Universitat Autònoma de Barcelona

Modeling and Simulation for Healthcare Operations Management using High Performance Computing and Agent-Based Model

Zhengchun Liu

Universitat Autònoma de Barcelona

Escola d'Enginyeria.

Departament d'Arquitectura de Computadors i Sistemes Operatius

Modeling and Simulation for Healthcare Operations Management using High Performance Computing and Agent-Based Model

Thesis submitted by **Zhengchun Liu** for the degree of Doctor of Philosophy by the Universitat Autònoma de Barcelona, under the supervision of Dr. Emilio Luque Fadón, done at the Computer Architecture and Operating Systems Department, PhD. in Computer Science.

Barcelona, June 2016

Modeling and Simulation for Healthcare Operations Management using High Performance Computing and Agent-Based Model

Thesis submitted by **Zhengchun Liu** for the degree of Doctor of Philosophy by the Universitat Autònoma de Barcelona, under the supervision of Dr. Emilio Luque Fadón, done at the Computer Architecture and Operating Systems Department, PhD. in Computer Science.

Supervisor	Author
Dr. Emilio Luque Fadón	Zhengchun Liu

Barcelona, June 2016

Abstract

Hospital based emergency departments (EDs) are highly integrated service units to primarily handle the needs of the patients arriving without prior appointment, and with uncertain conditions. In this context, analysis and management of patient flows play a key role in developing policies and decision tools for overall performance improvement of the system. However, patient flows in EDs are considered to be very complex because of the different pathways patients may take and the inherent uncertainty and variability of healthcare processes. Due to the complexity and crucial role of an ED in the healthcare system, the ability to accurately represent, simulate and predict performance of ED is invaluable for decision makers to solve operations management problems. One way to realize this requirement is by modeling and simulation. Armed with the ability to execute a compute-intensive model and analyze huge datasets, the overall goal of this study is to develop tools to better understand the complexity (explain), evaluate policy (predict) and improve efficiencies (optimize) of ED units. The two main contributions are:

(1) *An agent-based model for quantitatively predicting and analyzing the complex behavior of emergency departments.* The objective of this model is to grasp the non-linear association between macro-level features and micro-level behavior with the goal of better understanding the bottleneck of ED performance and provide ability to quantify such performance on defined condition. The model was built in collaboration with healthcare staff in a typical ED and has been implemented in a NetLogo modeling environment. In order to validate its adaptivity, the presented model has been calibrated to emulate a real ED in Spain, simulation results have proven the feasibility and ideality of using agent-based model & simulation techniques to study the ED system. Case studies are provided to present some capabilities of the simulator on quantitively analyzing ED behavior and supporting decision making.

(2) *A simulation and optimization based methodology for calibrating model parameters under data scarcity.* To achieve high fidelity and credibility in conducting prediction and exploration of the actual system with simulation models, a rigorous calibration and validation procedure should firstly be applied. However, one of the key issues in calibration is the acquisition of valid source information from the target system. The aim of this contribution is to develop a systematic method to automatically calibrate a general emergency department model with incomplete data. The proposed calibration method enables simulation users to calibrate the general model for simulating their system without the involvement of model developers. High performance computing techniques were used to efficiently search for the optimal set of parameters. The case study indicates that the proposed method appears to be capable of properly calibrating and validating the simulation model with incomplete data. We believe that an automatic calibration tool released with a general ED model is promising for promoting the application of simulation in ED studies. In addition, the integration of the ED simulator and optimization techniques could be used for ED systematic performance optimization as well.

Starting from simulating the emergency departments, our efforts proved the feasibility and ideality of using agent-based model methods to study healthcare systems.

Resumen

Los servicios hospitalarios de urgencias (SU) son servicios altamente integrados que gestionan las necesidades primarias de los pacientes que llegan sin cita previa y en condiciones inciertas. En este contexto, el análisis y la gestión de flujos de pacientes ejercen un papel clave en el desarrollo de las políticas y herramientas de decisión para mejorar la actuación global del sistema. Pese a esto, los mismos flujos de pacientes en un SU son considerados muy complejos debido a los diferentes caminos que pueden tomar los pacientes y a la inherente incerteza y variabilidad de los servicios de salud. Debido a la complejidad y al papel crucial de un SU en el sistema sanitario, la habilidad de representar, simular y predecir el rendimiento de un SU tiene un valor incalculable para quien toma decisiones para resolver los problemas de la gestión de las operaciones. Una manera a percatarse de las consecuencias es mediante el modelado y la simulación. El objetivo general de este estudio es desarrollar herramientas para entender mejora la complejidad (explicar), evaluar la política (predecir) y mejorar la eficiencia (optimizar) de unidades de SU. Las dos aportaciones principales son:

(1) *Un modelo basado en agentes para predecir y analizar cuantitativamente el complejo comportamiento de los servicios de urgencias.* E objetivo de este modelo es captar la asociación no lineal entre las funciones de nivel macro y el comportamiento a nivel micro con el objetivo de comprender mejor el cuello de botella del rendimiento de los SU y proporcionar la capacidad de cuantificar este rendimiento en una condición dada. El modelo fue construido en colaboración con el personal de asistencia sanitaria en un SU típica y ha sido implementado en el entorno de modelado NetLogo. Se proporcionan casos de estudio para presentar algunas capacidades del simulador que analizan cuantitativamente el comportamiento del SU así como el soporte a la toma de decisiones.

(2) *Una metodología de simulación basada en la optimización para el calibrado de los parámetros del modelo en condiciones de escasez de datos.* Para conseguir una alta fidelidad y credibilidad en la realización de la predicción y exploración del sistema actual con modelos de simulaciones se ha de aplicar en primer lugar una calibración rigurosa y un procedimiento de validación. No obstante, una de las cuestiones clave en el calibrado es la adquisición de información de una fuente válida para el sistema destino. El objetivo de este trabajo es desarrollar un método sistemático para calibrar automáticamente un modelo genérico de un servicio de urgencias con datos incompletos. El método de calibrado propuesto permite a los usuarios de la simulación calibrar el modelo genérico para la simulación de los propios sistemas sin involucrarse en el modelo. Las técnicas de computación de alto rendimiento se utilizaron para buscar el conjunto óptimo de parámetros de manera eficiente. Creemos que una herramienta de calibrado automático publicado juntamente con un modelo genérico de un SU es prometedor para la promoción de la aplicación de la simulación en los estudios de SU. Además, la integración de técnicas de simulación de un SU i optimación podrían también ser utilizada para la optimización sistemática de un SU.

A partir de la simulación de los servicios de urgencias, nuestros esfuerzos probaron la viabilidad y la idoneidad de la utilización del modelo de simulación y técnicas basadas en agentes para el estudio del sistema de salud.

Resum

Els serveis hospitalaris d'urgències (SU) són serveis altament integrats que gestionen les necessitats primàries dels pacients que arriben sense cita prèvia i en condicions incertes. En aquest context, l'anàlisi i gestió de fluxos de pacients exerceix un paper clau en el desenvolupament de les polítiques i instruments de decisió per millorar l'actuació global del sistema. Malgrat això, els fluxos dels pacients en un SU són considerats molt complexes degut als diferents camins que poden prendre els pacients o que pot tindre el mateix la inherent incertesa i variabilitat dels serveis de salut. Degut a la complexitat i el paper crucial d'un SU en el sistema sanitari, l'habilitat per representar, simular i predir el rendiment d'un SU té un valor incalculable per a qui pren les decisions per resoldre els problemes de gestió d'operacions. Una manera d'adonar-se de les conseqüències is mitjançant el modelatge i la simulació. La objectiu general d'aquest estudi és desenvolupar eines per entendre millor la complexitat (explicar), avaluar la política (predir) i millorar l'eficiència (optimitzar) d'unitats d'un SU. Les dues aportacions principals són:

(1) *Un model basat en agents per predir i analitzar quantitativament el complex comportament dels serveis d'urgències.* L'objectiu d'aquest model és captar l'associació no lineal entre les funcions de nivell macro i el comportament a nivell micro amb l'objectiu de comprendre millor el coll d'ampolla de rendiment dels SU i proporcionar la capacitat de quantificar aquest rendiment en una condició donada. El model va ser construït en col·laboració amb el personal d'assistència sanitària en un SU típic i ha estat implementat en l'entorn de modelatge NetLogo. Els resultats de la simulació han demostrat la viabilitat i la idoneïtat de la utilització del model de simulació i tècniques basades en agents per estudiar un SU. Es proporcionen casos de estudi per presentar algunes capacitats del simulador que analitzen quantitativament el comportament del SU així com el suport a la presa de decisions.

(2) *Una metodologia de simulació basada en l'optimització per al calibratge dels paràmetres del model en condicions d'escassetat de dades.* Per aconseguir una alta fidelitat i credibilitat en la realització de la predicció i l'exploració del sistema actual amb models de simulació, s'ha d'aplicar en primer lloc un calibratge rigorós i un procediment de validació. No obstant això, una de les qüestions clau en el calibratge és l'adquisició d'informació d'una font vàlida del sistema de destinació. L'objectiu d'aquest treball és desenvolupar un mètode sistemàtic per calibrar automàticament un model genèric d'un servei d'urgències amb dades incompletes. El mètode de calibratge proposat permet als usuaris de la simulació calibrar el model genèric per a la simulació del seu propis sistemes sense involucrar-se en el model. Les tècniques de computació d'alt rendiment es van utilitzar per buscar el conjunt òptim de paràmetres de manera eficient. Creiem que una eina de calibrat automàtica publicat juntament amb un model genèric d'un SU és prometedor per a la promoció de la aplicació de la simulació en SU. A més, la integració de tècniques de simulació d'un SU i optimització podrien també ser emprades per a la optimització sistemàtica d'un SU.

A partir de la simulació dels serveis d'urgències, els nostres esforços van provar la viabilitat i la idoneïtat de la utilització del model de simulació i tècniques basades en agents per estudiar el sistema de salut.

Acknowledgments

First and foremost, I would like to thank my advisor Emilio Luque. He has been very generous with his time, and taught me to be a better problem solver, writer, presenter and researcher. Just as important, he has helped form my own perspective on what problems are interesting and important. His indefatigable positivity and unshakable belief that good research can have lasting impact have shaped me profoundly. Finally, he has been a great person to work and socialize with for the last three years.

I would like to thank doctor Francisco Epelde of the Emergency Department of the Hospital Universitari Parc Taulì, for the meetings and data provided for the preparation of this research. Without their data, much of this research would not have been possible. My deep appreciation for Gemma Roque and all the staff at the Computer Architecture and Operating Systems Department, and the High Performance Computing for Efficient Applications and Simulation Research Group, especially to Dolores Rexachs, Manel Taboada, Remo Suppi and Eduardo Cesar Cabrera Flores(former colleague).

I am incredibly fortunate to have so many friends and colleagues, without whom I may never have completed this thesis. Special thanks to Joe Carrión and Cecilia Jaramillo, their kind help makes me feel like living in my own country. All of these are the precondition for the fulfillment of my research.

I would also like to thank all the people in the Oak Ridge National Laboratory for their hospitality during my stay in United States, especially to Kalyan S. Perumalla and Rochelle L. Womble.

Last but not least, I would like to THANK my parents and my wife Qi Liu, for their patience, wisdom, example, care, love, and support.

This research has been supported by the MINECO (MICINN) Spain under contracts TIN2011-24384 and TIN2014-53172-P, and has been partially supported by a grant from the China Scholarship Council (CSC) under reference number: 201306290023.

Barcelona, May 27, 2016, Zhengchun Liu

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Emergency Departments (EDs) serve as the primary gateway to the acute healthcare system, are struggling to provide timely care to a steadily increasing number of unscheduled visits [1]. In this thesis, we build high fidelity simulation tools to identify system bottleneck, quantitatively predict the benefit and cost of a policy, and discovery knowledge for a better understanding of the complex ED system. In particular, we use agent-based model and simulation techniques to model the interaction of ED components. We then applied high performance computing techniques to execute the model and analyze simulation results. This first chapter presents a broad overview of EDs, challenges for modeling and simulating EDs as well as the motivation for simulating an ED.

1.1 The Emergency Department

An ED, also known as an accident & emergency department (A & E), emergency room or casualty department, is a medical treatment facility specializing in emergency medicine. EDs are responsible for managing patients who present without prior appointment. They operate 24 hours per day, 365 days per year although staffing levels may be varied in an attempt to reflect patient volume. EDs are the gatekeeper of the hospital's different services, and also an important link between the community (primary care) and the hospital [2]. Due to the unplanned nature of patient attendance, the ED must provide initial treatment for a broad spectrum of illnesses and injuries, some of which may be life-threatening and require immediate attention. In addition, population aging in developed countries will challenge ED as old patients visit these health services more frequently and present special needs.

EDs are suffering from increasing stress in recent years, due to remarkable growth in demand, limited productivity, and reduced budgets which mostly led to overcrowding in EDs [3]. For example in Spain, all autonomous communities except Madrid have faced continuous cuts in the healthcare budget since 2010. As a consequence, patient congestion and long waiting times in EDs are the most common problems in public hospitals [4]. Moreover, as investigated in Ref. [5]: "patients are no longer prepared to accept poor quality service; they expect that services are well organized from a 'customer' perspective. That is, the service concept has shifted from optimizing use of resources to finding the tradeoff between quality of service for patients and operational efficiency for healthcare providers" [5]. For this purpose, ED managers must control problems related to process flow (patients and information), as well as internal restructuring reflected by resource pooling [6]. Moreover, an ED is a highly complex, emotionally charged work environment where inefficient operational decisions may lead to serious consequences, or even unnecessary deaths. Therefore, efficient management of patient flow in EDs has become an urgent issue for many hospital administrations while healthcare personnel are neither prepared nor trained to solve such problems.

Healthcare and ED management is concerned with the mission of improving the healthcare delivery system, i.e., organization, planning, coordination, staffing, evaluating and controlling of healthcare services. Their main objective is to provide affordable healthcare of the best quality. Since a key activity in emergency management is planning and preparation for unscheduled scenarios. If the right safety measures are implemented beforehand, harmful effects can be significantly mitigated. However, the prediction, explanation & optimization of the system performance are challenging for a complex system like an ED. Since performing experiments directly with an actual ED system is a time-consuming, irresponsible (putting patients in risk situation) and dangerous method, a virtual platform which mimic the behavior of a real ED for conducting experiments to study disordered system behavior is promising. For example, from a point view of healthcare operations managers, this virtual ED platform could help us make decisions based on information extracted from data, instead of guessing.

1.2 Simulation for Healthcare Operation

1.2.1 Modeling & simulation

Computer simulation based methods have enjoyed widespread use in healthcare system investigation and improvement in recent years [7]. From a computation theoretical perspective, simulation of a system can be defined as an *"imitation (on a computer) of a system as it progresses through time"* [8]. Normally, simulation is used for a dynamic as opposed to static analysis [9] of the healthcare system. Today, many researchers are interested in modeling and simulating the operation of EDs because it helps managers carry out different kinds of analyses, such as: (1) resource utilization (human, equipment, and space resources) for alternative resource scheduling and allocation policies, (2) finding the most influential factors affecting the performance of the system in a given situation, (3) exploring the interrelationship between agents' behavior and system performance under various scenarios, (4) estimate system robustness under unexpected situations (e.g., outbreak of infectious diseases) [10, 11, 12, 13, 14, 15]. An important characteristic of simulation modeling is that it allows us to evaluate various scenarios so that "what-if" analyses can be performed and improvement initiatives can be taken [16]. Once developed, the model can be used to simulate extended horizons to thoroughly understand the performance of the plan and resources required to implement it. In addition, modifications or alternatives to the plan can be quantitively evaluated to improve overall responsiveness.

In summary, simulation is the imitation of the operation of a real-world system. It allows users to reconstruct a more comprehensive representation of real-world system, so as to perform experiments for predicting, explaining and optimizing operations. It is used in many contexts, such as performance optimization, safety engineering, testing, training, education, scientific research, and video games.

1.2.2 Simulation method

With respect to modeling and simulation methods for studying healthcare systems, Discrete Event Simulation (DES), System Dynamics (SD) and agent-based model (ABM) are the three main approaches [17] (thoroughly review will be given in section 2.2). SD and DES focus on system-level behavior but they differ in how the system is modeled and how time is simulated [17]. The system dynamics (SD) models represent entities as continuous variables whose states change continuously with time [18], whereas discrete event simulation (DES) models contain individual components whose states only change at discrete moments in time [19, 20, 21]. In either case, the goal is to aggregate the system behavior and draw conclusions about how the system evolves over time under internal and external "forces". These techniques can provide valuable insight to problems in healthcare operations management and are ideally suited for many such problems. However, SD and DES requires excessive abstractions from actual system to form computer algorithms. This requirements may cause difficulty in explaining model concepts to healthcare professionals who are not trained in mathematical or computational disciplines.

In contrast to system dynamics and discrete event simulation methodologies, ABM focuses on modeling individuals, interactions between individuals, and in some cases, interactions with a physical or influential surrounding environment [22]. That is, compare with system-level model methods, an ABM is a more realistic modeling approach for many problems, especially problems in which there are multiple types of actors that interact in different ways. For these cases, it is very straightforward to model these actors as agents that have distinct sets of behaviors and characteristics, and no systemic level assumption is needed. They are thus more explainable than most SD and DES models because of their direct correlation to reality. This explainability is an important factor in gaining the confidence of healthcare professionals and ultimately having an impact. In addition, from the point view of the system as a whole, few assumptions need to be made because system behavior is determined by the activities at the individual level, i.e., no systemic abstraction needed. Normally, few assumptions to the model could bring high flexibility and adaptability. The simulation users, either to develop new models based on existing ABM or to conduct simulation experiments for prediction, could have more sufficient flexibility to customize and more straightforwardly apply changes to the system.

As in all methodologies, there are disadvantages to ABM as well. ABMs can become very complex when they incorporate a lot of details. When this happens, it becomes difficult to separate the actual effect of each input parameter in the model as well. In addition, ABMs can become computationally expensive, requiring excessively long computer run times for executing model and analyzing simulation results. Fortunately, this problem has been alleviated to some degree by high performance computing techniques, but it demands additional developmental resources that are not typically required by SD or DES models. In this study, agent-based modeling and simulation techniques were used, the following subsection 1.2.3 will further introduces its features and explain our reason of choosing agent-based modeling and simulation method.

1.2.3 Agent-based modeling and simulation

Although, there is not a general accepted definition, it can be said that an ABM (or "individual-based" modeling, as the approach is called in some fields) is a computational model of a population of agents (the system's components, agents and system components are equally used in this thesis) and their interactions, as well as the interaction of the former with the environment. This type of modeling methods is commonly used to analyze complex systems that are difficult to be tackled by classical or formal methods [22] such as differential equations. Since ABM is a typical bottom-up method, its direct simulation results are interaction information (i.e., interaction records between agents and, agent with environment). These result can be used to extract information to indicate system-level behavior. Having said that, an

ABM is comprised of autonomous, decision making entities called agents that have the ability to interact with each other and with the environment they stay in. Each agent has a relatively simple set of rules for how it responds to its environment and other agents. Thus, in the agent-based simulation, agents perform on behalves of their actual person (or machine, group of person, department) to make decisions.

An agent-based model is commonly used to search for explanatory insight into the collective behavior of agents that obeying simple rules and interacting in a shared environment. [23]. This explainability could be useful to indicate the system bottleneck and insight root-cause of the bottleneck. Furthermore, an ABM also provides flexibility for construction of models (or adding new feathers to an existing ABM) in the absence of knowledge about the global interdependencies [24]. M. Gul, and A. F. Guneri in Ref. [25] made a comprehensive review of ED simulation application studies, they stated that agent-based simulation (ABS) has a significant capability and is becoming an emerging methodology for ED simulation applications. They also found that the number of publications with ABS and ABS and others from 2011 to 2015 are steadily increasing compared to before 2011. Furthermore, the system analysis in many application domains is not only about accuracy in prediction, interpretability is also extremely important to have transparency in predictive modeling. That is, domain experts do not tend to prefer "black box" predictive models. They would like to understand how predictions are made, and possibly, prefer models that emulate the way a human expert might make a decision [26]. Consequently, the agent-based model is capable of answering questions such as "why the system behaves the way it does".

In summary, agent-based modeling and simulation is a new approach to modeling systems comprised of autonomous, interacting agents. Computational advances have made possible to execute a growing number of ABMs across a variety of application domains. As summarized in Ref. [27], applications range from modeling agent behavior in the stock market, supply chains, and consumer markets, to predicting the spread of epidemics, mitigating the threat of bio-warfare, and understanding the factors that may be responsible for the fall of ancient civilizations [27]. We also believe the imagination in Ref. [22] that in the future, virtually most of the computer simulations will be in the form of agent-based simulations because of the natural way that agent models can represent the actual system issues, and the close similarity of agent modeling to the predominant computational paradigm of object-oriented programming.

1.2.4 Work process

Starting from understanding how a real ED system work in reality, the process of building a simulator to mimic the target system is shown in Figure 1-1, and explained as followings:

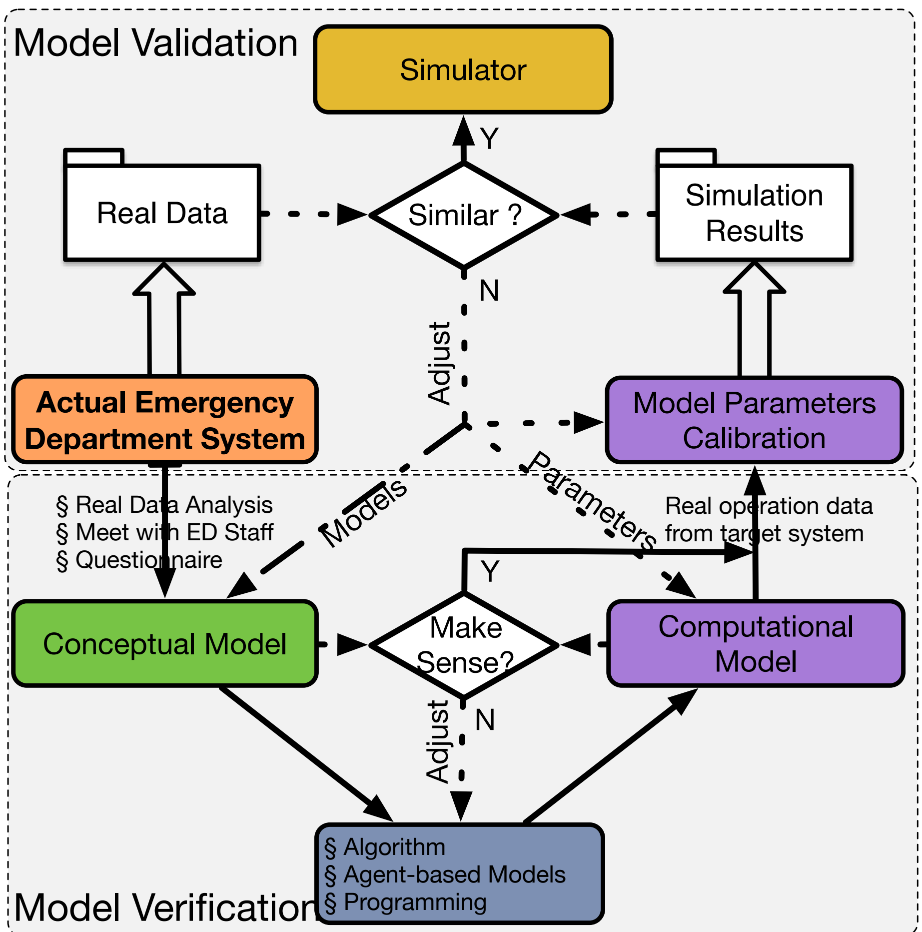


Figure 1-1: The steps of modeling and simulating an emergency department system.

