

# A Study on Dynamic Return and Build Strategies for Ambulances Based on an Approximate Hypercube Model

Wenfei Lei

Shanghai University, Shanghai 200000, China

## ABSTRACT

With the ageing population and the accelerating pace of urbanization, the demand for emergency medical services is experiencing sustained rapid growth. The traditional resource allocation model of ‘fixed stations plus return to base’ struggles to adapt to the dual time-varying characteristics of both resources and demand, resulting in an increasingly acute imbalance between supply and demand. This paper focuses on the dynamic return-to-base decision-making problem for ambulances that do not return to their original station upon completing a mission, and proposes a dynamic return-to-base strategy based on the Jarvis approximate hypercube model. This study advances ambulance return strategies from ‘static rule optimization to ‘dynamic online decision-making’, achieving an improvement from ‘single-objective’ to ‘multi-objective’ approaches, thereby providing theoretical and methodological support for the dynamic allocation of emergency resources.

## KEYWORDS

Return and Build; Dynamic Decision-making; Multi-objective.

## 1. INTRODUCTION

In recent years, with the rapid sociolect-economic development of China, the acceleration of urbanization and the deepening of population ageing, the number of emergency calls has shown a sustained and rapid upward trend. Demand for emergency services is entering a period of rapid growth, and the imbalance between supply and demand for emergency resources is becoming increasingly pronounced, making this a public welfare issue of widespread concern across all sectors of society. Against the backdrop of a surge in demand for emergency medical services, the traditional ‘fixed-station plus return-to-base’ model faces multiple challenges, including an insufficient number of stations, uneven distribution, wasted journeys, and disparities between urban and rural areas. The key to resolving this dilemma lies in shifting from a ‘static return-to-base’ approach to ‘dynamic deployment’—allowing vehicles to select the optimal return location based on real-time demand after completing a mission, rather than mechanically returning to their original station. The core of this approach is ‘post-mission location decision-making’, which requires ‘ex-post’ decision-making and constitutes a post-event resource reallocation strategy. In the fields of emergency medical services management and operations research, this is referred to as ‘Return and Build’. It involves determining the next course of action—whether to return directly to the hospital (‘Return’) or to proceed to a specific location to stand by (‘Build’) whilst en route—once an emergency call has been completed and the vehicle is released into an idle state, in accordance with dispatch instructions or predefined rules. Return and Build strategies are typically formulated based on demand forecasts for the current time period and the distribution of available ambulances. How to optimize the dynamic allocation efficiency of existing emergency resources through scientific methods has become a critical issue

requiring urgent resolution. This is precisely the core research question of this paper: advancing return strategies from ‘static rule optimization to ‘dynamic priority decision-making’.

## **2. THE CURRENT STATE OF RESEARCH ON THE ALLOCATION OF EMERGENCY MEDICAL RESOURCES**

### **(1) Current Research on Queuing Models**

In dispatch systems for emergency vehicles such as ambulances, the core of the queuing model lies in describing the dynamic equilibrium relationship between the ‘arrival of emergency calls’ and the ‘response of resources’. The dynamic dispatch process for ambulances that do not return to a fixed base, taking into account the time-varying response efficiency of system resources and demand, can be modeled and analyzed using queuing theory. Lv Chuanzhu explicitly pointed out that this strategy essentially constitutes a dynamic equilibrium problem of spatial queues [1]. The “Hypercube Queuing Model (HQM)” and its modified versions are currently among the mainstream research methods for the allocation and dispatching of emergency vehicles both domestically and internationally [2]. The Hypercube Queuing Model (HQM), proposed by Larson in 1974, has gradually become the mainstream paradigm in EMS research due to its ability to effectively capture spatial heterogeneity and server cooperation characteristics. It pioneered the introduction of spatial dimensions into traditional queuing theory, establishing a hypercube queuing analysis framework for urban emergency services. Based on spatially distributed queuing systems, it is used to analyse the state of an EMS with mobile servers that act as mutual backups; when a customer arrives and the highest-priority server is busy, the second-highest-priority server can be dispatched to serve the customer [3]. Golabian point out that the hypercube model is an effective tool for analyzing system states, capable of calculating steady-state probabilities to evaluate performance metrics such as customer loss rate, average waiting time and maximum waiting time. Golabian were the first to systematically define the ambulance return strategy problem and demonstrated that optimizing return strategies can significantly improve system coverage compared to traditional static return strategies. They employed a multi-universe optimizer and a genetic algorithm to select the optimal static return strategy from  $J!$  fixed priority lists. However, this study presents two areas for further development: firstly, the return strategy is determined once during the planning phase and cannot adapt to real-time fluctuations in demand; secondly, the definition of vehicle status approximates the location of busy vehicles using only a demand ratio, which may result in a loss of accuracy[4].

