction to the problem environment in a gradual manner as opposed to at one single point in time. This is accomplished through the partitioning of all tasks within the model into two complementary sets: *fixed*, denoting tasks which have been given as instructions and which a responder unit has begun; and *free*, denoting all tasks yet to be issued to a responder unit.

Employing the local search optimization procedure in *real-time* means that, at any given point in the response operation, the local search procedure is carried out over only the set of free tasks, adjusting their positions in the schedule in an attempt to find a solution of higher quality. At the outset of the operation, optimization is carried out over all tasks in the model. Towards the end of the operation, where the majority of tasks have been carried by responders and are now fixed in their positions in the schedule, optimization may involve only a handful of remaining free tasks. When a responder unit becomes available, their next task is chosen based on the best overall schedule found so far by the optimization algorithm. After this task is issued and becomes fixed, optimization continues considering the remaining fee tasks. Thus, a responder's schedule is not fixed at time $\tau = 0$, but rather continuously built as the response operation progresses. In this manner, the optimization model can be used in real-time as the event unfolds, regularly issuing instructions. This is in contrast with the usual offline approach, where the model issues a full schedule of instructions once, at the outset of the response operation.

It should be emphasized that by employing the optimization procedure in real-time, the common concern of algorithm computation time is no longer of direct relevance. Usually, the evaluation of an algorithm would concern both the quality of the solutions it suggests, and the time required to do so. In designing an algorithm, one would trade-off these two characteristics to achieve the right balance for the problem at hand. In our case, however, we continue to optimise over the current set of free tasks until a responder requires instruction. There is no benefit in pausing or terminating the optimization procedure before this, and so we are not concerned with trading off computation time for solution quality. An inefficient or slow algorithm will impact the quality of the proposed solutions, but this impact will be entirely encapsulated by the final solution quality observed upon completion of the response operation.

The continuous passing of instructions from the model to the response environment is complemented by the continuous feedback of information from the environment to the model in what we term *online* optimization. There, any changes in the environment which are relevant to the model are noted and passed back to the model as they are observed, to allow for the model to be updated and reflect the problem more accurately. This process of continuous two-way communication is illustrated in Fig. 1.

In the remainder of this Section, we describe the various changes in the response environment which can be updated within the model, and how these changes can be simulated for the purposes of experimental evaluation.

3.2. Solution space parameters

As described in Section 2.1, the decision problem modeled 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.

As the hospitals available for use in the response operation are unlikely to alter, we do not consider any dynamic changes to the set $\mathcal{H}$. Regarding the set of available responder units, we note that this can both increase and decrease as the response operation progresses. As discussed in Auf der Heide (2006), it is common for responders from areas neighboring 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 time-scale of problem scenarios considered in this paper we do not account for this possibility.

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, and in a sudden manner, if another incident were to occur nearby. Moreover, a decrease in the number of casualties can occur due to self presentation. An illustration of the dynamic nature of casualty and resource numbers is provided in

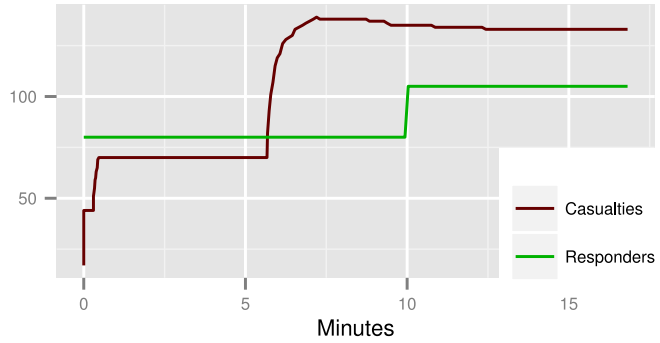

Fig. 2. Changes in the numbers of casualties and responders over the course of an example MCI response operation.

Fig. 2 which charts the corresponding set size over the first fifteen minutes of the problem which will be considered in Section 4.

In order to extend the model to allow for these changes, we require (a) a heuristic procedure to govern the re-assignment of tasks assigned to a responder when he/she leaves the set, and (b) a heuristic procedure to govern the assignment of tasks associated with a newly discovered casualty. The constructive heuristic procedure described in Section 2.3.2 can easily be employed for this purpose. Once tasks have been initially assigned in this manner, the local search procedure can go on to find higher quality allocations.

3.3. Objective space parameters

Regarding the objective values assigned to any proposed solution, a number of parameters upon which these values depend are subject to uncertainty, or are of a dynamic nature, or both. Specifically, the health parameter associated with each casualty will be subject to measurement error, known more commonly as under-triage or over-triage. In addition, health will evolve with time and so this evolution must be predicted, introducing further uncertainty. Schedule parameters will also be subject to uncertainty; both the time needed to travel from one destination to another and the duration of certain tasks must be estimated from the information available at that point in time, and will generally have some degree of error. Moreover, we allow for possible delays in the communication of such information from the problem environment to the model. The dynamic and uncertain nature of these parameters results in a schedule which evolves over the course of the response operation.

A simple example of an evolving schedule is given in Fig. 3. The illustration shows the schedule of a single responder, as viewed from the perspective of the optimization 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, tasks move from a free 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 6 minutes. Thus, at the 5 minute mark, a message is sent from the simulation to the optimization model notifying it that the true duration of task t_1 is 6 minutes. However, there is a delay of 2 minutes in this message reaching the optimization model. It is therefore not until the 8 minute mark that the optimization model is updated, with the duration parameter of task t_1 changed from the original estimate of 7 to the true value of 6. Similar behavior will occur with respect to the times taken to travel between task locations, as illustrated in the figure. The final section of the figure illustrates a scenario where a task, specifically task t_2, takes longer to complete than initially forecast. During the simulation and optimization of a full problem instance, such evolution of model parameters will clearly occur on a much larger scale.

In order to improve the optimization model with respect to addressing these challenges, a number of alterations must be made. In what follows we give details of these alterations, and describe the underlying simulation models which govern the uncertain and dynamic nature of these parameters.