1. The first step is to build a conceptual model based on knowledge from inter-

viewing ED staff and analyzing ED's real operation data. The conceptual model would clearly show how the system is operated in actual situation. Qualitative predictions could be made according to the conceptual model (by our intuition). To step ahead, algorithm and agent-based models of the system components are built based upon the conceptual model. The algorithm and ABMs are the "translation" of the conceptual model.

2. In the second step, agents' models and algorithms could be implemented as a computational model by using programming language. Then the computation model could be executed to make quantitative prediction. Since there may be unreasonable assumptions in building ABMs and mistakes in implementation, the computational model should be verified before stepping ahead. The verification will be done via comparing simulation results with analytical results from the conceptual model. Since the conceptual model could not carry out quantitative predictions, the comparison result is "make sense" or not (as shown in Figure 1-1). For example, more patients should cause longer patients' length of stay. The goal of verification is to ascertain whether the model implements the assumptions correctly.

3. After verification, the computational model could represent the general behavior/pattern of the ED system. However, actual systems are very different from each other, the computational model still needs parameters to set up for a target ED. The adjustment of parameters according to the comparison of simulation and actual data is known as model parameter calibration (A.K.A. tuning). If the actual data is not enough to set up all the model parameters, i.e., data scarcity, the calibration need to be applied to "search" the optimum value for parameters. In the end, if the similarity of simulation results and the actual data is close enough (i.e., the assumptions which have been made are reasonable with respect to the real system), the computational model can then be used as a tool to represent/mimic the target system.

The following chapters will detail all of these steps, and a validated simulator for a typical Spanish ED will be shown in the end. By using the simulator, some demos will be presented to show its potential use in reality.

1.3 Motivation

1.3.1 Operations management

Prediction, explanation & optimization are challenging for a complex system like healthcare systems. Here, a complex system is made up of an interconnected set of relationships between individuals, organizations and groups, and all of which have unique aims, motivations, beliefs and cultures. Given this complexity, it is very difficult to evaluate the cost and benefit of a major policy change through unilateral research or consultation processes, and efficient management of the system also becomes a big challenge. In the field of engineering, simulation can be used for dynamic as opposed to static analysis of healthcare system. Hence the first motivation is to build accurate models for representing healthcare systems, with the purpose of supporting decisions for efficient healthcare operations management.

Moving forward, since one motivation of modeling and simulating healthcare system is to support making better management decisions. Considering the fact that healthcare systems are large, dynamic, complex environment. In this circumstance, not only accuracy in prediction, interpretability is also extremely important to have transparency in predictive modeling. Therefore, different with conventional mathematic models or data-driven models, we are working on building a transparency model of the complex healthcare system to predict the system-level behavior from micro-level interactions, so as to see the forest through the trees, i.e., prediction and explanation.

1.3.2 A platform for studying healthcare-related problems

Since an ED is the main entrance to a healthcare system and it faces uncertainty everyday. Its efficiency and quality of service in ED have big influence on the whole healthcare system. Thus, due to the flexibility, adaptability and customizability an ABM could provide, the emergency department model could be used as a platform to study emergency department related problems, like bacteria propagation in ED. As the fact that humans have difficulty in understanding the complexities caused by the dynamic and systematic nature of certain problem [28, 29]. The bottom-up ED model could also be used to study disordered system behavior based on integration of first-principles model and data-driven model (based upon real operation data).

Moreover, the demographic development resulting shows that the population distribution has already changed considerably and will further change over the next decades. That is, the number of older people with chronic conditions will increase intensively. This expected increase of older people in society poses an immense challenge to the public health care system. The integrated care could be a promising concept in redesigning care to tackle this. However, changing from conventional public healthcare policy to integrated care requires a lot of feasibility studies, and major health policy changes often have wide-ranging impacts on our community. The evidence of process redesign interventions regarding their ability to improve quality of care must be interpretable for populations. Therefore, this simulation study could judge the feasibility and ideality of using agent-based model & simulation techniques to study healthcare system. Then, the developed framework could be used as a step towards building a full model of integrated care system. The final model will be able to represent a comprehensive tool to quantitatively evaluate prospective planned changes to the integrated care system for decision making, and open a wide field of possible simulation scenarios for a better understanding of the complex integrated care system.

1.3.3 The ultimate goal

In summary, the ultimate objective of this work is to propose an accurate model of EDs and design accessory tools to make the model accessible (as a platform) for ED related researchers. This model will be used to predict behavior of EDs under various conditions such as staffing change, physical resources resize and influx of patients (e.g., in influenza season). In summary, the presented work is from a long-term project which aims to develop a generic ABM of EDs for the purpose:

1. In the management of ED, the simulator can work as parts of decision support system to quantitively evaluate effects of proposals.

2. Make the ED simulator work as a platform to study ED related problems, for example, one of our researchers is using the presented simulator to study Methicillin-resistant Staphylococcus aureus (MRSA) transmission in ED.

3. Use the simulator to study disordered system behavior based on integration of first-principles model and data-driven model (based upon real operation data). There is a researcher in our group uses the presented simulator and novel analytical methods to discover hidden knowledge of the ED.

4. Prove the feasibility and ideality of using agent-based modeling and simulation to study healthcare system. Then, build a full model of public healthcare system with the framework developed in simulating ED. A researcher in our group is dedicating to using the framework developed in this study to build a full model of integrated care system.

5. Enrich the simulation-based healthcare study ecosystem. Provide series of tools to simplify the study of healthcare systems by using simulation techniques. For example, model and simulation framework, optimization methods and the integration of simulation and optimization, model parameters calibration methodology and model validation tools.

1.4 Thesis Contributions

The two main contributions of this thesis are: (1) a general agent-based model for simulating Spanish type emergency department, and (2) a systematic method for calibrating model parameters to simulate a specific emergency department with data scarcity.

1.4.1 A general agent-based emergency department model

Based upon some previous studies in our research group [30, 31, 32] which simulated an area in ED for non-urgent patients, this thesis presented a full simulation model for EDs (focused on Spanish type). As shown in Figure 1-1, start from understanding operation of actual systems, and end with a simulator for a specific ED. The ED model is developed to enhance understanding of ED's complexity, completely evaluate "what-if" scenarios and perform experiment with the system prior to making significant changes. The ED managers can use it to quantify the impact of proposed decisions on patient flow as well as system efficiency prior to implementation. This research was devised in close collaboration with experienced ED staff in the Hospital Universitari Parc Taulí (a University tertiary level hospital in Barcelona, Spain that provides care service to a catchment area of 500, 000 people, and attends more than 160,000 patients per year in the ED). With the flexibility and customizability the simulator provides, the presented model has been used by other researchers to study the Methicillin-resistant Staphylococcus aureus (MRSA) transmission in EDs in Ref. [33, 34], and as a sensor of EDs to provide data for knowledge discovery in Ref. [35]. Some case studies and demo applications carried out by using this simulator have been presented in Ref. [36, 37].

1.4.2 A systematic method for calibrating model parameters

Data scarcity is a common problem in setting up a general model to simulate a target system. For example, the duration of healthcare staff's service time is among the most common missing pieces of information because it is out of the scope of the information system. Therefore, we developed a systematic method to automatically calibrate a general ED model with incomplete data. The proposed calibration method enables simulation users to calibrate the general model for simulating their system without the involvement of model developers. With a few expectations, the proposed systematic method has been proved to be able to find the parameters for fitting the duration of service, with which the simulated results and the actual data were consistent. We believe that an automatic calibration tool released with a general ED model is promising for promoting the application of simulation in ED-related studies. Furthermore, a framework that integrates the simulation model and optimization method was developed for calibrating model parameters. Although it is originally designed to search optimal values for model parameters with incomplete real data, the framework could also be used for systematic performance optimization (with customized objective function and variable constraints).

1.5 Thesis Outline

This thesis, that gathers the work carried out by the author in the last three years of research, is composed by: (1) a general agent-based model of Spanish type EDs. (2) a systematic method to automatically calibrate a general ED model with incomplete data for simulating a target ED and, (3) several case studies carried out with the developed ED simulator to show potential use of the tool. This thesis interpolates material from several publications by the author. The chapter 3 is based on Reference [38] and one paper currently under peer-review. The chapter 4 is heavily based on another paper by the author which is currently under peer-review. Meanwhile, chapter 5 uses material from References [36], chapter 6 uses case studies in Reference [37]. Some material from each of these papers has also been incorporated into this introductory chapter 1 and literature review chapter 2. Furthermore, while I wrote the text of the thesis, naturally not all of the ideas and work presented are my own. Besides the presented background material, many of the results and ideas in this thesis have been developed through collaboration with various colleagues and former colleagues, in particular my supervisor Dr. Emilio Luque. The outline of this dissertation is illustrated in Figure 1-2 and detailed as follows.

The chapter 2 gives the state-of-the-art of the emergency department simulation and its application, commonly used modeling and simulation approaches, and model parameters calibration methods.

In chapter 3 our attention is focused on the development of a general ABM for

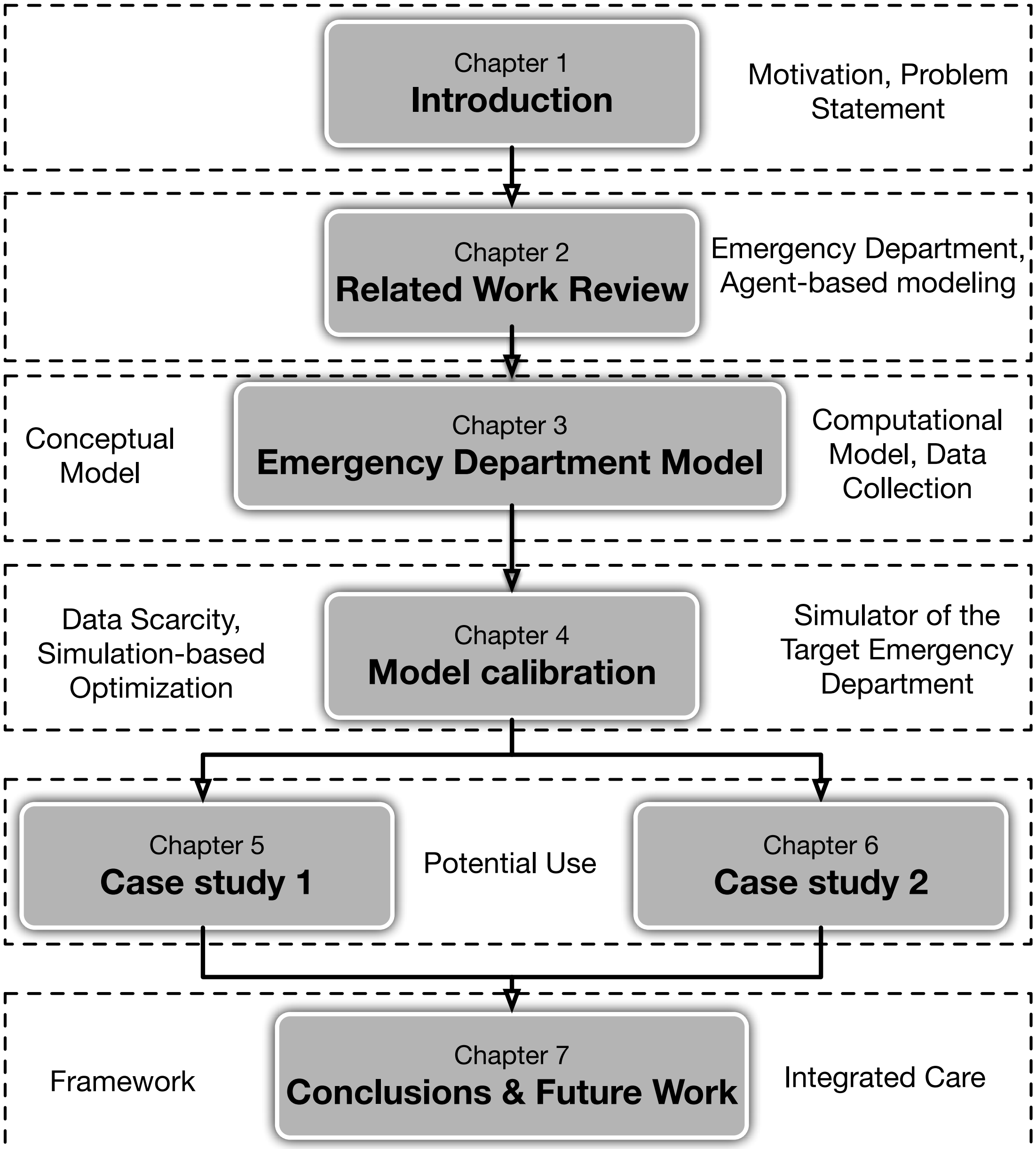


Figure 1-2: Thesis outline.

simulating EDs in Spain. In which, the section 3.1 describes the conceptual model of ED observed via analyzing real operation data and meeting with experienced ED staff. The full agent-based ED model is detailed in section 3.3. It contains two parts, subsection 3.3.1 describes the model of agents considered in ED, and subsection 3.3.2 describes the interaction model among agents as well as agents with the environment. The section 3.4 formulates the agent models into algorithm and implemented with programming language so that can be executed to emerge systemic behavior. When getting the computational model, in section 3.5 we present the methods to design experiments as well as efficiently execute the model for predicting and discovering hidden features. The simulation results that validated the model are presented in section 3.6.

In chapter 4, we proposed an automatic calibration tool for calibrating model parameters with incomplete real data from target system. Basically, data scarcity for setting up a model of complex system is common because many of the data (e.g., time of each interaction between patient and service-provider) is out the scope of the information system to monitor in real situation. The simulation-based optimization was used to find the optimal values of parameters with carefully proposed object function. The simulation-based optimization was conducted by using the APPSPACK [39, 40, 41] developed by Sandia National Laboratory. According to the practical requirements of evaluating a simulation-based objective function, an initial distance-based lookup mechanism was proposed to further accelerate the optimization process.

In chapter 5 and chapter 6 we give some case studies carried out by using the ED simulator in order to show its potential usage. Where, chapter 5 gives a process of using the simulator for decision support to deal with continuously increasing number of patient attending at ED (overcrowding). In which, the simulator shows its ability to clearly identify and quantify the bottleneck of the system. With a simulator, the proposed solutions by a policy makers could be evaluated with the simulator before implementing to the system. These evaluations could avoid improper changes to the system and putting patients in risk situation. In chapter 6, we give two examples on knowledge discovery with the purpose of better understanding the complex system.

Finally, chapter 7 closes the thesis with our conclusions and potential future contributions.

Chapter 2

Literature review

"If I have seen further it is by standing on the shoulders of Giants."

- Isaac Newton

High demand of healthcare services due to changes in population demography, technological and medical advancements, budget limitations has direct effect on medical staff and medical organizations in particularly hospitals [42]. In this chapter, concepts about what modeling & simulation is, its application on studying healthcare system, and characteristics are outlined. In particular, the state-of-the-art of modeling and simulation for Emergency Department (ED) is described.

2.1 Emergency Department Simulation

Modeling and simulation provides a quantitive way to analyze the behavior and predict the performance of an ED under designed scenarios. There have been fruitful efforts in developing simulation models for solving healthcare operations management problems [43], [44], [45]. Historically, the complexity of models was often limited by mathematical tractability [46]. That is: when differential calculus was the only approach we had for modeling, we had to keep models simple enough to "solve" mathematically and so, we were often need highly abstract concepts to keep a complex system as simple as possible or limited to modeling quite simple problems. Conventional simulation methods, for example the empirical data based queuing model and Markov chain, can provide the overall behavior of ED, e.g. [47] and [48]. However, the conventional simulation methods mostly focus on compute systemic key performance indicators. These systemic key performance indicators are either difficult to be comprehensive or require professional knowledge on mathematic, statistics and simulation to understand. Thus, conventional methods lack ability on explaining how the predictions are made and explainability is often important than sheer number.

With respect to the importance of interaction information that details how the system components behave over time. Boudreaux et al.[49] reviewed literatures on patient satisfaction study in ED. Contrary to popular belief, research has repeatedly shown that the systemic performance indicators such as actual waiting times and overall length of stay are relatively unimportant in determining satisfaction. What does seem to be important, however, is the patient's subjective experience of the waiting time. This means how satisfied the patient is with interpersonal interactions with ED physicians and nurses. Therefore, the detailed interaction information is more important to evaluate policies for improving Quality of Service (QoS). One technique that shows considerable promise for this requirement is agent-based modeling.

Simulation has long been used in healthcare system operation research, Rising et al. [50] are among one of the earliest publications on using computer modeling and simulation for improving healthcare service. The authors use the Monte Carlo simulation model for analyzing effects of alternative decision rules for scheduling appointment periods during the day to increase patient throughput and physician utilization. Hancock et al. [51] developed a computer-based simulator of hospital systems, which is used for predicting the size of nursing staff configurations under different scenarios.

Concerning the development of the computational model of ED, Paulussen et al. [52] describe a multi-agent based approach for patient scheduling in hospitals. In such a system, patients and hospital resources are implemented as autonomous agents in which the resource agents view the patients as entities to be treated, and the patient agents view the medical actions as tasks that need to be performed. The coordination of patients is achieved through a market mechanism. Patient agents negotiate with each other over scarce hospital resources, using state health dependent cost functions to compute bid and ask prices for time slots. Within this concept, stochastic processing times and variable pathways are considered. Unfortunately, the system does not take into consideration patient variety or the different kinds of healthcare staff. But in fact, the variety of patients and staff has great influence on the performance of ED.

As the use of modeling and simulation for studying EDs, Badri and Hollingsworth [53] developed an Emergency Room (ER) simulation model incorporating the major activities. The model allows the evaluation of "what if?" questions through changing the values of the variables. The ER simulation model determines the effects of changes in the scheduling practices, allocation of scarce resources, patient demand patterns, and priority rules for serving patients. In the study of Gove and Hewett [54], they examined the problem of capacity in hospitals and proved that: due to the complexity of the hospital and its departments, simulation was an ideal choice to study. Moreover, Diefenbach et al. [55] found that varying the number of beds, physical layouts, access to radiology and pathology services in the ED has an exponential effect on expression of the system. The simulation result under the change of system configurations can provide valuable reference for management decision making. Kuljis et al. [56] compared the healthcare system with business and manufacturing, and provided the feasibility of using modeling and simulation methods to improve the QoS in healthcare system. Shin et al. [57] addressed the resource allocation and scheduling problems by creating discrete-event simulations based on detailed models of system processes, and detailed models of resource characteristics and constraints. Their simulation model can help the operations manager at the department to find the best proposal via analyzing different "what-if" scenarios before implementation. This could help to enhance the quality of service that the ED is providing (e.g. reduce waiting times) since the manpower and resources can be well-allocated, and more patients are expected to be treated as a consequence.

Regarding the application of agent-based modeling & simulation (ABMS) approach for simulating EDs, Macal et al. [58] gave a tutorial on creating an agent-based model for the complex system, and they suggested that ABMS promises to have far-reaching effects in the future on how to use computers to support decision making. As for the reason for chosing ABMS approach for simulating ED, Escudero-Marin et al. [59] gave the reason why ABMS is better for modeling EDs than others. They also provided a general description of the possible potential use of ABMS in healthcare application. Laskowski and Mukhi [60] developed an agent-based model to simulate a number of EDs in an area, through which one can extract patient data from EDs of a city to examine patient diversion policies. Jones and Evans in Ref. [61] described the development of an agent based simulation tool that was designed to evaluate the impact of various physician staffing configurations on patient waiting times in the ED. The feasibility of their tool was evaluate at a single real hospital ED.

In relation to the difficulty arising from complexity and uncertainty in managing a big critical healthcare system, P. Barach et al. proposed a microsystem framework as a design concept in Ref. [62]. More specifically, they designed the microsystem framework for the role of understanding and supporting process in designing and redesigning clinical care. The microsystem in their work is a group of clinicians and staff, working together with shared clinical purpose to provide care for population of patients. There are several micro-systems co-existing within a larger organization such as a hospital. Thus, the challenge for the management of the large system is transferred to the management of several relatively independent micro-systems. In this way, the behavior of the large system will be the aggregate of these micro-systems. Their work highlighted the issues of managing a complex system due to the difficulty in understanding complexities as well as the decentralized solution.

Considering the importance of modeling and simulating the interaction between physicians and delegates in ED, M. Lim et al. compared two models with and without the consideration of agents interacting in an ED [63]. In their hospital ED model, comparisons between the approach with interaction and without showed physician utilization increase from 23% to 41% and delegate utilization increase from 56% to 71%. They stated that neglecting these relationships could lead to inefficient resource allocation due to inaccurate estimates of physician and delegate time spent on patient related activities and length of stay. Their work strengthens the importance of accurately modeling physician relationships and the roles in which they treat patients. Furthermore, [64] also discussed several reasons for using an agent-based modeling technique, especially compared to traditional approaches to modeling economic systems.

The previous studies in our research group (High Performance Computing for Efficient Applications and Simulation) at Universitat Autónoma de Barcelona mainly included creating the simulator of an area in ED [32] for non-urgent patients (labeled as area B in Figure 3-2), balancing between the budget and QoS, finding the optimal and sub-optimal resource configurations of ED to achieve better QoS with limited budget by using K-means methods and pipeline scheme [31][30]. Unlike the area for non-urgent patient, area A (as shown in Figure 3-2) is dedicated to critical patients. It is more complex and quite different with area B mainly because patients in this area usually have more complicated conditions and they cannot move by themselves. Consequently, the doctor, nurse and other auxiliary staff need move around ED to attend patients in the carebox. These cases lead to a greater amount of restrictions and interactions between the patients, healthcare staff and environment.