### **(2) Current Research on Ambulance Dispatch Issues**

Guigues introduced two new optimization models for ambulance dispatching. The first model is the ambulance selection problem, which is used to determine whether an ambulance should be dispatched to respond to an emergency call when it is received. The second model is the ambulance reallocation problem, which is used to determine whether an ambulance should be dispatched to a request in the queue once it has completed its current assignment. Based on these research problems, a rolling horizon method has been proposed, which combines the current decision—made as a first-stage decision—with a second-stage model representing the ambulance fleet’s capacity to serve future calls[5].

Bélanger argue that ambulance dispatching is a real-time decision-making process; when selecting which ambulance to dispatch in response to an incoming emergency call, the current state of the system (i.e. busy and idle ambulances) must be taken into account. Using a queuing approach, they employ a dispatching list format to define an ordered list for each zone within the coverage area, providing a hierarchy of ambulances to be selected whenever a call is received. They also propose a recursive simulation optimization framework to generate empirical estimates of ambulance availability and system performance, thereby addressing the problem of ambulance location and dispatching[6].

From the above literature review on queuing models and ambulance dispatch problems, it can be seen that the dynamic allocation and dispatch of ambulance resources primarily focus on two aspects: firstly, the dynamic dispatch of ambulances during the on-call phase and prior to rescue operations, optimizing pre-hospital ambulance resource dispatch decisions through real-time analysis of factors such as emergency demand and the number of available ambulances; secondly, real-time dynamic monitoring during the ambulance service process, re-routing based on traffic conditions, and the introduction of two-way monitoring capabilities—which record the ambulance’s route and the medical treatment process—to enhance emergency response efficiency. This paper focuses its research on the dynamic resource allocation of ambulances, specifically examining the scheduling and preparation process by which ambulances return to the emergency dispatch centre to stand by after completing a rescue mission. We propose to use an approximate hypercube model to evaluate the response capacity of the emergency station’s resources and demand. By combining this with a set of evaluation metrics for ambulance dispatch costs, we aim to establish an evaluation model for ambulance return priority that accounts for the time-varying nature of the emergency station’s response capacity, and to develop an ambulance return strategy that considers the time-varying response efficiency of the system’s resources and demand.

### **3. THE MODEL DEVELOPMENT FOR A DYNAMIC RECONSTRUCTION STRATEGY BASED ON AN APPROXIMATE HYPERCUBE MODEL**

The ambulance return-to-base model discussed in this paper focuses on the independent pre-hospital emergency care model currently prevalent in China. Under this independent model, emergency centers operate entirely separately from hospitals and possess independent legal status; they are responsible for dispatching and assigning resources for all stages of the process—from receiving emergency calls and maintaining vital signs at the scene to transporting patients to hospital. In the vast majority of practical scenarios, ambulance dispatch is currently based on the principle of proximity. Upon receiving a call from a caller, the dispatcher selects the emergency station closest to the patient based on geographical location to dispatch an ambulance for rescue, and transports the patient to the nearest hospital. After completing the rescue mission, the ambulance typically follows traditional return-to-base strategies, such as ‘returning to the home station’ or ‘returning to the nearest station’. In recent years, the independent model has gained increasing favour and attention due to its efficient operation. However, in practical application, emergency call demands are dispersed and unpredictable. Such return-to-base strategies can only ensure a rapid response to current demands; whenever a new emergency call arises, an ambulance from a more distant station must be dispatched to the scene. Furthermore, there is typically an imbalance between supply and demand across stations, leading to significant variations in response levels between them and adversely affecting overall emergency response times and efficiency. Although the return-to-base operations under this independent pre-hospital emergency care model are straightforward, their static and localized perspective often overlooks the dynamic spatiotemporal distribution of demand across the entire city’s EMS system. Consequently, ambulance resource allocation strategies have attracted increasing attention. Current research focuses on ambulance redeployment strategies, whereby ambulance resources are pre-allocated to high-demand areas to better meet call requirements—that is, the redeployment of idle ambulances between different stations. Although ambulance redeployment can improve the coverage of the pre-hospital emergency care system, it also incurs additional operational costs and increases the complexity of system decision-making and implementation; consequently, ambulance redeployment has not yet been widely adopted. If ambulance resources that have just been released and are in an idle state can be pre-allocated to areas of high demand, call requests can be met more effectively. Unlike the redeployment of idle ambulance resources between stations, dynamically returning ambulances that have completed rescue missions and entered an idle state to predicted high-demand stations is a ‘post-event’ resource allocation strategy. Upon completion of an emergency mission (where the patient has been delivered to hospital and handover completed), the vehicle is