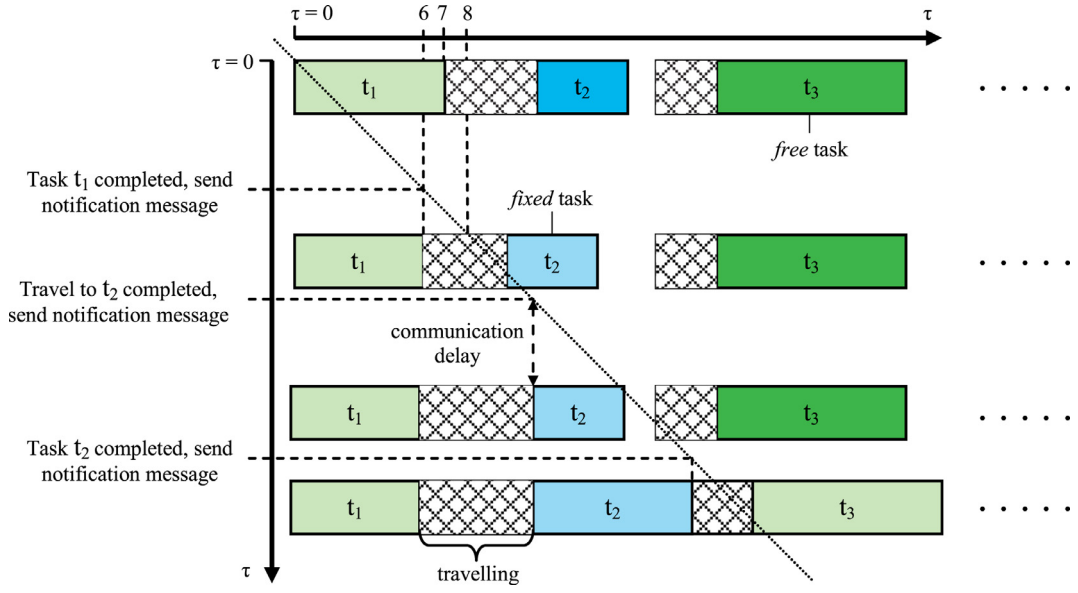

Fig. 3. The evolution in time of schedule parameters affecting a single responder's schedule. As time progresses, information is received by the model following communication delays, and is subsequently used in revising the relevant parameters.

Table 3
The probability of health assessment outcomes with error level ϵ.

True health state	$P(A[T1])$	$P(A[T2])$	$P(A[T3])$
$T1$	$1-\epsilon/2$	$\epsilon/2$	0
$T2$	$\epsilon/3$	$1-2\epsilon/3$	$\epsilon/3$
$T3$	0	$\epsilon/2$	$1-\epsilon/2$

3.3.1. Casualties and health

As discussed in Section 2.2, the health of a casualty is described through the discrete triage classification system with states $T1$, $T2$, $T3$ and *dead*. We wish to increase the realism of the triage assessment process by allowing for the fact that the classification assigned to some casualties may not accurately reflect their true health state.

Denoting by $A[Ti]$ the event that a triage assessment has led to a casualty being classified in state Ti, the probabilities of these events conditional on the true health state of the casualty are given in Table 3. The parameter $\epsilon \in [0, 1]$ allows for the degree of error to be modified, where $\epsilon = 0$ corresponds to completely accurate classification. The resulting probability distribution leads to unbiased errors, where the probability of under-triage is equal to the probability of over-triage in cases where both outcomes are possible. In practice, a significant bias towards the over triage of casualties has been observed (Frykberg, 2002). We do not model this systematic bias for two reasons. Firstly, to do so whilst also modeling imprecision in triage assessment would lead to difficulty in interpreting the results of the experimental analysis presented in Section 4.3, as it would not be clear if any observed effect was due to a lack of accuracy, a bias, or both. As such, removing bias allows us to focus on evaluating the effect of imprecision only. Secondly, while a bias has been documented, its precise nature has yet to be adequately described in quantitative terms.

In addition to allowing for errors in the triaging of casualties, we wish to allow for the dynamic nature of casualty health in the online model. The simulation of this dynamic behavior is achieved through using the same Markov chain model described in Section 2.2 which is used in predicting future variation in casualty health. By simulating the actual variations in health state of all casualties, we may now periodically update the model to reflect any such changes. This updating corresponds to another triage assessment being carried out. This is reflective of real MCI response operations, where casualties are regularly re-assessed and changes in health are noted. The frequency of any such triage operations is a variable of the model, which we will denote λ_{tri}, and will be adjusted in the experimental analysis of Section 4.2.

Another aspect of casualty behavior which may be captured via an online modeling approach is their tendency to *self-present* at hospital. Self-presentation is known to occur in MCI response, when casualties with less significant injuries (specifically, those in health state T3) may decide to leave the incident site and transport themselves to a hospital of their choosing. Self presentation leads to changes in the solution space, as any casualty who has transported themselves to hospital will no longer require any attention from responders and so can be removed from the casualty processing model. Self-presentation will also lead to changes in the objective space, with information regarding any self-presentation being used to update model parameters and allow for better prediction of solution quality. In particular, upon receiving notification that a casualty has arrived at a specific hospital it is possible to infer how long said casualty remained at their incident site before leaving. Denoting this observed data as x, we require that a probability density function $p(x|\theta)$ relating x to an unknown parameter θ is defined, together with a prior probability distribution on θ, $p(\theta)$.

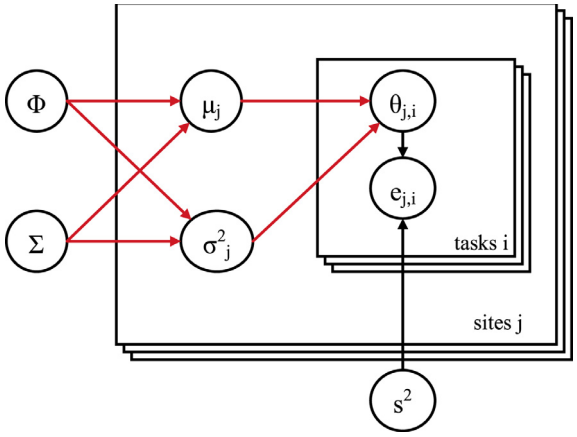

Fig. 4. A hierarchical model of task durations.

The posterior probability distribution of the unknown parameter θ can then be calculated through the usual application of Bayes rule:

$$p(\theta|x) = \frac{p(x|\theta)p(\theta)}{p(x)} \tag{3}$$

In this manner, estimates of the parameter θ will improve over the course of the response operation as relevant data is accrued, potentially correcting any inaccurate initial specifications. Note that this approach is generic and may be applied regardless of the form of the probability distribution describing the length of time a casualty will wait before leaving to self-present. However, cases leading to conjugate priors will enable the calculation of posterior probability distributions without any significant computational burden. As an example, we may consider self-presentation times to follow an exponential distribution. In this case, the parameter θ will be of one dimension and may be interpreted as the average rate at which casualties leave the incident site to self-present.

3.3.2. Task durations

As described in more detail in Wilson, Hawe, Coates, and Crouch (2012), the uncertain nature of task durations is encapsulated through a hierarchical model reflecting the different nature of incident sites in a multi-site MCI. We discuss the case of rescue tasks, noting that the model for treatment tasks (pre-rescue and post-rescue) is identical.

The true duration of a rescue task relating to casualty i at site j, $\theta_{j,i}$, is a random variable with mean μ_j and variance σ_j^2. These site-specific parameters are themselves considered to be random variables, with joint mean Φ and covariance matrix Σ. This hierarchical model, illustrated in Fig. 4, will allow for incident sites of varying severity (in terms of the durations of the tasks to be undertaken there) to be modeled. As in the case of modeling self-presentation times, a lack of empirical data prevents us from recommending a specific parametric model to describe these random variables.