In summary, compare with reviewed literatures and previous studies in our group, the main improvements and contributions of the ED model include: considered some more system components, introduced a new way to define and simulate the interactions between agents and state transition of the agents, and provided an easy-tuning general model to simulate Spanish type EDs. Moreover, with the motivation to promote the utilization of simulation for studying ED related problems, this study also thoroughly investigated the model parameters calibration issues for a general agent-based ED model.

2.2 Modeling Approach

For simulation purpose, there are two different possible fundamental ways to describe a system: black-box model (e.g., data-driven model) and white-box model (e.g., first-principle model). The black-box model ignores the actual mechanism of a system while investigating relationships between input and output parameters. For example, these relationships can be replicated by artificial neural network models which can be trained to replicate the behavior of the original system without a prior knowledge of the system. Bibi et al. [65] concentrated on the prediction of the effect of atmospheric changes, including pollutants, on ED visits via an artificial neural network with a back-propagation training algorithm and genetic algorithm optimization. With enough historical reference data to cover the dynamics of the target system, the artificial neural network model could be trained to represent the system behavior for interpolation prediction. But, it is not an extrapolation method and gaining insight into a black-box is a difficult undertaking.

However, the first-principles model, more specifically behavioral-driven approaches in this study, utilizes reengineering methods to capture details of system behavior from the interaction of system components. It is not as quick and easy to build, but they have many advantages. In terms of simulation, first-principle models provide extrapolation in addition to the interpolation provided by data-driven models. They also can be used for prediction, explanation and optimization.

Simulating healthcare processes is a sophisticated endeavor. Treatment processes and patient arrival patterns differ significantly in their statistical attributes and implicate a high degree of variability. In addition, there are several types of interconnected processes of medical staff involved that accompany a patient's journey through the healthcare facility. To deal with the challenge of model this sophisticate process, M. Thorwarth and A. Arisha [66] introduced a framework which delivers an algorithm that allows to implement multiple participant pathway modeling under the consideration of flexible resource allocation. But, that framework is designed specifically for an Irish ED.

In summary, as reviewed in section 2.1, there are massive Operational Research (OR) results in the reviewed literatures achieved by using a simulation approach. According to the comprehensive review in Ref. [17], System Dynamics (SD), Discrete Event Simulation (DES) and Agent-Based Simulation (ABS) are three most widely used simulation methods in operational research community. The main differences among these three simulation methods is the level of perspective [17]. SD is a top-down approach from the macro-level perspective; The DES is a process-oriented approach focused on workflow simulation and the ABS is a bottom-up one from an individual level. At the 2010 OR Society Simulation Workshop, there was a lively panel discussion [67]. From the discussion, we can see that actually both DES and ABS are widely used now. Neither have the absolute substituting capability in all application fields. The same workshop in 2011, as well as Ref. [68] discussed the challenges of DES and several typical features of a system with which the ABS is more appropriate [69]. There is no fixed rule to select a suitable approach for studying a specific system, a proper combination may halve the work with double the results. Based on this idea, [70] presented a multi-paradigm simulation method by using SD for simulations at a high abstraction level and ABS/DES at an individual level in a common simulation environment. Two examples were shown on how the new innovative technology can evaluate the "what-if" problems prospectively and how new ideas can be derived by parameter variations. Combining the reviewed discussions and simulation studies with our experience and objective, ABS was selected to simulate the EDs, the reason will be given in section 3.2.

2.3 Model Parameters Calibration

Model calibration is the task of adjusting an existing (general) model to a target system. M. Hofmann [71] introduced a formal approach to model calibration, within the frame of the presented formalism it is shown that the computational complexity of model calibration is NP-complete. The author addressed that, for huge model federations the complexity of parameter calibration could draw a serious line with respect to the validation of the federation and its cost-benefit ratio. This is mostly because in huge model of complex system, no single person has an overview of the whole simulation, and the interpretation of unexpected results is extremely difficult.

The model parameter calibration process can be easily formed as a simulation-based optimization process. Due to the complex behavior of the objective function, Evolutionary Algorithms (EAs) are often used to efficiently explore large parameter spaces. However, EA still takes a considerable amount of time because it requires a large number of simulation runs, and each run takes considerable length of time in simulation. To this end, M. Wagner et al. [72] proposed the use of complexification as it emulates the natural way of evolution, to improve the performance of EAs. This method has been used for parameter estimation of multi-agent based models. J. Zhong et al.[73] proposed an evolutionary framework to automate the crowd model calibration process. In the proposed framework, a density-based matching scheme is introduced. By using the dynamic density of the crowd over time, and a weight landscape to emphasize important spatial regions, the proposed matching scheme provides a generally applicable way to evaluate the simulated crowd behaviors. Besides, the authors also proposed a hybrid search mechanism based on differential evolution to efficiently tune parameters of crowd models. In Ref [74], J. Zhong et al. proposed another novel evolutionary algorithm named differential evolution with sensitivity analysis and the Powell's method (DESAP) for model calibration. The proposed DESAP firstly applies an entropy-based sensitivity analysis operation to dynamically identify important parameters of the model. Then, Powell's method is performed periodically to fine-tune the important parameters of the best individual in the population. Finally, in each generation, the evolutionary operators are performed on a small number of better individuals in the population. Their new search mechanisms are integrated into the differential evolution framework to improve search efficiency. However, all of these developed algorithms are mostly focused on solving agent-based crowd behavior model calibration problem with complete data, in which the system metrics and objective function are different from the requirement of tuning an agent-based ED model.

In contrast to traditional black box search methods which only consider the input and output of simulation model, M. Fehler et al. [75, 76] proposed a promising white box calibration approach, which uses the knowledge of the agent-based model to improve the tuning process. In which, the idea is to reduce the parameter space by breaking down the model into smaller sub-models. Each of the sub-models is then calibrated before merging them back to form the model. Thus, the main focus was put on the sensible introduction of heterogeneity into the model and the analysis of bottom-up model designs, to overcome the often infeasible trial and error step of parameter and micro level structure refinement. However, in this method, the division and fusion operations are difficult steps, and require additional knowledge about the model, and these knowledge may not be available for simulation users (non-developer). Moreover, the fusion operation has to merge calibrated sub-models into a calibrated higher model, which is not automatic.

In summary, although parameter calibration is critical and one of the key steps in modeling & simulation work and it can be easily formalized as a simulation-based optimization problem, to the best of our knowledge, such model parameter calibration problems under data scarcity have not been explicitly addressed in the literature. No literature was found providing an automatic calibration tool for simulation users to calibrate the general model for a new system without the involvement of model developers. Since model calibration is carried out via simulation-based optimization, model evaluation is computation expensive. High performance computing techniques were used in our work in order to search the robust and optimal model parameters in an acceptable time frame.

2.4 Summary and Opportunities

It can be seen from the review of literatures that modeling and simulation has been applied successfully to several applications of healthcare operations management. Among these modeling approaches, ABM provides new insight to problems by modeling individuals and the interactions between them. This perspective has facilitated analysis at both the individual and system levels, which is not typically possible by using other methods or without the support of high performance computing techniques.

However, most of these simulation based studies are oriented by specific requirements. That is to say that the simulator becomes an appropriative tool for a specific requirement and lacks scalability in application perspective. In addition, due to the complexity and criticality of ED systems, knowledge for full insight into the dynamics of the system would provide a great deal of help to efficient and optimum management. However, there is little work on knowledge discovery through simulation. Since ABM represents details of the system at individual level, discovering from micro-level interaction information is a way to gain insight into the emergent behavior of complex systems, and indicate the root-cause of disordered system behaviors.

Due to the nature of agent-based models, computationally expensive and complexity on model parameter calibration are two main challenges of using ABM. Nowadays, we are armed with the ability to execute compute-intensive models and analyze massive datasets. As indicated by Douglas A. Samuelson and Charles M. Macal, the age of agent-based model & simulation is coming [77].

Chapter 3

The model of emergency departments

Essentially, all models are wrong, but some are useful.

- George Edward Pelham Box

Hospital based emergency departments (EDs) are highly integrated service units to primarily handle the needs of the patients arriving without prior appointment, and with uncertain conditions. In this context, analysis and management of patient flows play a key role in developing policies and decision tools for overall performance improvement of the system. However, patient flows in EDs are considered to be very complex because of the different pathways patients may take and the inherent uncertainty and variability of healthcare processes. The agent-based model provides a flexible platform for studying ED operations as it predicts the system-level behavior from micro-level interactions, so as to see the forest through the trees. In this way, policies such as staffing could be changed and the effect on parameters such as waiting times and throughput could be quantified. The overall goal of this chapter 3 is to develop models to better understand the complexity, evaluate policy and improve efficiencies of ED units. This chapter details a general agent-based ED model for simulating Spanish ED. The presented model will be calibrated to emulate a real ED in chapter 4, simulation results have proven the feasibility and ideality of using agent-based model & simulation techniques to study ED system.

This chapter is structured as follows: section 3.1 describes the conceptual model of ED observed from real operation data and the involvement of expertise ED staff. In section 3.2, we further (in particular for ED system) describe the modeling approach we used in this study. The full agent-based model of ED is detailed in section 3.3, which contains two parts, subsection 3.3.1 describes the model of agents considered in ED, and subsection 3.3.2 describes the interaction model among agents as well as agents with the environment. The section 3.4 formulates the agent models into algorithm and implemented with programming language that can be executed to emerge systemic behavior, i.e., from conceptual model to computational model. Then, section 3.5 presents the way to design experiments as well as execute the model for predicting and discovering hidden features.

3.1 Conceptual model of emergency departments

Typical EDs have common interacting elements such as doctors (physicians), nurses, technicians, receptionists, beds, medical devices that are interconnected via flows of information and processes (registration, triage, diagnostic, discharge). All of these elements methodically interact with each other to produce diagnoses, treatments, and information. The ED studied in this research is focused on Spanish type. However, we are confident that the proposed methods as well as the simulation framework can be used for other EDs.

The conceptual model will be discussed with an order of patient arrival, flow in ED and discharge. This conceptual model was carried out by analyzing 4-year historical data provided by a typical ED, and conducting interviews with experienced ED staff. The analysis of historical operation data determines the nature of distributions followed by patient arrivals, service times, and related parameters to determine transition probabilities. Meanwhile, the participation of experienced ED staff helped us to establish a comprehensive understanding of the hospital ED and focus on considering significant features of ED in modeling & simulation.

3.1.1 Patient arrival

Patient arrival is the input of the ED model, which has direct influence on the system behavior. As discussed in section 1.3, The simulator is not only designed to imitate real situation but also for studying system response under unexperienced scenarios (i.e., extrapolation). Therefore, a precision and customizable patient arrival model which could reflect patients' arrival characteristic is crucial for the accuracy of imitating and predicting. Specifically, the model should reflect the general pattern of patient arrival, and it should be easy for users to customize for specific scenarios. In this study, the patient arrival model is composed of two parts: the arrival rate defined as number of patients arrive at ED per hour, and arrival patients' key characteristics (age and severity measured by acuity level (AL)).

With respect to arrival rate, after synthesized various opinions of expert ED staff, on-site observation, and four-year's actual data from the information system database of the Hospital Universitari Parc Taulí. We found that the arrival rate fluctuated significantly through the day, and the number of daily arrival patients was influenced by day of week and season. For instance, EDs get fewer patients in August, patient arrival reaches a maximum on Monday and minimum on Saturday in one week. Consequently, we modeled the patient arrival rate in a time interval of one week. The arrival rate model includes a table of normalized (proportion of weekly arrival) hourly arrival rate $R_{ar}[hour, day]$ in one week and the number of weekly patient arrives $N_{ar}[week]$. Accordingly, for a given *hour* in *day* and *week*, the number of patient arrival in a specific hour (R_{hour}) can be computed with $R_{hour} = N_{ar}[week] \times R_{ar}[hour, day]$. Subsequently, the arrival process in one hour was fitted by a non-homogenous Poisson process [78]. The probability density function of patient inter-arrival time can be expressed as follows, and a detailed emulation algorithm is shown in Algorithm 1:

$$f\left(k;\lambda\right) = \frac{\lambda^k e^{-\lambda}}{k!} \quad (k \geq 0) \tag{3.1}$$

$$\lambda = \frac{60}{N_{ar}[week] \times R_{ar}[hour, day]} \tag{3.2}$$

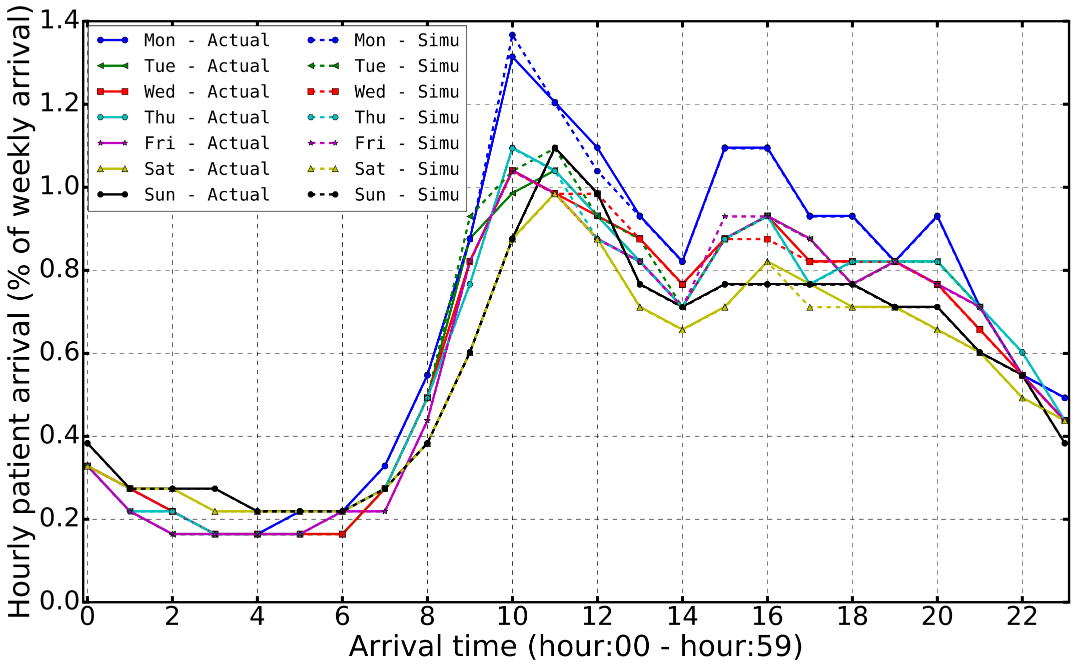
Where, e is Euler's number, $k!$ is the factorial of k, k denotes the inter-arrival time between the nth patient and the $(n+1)$th, and λ denotes the inverse of average arrival rate (number of patients per minute). Regarding implementation, in the first minute of each hour h $(0 \leq h \leq 23)$, we generate a random number k_1 from the Poisson distribution $X \sim Poi(\lambda)$ with parameter λ computed with Equation 3.2. The value of k_1 represents that the first patient will arrival at time $h : k_1 : 00$. Subsequently, when simulation time is up to $h : k_1 : 00$, a new agent will be created and send to registration waiting room. At the same time, we generate another random number k_2 which represents that the second patient will arrive at time $h : (k_1 + k_2) : 00$. In this way, the nth patient will arrive at time $h : (\sum_{i=1}^{n} k_i) : 00 \, (0 \leq \sum_{i=1}^{n} k_i \leq 59)$. The flow of patient arrival model is shown in Algorithm 1, this procedure will be called at every simulation time-step (iteration).

Algorithm 1 Patient arrival model algorithm

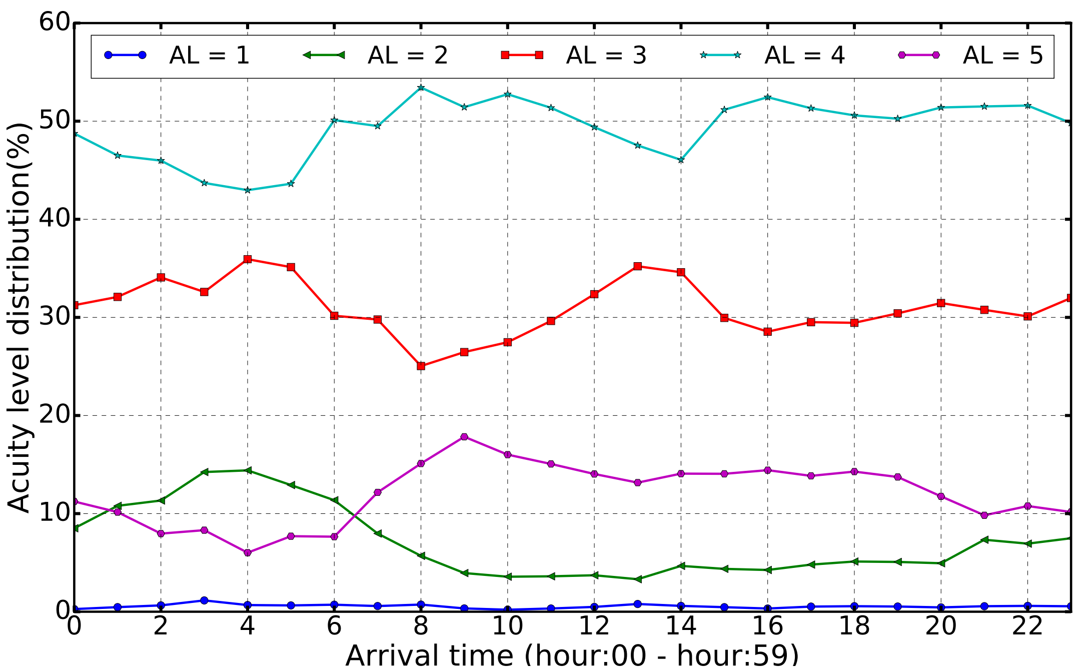
procedure PATIENTARRIVAL
    static $k \leftarrow 0$ ▷ define k as a static variable
    if $now.minute == 0$ **and** $now.second == 0$ **then**
        $\lambda \leftarrow$ **call** $computeLambda(datetime.now())$ ▷ implementation of Equation (2)
        $k \leftarrow random.poisson(\lambda)$ ▷ arrival time of the first patient in this hour
    end if
    if $now.minute == k$ **and** $now.second == 0$ **then** ▷ arrive at time hour:k:00
        call $createPatient()$ ▷ create patient; characteristics will be set
        $k \leftarrow k + random.poisson(\lambda)$ ▷ arrival time of the next patient
    end if
end procedure

With the above descried patient arrival model and real operation data of 2014, Figure 3-1a shows 12-month simulated results compared with actual data (broken versus solid lines). It is cleat that the simulation result is very close to actual data at an hourly rate. The weekly arrival (the sum of hourly arrival), simulation: 1829 versus actual: 1826 is close to the actual as well.

Regarding arrival patients' key characteristics, we considered their age and severity (measured by acuity level). Observing from actual data, we found that the severity of arrival patients also fluctuates significantly through the day but no significant



(a) Patients' hourly arrival rate in one week, actual vs. simulation.



(b) Arrival patients' acuity level distribution (hourly).

Figure 3-1: Patient arrival model: hourly arrival rate (quantified by percentage of weekly arrival rate), simulation vs. actual (average in 12 months), and arrival patients' acuity level distribution (extracted from 12-month actual data, 2014).

effects were found in longer time periods. The data illustrated in Figure 3-1b was retrieved from 12-month actual data of year 2014. Then in simulation, a Gibbs sampler (a Markov chain Monte Carlo algorithm) was used to obtain a sequence of values to represent patients' severity. With enough samples, these values will be able to approximate the distribution data shown in Figure 3-1b. Similarly, the age distribution was extracted firstly from actual data and then fitted with a Gibbs sampler from the distribution in simulation. In short, the model describes the pattern of patients' arrival, the parameters for characterizing actual behavior should be related with simulation scenarios, and simulation users can customize all the parameters.

3.1.2 Process in ED

This section gives a brief overview of the patient flow in ED. Typically, a patient enters the EDs through one of two ways: by themselves or by ambulance (as shown in Figure 3-2). Upon arrival, walk-in patients need to walk to the registration window and briefly give their personal information to the registration staff. After that, they have to stay in a waiting room wait for triage. Then they will go to the triage box and interact with the nurse once the information system assigns a triage nurse to the patient. Triage consists of a brief assessment of the patient's body condition and an acuity level will be assigned to the patient according to their severity afterwards. After triage, patients will wait in another waiting room before entering the diagnosis & treatment area. For those patients who arrive by ambulance, they are registered and triaged in the ambulance and could go to the second waiting room directly. The Spanish scale of triage is very similar to the worldwide Canadian Emergency Department Triage and Acuity Scale [79, 80]. The scale consists of 5 levels, with 1 being the most critical (resuscitation), and 5 being the least critical (non-urgent). The triage process also determines the order and priority with which the patient must be attended and the treatment area where they will be treated. The registration and triage service are first-come, first-served (FCFS) for all the patients, whereas entering the diagnosis & treatment area is acuity-level-dependent FCFS (patients with AL 1 have the highest priority).

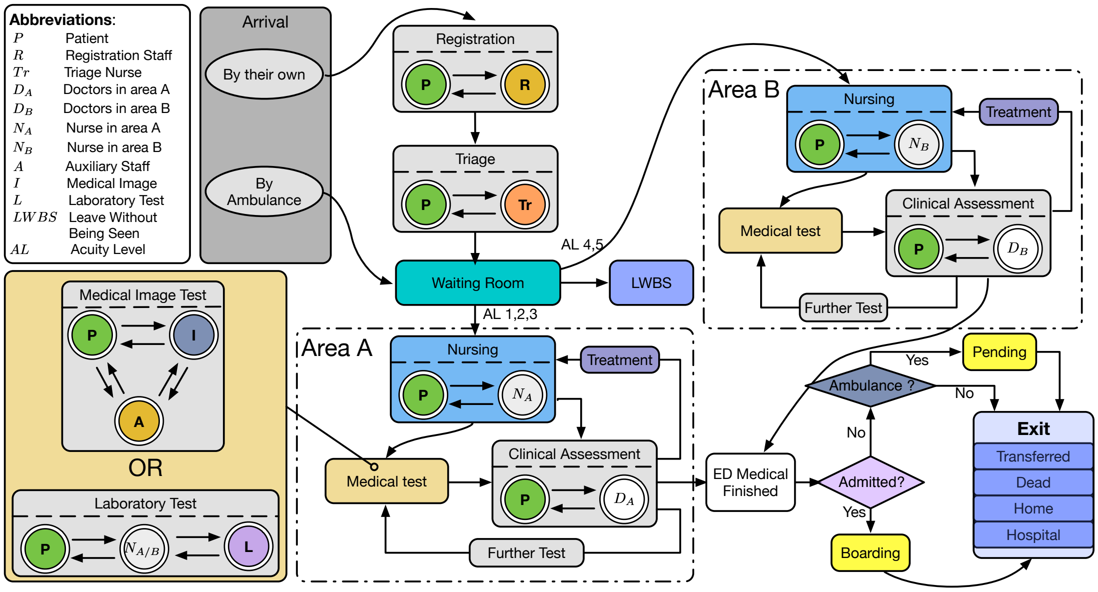


Figure 3-2: The emergency department operation process as well as interactions among its components. The group of two parallel lines with arrow stands for interaction. Patient flow is managed by emergency department information system, the whole process can be seen as a multi-class queuing system with probabilistic routing. That is, there are queues for each interaction because service providers are not always prepared to accept new patient (providers' service capability is not infinite).