released to an idle state, and the system must determine which emergency station the vehicle should proceed to for standby duty. This decision is termed the ‘return-to-base decision’, the essence of which is to enhance the response efficiency for future calls by optimizing ambulance resource allocation within a dual-time-varying environment of resources and demand. By pre-allocating ambulance resources based on predicted demand before an emergency arises, the system enables rapid response to calls once they materialize. The advantage of this approach lies in the system’s ability to allocate idle ambulances to high-demand stations based on predicted demand at each station, thereby ensuring that calls near high-demand stations are responded to swiftly and reducing overall response times. In light of these practical challenges, this paper proposes a new dynamic ambulance return-to-base strategy: as soon as an ambulance is released from an emergency mission and enters an idle state, it can be directed to its original station or another emergency station to meet future demand at that station, based on predictions derived from historical call data. To improve ambulance coverage, establishing a dynamic, global and forward-looking return-to-base decision requires the development of a mathematical model capable of quantifying such decisions. This model must assess how dispatching an idle ambulance to different alternative stations will affect the entire system’s future capacity to meet new demands. This paper will construct a dynamic ambulance redeployment priority assessment model to determine, upon the completion of an emergency mission (where the patient has been delivered to hospital and handover completed), which emergency station the vehicle should proceed to for standby duty once it is released as idle. The evaluation model comprises two dimensions: a multi-dimensional indicator system for assessing the resource and demand response capabilities of emergency stations, based on predictions from an approximate hypercube model, and a set of indicators for evaluating the cost of returning ambulances to base.

#### (1) Development of a Multidimensional Indicator System for the Demand-Response Capability of Emergency Medical Service Stations Based on an Approximate Hypercube Model and Real-Time Capability Assessment

This paper applies the Approximate Hypercube Queuing Model proposed by Richard C. Larson. Leveraging its robust capability for predicting system performance, it constructs a multidimensional indicator system capable of quantifying the future demand response capacity of emergency medical service (EMS) stations. As a classic spatial queuing model, this model enables the performance modelling of large-scale EMS systems with high computational efficiency. By predicting key indicators such as vehicle occupancy probability and average response time, it serves as an ideal tool for real-time decision support. To scientifically evaluate the demand response capacity of emergency medical service (EMS) stations in dynamic environments, the indicator system adheres to the following principles: 1. Combining predictability with real-time data: It considers both the future coverage potential based on the Approximate Hypercube Queuing Model and the real-time status of the station’s currently available resources. 2. Multi-dimensional quantification: It covers key capability dimensions such as spatial coverage, load balancing, and response timeliness. 3. Compatibility: All indicators can be directly obtained through Jarvis fixed-point iteration or system status. This framework encompasses multiple dimensions, including response efficiency, resource accessibility and system resilience, transforming the model’s predictive outputs (such as regional response times and vehicle utilization rates) into quantifiable evaluation metrics. Secondly, based on the cost structure involved in the return of ambulances to base, a comprehensive set of cost indicators has been established, comprising direct costs (fuel, wear and tear, personnel) and indirect opportunity costs (empty-run opportunities, idle resources, dispatch switching). The paper also explores methods for quantifying these various costs.