Uncertainty is introduced by generating unbiased estimates of each duration $\theta_{j,i}$, denoted by $e_{j,i}$, by sampling from normal distributions with mean $\theta_{j,i}$ and variance s^2. This variance, which determines the accuracy of duration estimates, is specified as problem input. Given these estimates and assumed distribution, the true duration of each task is known to follow the distribution $N(e_{j,i}, s^2)$. For example, given a variance $s^2 = 0.7$ and a simulated true task duration of $\theta_{j,i} = 4$, the estimated task duration will be simulated from a normal distribution $N(4, 0.7)$. A resulting value of, say, $e_{j,i} = 5.2$ would then be used by the optimization model when attempting to predict the true task duration $\theta_{j,i}$.

Table 4
Problem scenario time line.

Time	Event	Description
	Explosion, site i_1	A first set of casualties is created. The ambulance service conducts a triage sieve operation, noting the state of each located casualty regarding their task requirements.
00:00:00	Model initialization	Using the information gathered during the initial triage operation, the scheduling model is initialized.
00:04:00	Explosion, site i_2	A second incident occurs, producing a further set of casualties. A triage sieve operation begins.
00:10:00	Mutual aid arrival	A set of responders, namely 18 Ambulance responders and 7 SAR responders, arrive from neighboring areas to assist in the response operation.
00:15:00	Explosion, site i_3	A third and final incident occurs.

As it is known that $\theta_{j,i}$ will follow the distribution $N(e_{j,i}, s^2)$, the natural (most likely) estimate of $\theta_{j,i}$ will be $e_{j,i} = 5.2$. However, we allow for the model to make other estimates in order to reach a desired level of confidence that the true value will be equal to or less than the estimated value. That is, for a given confidence level $\psi \in [0, 1]$ we estimate the duration of tasks to be $\hat{\theta_{j,i}} = F_{j,i}^{-1}(\psi)$ where $F_{j,i}$ denotes the relevant cumulative distribution function. For example, setting $\psi = 0.8$ would result in conservative task durations estimates which will over-estimate the true durations with probability 0.8.

3.3.3. Travel times

Travel times are initially estimated by applying Dijkstra's algorithm (Skiena, 1990) to the road network which covers the problem environment. This procedure generates a route, following which a distance d travelled can be calculated. An estimate for the travel time using this distance is provided by the model of Kolesar, Walker, and Hausner (1975), as recently validated by Budge, Ingolfsson, and Zerom (2010). The function, denoted $KWH(d)$, gives an estimate of the median travel time. The median travel time is then estimated as

$$\hat{m} = KWH(d) = \begin{cases} 2.42\sqrt{d}, d \leq 4.13 \text{ kilometer} \\ 2.46 + 0.596d, d > 4.13 \text{ kilometer} \end{cases} \quad (4)$$

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 from rest to v_c. The values of these parameters are taken from the analysis of ambulance travel times in Calgary, Canada, presented in (Budge et al., 2010).

The actual travel time is simulated through applying a disruption to the transport network, resulting in an uncertain increase in the time needed to traverse each arc. This disruption procedure consists of randomly generating a multiplying factor for each road link, to be applied to the link's 'length' parameter. The level of disruption is controlled by a parameter denoted κ. The disruption model is described in more detail in Wilson et al. (2013b).

We are required to generate and update the probability distribution of the travel time associated with each journey, as defined by a location pair $A - B$. To do so, we assume travel times are independent and identically distributed according to a lognormal distribution, $X \sim logN(\nu, \rho)$, with an assumed, constant precision ρ, as discussed in Westgate, Woodward, Matteson, and Henderson (2011). Note the variation in travel times may arise from either variation in the route chosen by responders making the journey in question, or variation the time taken to traverse any specific route. Given the dynamic nature of the problem as discussed in Section 3, we employ the previously outlines Bayesian approach in revising the estimate of the unknown parameter ν as more travel time data becomes available. Given the assumed lognormal distribution and the associated conjugate prior distribution for ν, $\nu \sim N(\nu_0, \rho_0)$, we can calculate the posterior distribution following the observation of n data x_i, $\nu \sim N(\nu_n, \rho_n)$ where

$$\nu_n = \frac{\rho_0 \nu_0 + \rho \sum_{i=1}^{n} \ln(x_i)}{\rho_0 + n\rho} \quad (5)$$

and

$$\rho_n = \rho_0 + n\rho. \quad (6)$$

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

4. Evaluation and analysis

In this section we report the results of several computational experiments analyzing the performance of the described model. In particular, we aim to answer the following key questions:

1. To what extent are pre-computed static schedules applicable in dynamic problems?
2. Can the scheduling methodology cope with solution space dynamics as well as non-predictive methods can?
3. How sensitive is the model to underlying variation and uncertainty in objective space parameters?

In order to do so we first identify problem characteristics with potential to influence the answers to these questions, and vary these in a comprehensive experimental design to produce a large, space filling data set. This data is then analyzed through the fitting of linear regression models in order to identify key relationships between problem parameters and performance.

4.1. Problem scenario

The following elements of the problem scenario are held constant through all variants used throughout the set of computational experiments:

- Incident sites $\mathcal{I} = \{i_1, i_2, i_3\}$ – their location, time of event and resulting set of casualties;
- Responders $\mathcal{R}$ – including initial location and arrivals through mutual aid (including time(s) of arrival);
- Hospitals $\mathcal{H} = \{h_1, h_2, h_3\}$ – their location, initial occupancy levels, maximum capacities and specialist treatment facilities.

The sequence and timings of events occurring in the problem scenario is given in Table 4.

Each of the three incident sites results in a set of seventy casualties with an identical profile in terms of their initial state (their health level and whether or not they are trapped). Responders are given initial locations which correspond to one of the three hospitals' locations in Fig. 5 (for Ambulance, MERIT and HART responders) or, for SAR responders, one of several fire stations in the area (not shown).

In terms of distributional assumptions, we use normal distributions in the hierarchical model of Fig. 4. To simulate task durations, we use a common value for the problem level mean of task durations and covariance. That is, for each task type the value of $\Phi = (\bar{\mu}, \bar{\sigma^2})$ has fixed $\bar{\mu}$ and covariance Σ, while $\bar{\sigma^2}$ will be altered in the experimental design (see Section 4.2). For rescue tasks, $\bar{\mu}$ is set to seven minutes. For pre-transportation stabilizing treatment, $\bar{\mu}$ is set to three minutes. Finally, for pre-rescue stabilizing

Fig. 5. Three incident sites and three hospitals in central London, as part of the test problem environment.

treatment, $\bar{\mu}$ is set to five minutes. We emphasise that these values are example only, and have been chosen to differentiate between task types. The covariance for all task types is

$$\Sigma = \begin{pmatrix} 1 & 0.5 \\ 0.5 & 0.5 \end{pmatrix}. \tag{7}$$

We assume that the times which casualties will wait at incident sites before leaving to self-present at a hospital are independent and identically distributed following an an exponential distribution, $\tau_c^{sp} \sim \exp(\lambda_{sp})$. Given this assumed distribution, the model estimate of parameter λ_{sp} can be continually revised following the general Bayesian strategy outlined in Section 3.3. Specifically, we note that the Gamma distribution acts as a conjugate prior for the parameter,

$$\lambda_{sp} \sim \text{Gamma}(\alpha, \beta). \tag{8}$$

Here, the hyperparameters α and β are set to reflect the initial estimate of the self-presentation rate. Upon observing a waiting time τ_c^{sp}, the posterior distribution is updated to be

$$\lambda_{sp} \sim \text{Gamma}(\alpha + 1, \beta + \tau_c^{sp}), \tag{9}$$

and the parameter of interest is estimated as $\tilde{\lambda}_{sp} = \text{E}(\lambda_{sp})$. Without sufficient data regarding actual times spent waiting before self-presenting in real MCIs, it is not possible to verify to what extent an exponential distribution is a realistic choice in modeling this process. However, given its common use in queuing models, which are of a similar nature, it appears to be an appropriate choice and will allow for an initial analysis of the effect of self-presentation in an online model.