With regard to the treatment area, in most Spanish EDs, there are two treatment areas (labeled as A and B in this study, see Figure 3-2) which operate independently to provide the diagnosis & treatment service. Area A is for those patients with acuity levels 1, 2 and 3, while area B, also called fast track, in the ED is a dedicated stream of resources to process lower acuity patients with levels 4 and 5 more quickly. Occupied by the most urgent patients, the area A is made up of careboxes. A carebox is a small room contains essential medical equipment and supplies that could be used for patients' treatment. Patients attended in area A will stay in their own carebox during all the diagnosis & treatment phase, transporting will be done by auxiliary staff if necessary. In area B, there are several attention boxes in which doctors and nurses interact with patients, and a large waiting room in which all patients will remain while not having interaction with the ED staff.

The patient flow as well as interactions (the group of two parallel lines) in general EDs are demonstrated in Figure 3-2. It is worth noting that the process in Figure 3-2 is a generic routing from all patients. Every patient who comes through the door is an unknown, with a condition that unfolds over time in a functionally non-deterministic way. Theoretically speaking, no two paths through this "system" are the same for any two patients. It can be seen as a multi-class queuing system with probabilistic routing. Accordingly, from a patient's perspective, he/she is either receiving service (interacting) or waiting for resources (healthcare staff or physical resources like beds, testing equipment) becoming available.

With respect to the movement and spatial constraint, such as in area A, when a medical imaging test is assigned (e.g., X-rays, CT-scan, MRI-scan.), the auxiliary technician will transfer the patient to the facility area and accompany the patient throughout the test process. However, in area B, patients will move to the corresponding places by themselves when get notified by information system.

3.1.3 Discharge

Normally, there are four possible destinations when patients finish their treatment: admission to hospital, going back home, being transferred to another hospital or death. If a patient is admitted to hospital and there is no bed available in the hospital, they have to stay in ED keep occupying all the service. That is, patients admitted to hospital often occupy ED beds (also referred as 'access block' or boarding) due to the unavailability of beds in hospital. As a consequence, hospital bed occupancy will strongly affect patients' length of stay (LoS) and throughput (the number of attended patients per day) of ED. The results in Ref. [81] show that increased hospital occupancy is strongly associated with ED LoS for admitted patients. It prevents EDs from serving new patients in a timely manner and results in longer LoS as well as a percentage of patients who left without being seen. Moreover, patient discharge is often delayed because staff are tied up with more urgent patients [82]. This indicates that staffing and scheduling have a widespread effect on all areas of the ED. Therefore, although our work was focused on modeling EDs, the model of bed availability in the hospital should be carefully considered as well. Based on the actual historical boarding time from the database of ED information system, a Poisson distribution is used to fit the bed availability pattern in hospital.

Similarly, for those patients who will go home or transfer to another hospital, some of them will request the ambulance service. Since the ambulance service is provided by a service center, it is common that there is a delay from requesting until becoming available.The same as bed occupancy in hospital, the response time of the ambulance service also affects ED's behavior because patients keep using physical resources of the ED during waiting. Having said that, it is also necessary to include the model of ambulance response time as part of the ED model to analyze the degree of the impact on an ED. Based upon 4-year's real data, a Gamma distribution $X \sim \Gamma(k, \theta)$ was used to fit the length of response time. A detailed ambulance response time model can be found in Ref. [36].

In short, patients that have been discharged from the ED either leave immediately or undergo another waiting phase: boarding. While patients remain blocked or boarded in an ED bed, they prevent other patients from starting treatment, which might lead to ED overcrowding. Therefore, discharge behavior should be carefully considered in an ED simulation.

3.1.4 Door-to-doctor time and Leave Without Being Seen

Regarding to length of waiting before seeing a doctor (also known as door-to-doctor time), according to the public healthcare regulations in Spain [79], for those patients triaged as AL-1, 98% of them should be attended immediately, 85% of the patients with AL-2 must be attended immediately by a nurse and within 7 minutes by a doctor, and for AL-3, 80%, within 15 minutes; AL-4, 75%, within 30 minutes; AL-5, 70%, within 40 minutes [79]. If there is no free space for upcoming patients, they will be attended in the corridor temporarily (considered as a virtual carebox in this model). However, the service capability of the healthcare staff is limited, some patients, especially those triaged as low acuity level may face long waiting time and they may leave without being seen (LWBS). According to Ref. [83], it is common to see patient LWBS rates above 6% due to physician unavailability. Accordingly, we consider LWBS as a system performance indicator in evaluating proposed changes to the system.

Patient leaving without being seen is a crucial metric for efficiency and effectiveness of public EDs. As investigated in Ref. [84] and [85], a patient's decision to LWBS is influenced by many factors. In this model, we assume that each patient's decision to LWBS only depends on their waiting time length. Thus, after some questionnaires and on-site interview, we use an acuity-level-dependent Triangular distribution ($X \sim Triangular(a, b, c)$) to fit the length of time that patients stated they were willing to wait before LWBS, the parameters a, b and c are related with patient's age, acuity level and gender. The simulator implementation also provides interface to simulation users to define a sub-model for patients' tolerance of waiting.

3.1.5 Healthcare Staff Behavior

Healthcare staff in an ED are service providers, admission staff and triage nurses are easy to model because they provide service to patient one by one with First-Come-First-Serve (FCFS) order. Doctors and nurses in the treatment zone are more complicated because they need to take care of several patients simultaneously with considering patients' acuity level, and for each patient, they need to provide different services under different physical conditions of a patient. Modeling these services is almost equivalent to model these different interactions.

In real situations, the Information System(IS) drives the patient flow in ED, it works as a central task dispatcher. For example, when a doctor assigns a test service for a patient, the IS will allocate one auxiliary to help the patient move to the corresponding test room, when finished, the auxiliary sends the patient back to their carebox. When the test service provider gets the test result, they will send the result to IS, and IS will forward to the corresponding doctor's task-list. The doctor will perform the task according to predefined order rules. The central IS helps to cooperate these multitasks which drive all the healthcare staff forward. The model of the healthcare staff can be abstracted as a passive task executor as shown in Figure 3-3, i.e., each healthcare staff has a task-list and the IS post tasks to their task-list.

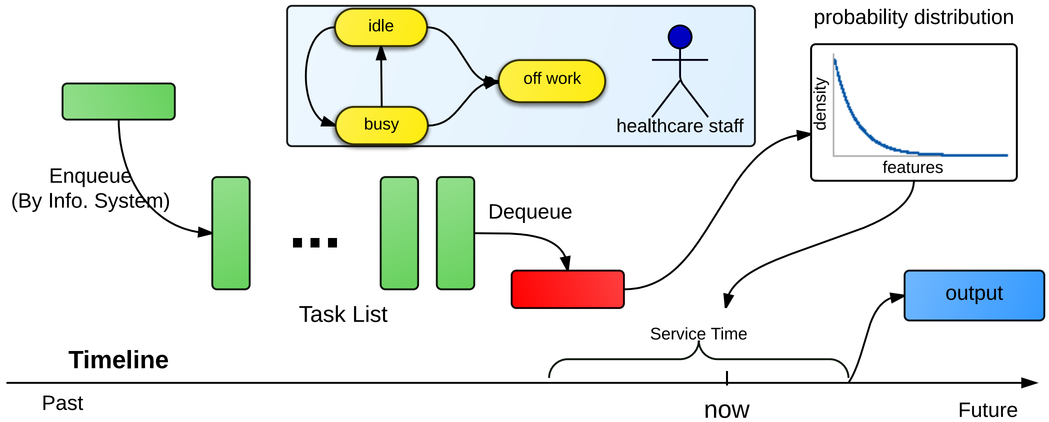


Figure 3-3: Model of service providers. Same as their work in real ED. The service providers were modeled with a task list, the information system detaches and pushes tasks to the corresponding list. Service providers keep checking their own list, whenever the list has task, they pop, move to corresponding place and perform it.

3.2 Modeling approach

Conducting a valid simulation is both an art and a science. One of the main challenges when developing a general simulation model is to keep a model as simple as possible whilst including all the key system information to achieve the objectives of simulation. One feasible way to do this is through the following three steps: (1) survey multitude real models; (2) analyze the concept structures of these real models; and (3) abstract and generalize from these real models to develop a reusable generic pattern model.

Regarding modeling approaches, as we reviewed in section 2.2, Discrete Event Simulation (DES), System Dynamics (SD) and ABMS are the three main approaches used when simulating a complex system such as ED. There is a large body of literature describing the use of DES models in ED studies, whilst there is considerably less literature on the use of ABMS for this purpose. As healthcare systems are based on human actions and interactions, combined with our experience and requirement, it can be more properly to model with ABMS [59]. ABMS models can offer ways to provide a deep insight view and to generate hypotheses about system behavior by representing this as a result of the interaction between the agents.

At the heart of the difficulty to manage a complex system is the fact that humans have difficulty in understanding the complexities caused by the dynamic and systemic nature of certain problems [62, 28]. When we face with a complex system, it is difficult to model all its functionality directly at systematic-level because there are large number of factors that can affect the result and need to consider. A good way to model is by using a bottom-up-modeling approach. Starting from the bottom system components (model of system components' behavior, interactions between components), the execution of the simulator will cause a large amount of interactions between these agents, and then these interactions will emerge the functionality of the ED indirectly. As shown in Figure 3-4, it works by modeling the agents, their

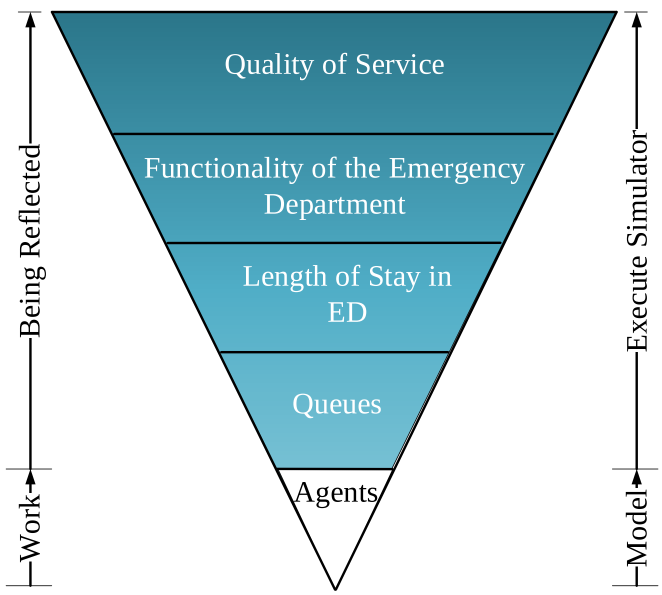


Figure 3-4: Bottom up modeling approach. Systemic level behaviors are reflected/emerged from the execution of bottom level simulation models.

behaviors and interactions between them. Then, when executing the model, the state of agents will be changed by their interactions, and the queue for the interactions and their length of time on interacting will emerge, by such analogy, the functionality of the ED will emerge indirectly through the execution of bottom level models. Furthermore, the quality of service can be evaluated through the results of different simulation scenarios.

Agent-based simulation is an approach to model systems comprised of individual, autonomous, interacting "agents". The interacting is a key characteristic since that is the smallest element which would emerge the functionality of the system. Such interaction data has incredible potential to address complex features and dynamics of the objective system. Agent-based modeling offers ways to model individual behaviors more easily and to see how behaviors affect others in ways that have not been available before [68]. It is commonly used to study complex systems since it shows behavior more like our idea of how the real system works, and such models can be constructed even if we do not understand the systematic behavior. That is, we only need to know how an individual agent behaves in system and how it interacts with others, then we can set the model to an initial configuration and watch it evolve over time [86]. Furthermore, in the micro-level, the spatial agent-based simulator is not a design for any specified application. Instead, it is a behavior simulator to simulate interaction among the smallest components of the ED system. Thus, it is customizable and adaptable in nature for different ED simulation requirements. Therefore, with respect to simulation model reusability, further characteristics can be easily added to the agent models to study additional problems, such as disease transmission and bacterial infections in EDs.

In summary, the reasons why ABS was selected to model an ED in this study include: (1) in an ED system, agents have dynamic relationships with other agents. For example, patients have dynamic relationships with sanitary staff, doctors have dynamic relationships with nurses and medical test rooms. These dynamic relationships are important to consider and, by their nature, well suited to be modeled as part of agent-based model. (2) The agents have a spatial component to their behaviors and interactions, i.e., most of the agents in ED need to move around and the spatial location is one of the key states which determines their potential interacting object and state transferring. (3) A large numbers of agents, agent interactions and agent states are important for information extraction. In an ED, services are provided via multiple interactions, patients pass through ED with a series of non-deterministic interactions. These interactions can deeply reflect the functionality of the target system. (4) Model reusability. Agent-based model directly represents behavior of the system components, so it can provide the all-side meta-data needed for analyzing the macro-level behavior of the system.

3.3 Agent-based model of emergency departments

The section 3.1 gives an overview of the process in ED. It is clear that the care services in ED are carried out by interactions between patient and components of ED (human and material resources). As Carmen et al. indicated in Ref. [87]: "*EDs are highly complex environments: patient arrival rates vary over time, patient care paths depend on urgency and pathology, resources may or may not be suitable for treating all patient types, urgent patients typically get priority over (and may even preempt resources from) non-urgent patients, patients who need to be admitted often occupy ED beds.*" [87]. However, the study of complex systems poses unique challenges: some of our most powerful mathematical tools, particularly methods involving fixed points and attractors are of limited help in understanding the development of complex system [88]. The conceptual model described in section 3.1 provides a brief understand of the complex operations in EDs. With this understanding, qualitative predictions can be yielded via intuition. Since quantified performance measurement and evaluation of system are more important for continual improvements in the system, the conceptual model must be formulated into algorithm and implemented by programming languages. Then the model can be executed to quantify predictions and discover the hidden features.

As the reason of choosing agent-based modeling & simulation for studying ED system given in section 3.2. The Agent-Based Models (ABMs) are a class of computational models for simulating actions and interactions of agents (individual or collective entities such as medical imaging test-room) with a view to assessing their effects on the system as a whole. ABMs arise in computer experiments, in which each agent is guided by a set of programmed rules with proper transition probability, and capable of acting independently. In the case of studying EDs, the bottom-up method is a good choice because healthcare staff relies heavily on the interaction with patients [6], and the bottom-up method is straightforward on exploring these interactions. In addition, the individual based simulation approaches are capable of gaining insight into the complex care process to understand the root cause of phenomena. An ABM is composed of behavior model of agents and interaction model among agents. An ABM of ED will be discussed in subsection 3.3.1 and subsection 3.3.2.

3.3.1 Design of Agent Models

To formally study complex systems with agent-based modeling and simulation techniques, a way to precisely define the agents and their interactions must be provided first [88]. The actions of agents usually depend on the signals they receive. That is, the agents could be formulated as an IF/THEN structure: IF [signal vector x is present] THEN [execute act y] [88]. If an agent is busy with an interaction while a signal is being presented, the presented signal will be pushed into its task queue. This behavior model structure is: (1) easy to abstract from agents' actual behavior in the real system, (2) in accord with KISS principle (keep it simple, stupid), and (3) easy to be converted to programming language. The following sub-sections 3.3.1 - 3.3.1 will detail the behavior rule of all the agents in ED in the IF/THEN structure. For convenience, we defined notations in Table 3.1 to represent a specific group of agents.

Table 3.1: Notations used in this model.

Notations	**Description**
i	ID of agent, it is an unique ID among all the agents in the model.
N_i^{CB}	a set of careboxes under responsibly of nurse i.
N_i^P	a set of patients (mainly for patients in area B) under responsibly of nurse i.
D_i^P	a set of patients under responsibly of doctor i.
IS	The information system in ED, a system for communicating and coordinating among staff, patient and test-room, also treated as an agent in this model.

Patients

As the leading role of an ED, patients in ED are guided by information system (IS), i.e., going to the corresponding place when they get notified. During all the process in ED, the patient alternates between two states: receiving treatment or waiting (i.e., waiting for a doctor, nurse, medical testing service/result). A patient's behavior is the same in all stage except diagnosis & treatment in area A. The patients in area A are solely guided by service providers, i.e., doctors, nurses and auxiliaries. They will stay in their carebox when there is no interaction with service providers. Therefore, the LoS in ED is the sum of all activities (meeting with doctor, medical tests, and having rest to wait for drug therapies to take effect) they have to attend to, and time on waiting for resources (including test rooms, doctors, nurses and auxiliaries) to become available. The behavior rule of patient is given in Table 3.2. It is worth noting that the time to wait until drug therapies take effect (t_{drug}) is significant because it is usually the longest part of LoS. In this study, we fitted the t_{drug} with acuity level dependent random distributions. The parameters of the distributions depend upon patient's acuity level and will be calibrated in model tuning process.

Table 3.2: Behavior rules of patients.

IF	**THEN**
notified by IS (before entering treatment area).	go to the corresponding place in the notification.
no requests from IS (before entering treatment area).	keep staying in waiting room.
no interaction requested by healthcare staff (nurse, doctor or auxiliary).	keep staying in carebox (for patients in area A).
no requests from IS or healthcare staff.	keep staying in waiting room (for patients in area B).
notified by IS (in area B).	go to diagnosis room or medical image test-room as indicated in the notification.
needs additional help.	ask nurse through IS (the IS will notify the corresponding nurse).

Registration staff and triage nurse

The abstract behavior of triage nurses and registration staff are similar, we describe their model together in this section. The service time duration for triage nurses and registration staff is not significant with different patients. In our model, we considered the duration of service time simply on the basis of the experience (junior or senior) of registration staff and triage nurse. The behavior rule of registration staff and triage nurse are detailed in Table 3.3.

Table 3.3: Behavior rules of registration staff / triage nurse.

IF	**THEN**
time to work.	interact with colleague in previous shift, take over materials from them.
no patient in front of the desk/window.	keep waiting for patient (IDLE).
one patient is waiting in front of the desk.	interact with patient for registration/triage.
shifting of duty time is up.	accomplish work at hand, interact with colleague in following shift, hand over requested material.

Doctor

As service provider, doctors in different areas behave differently. The distinct difference is the movement requirement for interaction. Doctors in area A have to walk around the area in order to interact with their patients in careboxes because patients in area A are not allowed to move by themselves. The duration of time it takes for the doctor to move (t_{move}) is important to consider as it is not constant and significant when regarding system efficiency. However, doctors in area B sit in their office waite for patient to come in. Considering that the waiting room is not far from doctor's room in area B, the time it takes for the patient to move is constant and negligible. Detailed behavior rules of doctors are shown in Table 3.4.

Similarly, doctors in area B have the same behavior rules except moving to the corresponding careboxes. Regarding the duration of time for each interaction (service time), it depends on many factors. According to the real behavioral data and findings

Table 3.4: Behavior rules of doctors, rules specified with *in area A* means doctors in area A, otherwise applies to area A and B.

IF	**THEN**
time to work.	interact with doctor in previous shift, take over patients from them.
no task assigned by IS (task queue is empty).	stay in their office (IDLE).
IS notifies a new patients in carebox i (in area A) / A new patient comes into office (in area B).	move to carebox i (in area A), perform first-interaction, make treatment plan.
IS notifies: the test report for one of the patients in set D_i^P is ready to review.	review medical test report, walk to the carebox (in area A) if necessary, and make follow-up treatment plan (do more test, drug therapy, discharge or admit to hospital).
scheduled drug therapy time of any patient in set D_i^P is up.	walk to the carebox (in area A), check effect of drug therapy, and make follow-up treatment plan.
shifting of duty time is up.	accomplish work at hand, interact with doctors in following shift, hand over all the patients in D_i^P.

in queue theory [89], the duration of service time for a specific worker could be fitted with exponential distribution ($Exp(\lambda_{st})$). The parameter λ_{st} is acuity level-, service type- and service provider's experience- dependent in this model. Equation 3.3 expresses the sub-model of service time.

$$T_{st} = Exp(\lambda_{st}) + t_{move}, \quad \lambda_{st} = \gamma_n \cdot f(s, sp, al) \tag{3.3}$$

Where, T_{st} is the service time for one interaction, s denotes the type of service (purpose of interaction), sp is the experience of the service provider (doctor, nurse, test-room), say, junior or senior; al represents the acuity level of patients; t_{move} is the time takes on movement which depends on the location of the patient's carebox. t_{move} is set as zero in area B. Some on-site interviews and staff experience show that the first meeting with a patient takes significantly longer time than follow-ups and most patients (especially in area A) need several meetings with their doctor. Therefore, from a doctor's perspective, it is better to consider the service time model separately for the first interaction and follow-ups. To avoid an over-complex model, we added a proportionality coefficient parameter (γ_n) for the service time in Equation 3.3, i.e., same statistical model, different scale. The γ_n thus represents the proportionality coefficient for the first meeting with patient and follow-ups (e.g., 1.0 for first meeting and 0.7 for follow-ups). The function is identified and calibrated with real operation records of the target ED. Note that the service model described by Equation 3.3 is also applied for nurses but with different parameter values. It can be seen that the duration of time for interaction is determined by the service provider, yet with patient's characteristic used as parameters. Regarding the treatment plan made by the doctor, it is based on the routing probabilities retrieved from ED patient records, and depends on the patient's characters.

Nurse

Similar to doctors, the nurses in area A have to move to patient's carebox to provide service and the duration of time on moving is crucial to be considered. Regarding the service time of nurse, we use the same model as shown in Equation 3.3 but with different parameters. The IF/THEN behavior rules of nurse are detailed in Table 3.5.

It can be seen from Table 3.5 that all the behaviors are driven by the information system, and we assume that nurses always behave according to the rules. Thus, the uncertainties of nurses' behavior are mainly due to the uncertainties of the patients' condition.