## 4. MULTI-OBJECTIVE DECISION-MAKING MODEL FOR AMBULANCE DEPLOYMENT

The multi-objective decision-making problem regarding the return of ambulances to base can be described as follows: once an ambulance has completed an emergency mission, the system must select one of all feasible standby locations as the return destination, thereby maximizing the system's dynamic responsiveness to future demand whilst minimizing the service risks associated with vehicle repositioning. A unified, commutable 'return priority scoring model' has been developed. This model provides a clear priority score for each ambulance awaiting return to base, offering direct data support for dispatch decisions. By applying a combined multi-objective decision-making analysis method, the two major indicator systems are weighted in combination. This integration forms a unified return-to-base priority scoring model, providing a quantitative basis for dispatch decisions. First, multiple indicators were extracted to quantify the advantages of constructing stations in each area. The Analytic Hierarchy Process (AHP) and TOPSIS were used to score each area, and stations requiring construction were selected based on these scores. Subsequently, a vehicle allocation model was developed using integer programming, taking into account factors such as call demand in each area, distance from the station, and available resources. This paper constructs a station construction priority evaluation framework based on an approximate hypercube model and employs the entropy-weighted TOPSIS method to comprehensively rank candidate sites. However, this model is based on a core assumption: that future emergency demand and real-time traffic conditions will fluctuate within expected ranges. Whilst this assumption is reasonable under conventional operating scenarios, the emergency response system inherently operates in a highly uncertain environment—mass incidents may cause an instantaneous surge in demand, major traffic accidents may drastically extend response radii, and traffic congestion may render previously viable return routes ineffective. As highlighted in research on the optimization of emergency resource allocation under uncertain conditions, large-scale emergencies often result in critical resource shortages in affected areas. Conducting emergency response under time-critical, supply-constrained and uncertain conditions is a difficult and complex process, requiring a comprehensive consideration of equity, reliability and timeliness, whilst also balancing costs. Therefore, a return-to-base decision-making model with practical reference value must, whilst pursuing 'routine efficiency', also ensure 'extraordinary resilience'.

## 5. CONCLUSION

This paper addresses the imbalance between supply and demand of emergency medical resources under the background of population aging and urbanization. It systematically investigates the dynamic return-to-base decision problem for ambulances after completing a mission, proposes a dynamic repositioning strategy based on the approximate hypercube queuing model, and develops a multi-objective return priority evaluation model. Through theoretical analysis and model construction in the preceding chapters, the main contributions of this paper are as follows:

- (1) Problem reformulation and strategic innovation: The research breaks through the traditional static "fixed station plus return to original station" mindset, upgrading the return decision from "passive post-mission return" to "proactive dynamic pre-allocation." This transformation advances return strategies from a single objective (returning to the nearest station) to a multi-objective framework that balances response efficiency, operational cost, and system resilience.
- (2) Model development and quantitative evaluation: Based on Jarvis's approximate hypercube queuing model, a two-dimensional indicator system is established, comprising demand-response capability (spatial coverage, load balancing, response timeliness) and return costs (direct costs plus opportunity costs). This system provides a quantifiable basis for dynamic return decisions.
- (3) Multi-objective decision-making framework: Using the Analytic Hierarchy Process (AHP) and TOPSIS method, comprehensive priority scores are calculated for candidate standby locations.

Combined with integer programming for vehicle allocation, a complete theoretical loop from "prediction – evaluation – decision" is formed.

(4)Balancing routine efficiency and extraordinary resilience: While pursuing operational efficiency under normal conditions, the paper acknowledges that the current model relies on the assumption that future demand and real-time traffic conditions fluctuate within expected ranges. It highlights the need to account for system resilience under extreme events (mass casualties, major accidents, severe congestion), thereby opening avenues for incorporating uncertainty optimization in future work.

In summary, the dynamic return strategy and multi-objective decision model proposed in this paper can effectively improve the responsiveness of emergency resources under spatiotemporally varying conditions, providing theoretical and methodological support for dynamic dispatch in urban pre-hospital emergency medical systems. Future research may further integrate real-time traffic data, uncertainty scenario simulation, and multi-agent reinforcement learning to enhance model adaptability and robustness under extreme circumstances.

## REFERENCES

- [1] Lv Chuanzhu, Huang Hang, Zhang Wei. Analysis of Performance for Ambulance Emergency Response Based on Hypercube Queuing Model[J], 2012.
- [2] Tapia C A B. Emergency Medical Service Ambulance System Planning: History and Models[D]. University of California, Santa Barbara, 2017.
- [3] Richard C. Larson,A hypercube queuing model for facility location and redistricting in urban emergency services,Computers & Operations Research,Volume 1, Issue 1,1974,Pages 67-95,ISSN 0305-0548,
- [4] Golabian, H, Arkat, J, Tavakkoli-Moghaddam, R. et al. A multi-verse optimizer algorithm for ambulance repositioning in emergency medical service systems. J Ambient Intell Human Comput 13, 549–570 (2022).
- [5] Guigues, Vincent, et al. “New Heuristics for the Operation of an Ambulance Fleet under Uncertainty.” ArXiv.org, 2024.
- [6] V. Bélanger, E. Lanzarone, V. Nicoletta, A. Ruiz, P. Soriano,A recursive simulation-optimization framework for the ambulance location and dispatching problem,European Journal of Operational Research,Volume 286, Issue 2,2020,Pages 713-725.