Finally, we simulate the gradual discovery of casualties at incident sites by randomly generating a time at which they are discovered and added to the model. These times are generated according to an exponential distribution, parametrized so that the average time for a casualty to be discovered is equal to ρ.

4.2. Experimental design

Considering the two models described in this paper, namely the pre-computation model M_{pc} and the online model M_o, we wish to evaluate performance in problems exhibiting different degrees of dynamic behavior. Prior to consideration of the online model M_o, it is of interest to evaluate the performance of the pre-computation model M_{pc} in a partially dynamic environment. That is, we wish to investigate to what extent a pre-computed, optimized schedule will lead to high quality solutions when applied to a more realistic problem scenario than was considered in Wilson et al. (2013a). Following this, the online model M_o will be evaluated in fully dynamic problem environments. These analyzes complement that presented in Wilson et al. (2013a), where the model M_{pc} was evaluated in static problem environments. In each case, summarized in Table 5, evaluation consists of comparing the performance of the full, local search based scheduling methodology with that of the constructive heuristic approach.

Table 5
Model and problem pairs analyzed.

Model	Problem		
	Static	Partially dynamic	Fully dynamic
Pre-computed, M_{pc}	Wilson et al. (2013a)	4.3.1	n/a
Online, M_o	–	–	4.3.2

The problem scenario defined in Section 4.1 is used in all problem instances to be considered in the experimental procedure. Individual problem instances may vary in the nine dimensions described in Table 6. For some parameters, namely λ_1 and λ_5, the choice of range is a natural one. For other parameters a judgement has been made regarding feasible levels; for example, we consider it unlikely that the communication delay described in Section 3.3 will be larger than five minutes.

The frequency of triage has been allowed to vary from every minute to every 20 minutes. This includes the rate of once every 15 minutes, currently used in practice. The variance in task durations takes values from 0, corresponding to all task durations being identical, to 3, which would lead to 95 percent of task durations to be within a range of $+/-$ 3.46 minutes. The maximum delay in communicating information, denoted ν from the response environment to the optimization model has been set to 5 minutes, reflecting the fact that model communication technology will prevent any more significant delay from occurring. We have considered a rate of casualty discovery ranging from an average of ten per minute to one every ten minutes, reflecting the wide variety of incidents and the corresponding difficulty of search and rescue operations. For the average time a casualty will wait before leaving the site to self-present, a range of between 5 and 20 minutes has been considered.

All experiments follow a Sobol sequence of 500 points in the 9 dimension experimental design space, as constructed using the R package randtoolbox (Dutang & Savicky, 2013). This provides a set of points in the experiment parameter space which is 'space filling', in the sense that points are evenly distributed around the space. In contrast with an experimental design which places points only at the edges of the parameter space (i.e., where parameters are set at the end points of their ranges), a space filling design will allow for non-linear relationships between the parameters and the response to be identified in the analysis of the data.

4.3. Results

In this section we report the results of the computational experiments defined in Section 4.2.

4.3.1. Pre-computed scheduling in partially dynamic problems

The applicability of the static scheduling model to dynamic environments may now be evaluated through employing the simulation routine described in this paper. 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. By means of comparison, the same problem setup was addressed using the constructive heuristic defined in Wilson et al. (2013a), designed to replicate how decisions would be made in reality when faced with an evolving problem. Descriptive statistics of these experiments are provided in Table 7. Fig. 6 shows

Table 6
Experiment design parameters, denoting (a) the coefficients associated with each parameter when fitting linear regression models to the simulated data, and (b) the ranges considered for each parameter.

Parameter	Regression coefficient	Range	Description
Triage assessment error, ϵ	β_1	[0, 1]	Error in the triage classification process.
Triage frequency, λ_{tri}	β_2	[1, 20]	Time between each triage assessment of any given casualty.
Task duration variance, σ^2	β_3	[0, 3]	Inherent variation in the durations of all tasks.
Task duration assessment error, s^2	β_4	[0, 2]	Error in the estimation of durations of all tasks.
Task duration confidence, Ψ	β_5	[0.1, 0.9]	Level of confidence required that task duration estimates will not be short.
Communication delay, ν	β_6	[0, 5]	Average wait between a temporal event being recorded and the optimization model being notified.
Road network disruption, κ	β_7	(0.5, 2]	Extent to which the road transport network is disrupted.
Casualty discovery rate, ρ	β_8	[0.1, 10]	Average time taken to locate a casualty following an incident.
Casualty self-presentation rate, λ_{sp}	β_9	[5, 20]	Average time an eligible casualty will wait at scene before leaving to self-present.

Table 7
Descriptive statistics of final objective values across all partially dynamic problem instances, for both search and constructor solution methods.

	f_1				f_2			
	Mean	Min	Median	Max	Mean	Min	Median	Max
M_{pc} search	20.82	9	21	33	31124	18265	30768	51365
M_{pc} constructor	17.63	7	18	29	49808	33172	49295	76954

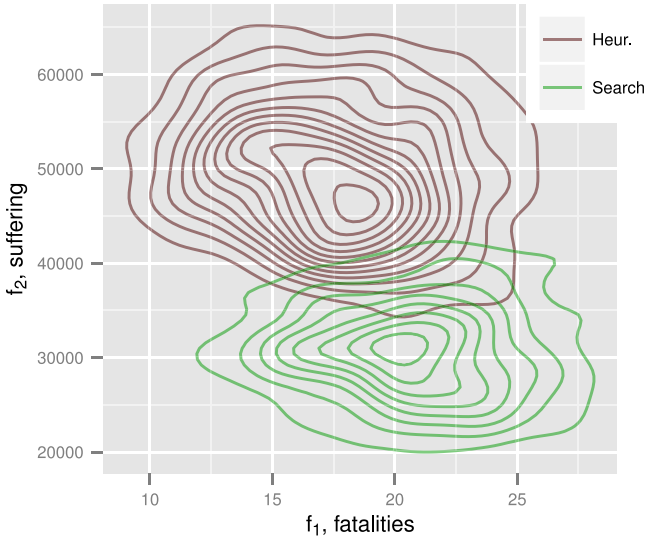

Fig. 6. Density plots of final objective values obtained by constructive and search methodologies, pre-computed case.