Auxiliary technician

Auxiliary technicians work in area A to assist patients to move around for medical testing. The behavior of auxiliary technicians is simple but crucial to consider because their behavior also has significant impact on system performance. For example, a shortage of technicians will result in delays in patients' tests, and further affects the efficiency of test-rooms. This chain reaction will finally affects the throughput of the ED system. Different from doctors and nurses, all the technicians share one task queue, i.e., when there is a request and if there are idle technicians, one of them will

Table 3.5: Behavior rules of nurses, rules specified with *in area A* means nurses in area A, otherwise applies to area A and B..

IF	THEN
time to work.	interact with nurse in previous shift, take over patients from them.
no task assigned by IS (task queue is empty).	stay in the nurse room.
doctor assigned laboratory test to one of the patients in set N_i^P.	walk to the patient (to carebox N_i^{CB} in area A), taking sample from patient.
drug therapy assigned to one of the patients in set N_i^P by doctor.	go to the pharmacy, take pill and then walk to the place of patient for treatment.
IS notifies an additional-help call from patient in set N_i^P.	go to the patient (to carebox N_i^{CB} in area A).
Periodic checking time is up.	Check every patient's body condition in set N_i^P.
doctor discharged one patient in set N_i^P.	help patient leaving ED.
shifting of duty time is up.	accomplish task at hand, interact with nurses in following shift, hand over all the patients in set N_i^P.

take the task, otherwise, the task will be pushed into the technician group's task queue. Therefore, they have only one IF/THEN behavior rule, i.e., going to perform the task when get notified by IS.

Medical image test-room

Medical imaging is the service and process of creating visual representations of the interior of a body for clinical analysis and medical intervention, as well as visual representation of the function of some organs or tissues. Although the medical image test-room is comprised of equipment and technicians, we model this unit as a single agent to ignore unnecessary complexity. These agents will interact with patients and auxiliary technicians. There are several kinds of medical imaging services, such as CT-scan, B-scan ultrasonography, X-Ray and MRI-scan. Since the processes are similar, the significant difference is the service time from point view of modeling and simulation. The type of test was determined by doctor based on patient's character (acuity level, previous test). In our model, the service time model was fitted with exponential distributions based upon real data analysis and findings in queue theory [89]. The IF/THEN behavior rule of medical imaging test-room is described in Table 3.6.

Table 3.6: Behavior rules of medical image test-room.

IF	**THEN**
no patient waiting outside.	waiting for patient (IDLE).
patient with auxiliary staff waiting outside, and test-room is ready.	interact with patient and accompanied auxiliary staff.
physical test finished.	process test results, and send to the corresponding doctor through IS.

Laboratory test-room

The laboratory test-rooms receive patients' samples (e.g., blood) taken by a nurse, and analyze one by one by machine on a FCFS policy. Normally, there are several machines which can process samples simultaneously. Here, each of the machines is considered as a whole as an agent. The type of processing is a parameter of the agents customized by simulation users. For each machine, a maintenance service is required every 24 hours, and this process takes up to one hour, during which samples cannot be processed.

Regarding process time of different kinds of analyses, similar as medical image test in section 3.3.1, they are separately fitted with exponential distribution. The parameters of the distribution are based on the specification of machines and carefully calibrated based upon real data. The behavior rules are detailed in Table 3.7.

3.3.2 Interaction Model

The functional behavior of any system can be specified by a state machine (also called an object) [90]. In this research, to model the interactions between the agents and their states, the Finite State Machine (FSM) was used. The subsection 3.3.1 describes all the model of agents in ED. Once they are interconnected by information flow, they fulfill the purpose of providing healthcare service. Hence, agent modeling

Table 3.7: Behavior rules of laboratory test-room.

IF	THEN
no sample in the queue.	waiting for sample (IDLE).
new sample(s) waiting in the queue, and there are free analyzing machine(s).	detach sample(s) to free machine(s).
machine(s) completed the analysis.	catch results and send to the corresponding doctor through IS.
daily machine maintenance time is up.	start maintaining when machine completes current task.

is just one side of an agent-based model & simulation system, the interaction model which connects all agents to form a vivid system is another side. Although human behavior and interactions among people are among the most complex systems that exist, hospitals have strict behavior policies. It is thus reasonable to assume that all the agents in EDs behave regularly in reality. The interaction happens among agents was illustrated in Figure 3-2 (the group of two parallel lines with arrow between two agents). There are one-to-one interactions (e.g., doctor with patient), one-to-n interactions (like information system with patients), and triangular interactions (e.g., test-room, patient and auxiliary staff). From the point view of simulation which uses time as key indicator of the system, the duration of the time it takes for an interaction is significant. The model for fitting the time duration of interactions has been described in the agents' model.

The state of agents are presented by the value of their state variables and each state variable has a set of possible values. Based on the actual situation, the transition of agents' state is caused by interaction with other agents or in some cases with time elapse. Thus, the value of the state variables are changed by one of their behaviors or time elapsing (as illustrated in Equation 3.4):

$$Y_{K_i} = f(B_j, T)(0 \leq i \leq m, 0 \leq j \leq n) \tag{3.4}$$

where B_j represents the corresponding behavior with other agents, it is an element of the behavior set B. T represents the elapsing of the time because sometimes the state of the agents, e.g., patients' body condition after medicating, can change with time goes on without any interactions.

As shown in Figure 3-5, the state machine accepts commands and produces outputs, which means that when the agents interact with other agents and/or with the time elapsing (accept input), the value of one or several variables will be changed (because of the outputs produced). Any one of the variables' value changing will represent the state transition.

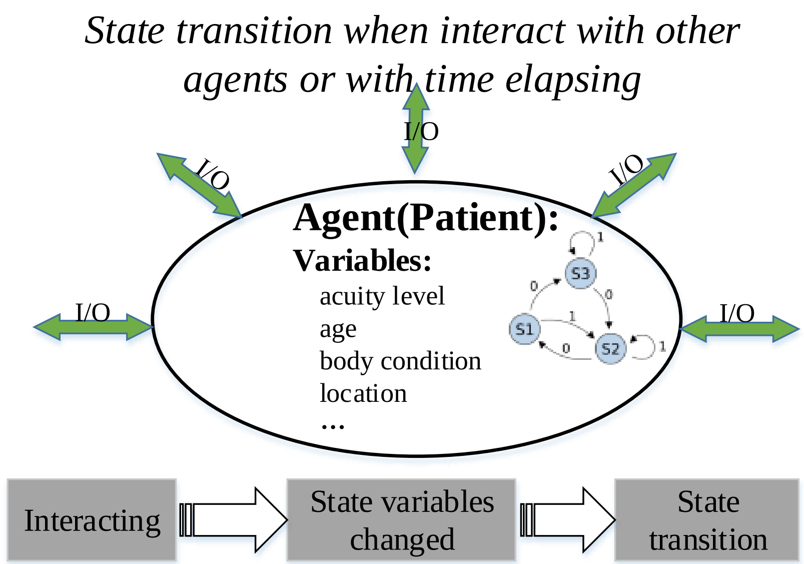


Figure 3-5: A typical patient's conceptual state transfer model.

Therefore, the set of one kind of agents' states is the cartesian produce of each state variable's possible value set V_i (see (Equation 3.5)). The state set of a specific agent in this type is a subset of S, which is determined by the specific configuration of the agent.

$$S = \{S_0, S_1, S_2, \dots, S_t\} = \{V_1 \times V_2 \times V_3 \times \dots \times V_n\} \qquad (0 \leq t \leq \prod_{i=1}^{n} K_i) \tag{3.5}$$

The models of agents described in subsection 3.3.1 are defined from single agent's

Table 3.8: A part of a patient's interaction log.

State	Source State	Destination state	Input
...	...	...	...
S_t	Waiting for service (free carebox).	Waiting for service (Doctor's diagnosis).	Notice from IS with a free care box.
S_{t+1}	Waiting for service (doctor's diagnosis).	Accepting Service(meet with doctor).	Doctor arrive at patient's carebox.
S_{t+2}	Accepting Service(meet with doctor).	Waiting for service (X-Ray test service).	Doctor order X-Ray test for patient.
S_{t+3}	Waiting for service (X-Ray test service).	Accepting Service(X-Ray test service).	X-Ray service available.
S_{t+4}	Accepting Service(X-Ray test service).	Waiting for service (Doctor's review of the test result).	X-Ray service finished.
...	...	...	...

perspective. To accurately represent a "live" agent in simulation, besides behavior rules, each agent has its own state variables for determining their current state. Although the IF/THEN behavior rules of an agent are generalized and do not represent specific actions with their interaction objects, the combination of the value of their state variables and the generalized behavior could specify the real action. For example, if the value of patient's state variable indicates that a patient stays in the first waiting room (i.e., *location = 1st waiting room*), it is sure that they are **waiting** for **triage service** instead of other services. Table 3.8 gives a part of one patient's state transition, although several states are the same as *Waiting for service*, the value of its state variables will determine the specific service the patient is waiting for.

Above all, it is feasible to deal with huge amount of agent states by means of defining agents through state variables. At the same time, it will be easy to add/remove states simply by adding/removing elements in the set of state variables and their corresponding behavior. For the study of other ED related problems, for example, the study of MRSA propagation in ED, some new state variables and their possible values could be easily added to gain insights on how the parameters evolve over time (e.g., routes of bacteria transmission). With the same approach, further functionality of the research object will emerge from these new states.

3.4 Model Implementation, data collection and information extraction

3.4.1 Model Implementation

Above section 3.3 details agent models as well as their interaction. However, the nonlinearities and interactions among agents over time and space can lead to such complexity that it is only possible to understand through simulation [91]. The full model has been implemented in NetLogo [92] simulation environment, which is an agent-based programming language and integrated modeling environment. The same as it in reality, in the implementation, when one interaction is accomplished, agent will return to an inactive state and check their task list (enqueue by IS). The next task will be chosen from the list according to the priority policy. With this mechanism, agent models described in subsection 3.3.1 can be easily translated to NetLogo programming language with a state machine structure. Since the systemic key performance indicators (KPIs) are extracted from detailed interaction information among all the agents, a proper way to collect these atomic interacting data should be carefully designed.

3.4.2 Atomic Data Collecting

An agent-based model is comprised of individual components of the system as well as their interactions. The simulation results generated directly by simulation are atomic data about the state and interaction among agents. This atomic data is the source of knowledge for a better understanding of the complex system. Therefore, simulation output from the agent-based simulator is subjected to extract the system-level behavior information. However, one reason why emergent behavior of a complex system is hard to predict is that the number of interactions between components increases exponentially with the number of components. The interaction data and the state information of the system are thus massive. In the data collection process of a real system, we tend to collect more data than needed in order to cover as much information as possible. Whereas in simulation, data monitoring should be focused on the goal of analyzing because the simulation process is reproducible (using the same model configuration and random seed). To this end, we designed a customizable data monitoring layer between the micro-level behavior simulator and the data processing layer from a point view of "sensor". Through this customizable layer simulation users could customize their data collecting according to their analysis requirements without accessing to the source code. Moreover, some of the sensors also provide simple data processing methods (filter or reducer) to carry out basic analysis (e.g., maximum, minimum, average, median value, standard deviation) in order to reduce the size of micro-level simulation results without affecting final knowledge discovery (e.g., in cross-scenario analysis cases). In summary, we designed "sensors" along with agents. These "sensor" can monitor all the behavior of agents and the user can enable/disable these "sensors" according to their analysis requirement to avoid collecting redundant data which are not interested to the user.

This customizable data monitoring layer is implemented in two parts: an independent application with graphical user interface which enables simulation users to customize data-collecting behavior (shown in Figure 3-6), and a data-collection program along with the simulator to record and write data to comma-separated values (.csv) files. The two parts communicate through a configuration file before simulating. Since this layer is considered as a set of sensors to monitor states of agents and environment (e.g., waiting room), all of which are customizable, e.g., enable/disable, sampling frequency (for continuously monitoring passive agents' state such as the test-room occupancy).

Regarding the nature of an agent-based simulator, we classified the raw simulation data in two categories: environment state and interaction information. The state information, e.g., test-room occupancy and number of patients in waiting room, are sampled in a given time interval (sample frequency, e.g., output the length of test-room queue every 5 minutes for indicating its occupancy). The interaction in-

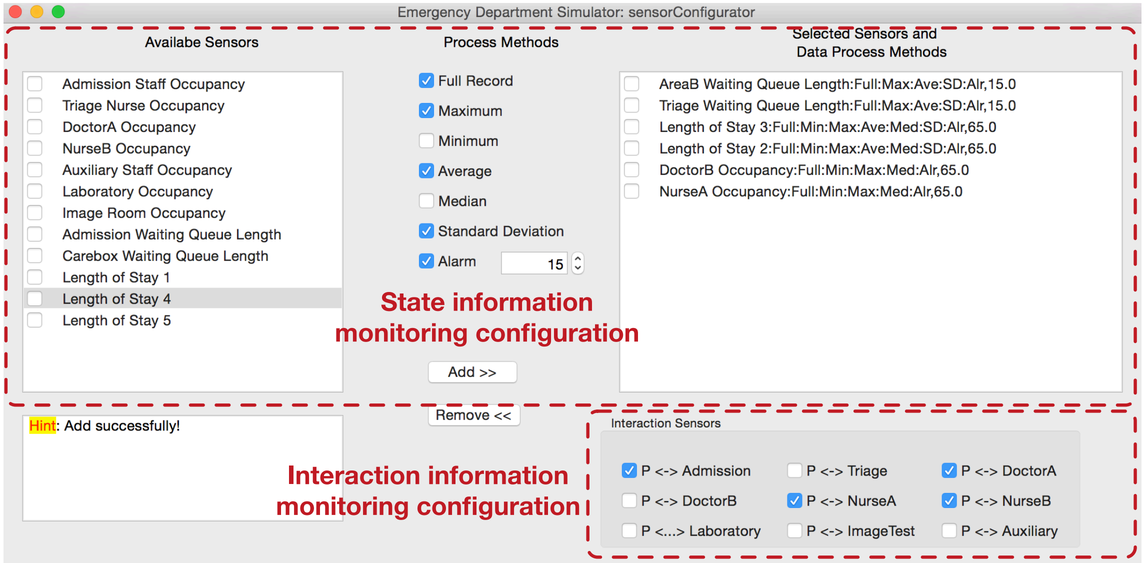


Figure 3-6: Interface of application for configuring the micro-level interaction information monitoring.

formation contains records of all the interaction among agents and this information is recorded as five Ws (Who, What, When, Where, Why) and one H (How long it takes). Figure 3-7 shows part of the interaction records.

1	who	what	when(minute)	where	why	how long(second)
86179	(doctorb 76) and (patient 16279)	first-visit	70446	doctorB's room	default	1200
86180	(doctorb 74) and (patient 16283)	first-visit	70447	doctorB's room	default	900
86181	(nursea 80) and (patient 16158)	go-home	70447.5	carebox	default	150
86182	(doctorb 75) and (patient 16277)	first-visit	70448	doctorB's room	default	210
86183	(doctorb 78) and (patient 16222)	treatment-finished	70449	doctorB's room	default	1320
86184	(doctora 69) and (patient 16211)	test-result-review	70449.5	carebox	default	330
86185	(doctorb 73) and (patient 16281)	first-visit	70449.5	doctorB's room	default	1290
86186	(admission 1) and (patient 16285)	admission	70451.5	admission desk	default	300
86187	(doctora 67) and (patient 16199)	test-result-review	70451.5	carebox	default	120
86188	(nursea 80) and (patient 16199)	laboratory test	70453.5	carebox	default	1080
86189	(nursea 84) and (patient 16211)	go-hospital	70455	carebox	default	1290
86190	(doctora 69) and (patient 16262)	test-result-review	70455.5	carebox	default	450
86191	(doctorb 77) and (patient 16154)	treatment-finished	70455.5	doctorB's room	default	510
86192	(doctora 66) and (patient 16033)	test-result-review	70456.5	carebox	default	300
86193	(doctorb 72) and (patient 16247)	test-result-review	70457	doctorB's room	default	360

Figure 3-7: Agent interaction records.

The interaction data is the most basic information on reflecting system behavior. They are obtained through putting a device ("sensor") on each of the individuals in the ED, and these "sensors" could monitor individual's activities as deep as a user required. Although these data are massive, theoretically all indicators about the system behavior could be retrieved. Different with blackbox simulation models, these basic interaction information not only emerges interesting KPIs but also explains the root-cause of the simulation scenario (e.g., case in subsection 6.2.1).

3.5 Experiment design and model execution

3.5.1 Scenario Design

The term "scenario" in this thesis represents a set of parameter value which specifies an ED. In our simulator, only the model (agent model and interaction model) was implemented in NetLogo, the value of parameters which characterize the model are in configuration files. Therefore, simulation users could design their experiments in a form of configuration files without accessing the source code. Thus in a decision support process, users could analyze their requirement in macro-level, then design experiments in micro-level (parameters to specify agents' model). Finally the macro-level systemic KPIs predicted by simulation will be extracted from micro-level interaction data. The basic principle is thus to propose different scenarios and then predict the behavior of ED under scenarios by using the individual behavior simulator. The state of the ED is then assessed by using selected KPIs retrieved from the simulation data. If the ED reaches or is in a degraded or critical state, corrective solutions are proposed, and then new simulations will be performed to verify the effectiveness of solutions. This process is repeated until the ED returns to a normal state as expected. In this case, the corrective solutions are listed and allow the formulation of the corrective rules associated with each situation.

3.5.2 Model Execution in Cluster

Armed with powerful computation and memory resources, the ability to test a large number of scenarios for "what-if" analysis in a short time period has made simulation a widespread tool in decision supporting and operation research. Agent-based simulation is computationally expensive. This sub-section describes a framework to efficiently execute the model that was implemented in NetLogo environment. Due to the inherent nature of patient flow with characteristics such as stochastic arrivals, stochastic service times, and uncertainties in patient routings, the agent-based ED model has stochastic characteristics by nature. Consequently, results from a single execution may not be statistically reliable, so repeating one scenario with different random seeds becomes necessary and challenging as well. High performance computing techniques are prominent means of leveraging the complex simulation power of ABMs. Since there is no data dependency among difference scenarios or repetitions of the same scenario, master-worker execution model becomes a good choice. However, execution time of different scenarios or even repetitions of the same scenario are different, load balance should be carefully considered. Given this, we designed an execution engine to launch and execute the agent-based model on cluster with NetLogo. Master-worker mechanism was used to achieve load balance. The execution model is illustrated in Figure 3-8, in which MPI (Message Passing Interface) is used to manage and communicate among master and worker processes.

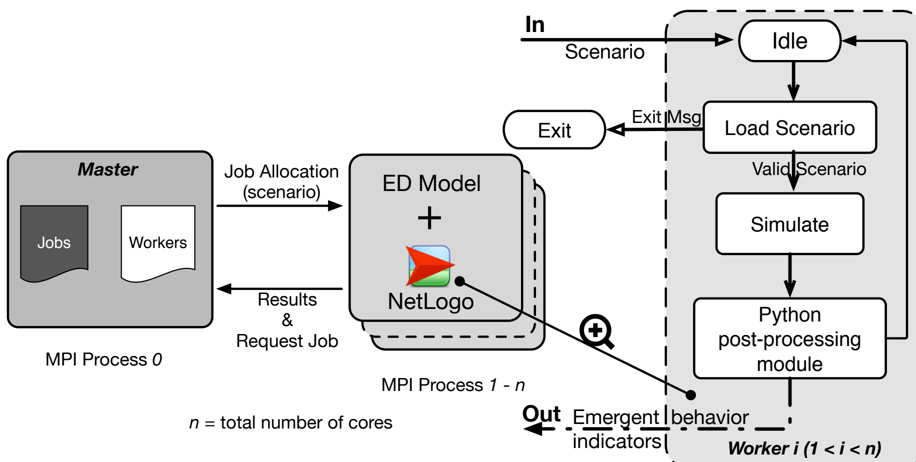


Figure 3-8: The master-worker execution framework for agent-based models on a cluster. Atomic data will be analyzed natively in the same node, and only systemic information will be send back to master. NetLogo controlling API (released alone with NetLogo) was used to invoke NetLogo and initialize the agent-based model so as to avoid loading NetLogo for each execution.

As the meaningful information (systemic behavior indicators) is extracted from interaction records. In addition that the size of interaction data increases exponentially with the number of agents, it is better to process data in the same computing node which simulates the scenario to avoid network crowding. More specifically, as illustrated in Figure 3-8, the number of workers is equal with the total number of CPU cores, and the master will be assigned to one of the nodes. When one scenario finishes its execution, the data analysis program will be launched to analyze the simulation data locally on the same computing node. Then the processed macro-level KPIs will be sent back to the master. Considering that the launch of NetLogo takes around 30 seconds, the NetLogo controlling API (comes with NetLogo.jar from released version) was used to invoke and control NetLogo. That is, when a new execution is assigned by the master, instead of reload the NetLogo environment, the controlling API will initialize the model as well as NetLogo environment with parameters in the scenario. Thus, NetLogo will only need to be launched once for each worker. Specifically, MPI library and C programming language were used to distribute and balance simulation task, Java was used to take full control of NetLogo (via controlling API), and Python was used for simulation data analysis. Furthermore, this model execution framework is one way of addressing the computational complexity of such systems, which is also useful for model parameter calibration and simulation based optimization. Although the presented framework was designed to speed up our simulation study, we are confident that it also has the potential to be used by for other ABMs that are implemented with NetLogo.

3.5.3 Warm-up Simulation

The simulation model of ED is a typical steady-state system that neither has obvious starting events nor requires stopping at a certain time during the simulation process. In order to remove the initial bias, we preset a period of warm-up time to be the transient period. To determine the proper length for warming up, we first used the linear regression approach suggested by [93] to identify the end of transient state. This approach uses the least-squares method to determine if the linear regression slope coefficient is close enough to zero for a specific range of observations, i.e., the transient state finishes when the slope is close to zero. Then based upon the transient period length (normally less than 100 hours), the system has been set at a big margin of one-week (168 hours) as warm-up period. The data collecting layer will start to record all the requested atomic data (i.e., enable sensors according to user's configuration) after warm-up period.

3.5.4 Replication

In view of the agent-based model, agents' behavior is fitted with probability distributions and most of the KPIs, e.g., LoS, door-to-doctor time are statistical indicators. Therefore, to make KPIs statistically reliable, a minimum sample size should be guaranteed. Here, we use LoS as an example to illustrate the method for computing the minimum sample size.

In probability theory, Chebyshev's inequality guarantees that in any probability distribution, "nearly all" values are close to the mean [78] . Specifically, let X be a random variable with finite mean μ and variance σ^2, then for any given positive ϵ we have:

$$Pr\left(|X-\mu| \leq \epsilon\right) \geq 1 - \frac{\sigma^2}{\epsilon^2} = 1 - \frac{{\sigma_r}^2}{n\epsilon^2} \tag{3.6}$$

Where, X is the random variable, σ^2 is the variance $var(X)$, $\mu = E(X)$ represents the expectation of X, ϵ denotes the absolute error, ${\sigma_r}^2$ is the theoretical variance extract from real data, and n is the sample size. More specifically, if we want to convince that the probability of the average value lies inside the interval $(\mu-\epsilon,\, \mu+\epsilon)$ is no less than α (confidence greater than α), so as to use average of samples to represent μ. The minimum size of sample n could be calculated by Equation 3.6 with given σ and ϵ. Take the evaluation of patients' LoS as an example, with the patient arrival model described in subsection 3.1.1, 10% absolute error, 95% confidence, the minimum sample size as well as simulation time are shown in Table 3.9 by category of patients with the same acuity level. The statistical information (${\sigma_r}^2$) was retrieved from about 100,000 valid patient records in 2014.

In order to meet the sample size requirement, if the simulation time for actual analysis is greater than the minimum request, it is reasonable to simulate once without repetition. Meanwhile, if actual analysis requires shorter simulation time period,

Table 3.9: Minimum sample size to evaluate patient's length of stay (LoS, Relative error: 10%, confidence: 95%.). The statistical information was retrieved from about 100,000 valid patient records in year 2014.

Items \ Acuity Level	1	2	3	4	5
Mean (LoS, minute)	520.27	437.39	617.96	166.43	116.40
Standard Deviation	493.88	734.57	806.79	182.055	94.24
Minimal sample size (number of patient)	901	2820	1704	1196	655
Percentage of patients arrival	0.55	7.10	31.00	49.30	12.05
Min. simulation time (day)	410	99	13	6	13

repetition the same scenario with different random seed is a must. The actual repetition times can be computed with minimum sample size requirement divided by the total arrival patients in the scenario separately for each acuity level and get the maximum times. It is clear to see from Table 3.9 that, evaluating behavior of patients with acuity level 1 requires the longest simulation time because the proportion of patients arrive with acuity level 1 is fairly small (about 0.55 %). While for patients with acuity level 4, which makes up the largest proportions, requires the shortest simulation time. Having said that, if we want to evaluate LoS for patients with acuity level 1 to 5, we have to simulate at least 410 days (the maximum in five acuity levels). If the actual simulation time (t_{simu}) is shorter than 410 days, we should repeat the same scenario with different random seeds for at least $\lceil 410/t_{simu} \rceil$ times. However, if only assess LoS of patients with acuity level 4 and 5 (the area B), simulating for 13 days would be enough for the minimum sample size requirements. To evaluate other KPIs, e.g., occupancy of sanitary staff and medical test equipment, the same procedure can be used to compute the sample size requirement.

3.6 Simulation Results

The model described above has been calibrated based on real data of 2009 - 2011. This section will illustrate the cross-validation results. That is, parameters for specifying patient arrival model (described in subsection 3.1.1) are retrieved from real data (in

2014), weekly arrival rates throughout the year are shown in Figure 3-9.

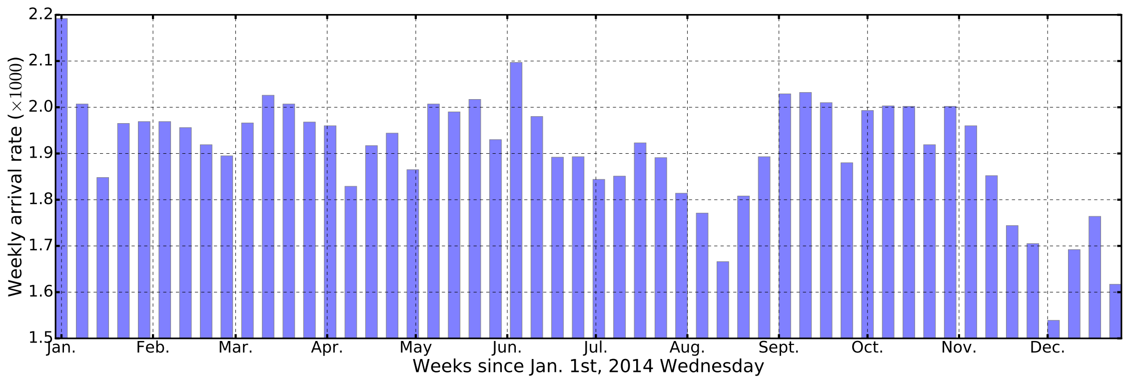
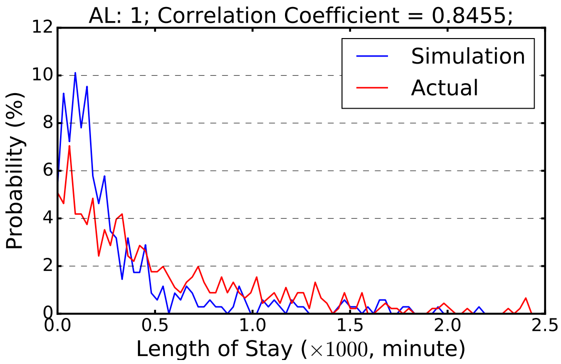


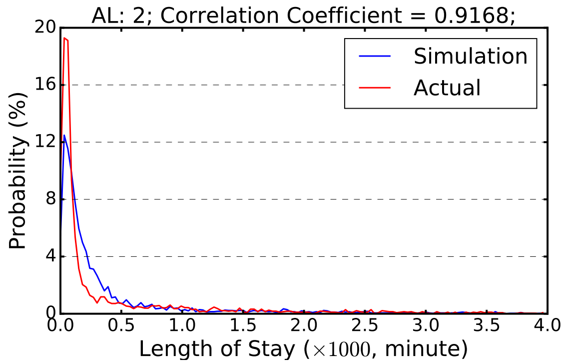
Figure 3-9: Patient weekly arrival rate, extracted from one-year actual data of the Hospital Universitari Parc Taulí. To emulate the ED of the Hospital Universitari Parc Taulí for model validation, this data will be used as (input) parameters to specify patient arrival pattern described in subsection 3.1.1.

The ED resource configuration, which includes human resources (as well as their shift), equipment, beds (as well as its layout) are specified as the same as real system. With patient arrival data shown in Figure 3-9 as input, we simulate to imitate the real operation in 2014. The patients' LoS, which can represent overall behavior of patient as well as EDs, was used as the main indicator of the system behavior. According to Table 3.9, two repetitions were applied to this scenario (the simulation scenario has 365 days). The simulation was carried out on an 8-node cluster with total number of 512 AMD Opteron^{TM} Processor 6262 HE cores, and 2TB RAM. Since there are only one scenario with two repetitions, only three CPU cores (one for master and two workers) are used. To validate simulation results, the LoS retrieved from real data and emergent behavior of simulation models are compared in statistical way (histogram). The comparison results are illustrated in Figure 3-10, in which each figure represents LoS distribution of patients with the same acuity level. The results for validating patient arrival model were demonstrated previously in Figure 3-1a. It is clear that the proposed patient arrival model could fit well with the actual patient arrival.

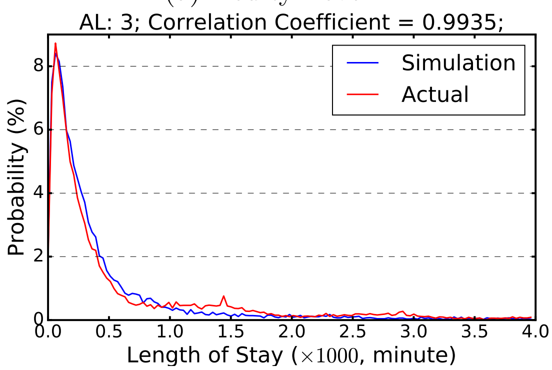
It can be seen from Figure 3-10 that, as a result of the small number of patients attending with acuity level 1 and 2 in real situation (about 0.55 % and 7.1 % re-



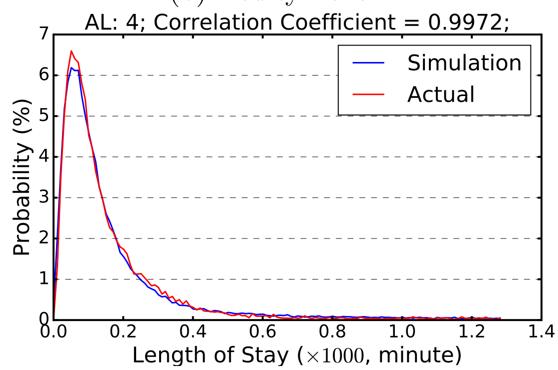
(a) Acuity Level 1



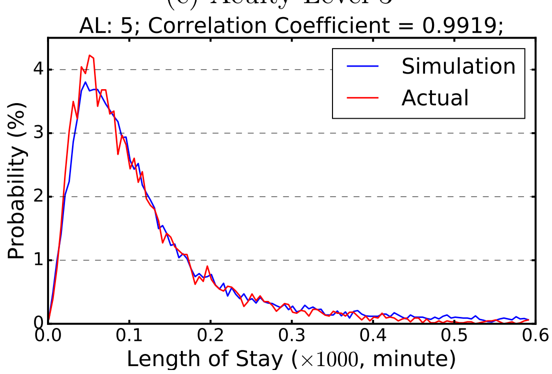
(b) Acuity Level 2



(c) Acuity Level 3



(d) Acuity Level 4



(e) Acuity Level 5

Figure 3-10: A set of simulation results about distribution of patients' LoS with each acuity level; each figure is the comparison of histogram of patients' LoS extracted from real data against simulation results. The statistical interval widths are: 30 minutes for acuity level 1, 2 and 3; 10 and 5 minutes for acuity level 4 and 5 respectively.

spectively), the fitness for these two levels are not as good as others. Based on the correlation coefficient quantitive index, simulation results and statistical characteristics of empirical data are very close.

3.7 Discussion

The Emergency Department (ED) is a complex, stochastic environment which has time-dependent behavior. Making changes to an ED's resources is a challenging activity since it has significant impact on its performance. Incorrect decisions may lead to serious consequences on the quality of service and cause unnecessary deaths. Simulation enables organizations to make better decisions by letting them see the impact of changes before implementing.

This chapter presented an agent-based model of EDs which could be used to settle problems such as prolonged waiting times, inefficient use of ED resources, and unbalanced staff scheduling. This is useful for designing empirically grounded agent-based simulations, and for gaining direct insight into observed dynamic processes. In this model, policies such as staffing, human factors such as sanitary staff behavior, inexperienced cases such as a flu outbreak could be set up and their effects on system performance such as waiting time and throughput could be quantified. With the amount of adjustable parameters, the simulator is customizable to simulate a variety of scenarios. The presented simulator is currently working as a platform to study Methicillin-resistant Staphylococcus Aureus (MRSA) transmission in EDs and as an experimental platform of EDs to provide data under various scenarios for knowledge discovery.

Furthermore, starting from simulating the EDs, our efforts proved the feasibility and ideality of using an agent based modeling & simulation techniques to study healthcare systems. The cross-validation results showed that the developed ED simulator can accurately represent the emergent behavior of the complex ED system. Some demo application results previously presented in conferences proved that the simulator is ready to work as part of decision support system (in Ref. [37, 36]). The framework developed in our work is a step towards building a full model of integrated care system. It opens a wide field of possible simulation scenarios for a better understanding of the integrated complex system. These simulation scenarios are crucial prerequisites for improving the coordination and integration of care and increasing the efficiency of resource allocation. In addition, more healthcare subsystem, such as EDs in an area, out-patient service unit, and in-patient unit could be simulated and connected together to allow for the assessment of ambulance and patient redirection policies.

A precise model of ED is the base for further study, another work direction could be to combine the simulation model with optimization algorithms to find the optimal (and sub-optimal) of design parameters to optimize the performance of the simulated system.

Chapter 4

Model Calibration

The First Rule of Program Optimization: Don't do it. The Second Rule of Program Optimization – For experts only: Don't do it yet.

- Michael A. Jackson

To tackle the problem of efficiently managing increasingly complex systems, simulation models have been widely used. This is because simulation is safer, less expensive, and faster than field implementation and testing. To achieve high fidelity and credibility in conducting prediction, explanation and exploration of the actual system with simulation models, a rigorous calibration and validation procedure should firstly be applied. However, one of the key issues in calibration is the acquisition of valid source information from the target system.

The aim of this chapter is to develop a systematic method to automatically calibrate a general emergency department (ED) model with incomplete data. The proposed calibration method enables simulation users to calibrate the general model for simulating their system without the involvement of model developers. High performance computing techniques were used to efficiently search for the optimal set of parameters. The case study indicates that the proposed method appears to be capable of properly calibrating and validating the simulation model with incomplete data.

4.1 Introduction

With the rapid growth of computational techniques, computational thinking brings researchers and practitioners into a new dimension to traditional modeling and simulation tasks. That is, the computational science transforms observed complex phenomena into conceptual models. Then the models are formulated into algorithms that can be executed to yield predictions and estimate hidden parameters. These predictions can be compared to the observations, revealing to what extent the model is an accurate description. Although data-driven simulation are mostly designed for prediction, the simulator should firstly be able to imitate the real system. Generally, a simulator of a specific system is comprised of the following: input (X), the model or transformation function ($f(X)$), and output (Y). For an accurate simulator, when we put the same input as it in a real system, the output of simulator should be close enough to the output of the actual system. Since $f(X)$ is based on abstractions, idealization, and many disputable assumptions, the model must be fine-tuned according to some historical input-output samples from the target system in order to get trustworthy simulation results.

The ED is a typical complex system, which serves essential needs in society, delivering emergency health care and simultaneously acting as a safety net provider [94]. In recent years, simulation has emerged as an increasingly effective tool to study ED related problems and support making decisions to efficiently manage the complex ED system. While these simulation models can be advantageous to engineers, the models must be calibrated and validated, i.e., the model should first be able to accurately imitate the real system. Advances in computational technology, along with the increased complexity of system design and management have created an environment in which microscopic simulation models have become useful tools for managing complex system. Among which, the Agent-based Model (ABM) is one of the most important tools for exploring emergent behavior (a phenomenon that describes the behavior of a system, which cannot be explained alone by the sum of its parts [72]) mostly because it can provide a way to see the forest through the trees and, insight is often more important than sheer numbers [23, 26, 25].

As described in chapter 3, the agent-based simulation models encompass numerous independent parameters to describe individual behavior of the system components. Reliable and complete real data from target system is obviously the precondition for setting up an accurate simulator. Unfortunately, many of the parameters are either unavailable in historical data or difficult to measure in real situation, yet they can have a substantial impact on the model accuracy. Thus, when existing real data was incomplete to allow direct estimation of the model parameters, a calibration process (also known as tuning) has to be conducted to indirectly estimate proper values for those unknown parameters. However, the calibration of model parameters for an ABM is a big challenge for standard calibration techniques, due to the large parameter search-space, long simulation run times, uncertainties in the structural model design and different observation levels upon which the model needs to be calibrated [75]. Given this, the model parameter calibration problem can be formulated as a stochastic programming problem whose objective function is an associated measurement of an experimental simulation. Nevertheless, the objective function is typically (a) subject to various levels of randomness, (b) not necessarily differentiable, and (c) computationally expensive to evaluate due to the complexity of the model. To the best of our knowledge, limited research has been conducted on this thorny and critical problem of estimation in the face of data scarcity.

Accordingly, conventional calibration, which is done manually by using the trial-and-error method, is time consuming and tedious. A systematic method to automatically search for the optimal value of model parameters is promising. The simulation-based optimization is an emerging field which integrates optimization techniques into simulation analysis. The primary goal of simulation-based optimization is to optimize the performance of a system through simulation. More specifically, it is a way to find the optimal set of parameters for a given criterion. Then the optimal parameter set will enable the model to achieve a specific function optimally or that the results of the simulation are close enough to actual data. Therefore, if we set the model input the same as reality, consider the unknown model parameters as variables, and the similarity between simulation output and actual system output as objective, the optimization is a model calibration process. When some of the model parameters are missing and impossible to get from real system, this optimization process will be able to find the optimal values for setting up the model. Thus, the precondition for the calibration process is a set of reliable input-output pairs from the target system.

In this chapter, we will address a critical step in simulating a complex system - the systematic model calibration and validation in the face of data scarcity. More specifically, calibrate a general agent-based model of ED to simulate the ED of the Hospital Universitari Parc Taulí (a university tertiary level hospital in Barcelona, Spain) with incomplete data (missing duration of key services). The simulation-based optimization was conducted by using an existing tool [39, 40, 41] developed by Sandia National Laboratory. According to the practical requirements of evaluating a simulation-based objective function, an initial distance-based lookup mechanism was proposed to further speed up the optimization. The rest of this chapter is structured as follows: an overview of the EDs as well as the general model (detailed in chapter 3) will be described in section 4.2. The section 4.3 details the proposed automatic calibration method and the results we got in the case study. At last, section 4.4 draws the conclusions.

4.2 The Agent-based Emergency Department Model

Typical EDs have common interacting elements such as doctors (physicians), nurses, technicians, receptionists, beds, medical devices that are interconnected via flows of patients, information and processes (registration, triage, diagnostic, discharge). The EDs studied in this research are focused on the Spanish type. This section gives the brief introduction of the system as well as the general model. Firstly, subsection 4.2.1 gives a brief introduction of the system and model, it is a summary of the model presented in chapter 3 in particular for explaining model parameters calibration. Then subsection 4.2.2 describes the parameters which are impossible to get from real data and need to be calibrated.

4.2.1 General process and model overview

As shown in Figure 4-1, typically, a patient enters the ED through one of two ways: by themselves or by ambulance. Upon arrival, walk-in patients need to walk to the registration window, briefly give their personal information to the registration staff. After that, they have to stay in a waiting room until triage. Once the information system assigns a triage nurse to the patient, they will go to the corresponding triage box and interact with the nurse. Triage consists of a brief assessment of the patient's body condition and an acuity level will be assigned to the patient according to their severity. Then, patients will wait in the second waiting room before entering the diagnosis & treatment area. For those patients who arrive by ambulance, they are registered and triaged in the ambulance, and thus go to the second waiting room directly. As described in subsection 3.1.2, the Spanish scale of triage is very similar to the worldwide Canadian Emergency Department Triage and Acuity Scale [79, 80]. The scale consists of 5 levels, with 1 being the most critical (resuscitation), and 5 being the least critical (non-urgent). The triage process also determines the order and priority with which the patient will be attended and the treatment area where they will be treated. The registration and triage service are first-come, first-served (FCFS) for all the patients, whereas entering the diagnosis & treatment area is acuity-level-dependent FCFS (patients with acuity level 1 have the highest priority).

With regard to the treatment area, as shown in Figure 4-1, in most Spanish EDs, there are two treatment areas (labeled as A and B in this study), which operate independently to provide a diagnosis & treatment service. Area A is for those patients with acuity levels 1, 2 and 3, while area B is a dedicated stream of resources to process lower acuity patients with levels 4 and 5 more quickly. Area A is made up of careboxes, a carebox is a small room which contains essential medical equipment and supplies that could be used for patients' treatment. Patients attended in area A will stay in their own carebox throughout the diagnosis & treatment phase, and transporting should be done by auxiliary staff. In area B, there are several attention boxes in which doctors and nurses interact with patients, and a large waiting room in which

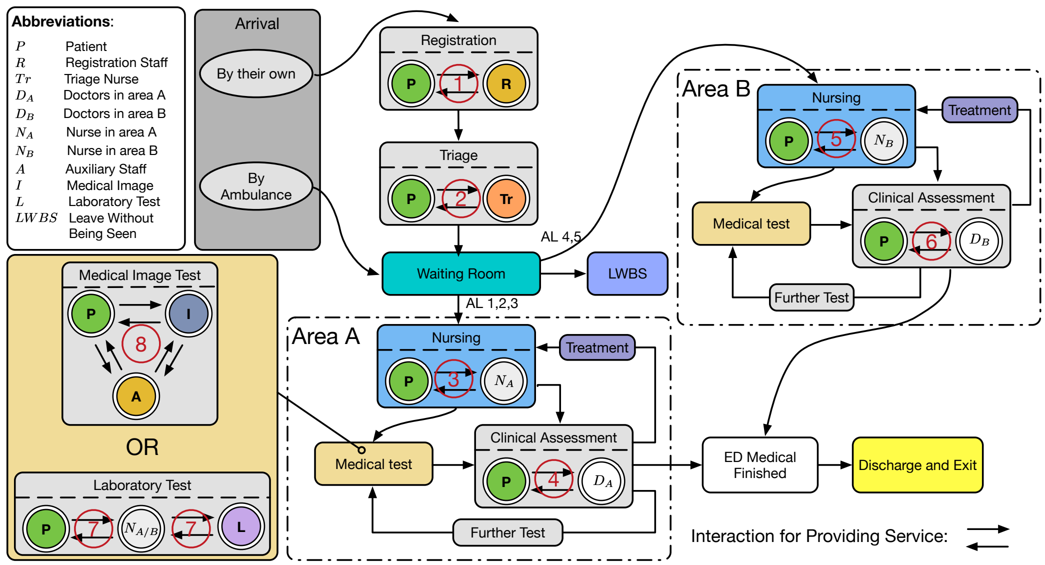


Figure 4-1: Diagram of patient flow through the emergency departments. Eight service processes, marked by the circled number, drive all aspects of patient flow. Most of the services are interdependent, the duration of service is different for each service. Note: Area A and area B are designed for urgent and non-urgent patients separately, they have different groups of staff and work independently.

all patients will remain while not having interaction with the ED staff. Note that the doctors and nurses are specified for different areas, their behavior is different, but medical image test-room and laboratory testing services are shared by area A and B.

In the diagnosis & treatment phase, once the patient has got a free space in the treatment area, the doctor will have an interaction with the patient, then the doctor makes one of the following decisions: (1) a patient needs to receive an imaging test (e.g., X-ray, B ultrasound); (2) assign laboratory tests (e.g., blood test, urinalysis); (3) discharge the patient and; (4) make a drug therapy plan. If testing was assigned, and when the results become available, the patient needs to have an interaction with the same doctor who conducted the consultation in order to receive a reassessment with their test results. Notably, as shown in Figure 4-1, some patients need to repeat the consultation-test-reassessment/treatment more than once. In summary, as marked by circled number in Figure 4-1, there are 8 different types of service (provided by different providers). As described in chapter 3 and previous studies in Ref. [38, 36, 37], the ED was modeled as a pure spatial agent-based model. It is formed entirely from the rules governing the behavior of the individual agents which populate the system, no higher-level behavior is modeled. Thus, the system behavior emerges as a result of micro-level actions and interactions. The full model has been implemented in the NetLogo [92] simulation environment, which is an agent-based programming language and an integrated modeling environment. The work proposed in this chapter, on calibrating the general model to imitate a real ED, was challenged by the fact that the parameters for characterizing the service-time distributions (eight in total marked by circled number in Figure 4-1) are not directly obtainable from historical data or a real system.