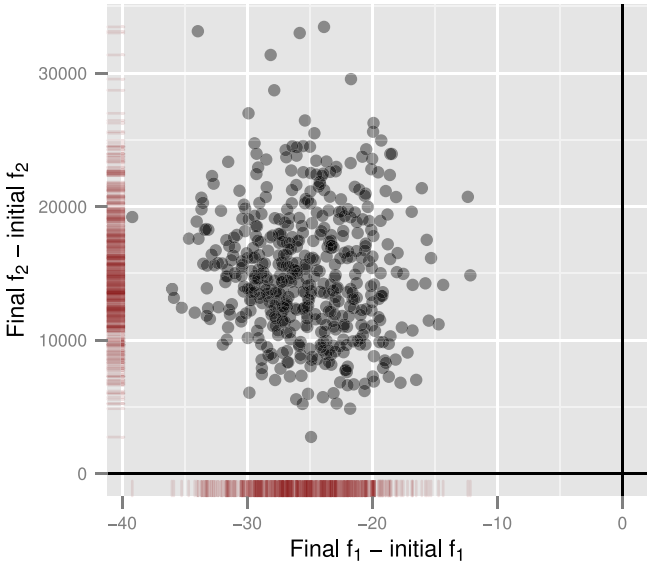

Fig. 7. Final simulated values minus initial estimated values of the pre-computed search procedure, where marginal distributions are shown in red. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

the joint distribution of objective values as contour plots for both cases, where we label the constructive heuristic method as 'Heur'.

The differences in objective values observed when comparing the expected values at the end of the search process with the simulated values are presented in Fig. 7.

It is of interest at this stage to consider the performance of the static search methodology in more detail. To do so, linear regression models relating each objective measure to the varied problem parameters were fitted in R by 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 their 95 percent confidence intervals, in Tables 8 and 9. We note that coefficient β_0 denotes the model intercept, while β_i is the coefficient of the term denoting the initial solution value generated by the constructive heuristic method.

Table 8
Regression coefficients, f_1, pre-computed case.

	Mean	Standard error	2.5 percent	97.5 percent
$\hat{\beta}_0$	16.43	2.674	11.17	21.68
$\hat{\beta}_i$	0.2177	0.0503	0.1188	0.3165
$\hat{\beta}_7$	−6.660	2.282	−11.14	−2.177
$\hat{\beta}_{7^2}$	1.528	0.9030	−0.2464	3.302

4.3.2. Online scheduling in fully dynamic problems

We now consider the fully dynamic problem, with variation in both solution space and objective space parameters. As described in Section 3.1, we propose that when applying the online model in real-time tasks should be issued to responders as soon as they become free, and that the appropriate task to issue can be found by consulting the best solution schedule found by the local search

Table 9
Regression coefficients, f_2, pre-computed case.

	Estimate	Standard error	2.5 percent	97.5 percent
$\hat{\beta}_0$	28740	2392	24041	33443
$\hat{\beta}_i$	1.220	0.1078	1.008	1.431
$\hat{\beta}_1$	−1101	605.4	−2291	88.47
$\hat{\beta}_6$	663.5	13.44	399.3	927.6
$\hat{\beta}_7$	−26230	2624	−31384	−21074
$\hat{\beta}_{7^2}$	8141	134.4	6101	10182

algorithm by that point in time. It is possible, however, that it is instead beneficial to wait a short period of time before issuing instructions to responders in the hope that the solution algorithm will find a schedule of higher quality. However, any such delay in responders completing tasks will also have a negative impact on the quality of the overall response operation. To assess the trade-off between these two factors, we conducted an experiment whereby the local search algorithm was allowed 1, 2, 3, 4 or 5 minutes to search for a solution schedule at the start of the response operation, after which point tasks began to be issued to responders. The distributions of the quality of the resulting solutions, as expressed by their percentage improvement over the solution produced by the constructive heuristic, are presented in Figs. 8 and 9 as box plots.

In Fig. 8 it can be seen that in the case of the fatalities objective f_1, any benefit brought through the optimization process is not enough to counteract the penalty of delaying action for any value of search time considered. In contrast, as indicated in Fig. 9, this is not the case for the suffering objective f_2, which shows moderate improvement for all search times considered. Thus, the lexicographic ordering of these objectives suggests that the optimal policy (for the example considered) is to issue instructions as soon as responders become available, rather than waiting for a period of time to allow the search algorithm to locate higher quality solutions.

Having established the benefits of immediately issuing instructions, we now consider the performance of the online model in fully dynamic problem instances. As in Section 4.3.1, both the scheduling and constructive heuristic methodologies were employed. The results are summarized in Table 10 and Fig. 10.

Further linear regression models were fit to the data corresponding to the scheduling approach. Coefficient estimates are provided in Tables 11 and 12.

In order to better appreciate the resulting trends, graphical residual and component plots of the most significant predictors of objective f_2 are given in Figs. 11–13.

4.4. Discussion

The results given in Section 4.3 allow us to answer the questions posed in Section 4.

4.4.1. To what extent are pre-computed static schedules applicable in dynamic problems?

Considering the results of the pre-computation case (Table 7 & Fig. 6), we note that the local search approach outperforms the constructive approach in terms of objective f_2. However, in the case of objective f_1 it is the constructive methodology which results in best average performance. In contrast, the results presented in Wilson et al. (2013a) suggested that the pre-computed approach could lead to improvements in both objectives. These results highlight a shortcoming of the search methodology 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

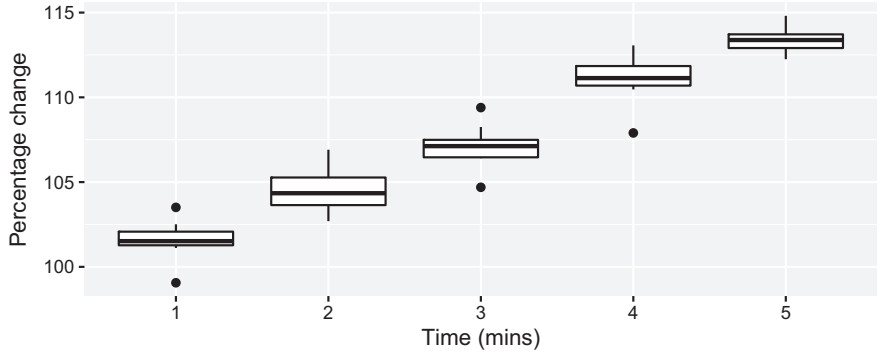

Fig. 8. Percentage change in f_1 as search time increases (box plots).