4.2.2 Model parameters

In this study, Hospital Universitari Parc Taulí in Catalonia, is the target system to imitate. It is a university tertiary level hospital in Spain that provides care service to a catchment area of 500,000 people, and attends more than 160,000 patients per year in the ED. In order to calibrate our general model to imitate the Hospital Universitari Parc Taulí, we requested 12 months (Jan. 1st, 2014 - Dec. 31st, 2014) historical operation data from the information system's database. The missing values and invalid records has been carefully handled. Regarding that August is holiday period, lots of people go on vacation, accordingly the configuration of ED is different (e.g., fewer staff or fewer senior staff), operation records of August is discarded for this calibration study.

The ABM requires numerous parameters to characterize the behavior and features of each agent. Some of them can be retrieved directly from actual operation data of the target ED system, such as the patients' features, the number of medical testing, the number of treatment processes and the number of doctor interactions with one patient. However, the service time information was not recorded by the information system (out the scope of an information system). Thus, the parameter for the entire service-time models could not be determined directly with the real data. As illustrated in Figure 4-1, there are 8 service processes (marked by a circled number), all the service are carried out by interacting between agents. According to the research findings in queue theory [89], exponential distribution is typically used to make mathematically simplifying assumptions. Given this and empirical data from ED staff, an exponential distribution was used to fit the duration of each type of service, but the parameters for these distributions should be calibrated in accordance with the target system.

More specifically, the service time is defined as the interval spent actually on receiving service (i.e., the time differences when the services started and ended). Note that, in the agent-based ED model, the duration of the service mentioned in this chapter only represents the time spent actually on interacting, waiting time is excluded because the waiting time is an emergent property of the system. In principle, the time for medical imaging test is composed of two parts, the interaction between patient and the test-room technicians, and time for processing testing results. Given that the second part is determinable, the key is to calibrate the duration for the interaction. Similar to a laboratory test, which is composed of two parts, taking samples by nurse and analyzing samples by machine. The second part is easy to obtain from the machine's specification thus only the duration of interaction for taking sample needs calibration.

Accordingly, we have the model input (arrival patient and their features), output (systemic performance indicator such as length of stay), and part of the model parameters retrieved directly from real data. With respect to the unknown parameters, empirical information such as boundary constraints, typical value can be obtained from experienced staff. Although the empirical information is not accurate, it can dramatically reduce search space-size. Table 4.1 lists all the parameters to be calibrated, as well as their boundary constraints. Thus the task is to search for an optimum set of parameters which can lead to good (acceptable) fitness between the simulation results and actual data.

In summary, due to data scarcity, although the distribution of specific service duration cannot be fitted by such standard techniques as maximum likelihood estimation, we had some other time stamps which enable us to derive an indirect approach to estimate the service-time distribution parameters.

Table 4.1: The parameters to be calibrated for the general agent-based model of emergency departments, in order to imitate the emergency department of Hospital Universitari Parc Taulí. Note: **LB** and **UB** denotes Lower and Upper Boundary respectively, **TV** represents the Typical Value; all the units of time are in minutes. The **Identity** column corresponds to the circled numbers in Figure 4-1 denote the type of service.

ID	Notation	Description	LB	UB	TV
1	$T_{service}^{register}$	the parameter for registration service-time distribution model.	2	15	5
2	$T_{service}^{triage}$	the parameter for triage service-time distribution model.	5	20	10
3	$T_{service}^{nurseA}$	the average duration of service of nurses in area A.	8	30	16
4	$T_{service}^{doctorA}$	the average duration of service of doctors in area A.	8	30	18
5	$T_{service}^{nurseB}$	the average duration of service of nurses in area B.	5	20	12
6	$T_{service}^{doctorB}$	the average duration of service of doctors in area B.	5	20	15
7	$T_{service}^{imaging}$	the average duration for taking medical imaging.	20	40	25
8	$T_{service}^{lab}$	the average duration for taking laboratory test sample.	10	30	15

4.3 Calibrating Model Parameters Under Data Scarcity

Calibration traditionally conceptualized as an step in model validation. It involves systematic adjustment of model parameters so that model outputs can accurately reflect the actual system behavior. To calibrate a model, three important issues need to be addressed. The first issue is to select significant metrics to represent the emergent behavior of the target system, and specify a general and effective fitness function to measure the distance between a simulated scenario and the real situation. The second issue is to reduce the computation time because exhaustive search in parameter space is expensive (exponential growth with the number of parameters). The third issue is to obtain robust solutions for avoiding the over-fitting problem. That is, the calibrated model is not only able to fit historical dataset (dataset for calibration), but also able to predict reliable result with new input data. Due to the fact that all the services in an ED are interdependent, it is unreasonable to characterize parameters one-by-one or evaluate fitness process-by-process. To address this issue, one way is to consider all unknown parameters as a set, then simulate with the set and evaluate the similarity of system metrics as a whole. That is to say, a full simulation has to be carried out to evaluate one set of parameters, changes to any of the parameters will result in one new simulation scenario. The following subsections 4.3.1 - 4.3.5 will detail all the issues and processes on calibrating the model parameters under data scarcity. The calibration results and discussion are given in subsection 4.3.6.

4.3.1 Problem formulation

The goal of simulation is to imitate the behavior of a real system so as to accurately predict system behavior under unknown scenarios. Due to data scarcity, some of the model parameters are difficult to obtain directly from actual data, we thus have to tune these parameters indirectly with the goal of producing similar macroscopic behaviors as it in real situation. Thus, the calibration process of agent-based model of ED is defined as: *Given an agent-based model, a setting of parameter X to be calibrated, the task is to find the global optimal X^* that minimizes the fitness function.* From an optimization point of view, the calibration can simply be expressed as:

$$
\begin{aligned}
&Minimize \quad f_{fitness}\left(p_1, p_2, \cdots, p_n\right) = K\left(actual, simulation\right)\\
&Subject \quad to:\\
&p_1, p_2, \cdots, p_n \quad \textit{make sense in real situation}
\end{aligned}
$$

Where, $K\left(actual, simulation\right)$ is a function to evaluate the similarity between simulation results and actual data. The $\{p_1, p_2, \cdots, p_n\}$ is the set of parameter values (also called scenario in this study). In this study, the value of parameters represent the ratio to the typical value in Table 4.1. However, there are two main challenges in solving this global optimization problem. One is that the condition – *make sense* is difficult to describe in the optimization model because these parameters represent the behavior of a physical system (rather than sheer numbers). The other challenge is that the fitness function is non-convex, it has a very complex response surface, and it is computationally expensive to evaluate. However, if we decompose the condition, i.e., only consider the lower/upper bounds of the parameters (main part of the condition), though the optimization process is not an absolute global optimization problem because the best fitness parameters may not make sense in reality, it becomes solvable. Therefore, a solution gives the best of both worlds is: searching for the local minimum points under boundary condition, then manually checking if the solution makes sense in reality. Considering the over-fitting problem and model validation, a systematic method to calibrate and validate a general model is illustrated in Figure 4-2.

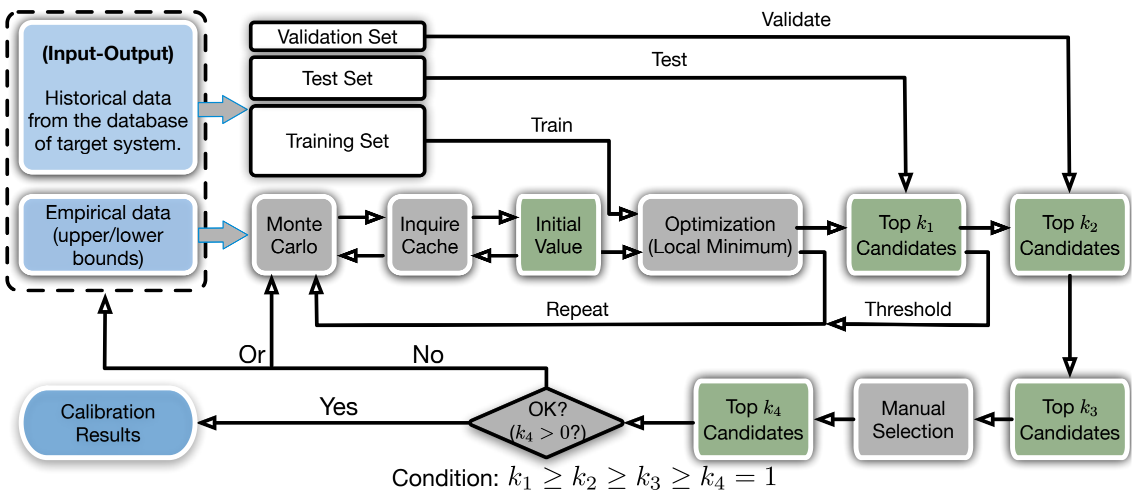


Figure 4-2: The systematic model calibration and validation process. The k_2 and k_3 is what left after applying threshold selection on test and validation datasets separately. The cache checking modular is designed to avoid duplicate optimization from close starting points. The manual selection is designed for experienced ED staff, to eliminate some solutions that could result in good fitness but makes less sense in reality.

It is worth noting that, the proposed method was focused on solving the practical problem. The method is based on the assumption that the calibration result is acceptable with a certain margin of error. That is, the proposed method cannot guarantee finding the theoretical global optimum point, but it can find an acceptable point in a practical application. As shown in Figure 4-2, the reference data (actual input-output pair) was divided into three parts for training, test and validating respectively. The key difference between test and validation processing is the feedback, i.e., simulator performance on validation tests will not affect the calibration process, while performance on test sets will affect the adjustment of parameter set and the

Monte Carlo scheme (e.g., boundary constraints). The Monte Carlo method (under the boundary constraint shown in Table 4.1) is used to generate initial value for the optimization solver. To make the calibration process more automatic, the proposed method will try to find a certain number of local minimum points, then gradually eliminate on test and validation datasets, and only provide several candidates to do manually checking. Thus, the simulation users only need to be involved in the calibration at the end. More specifically, the k_1 local optimum points will be evaluated on testing and validation datasets in sequence, a fitness threshold should be fulfilled and top $k_i (i = 2, 3)$ candidates will be selected, if there are more than one candidate left after evaluating on validation datasets, the top k_3 candidates will be manually checked by experienced ED staff and one which makes the best sense in reality will be chosen as the final solution. If any of $\{k_2, k_3, k_4\}$ is zero (none can pass the threshold or does not make sense in reality), the calibration process will either return to Monte Carlo to search more local minimum or re-divide the historical datasets for training and testing, and start over again. This depends on k_1 and overlap ratio of optimum points (assessed with inter-distance detailed in subsection 4.3.4, as shown in Figure 4-5a).

4.3.2 Evaluation metrics

As an agent-based model, the individual's behaviors, e.g., behaviors of a single patient, are highly dynamic and stochastic, matching these behaviors individually is usually unfeasible and unnecessary. Namely, the similarity between simulator and actual ED should be evaluated in a systematic manner rather than get entangled in each of the agents. Thus, to compare behavior of two complex system, the selection of system key performance indicators (KPIs) is crucial, and two issues must be addressed. On the one hand, the selected KPIs should be able to significantly reflect the macroscopic behavior of the target system. On the other hand, it should be possible to retrieve from historical data, and the historical data should be convincing for the KPIs. Ref. [95] and [96] listed various metrics by which ED operations can be measured. Among which, the LoS (length of stay), LWBS (percentage of patients who leave without being seen), door-to-diagnostic evaluation by a qualified medical professional (arrival time to provider contact time, also known as “door-to-doctor" time) and ambulance diversion (amount of time ambulances are diverted away from the ED) are commonly used. All of those metrics are possible to extract from the agent-based simulator. Given this, and considering that patient-centered records are the real data we have, the records include the time stamp of patients' arrival and discharge, thus the patients' length of stay in the ED could be retrieved. Moreover, the LoS is comprised of all the time on service and waiting/pending. It is one of the composite indicators which is able to indicate patients' flow as well as the system's efficiency. Thus LoS was used as the setting of metrics for system performance in this work.

Furthermore, patients' LoS is one of the aggregate behaviors of the ED system, when comparing simulated LoS with actual LoS, the absolute difference of their average cannot fully represent their differences because the same average may come from quite different distributions (e.g., uniform versus exponential distribution). In view of this, we analyze the actual LoS distribution by using a histogram. For each of the simulation outputs, we perform the same analysis. Thus, we will get two distributions and the goal is to measure the similarity between them, and the similarity will be used to evaluate the similarity between actual system and simulation results.

4.3.3 Fitness function

In view of the above–mentioned facts, a proper method has to be applied to measure the similarity between actual LoS distribution and the simulated one. It is about comparing statistical characteristics of empirical data against emergent behavior of simulation models. In probability theory and statistics, the Jensen–Shannon Divergence (JSD) is a popular method of measuring the similarity between two probability distributions [97, 98, 99]. Considering that patients in ED are classified in five categories (acuity level, also known as emergency severity index) according to their severity. Patients with different acuity levels have different routes and priority in receiving service. Their LoS are quite different on average. Accordingly, it is more reasonable to evaluate patients' LoS separately due to their acuity level. Moreover, the number of patients with different acuity level is quite different, it is about 1 %, 8 %, 32 %, 44 % and 15 % respectively from acuity level 1 to 5. According to the law of large numbers, when sample size is not big enough, the statistical information would be less accurate. Given this, as defined in Equation 4.1, we used a weighted average to calculate the overall fitness with JSD of the 5 categories of patients. Proper weights could be determined by sample size and the standard deviation of actual LoS.

$$f_{fitness} = \sum_{j=1}^{5} W_j D_{JS}^{j} \tag{4.1}$$

$$D_{JS}^{j} = \frac{1}{2} D_{KL}\left(P||Q\right) + \frac{1}{2} D_{KL}\left(Q||P\right) \tag{4.2}$$

Where, D_{JS}^{j} represents the Jensen–Shannon Divergence (JSD) similarity on LoS of patients with acuity level j, W_j is the weights according to patient category (acuity level) and $\sum_{j=1}^{5} W_j = 5$ (there are five patient categories), and D_{KL} denotes the Kullback–Leibler divergence (D_{KL}), which is defined as:

$$D_{KL}\left(P||Q\right) = \sum_{i=1}^{n} P\left(i\right) \log_2 \frac{P\left(i\right)}{Q\left(i\right)}, \quad D_{KL}\left(Q||P\right) = \sum_{i=1}^{n} Q\left(i\right) \log_2 \frac{Q\left(i\right)}{P\left(i\right)} \tag{4.3}$$

Where, $Q(i)$ is the frequency/probability of LoS located in ith interval extract from simulation results, and $P(i)$ denotes the same information extracted from real data. Having shown that, the range of $f_{fitness}$ function value will be 0.0 to 5.0, The lower it is, the closer the difference between simulation and actual will be.

As described in subsection 4.2.2, parameter constraint is defined by boundaries, although each of the parameters is guaranteed to fulfill the boundaries constraint, the combination of parameters may become unreasonable for the model. This case may occur either in the initial value set generated by the Monte Carlo method, or an evaluation scenario requested by the optimization solver. According to our primary experiments, some parameter sets created by optimization algorithm may cause ED saturation, i.e., patients waiting in any of the waiting room increases day-by-day. For example, the number of patients waiting to enter the treatment area is greater than daily arrival. These scenarios cannot result in good fitness because it is not a valid case. Since the complexity of an agent-based model is proportional to the number of agents in the simulation environment, system saturation will result in much longer simulation time. Give this, when the system is saturated, it is better to terminate the simulation evaluation and return the worst fitness evaluation as a penalty.

Furthermore, the patient leaving-without-being-seen (LWBS) is a common phenomenon and a crucial metric to EDs, which has been carefully considered as a possible decision patients may take in the model [83, 84, 85] . As the real data does not include the LWBS records, the final tuned simulator should not have patients who LWBS (equivalent to those patients did not go to ED). However, our primary results showed that some of the parameter set (either generated by Monte Carlo or created by optimization solver) resulted in LWBS. Instead of discarding the evaluations that has LWBS, which may result in lots of failure in optimization and waste lots of computing time, we added LWBS to the objective function as a part of the penalty (i.e., the optimization solver should be allowed to make mistakes). Our final experiments indicated the effectiveness of considering LWBS in fitness function. As shown in Figure 4-4, most of the initial values that lead to LWBS could converge in less than ten iterations. In summary, the final fitness function could be defined as:

$$F_{fitness}\left(P\right) = \begin{cases} f_{fitness}(P) + \lambda R_{lwbs} & (simulation\ succeed) \\ F_{max} & (system\ saturated) \end{cases} \tag{4.4}$$

Where, $P = \{p_1, p_2, \cdots, p_8\}$ denotes a parameter set from the Monte Carlo method or the optimization solver, R_{lwbs} is the ratio of patients leave-without-been-seen (range from 0 to 1.0), λ is an adjustable parameter which represent the weight of LWBS. F_{max} is the maximum penalty to the solver, which is the maximum of $F_{fitness}$ in the first case (simulation succeed). Given this, if we set λ as 5.0, that is to say, the D_{JS} similarity and LWBS have the same weight on the fitness evaluation, the value of $F_{fitness}$ will be between 0 to 10. The lower it is, the closer it will be to actual data.

4.3.4 Optimization method

As described in subsection 4.3.1, the calibration process can be formulated as a series of local minimum searching problems. There are many ready-made methods for searching local minimum value of a given fitness function. However, as explained in section 4.1, different from a pure mathematic problem, the simulation is just such a problem for which it is hard to formulate the relationship between inputs and outputs. Thus the objective function has some special character, e.g, non-convex, non-differentiable, computationally expensive. There are also some optimization methods for finding the minimum of a function of several variables without calculating derivatives. For example, Powell's method [100], which is an algorithm proposed by Michael J. D. Powell for finding a local minimum of a function. The function need not be differentiable, and no derivatives are taken. However, due to the nature of Powell's method, it is almost impossible to parallelize (parallel asynchronous versions [101] have strict condition to objective function). Since each of the fitness function evaluations needs considerable computation time, Powell's method results in very long computation time. According to our tests, it takes around 50 hours to find the closest local minimum point with a given initial value. It is fairly unacceptable for our calibration because it needs to find a considerable number of local minimum points.

Given this, a parallel optimization method is crucial for our requirement. The APPSPACK [39, 40, 41], developed by Sandia National Laboratories, implements an asynchronous parallel pattern search method that has been specifically designed for problems characterized by expensive function evaluations. The framework enables parallel operations using Message Passing Interface (MPI), and allows multiple solvers to run simultaneously and interact to find solution points. While considering our practical requirements and initial experiments, further optimization could still be conducted to speed up the calibration process. Given that the parameters to be calibrated represent the behavior of a practical agent, it is reasonable to assume that slight changes to parameters would not lead to a big difference in outputs. Considering that searching for local minimum is computationally expensive (hours for one process), we cached the initial values by Monte Carlo, as well as the local minimum found by APPSPACK as a pair (initial-optimum pair) to collection $C_p = \{(init, opt)_i\}$, thus when Monte Carlo generates a new set of initial values for finding other local minimum points, we firstly check the distance (d) between the new initial value and each of the initial values in collection C_p (as shown in Figure 4-2, the *Inquire Cache* step). The process is explained as follows:

$$if \quad \exists P^{\circ} \in C_p: \quad d = \sqrt{\sum_{i=1}^{n} |P_i^* - P_i^{\circ}|^2 / n} < \varepsilon \quad then: \quad f(P^*) := f(P^{\circ}) \tag{4.5}$$

Where, P° is the initial value sets of one pair (initial-optimum) in collection C_p. P^* is the new initial value generated by the Monte Carlo method, n is the number of parameters in p_i, and ε is the tolerance. Therefore, as shown in Equation 4.5, if the new initial value set is close to any of the solved pair (overlapped), it will be discarded and call Monte Carlo to generate a new initial set. If there are considerable number of overlapped initial value sets found (searching space is well covered), k_1 in Figure 4-2 should be considered as reduced. This mechanism could avoid some duplicated optimization, especially in small search-space. A similar cache mechanism is also applied for fitness function evaluation (each one takes around 15 minutes), all the scenario (a set of parameter values) to fitness pair (scenario - fitness pair) among all the optimization processes (which start with different initial value) were cached to a collection $C_s = \{(scenario, fitness)_k\}$. Thus, for a new scenario created by APPSPACK, procedure 4.5 (in C_s instead of C_p) is performed before invoking simulation. If the new scenario is close to any of the scenario that has been evaluated before, then the function returns the fitness directly, thus no simulation need to be invoked. Since there are several repetitions for one evaluation process, and generally there are hundreds of evaluations per each optimization, and many independent optimization processes needed for the calibration, this global cache mechanism could save considerable time. The experiments showed that around 10 % of fitness function evaluations were from cached value.

4.3.5 Design of experiment

As illustrated in Figure 4-2, the real dataset was divided into three subsets for training, testing and validating separately. To this end, the 11-month historical data from the ED information system database (Jan. - Dec. 2014, excluding August) has been randomly divided into three parts. More specifically, six months for training (training set), three months for testing (test set), and two months for validation.

Considering that the patients' LoS are statistics on patients who attended the ED. Due to the statistical nature of this model, the sample size should be guaranteed in order to provide reliable LoS. The minimum number of patients for retrieving LoS depends on deviation of LoS, confidence interval as well as margin of error, and could be determined by Chebyshev's inequality [78]. Therefore, multiple runs must be conducted for each scenario in order to reduce stochastic variability and average performance metrics will be used for evaluating the fitness by Equation 4.4. More specifically, the number of simulation replications are determined by deviation of LoS from the real dataset, and the simulation time. Namely, shorter time simulation will require more replications in order to meet the sample size requirements. In this study, according to the statistic characteristics of LoS in real dataset, 4 random seeded runs were performed for each scenario in training dataset, 8 replications were performed for each scenario on testing dataset, and 12 replications for validation dataset.