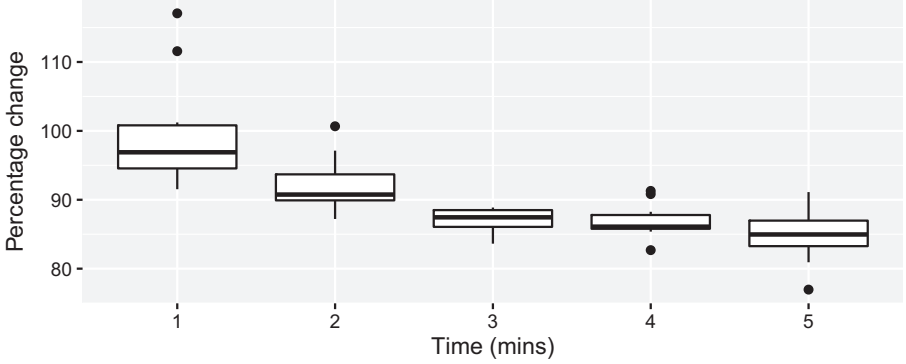

Fig. 9. Percentage change in f_2 as search time increases (box plots).

Table 10
Descriptive statistics of final values across all fully dynamic problem instances, for both search and constructor solution methods.

	f_1				f_2			
	Mean	Min	Median	Max	Mean	Min	Median	Max
M_o search	19.49	11	20	28	29008	18920	28784	44789
M_o constructor	21.50	10	21	31	31663	21793	31329	43168

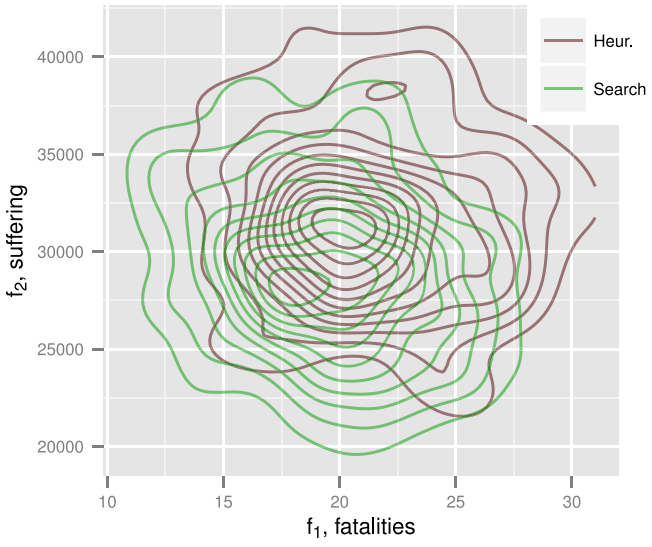

Fig. 10. Density plots of final objective values obtained by constructive and search methodologies, online case.

Table 11
Regression coefficients, f_1, online case.

	Mean	Standard error	2.5 percent	97.5 percent
$\hat{\beta}_0$	24.34	1.495	21.41	27.28
$\hat{\beta}_1$	−0.9790	0.2681	−1.506	−0.4522
$\hat{\beta}_3$	−0.3027	0.1803	−0.6570	0.0517
$\hat{\beta}_5$	2.477	0.6794	1.142	3.812
$\hat{\beta}_6$	0.1842	0.1209	−0.0534	0.4218
$\hat{\beta}_7$	−6.597	2.352	−11.22	−1.976
$\hat{\beta}_{7^2}$	1.764	0.9300	−0.0637	3.591

Table 12
Regression coefficients, f_2, online case.

	Estimate	Standard error	2.5 percent	97.5 percent
$\hat{\beta}_0$	33285	1596	30149	36420
$\hat{\beta}_3$	295.9	193.2	−83.68	675.4
$\hat{\beta}_4$	1209	287.1	644.8	1773
$\hat{\beta}_6$	286.6	129.5	32.11	541.1
$\hat{\beta}_7$	−8168	2518	−13116	−3219
$\hat{\beta}_{7^2}$	1697	995.9	−259.9	3654
$\hat{\beta}_8$	106.0	58.31	−8.55	220.6

different from those realized upon completing the simulation, as shown in Fig. 7. Specifically, the number of fatalities is typically over-estimated whereas the level of suffering is typically under-estimated. 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 4.3.1 (Tables 8 and 9) 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. This suggests that the proposed

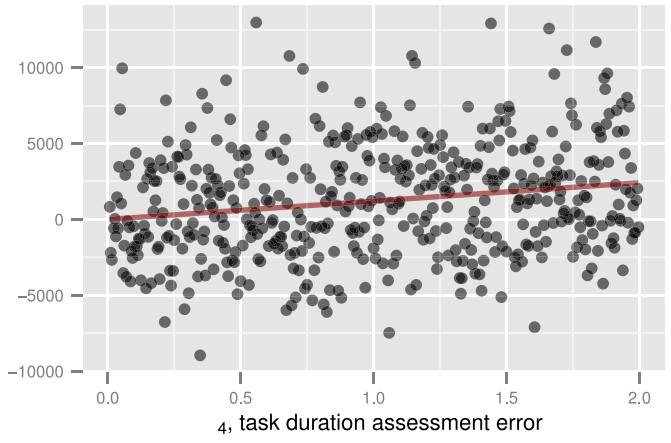

Fig. 11. Residual and component plot for task duration assessment error effects, $\hat{\beta}_4 s^2$, with the fitted trend shown in red. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

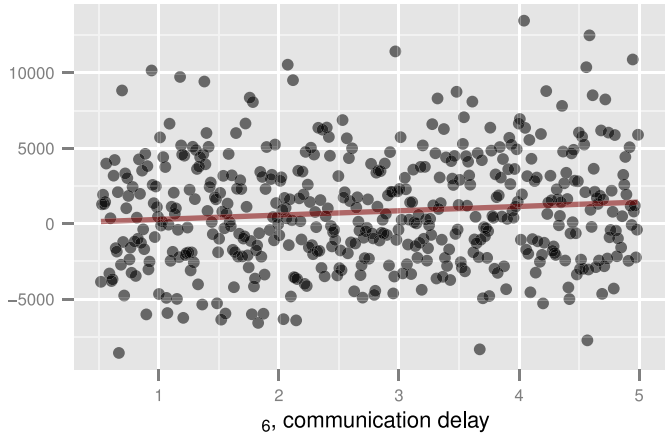

Fig. 12. Residual and component plot for communication delay effects, $\hat{\beta}_6 \nu$, with the fitted trend shown in red. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

model is relatively robust to any misspecification of task duration distributions. In contrast, errors in the triage assessment of casualties do have a significant and negative association with suffering, as does a delay in communication. 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. 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 behavior in the objective space. Were such behavior to exist in the solution space, the pre-compacted approach would lead to

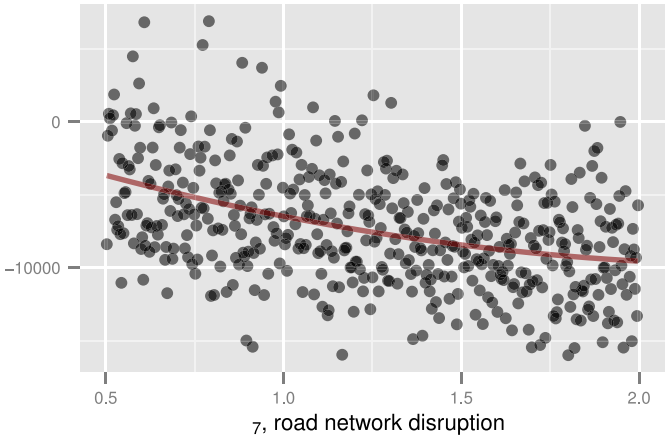

Fig. 13. Residual and component plot for road network disruption effects, $\hat{\beta}_7\kappa + \hat{\beta}_7^2\kappa^2$, with the fitted trend shown in red. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

schedules which would quickly become irrelevant as the response operation progressed.

4.4.2. Can the search-based solution methodology cope with solution space dynamics as well as non-predictive methods can?

In the online case, the local search solution methodology results in improved performance in both objectives, on average (Table 10 and Fig. 10). The difference in the bivariate distributions is shown to be statistically significant under a Hotteling's T2 test ($p \approx 0$). We can therefore conclude that the search based methodology equipped for online use in the manner described in this paper can deliver improved performance over the alternative heuristic approach.

4.4.3. How sensitive is the model to underlying variation and uncertainty in objective space parameters?

Considering the regression models fitted (Tables 11 and 12) we see that, in comparison to the pre-computed case, a larger number of relationships between parameters and objective values were identified in the online case. 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 fact that the model only allows for the health of casualty to change when they are in a hazardous area. As the majority of casualties will be removed from such areas relatively quickly, there may only be limited scope for health to change. As such, increasing frequency of triage will have limited scope to impact the quality of the response.

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, although the associated parameter estimate is lower than in the pre-computed case (see Table 9). This suggests that the online solution 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. 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.

5. Conclusions

The dynamic and uncertain nature of mass casualty incidents represents a significant challenge to the design of a robust and effective optimization model. In this paper we have described the extension of such a model, which employs a task scheduling representation of an MCI response operation, to better cope with this volatility. In particular, the model has been extended for use in an online manner, allowing for communication between model and problem environment to be carried out continually as the response progresses. This has resulted in the removal of several common assumptions made in such models, as highlighted in Galindo and Batta (2013).

Through the design and analysis of several computational experiments, the performance of the model in a simulated environment has been assessed. It has been shown that the extension of the model from its initial 'static' design to the online case has resulted in significant improvement in terms of both expected fatalities and the suffering of casualties. Statistically significant associations between parameters and model utility have been identified, highlighting the importance of fast communication between problem environment and model and the accurate estimation of task durations. Of equal value is the lack of such associations found in other parameters, where it may be natural to assume one would exist, such as the frequency of triage.

5.1. Further work

The computational burden arising from evaluating the proposed methodology through simulation has placed a practical limit upon the number of scenarios which could be considered and comparisons which could be made. Further work could, therefore, focus on extending the simulation study presented in Section 4.3. One study which would be of value would be the application of the online model M_o to partially dynamic problem scenarios. This would then enable a direct comparison of the models M_{pc} and M_o to be made, helping to demonstrate the superiority of the latter.

In the process of designing the simulation of the MCI response operation, we have made a number of decisions regarding the nature of probability distributions and their parameters. Unfortunately, due to the inherently low frequency of MCI events and the limited data collection which occurs during them, it has not been possible to base these decisions on the results of statistical models. While the assumptions made have enabled a valuable exploration of the sensitivity of the optimization model to the dynamic and uncertain nature of MCI response operations, it would be of value to examine how robust the results generated are to violations of these assumptions. For example, the sensitivity of the model to changes in the assumed transition probabilities of the Markov chain described in Section 2.2 could be assessed.

As discussed in Galindo and Batta (2013), a Bayesian approach to processing information in the dynamic environment of disaster response may be applicable. While the model presented in this paper 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 hierarchical 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.

As discussed in Section 4.4, the method used in predicting fatalities, as originally described in Wilson et al. (2013a), results in an average over-estimation when compared with the actual value resulting from the simulation. A more accurate and robust calculation would clearly improve performance, as it would enable the consequences of decisions to be predicted to a higher degree. This would in turn allow the benefits to optimization over the remaining time horizon to be fully realized.

The simulation of volatility presented in this paper 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 Hawe, Coates, Wilson, and Crouch (2012), 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.

Acknowledgments

The authors gratefully acknowledge the funding provided under the UKÆs EPSRC grant EP/G057516/1. Furthermore, the authors thank practitioners from Cleveland EPU, Tyne & Wear EPU, Co. Durham & Darlington CCU and Government Office for the North East. We would also like to thank 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. The data presented in this paper are available from http://dx.doi.org/10.15128/kk91fk84m.

References

Amiri, M., Zandieh, M., Yazdani, M., & Bagheri, A. (2010). A variable neighbourhood search algorithm for the flexible job-shop scheduling problem. *International Journal of Production Research, 48*, 5671–5689.

Auf der Heide, E. (2006). The importance of evidence-based disaster planning. *Annals of Emergency Medicine, 47 (1)*, 34–49. doi:10.1016/j.annemergmed.2005.05.009.

Barbarosoglu, G., & Arda, Y. (2004). A two-stage stochastic programming framework for transportation planning in disaster response. *Journal of the Operational Research Society, 55 (1)*, 43–53. doi:10.1057/palgrave.jors.2601652.

Barbarosoglu, G., Ozdamar, L., & Cevik, A. (2002). An interactive approach for hierarchical analysis of helicopter logistics in disaster relief operations. *European Journal of Operational Research, 140*, 118–133. doi:10.1016/S0377-2217(01)00222-3.

Bozorgi-Amiri, A., Jabalameli, M., Alinaghian, M., & Heydari, M. (2012). A modified particle swarm optimization for disaster relief logistics under uncertain environment. *International Journal of Advanced Manufacturing Technology, 60*, 357–371. doi:10.1007/s00170-011-3596-8.

Budge, S., Ingolfsson, A., & Zerom, D. (2010). Empirical analysis of ambulance travel times: The case of Calgary emergency medical services. *Management Science, 56*, 716–723. doi:10.1287/mnsc.1090.1142.

Chang, M.-S., Tseng, Y.-L., & Chen, J.-W. (2007). 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. doi:10.1016/j.tre.2006.10.013.

Chiu, Y.-C., & Zheng, H. (2007). 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. doi:10.1016/j.tre.2006.11.006.

Cotta, C. (2011). Effective patient prioritization in mass casualty incidents using hyperheuristics and the pilot method. *OR Spectrum, 33*, 699–720. doi:10.1007/s00291-011-0238-3.

Dutang, C., & Savicky, P. (2013). randtoolbox: Generating and Testing Random Numbers. R package version 1.13.

Engelmann, H., & Fiedrich, F. (2007). Decision support for the memebers of an emergency operation centre after an earthquake. In *Proceedings of the 4th international ISCRAM conference*.

Englemann, H., & Fiedrich, F. (2009). DMT-EOC û a combined system for the decision support and training of EOC members. In *Proceedings of the 6th international ISCRAM conference*.

Fiedrich, F. (2006). An HLA-based multiagent system for optimized resource allocation after strong earthquakes. In *Proceedings of the 2006 winter simulation conference*.

Fiedrich, F., Gehbauer, F., & Rickers, U. (2000). Optimized resource allocation for emergency response after earthquake disasters. *Safety Science, 35*, 41–57.

Frykberg, E. R. (2002). Medical management of disasters and mass casualties from terrorist bombings: How can we cope? *Journal of Trauma-Injury Infection & Critical Care, 53*, 201–212.

Galindo, G., & Batta, R. (2013). Review of recent developments in OR/MS research in disaster operations management. *European Journal of Operational Research, 230*, 201–211.

Gong, Q., & Batta, R. (2007). Allocation and reallocation of ambulances to casualty clusters in a disaster relief operation. *IIE Transactions, 39 (1)*, 27–39. doi:10.1080/07408170600743938.

Group, A. L. S. (2011). *Major incident medical management and support: The practical approach at the scene* (3rd ed.). Wiley-Blackwell.

Hawe, G. I., Coates, G., Wilson, D. T., & Crouch, R. S. (2012). 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)*.

Jiang, Y., Yuan, Y., Huang, K., & Zhao, L. (2012). Logistics for large-scale disaster response: Achievments and challenges. In *Proceedings of the 45th hawaii international conference on system sciences*. doi:10.1109/HICSS.2012.418.

Jotshi, A., Gong, Q., & Batta, R. (2009). Dispatching and routing of emergency vehicles in disaster mitigation using data fusion. *Socio-Economic Planning Sciences, 43*, 1–24. doi:10.1016/j.seps.2008.02.005.

Kolesar, P., Walker, W., & Hausner, J. (1975). Determining the relation between fire engine travel times and travel distances in New York City. *Operations Research, 23*, 614–627. doi:10.1287/opre.23.4.614.

Lee, Y. M., Ghosh, S., & Ettl, M. (2009). Simulating distribution of emergency relief supplies for disaster response operations. In *Proceedings of the 2009 winter simulation conference (WSC)*. doi:10.1109/WSC.2009.5429246.

Lodree, E. J., & Taskin, S. (2009). Supply chain planning for hurricane response with wind speed information updates. *Computers & Operations Research, 36*(1), 2–15. http://dx.doi.org/10.1016/j.cor.2007.09.003. Part Special Issue: Operations Research Approaches for Disaster Recovery Planning.

London Emergency Services Liaison Panel (2015). Major incident procedure manual. http://www.met.police.uk/leslp/docs/major-incident-procedure-manual-9th-ed.pdf, Accessed 01.02.16.

London Assembly (2006). Report of the 7 July review committee. http://news.bbc.co.uk/1/shared/bsp/hi/pdfs/05_06_06_london_bombing.pdf, Accessed 01.02.16.

Mete, H. O., & Zabinsky, Z. B. (2010). Stochastic optimization of medical supply location and distribution in disaster management. *International Journal of Production Economics, 126*, 76–84. doi:10.1016/j.ijpe.2009.10.004.

Mysore, V., Gill, O., Daruwala, R., Antoniotti, M., Mishra, B., & Saraswat, V. (2005). Multiagent modeling and analysis of the brazilian food poisoning scenario. In *Proceedings of the agent 2005 conference on generative social processes, models, and mechanisms*.

Ozdamar, L., Ekinci, E., & Kucukyazici, B. (2004). Emergency logistics planning in natural disasters. *Annals of Operations Research, 129*, 217–245.

Paton, D., & Flin, R. (1999). Disaster stress: an emergency management perspective. *Disaster Prevention and Management, 8*, 261–267. doi:10.1108/09653569910283897.

Rawls, C. G., & Turnquist, M. A. (2010). Pre-positioning of emergency supplies for disaster response. *Transportation Research Part B: Methodological, 44*, 521–534.

Rolland, E., Patterson, R. A., Ward, K., & Dodin, B. (2010). Decision support for disaster management. *Operations Management Research, 3*, 68–79. doi:10.1007/s12063-010-0028-0.

Skiena, S. (1990). Dijkstra's algorithm. In *Implementing discrete mathematics: combinatorics and graph theory with mathematica* (pp. 225–227). Reading, MA: Addison-Wesley.

Taskin, S., & Lodree, E. J. (2011). A Bayesian decision model with hurricane forecast updates for emergency supplies inventory management. *Journal of the Operational Research Society, 62*, 1098û1108.

Tatomir, B., & Rothkrantz, L. (2006). Ant based mechanism for crisis response coordination. *Lecture Notes in Computer Science, 4150*, 380–387.

Wallace, W. A., & de Balogh, F. (1985). Decision support systems for disaster management. *Public Administration Review, 45*, 134–146.

Westgate, B. S., Woodward, D. B., Matteson, D. S., & Henderson, S. G. (2011). Travel time estimation for ambulances using Bayesian data augmentation. *Technical Report*. Cornell University.

Wex, F., Schryen, G., & Neumann, D. (2011). Intelligent decision support for centralized coordination during emergency response. In *Proceedings of the 8th international ISCRAM conference*.

Wex, F., Schryen, G., & Neumann, D. (2012). Operational emergency response under informational uncertainty: A fuzzy optimization model for scheduling and allocating rescue units. In *Proceedings of the 9th international ISCRAM conference*.

Wilson, D. T., Hawe, G. I., Coates, G., & Crouch, R. S. (2012). 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)*.

Wilson, D. T., Hawe, G. I., Coates, G., & Crouch, R. S. (2013a). A multi-objective combinatorial model of casualty processing in major incident response. *European Journal of Operational Research, 230*, 643–655. doi:10.1016/j.ejor.2013.04.040.

Wilson, D. T., Hawe, G. I., Coates, G., & Crouch, R. S. (2013b). Scheduling response operations under transport network disruptions. In *Proceedings of the 10th international ISCRAM conference*.

Yi, W., & Kumar, A. (2007). Ant colony optimization for disaster relief operations. *Transportation Research Part E: Logistics and Transportation Review, 43*, 660–672. doi:10.1016/j.tre.2006.05.004.

Yi, W., & Ozdamar, L. (2007). A dynamic logistics coordination model for evacuation and support in disaster response activities. *European Journal of Operational Research, 179*, 1177–1193. doi:10.1016/j.ejor.2005.03.077.

Yuan, Y., & Wang, D. (2009). Path selection model and algorithm for emergency logistics management. *Computers & Industrial Engineering, 56*, 1081–1094.

Zhang, J.-H., Li, J., & Liu, Z.-P. (2012). Multiple-resource and multiple-depot emergency response problem considering secondary disasters. *Expert Systems with Applications, 39*, 11066–11071.

Zhang, X., Zhang, Z., Zhang, Y., Wei, D., & Deng, Y. (2013). Route selection for emergency logistics management: A bio-inspired algorithm. *Safety Science, 54*, 87–91.