The calibration was carried out on an 8-node cluster with total number of 512 AMD Opteron$^{\text{TM}}$ Processor 6262 HE cores, and 2TB RAM. All the nodes works in master/worker way, i.e., each one of the node (worker) runs the parallel version of APPSPACK to find the local minimum start from an initial value given by the master. The APPSPACK evaluators, which takes input (the parameter set) and returns fitness, was implemented with Python programming language. In the evaluator, the NetLogo controlling API (comes with NetLogo.jar from released version) was used to invoke and control NetLogo by another Java program running on the Java Virtual Machine. That is, for one fitness evaluation, the Python program will first read the value of variables and invoke several processes (the same as number of repetitions) to evaluate fitness with the same parameter but different random seeds, then each of the processes will call a Java program via system call with parameters as arguments. At the last step, the Java program will initiate the model in NetLogo via NetLogo controlling API and start the simulation. When all the simulations with the same parameter have finished, the program will return to Python, and a post-processing function will be called to analyze the system metrics in order to calculate the fitness value (via Equation 4.4). The data flow of the fitness function optimization with initial value given by Monte Carlo was shown in Figure 4-3.

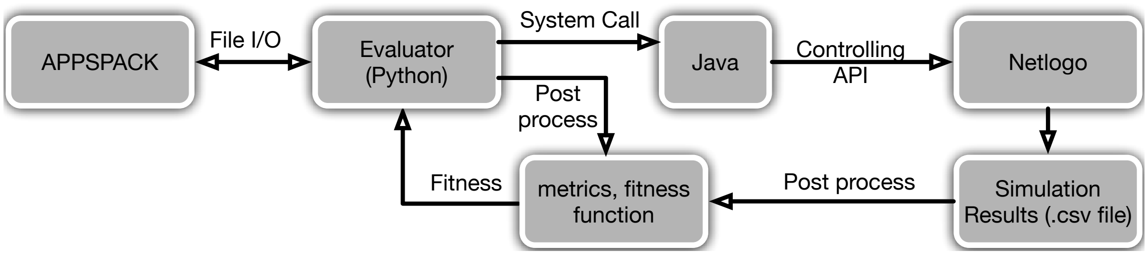


Figure 4-3: Data flow in optimization experiments.

4.3.6 Results and discussion

Use Equation 4.4 as fitness function, the iterations of optimization on training dataset with different initial value is shown in Figure 4-4.

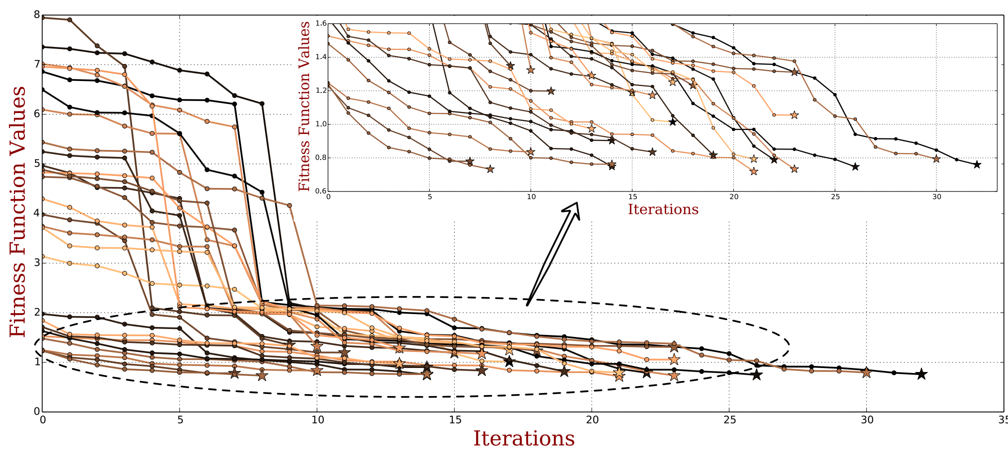
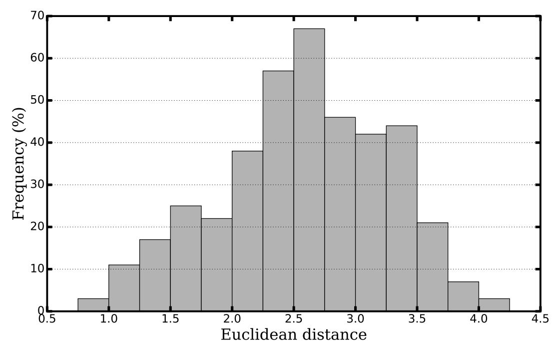
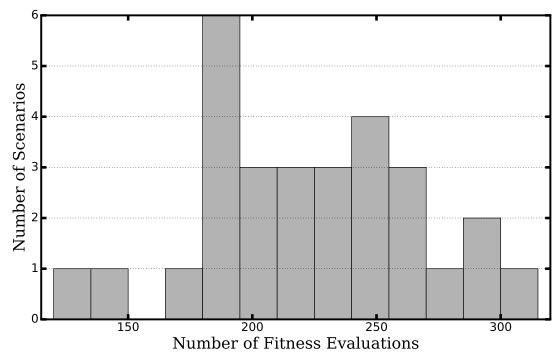


Figure 4-4: Fitness optimization on training dataset with different initial value, fitness values versus iterations. One broken line represents one optimization process with a given starting point from boundary constrained Monte Carlo.

It is clear to see from Figure 4-4 that different initial values resulted in a different number of iterations. Most optimums (the converged fitness values) are in the same level (i.e., no significant global minimum). Some initial values have caused high LWBS, i.e., their initial fitness is greater than 5 (maximum of $f_{fitness}$ part in Equation 4.4 is 5.0), and drop to normal after several iterations. Most optimization processes completed in less than 20 iterations. To analyze the location of local optimum points we found, Figure 4-5a shows the distribution of Euclidean distance between optimums points (there are $k_1\,(k_1-1)\,/2$ distance, where k_1 is the total number of local optimum points found, the same k_1 as it in Figure 4-2).



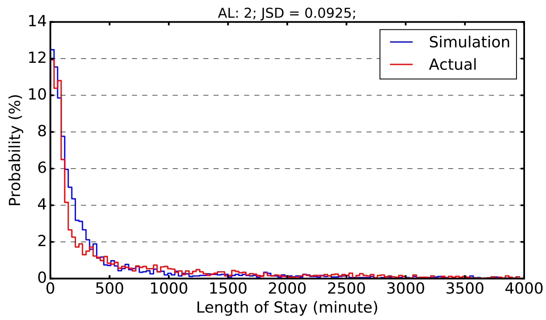
(a) Euclidean distance between optimum points.



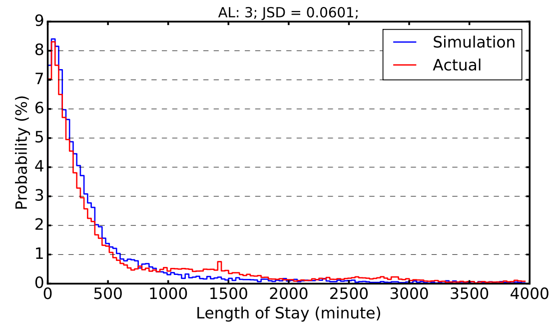
(b) Number of fitness evaluation distribution.

Figure 4-5: Training process analysis. The distribution of the number of fitness evaluations needed in finding local minimum points starting from different initial values, and the distribution analysis of distance between optimal points.

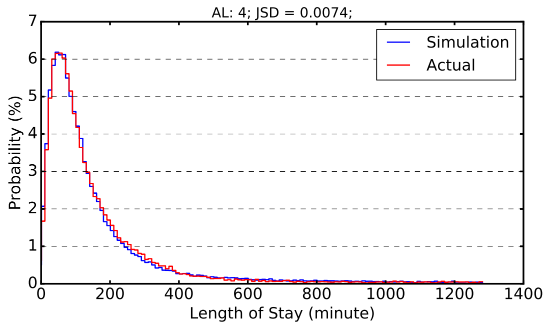
From Figure 4-5a, it is clear that most optimum parameter sets (note that in order to make all the parameter value for APPSPACK on a similar scale, here the parameter values represent the ratio to the typical values in Table 4.1) are far from each other (due to the initial value control by Equation 4.5), while there are some optimums, carried out by different initial values, converged to the same point (distance could not be zero because of the random nature of the simulator and the tolerance setting in APPSPACK). According to the search scheme of APPSPACK [39], each iteration requires many fitness function evaluations in several directions, the number of fitness evaluations has direct influence on optimization time. Figure 4-5b shows the distribution of the number of fitness evaluations. It is worth noting that, in each function evaluation, there are several replications on simulation with different random seeds, i.e., 4, 8, 12 for training, testing and validation separately. In this study, when $k_1 = 30, k_2 = 10, k_3 = 5, k_4 = 1$, the total time taken on the calibration is about 60 hours with the above-mentioned cluster. By following the process illustrated in Figure 4-2, one set of parameter value was selected manually from the k_3 candidates. With the selected parameter set and input (patient arrival) from validation dataset, the comparison (actual data versus simulation) of patients' LoS distribution, classified by patient's acuity level, are illustrated in Figure 4-6. Considering that the validation dataset is composed of two-month's real data and, there are very few patients (less than 1 %, about 160 patients in two months) triaged with acuity level 1, the sample size is not enough for statistical comparison, thus the LoS distribution of patients with acuity level 1 was not shown in Figure 4-6.



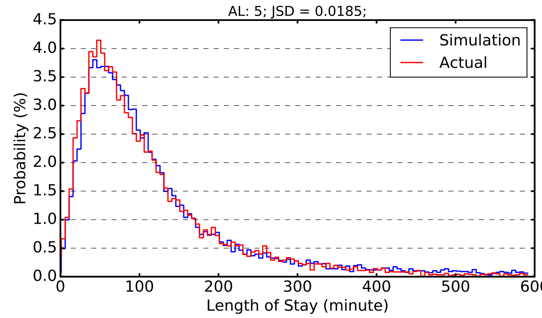
(a) Acuity Level 2



(b) Acuity Level 3



(c) Acuity Level 4



(d) Acuity Level 5

Figure 4-6: The comparison of model prediction results (patient length of stay distribution) on the validation dataset. Results about patients with acuity level 1 is not illustrated here because very few patients (less than 1 %) attend to ED with acuity level 1, the sample size (in two months) is not enough for statistical comparison. The JSD denotes Jensen-Shannon Divergence. Note: the statistical interval widths are: 30 minutes for acuity level 2 and 3; 10 and 5 minutes for acuity level 4 and 5 respectively.

Simulation results in Figure 4-6 demonstrate that the proposed framework is effective to calibrate the model parameters. As a result of the small number of patients attend with acuity level 2, the fitness (the D^j_{JS} in Equation 4.2) of patients with acuity level 2 (Figure 4-6a, $D^2_{JS} = 0.0925$) is not as good as the others. Since calibration process happens only once in the simulation, 60 hours is acceptable and further speedup can be reached via executing on cluster with more computing nodes.

4.4 Discussion

An Emergency Department (ED) is a complex, stochastic environment, which has time-dependent behavior. Advances in computational technology give us the ability to simulate complex models and analyze massive datasets. Given this, simulation has become an effective method to improve policies on operational, tactical and strategic decisions for EDs. However, the difficulty in collecting reliable and complete data can subsequently lead to invalid simulation results. To this end, this chapter proposed a systemic method to calibrate and validate a general model to imitate an actual ED under data scarcity (missing duration of service). Our final results indicated that the proposed approach can find the model parameters accurately within an acceptable time frame. With the parameter value we found, the general agent-based model of EDs can carry out accurate predictions. Although our work was focused on calibrating an ED model, we are confident that the proposed method could also make some contribution to calibrating other computationally expensive simulation models.

There are a number of limitations to our study, including the use of exponential distribution for fitting all the duration of service. Although it was commonly used in the conventional queue theory method, further research should be carried out to consider in more detail about the features of service type. Another limitation is the selection of system KPIs for calculating fitness. In our method, we only considered two indicators, i.e., patients' length of stay and leave-without-being-seen. Although both are commonly used in emergency medicine literature, further indicators such as door-to-doctor time and patients' length of waiting time should be investigated in future improvements. Furthermore, the proposed method has only been tested in one institution though no institution-specific assumption has been made, one of our future studies will apply the method on another ED.

In summary, with a few expectations, the proposed systematic method has been proved to be able to find the parameters for fitting the duration of service, with which the simulated results and the actual data were consistent. The duration of healthcare staff's service time is among the most common missing pieces of information because it is out of the scope of the information system. Moreover, an automatic calibration tool released with a general ED model is promising for promoting the application of simulation in ED studies. This tool will enable the simulation users, e.g., ED manager, to calibrate parameters for their own ED system without the involvement of model developers.

Furthermore, the integration of the ED simulator and optimization techniques (shown in subsection 4.3.4 and subsection 4.3.5) could also be used for systematic performance optimization. For example, with constraints (budget, place or quality of service guarantees) and design parameters, the proposed simulation-based optimization workflow could be used to find the optimal (and suboptimal) design parameters to achieve best system performance.

Chapter 5

Case Study: Decision Support

The best time to plant a tree was 20 years ago. The second best time is now.

- Chinese Proverb

Concerning the capability of an emergency department (ED) simulator, we considered two aspects. On the one hand, a well designed simulation model allows ED managers to answer important "what-if?" questions without making physical changes to the system and potentially putting patients in danger. On the other hand, the simulation model can provide more detailed information about the ED behavior to better understand the root-cause of system performance, i.e., explainability. In this chapter, we demonstrate two case studies on decision supporting by using the ED simulator to deal with ED system overcrowding. The first overcrowding problem is caused by the increasing patient arrival, e.g., flu pandemic. In the second case study, we tried to solve an overcrowding problem from investigating the effects of a connection with ED, i.e., ambulance service for discharging. The explainability of the simulation model will be discussed in chapter 6 with two case studies.

5.1 Experimental Condition

In this chapter, some experiments were designed to show the capability of an ED simulator of which information can provide. We simulate one scenario for 720 hours (about one month), the simulation was allowed a warm-up period of 168 hours (discussed in subsection 3.5.3), then observations were made during the following 720 hours. To make the results statistically reliable, this study require the execution of a big amount of parametric simulations (same model, different parameter value configuration), a 10-node cluster was used for model execution and, the execution task was assigned as one core for one scenario. The scenario organization and the execution framework were described in section 3.5.

5.2 ED Resources Configuration

As a service provider, the ED is characterized by resource (human and equipment) configuration. Table 5.1 lists all the value of the parameters for characterizing the resource of an ED to be simulated in this chapter. It is the current configuration of the ED, that will be used as baseline of the proposed changes to the system to solve overcrowding problem.

Table 5.1: Quantitative representation of the simulated emergency department. Annotation: n represents number of items

Label	**Interpretation**	n	**Label**	**Interpretation**	n
jA	junior admission	3	sD_A	senior doctor in area A	4
sA	senior admission	2	jD_B	junior doctor in area B	2
jT	junior triage nurse	3	sD_B	senior doctor in area B	5
sT	senior triage nurse	2	jD_A	junior doctor in area A	2
jN_A	junior nurse in area A	5	Tr_{in}	internal test room	6
sN_A	senior nurse in area A	5	Tr_{lab}	laboratory test room	4
jN_B	junior nurse in area B	4	Cb	carebox of area A	60
sN_B	senior nurse in area B	4	C_B	chair of area B(capacity)	60
$Auxi$	auxiliary nursing staff	3			

5.3 Experimental Input

In subsection 3.1.1 we described a data-driven patient arrival model in ED. In the first case study simulations , we will keep the distribution of patient's characters as the same as shown in subsection 3.1.1, whereas increase $N_{ar}[week]$ (via adding daily arrival) in an interval as a case study of flu pandemic to study ED's tolerance/robustness. In the second case study, the complete model shown in subsection 3.1.1 will be used.

5.4 Case study of decision support to deal with steady increase patients

As for the capability of a simulator for supporting decision making, from the point view of an ED operations manager, this subsection provides a case study to solve an overcrowding problem arises when the number of arrival patients continues to increase (say a flu in its coverage area). The original ED resources configuration was shown in Table 5.1, i.e., baseline. By using the simulator, one possible process for an ED manager to make decisions to deal with increasing patient arrival is shown in Figure 5-1. More specifically, based on some prediction (e.g., number of patients will attend in the next couple of days), the manager can quantitively predict the ED's performance via simulation. Then, Key Performance Indicators (KPI) represent systemic behavior could be predicted and system bottleneck could be indicated. With this bottleneck information, the manager can propose several alternative decision rules, and the manager can verify and evaluate these solutions via simulation. Through this way, the decision makers can quantitatively find the bottleneck and know the benefit and cost of a proposal solution without the commitment of any physical resources or interruption of the system.

According to the operation data from Hospital Universitari Parc Taulí, the normal arrival patient is about 329 per day ($N_{ar}[week] = 2,303$). Here we designed some experiments to investigate the system behavior. Specifically, we keep patient arrival

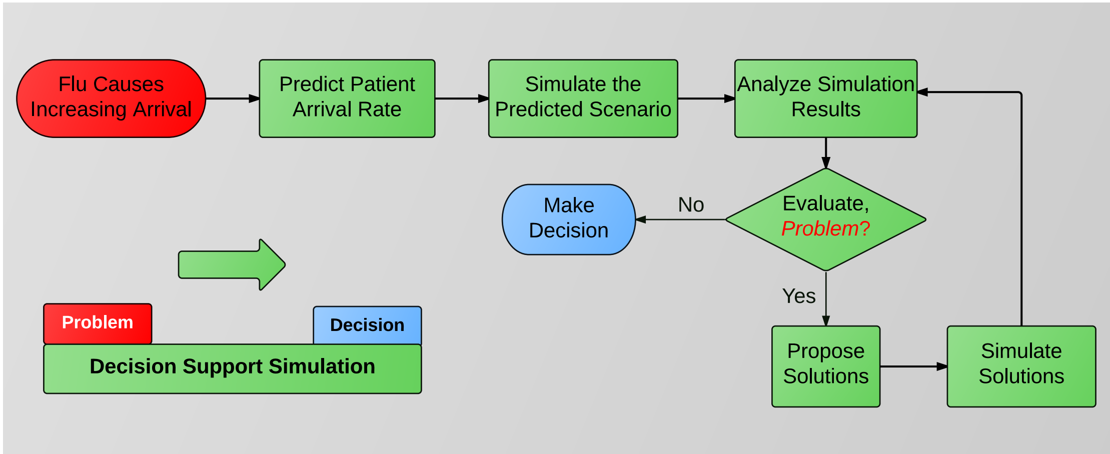


Figure 5-1: Simulation supported decision making process, quantify the cost and benefit of proposal without a real deployment.

model the same as it shown in subsection 3.1.1 whereas increase daily arrival number. The patients' average LoS and some key state information of the ED was shown in Table 5.2. It is worth to note that, we used resource occupancy/utilization (defined as the percentage of time that an agent actually spend on providing service) as indicator to explore the system's bottleneck, Table 5.2 - Table 5.4 only list the potential bottleneck. Theoretically speaking, behavior of any system components could be retrieved from the simulation data because the model was built from the individual level and simulates all the interactions.

Table 5.2: LoS and ED resources utilization with increasing daily arrival patient

Arrival	Average LoS by acuity level(hour)					ED resources utilization(%)				
	1	2	3	4	5	Tr_{lab}	N_A	D_A	D_B	N_B
361	10.83	10.30	9.79	3.01	2.81	70.51	40.57	67.94	53.95	43.68
397	10.84	10.90	10.41	3.43	3.81	81.39	46.31	78.29	62.05	50.27
416	11.66	11.28	10.69	3.59	4.12	83.64	48.01	80.59	64.23	52.16
436	11.87	11.73	11.31	3.78	5.28	86.75	50.01	84.50	66.84	54.17
456	11.71	12.09	11.85	3.98	**8.94**	**91.32**	51.85	87.19	69.80	56.27

It can be seen from Table 5.2 that, with an increasing number of arrival patients, some of the resource utilization increases dramatically. When daily arrival patients increases to 456 (140 % of normal) per day, the ED becomes saturated and faces serious overcrowding problem in area B. The LoS of patient increases a lot, especially patients with acuity level 5 in area B (more than three times to normal). The resource utilization could clearly show the bottleneck. By exploring the resource occupancy, it is clear that the laboratory test service becomes the bottleneck with the increasing number of patients arrival at ED. This is because the laboratory test service is saturated and patients with acuity level 5 have the lowest priority to get the resources. In this case, one of the most straightforward solution is to add laboratory test service capability. Here we assume that the ED manager proposes a solution to add two more technicians to the laboratory test room (i.e., a new scenario with $T_{r_{lab}} = 6$). To test and evaluate this proposal, we simulate this new resource configuration by modifying the ED resources configuration in Table 5.1, the result was shown in Table 5.3 (the first row).

Table 5.3: LoS and ED resources utilization with two more laboratory technicians.

Arrival	Average LoS by acuity level(hour)					ED resources utilization(%)				
	1	2	3	4	5	Tr_{lab}	N_A	D_A	D_B	N_B
456	11.58	11.90	11.70	3.65	**3.17**	**60.67**	51.99	87.19	69.47	56.65
476	12.54	12.70	14.33	3.80	3.57	64.19	55.04	92.30	73.01	59.42
496	13.23	12.90	**33.93**	4.02	4.16	66.37	56.90	**96.06**	76.32	62.25

From Table 5.3, we can see that after adding two more technicians to the laboratory test room, the laboratory utilization decreased dramatically and LoS of type 5 patient reduced to normal. But when the patient arrival rate keep rising, say 150% times of normal rate, patients with acuity level 3 in area A were seriously saturated because they have the lowest priority to get access to resources. Through resource utilization information, it is easy to find the bottleneck, i.e., the doctors in area A were saturated. Having said that, one of the straightforward solutions should be to assign more doctors to area A. Then the decision maker could simulate this proposal solution (e.g., $T_{r_{lab}} = 6$, $sD_A = 5$, and $jD_A = 3$) to evaluate its effect. The simulation results were shown in Table 5.4.

Through simulating the proposal solution, we can see that the solution can solve the overcrowding problem a certain degree. Moreover, as shown in Table 5.4, if the patient arrival keeps increasing, the doctor in area B may become the next latent

Table 5.4: LoS and ED resources utilization with two more doctors added to area A

Arrival	Average LoS by acuity level(hour)					ED resources utilization(%)				
	1	2	3	4	5	Tr_{lab}	N_A	D_A	D_B	N_B
496	10.89	11.01	**11.07**	3.98	4.15	66.73	57.50	**71.84**	75.79	61.58
516	11.12	10.86	11.20	4.13	4.79	68.75	58.67	72.99	78.80	64.30
535	11.26	11.31	12.54	4.36	5.82	71.39	60.65	76.00	82.52	67.14

bottleneck.

In summary, with an ED simulator, operations managers could clearly identify the system bottleneck and quantify the cost and benefit their proposal without the commitment of any physical resources or interruption of the system

5.5 Case Study of the Influence of the Response Time of Ambulance Service

5.5.1 Ambulance for departure

There are four typical destinations when a patients is discharged: (1) go home, (2) to hospital, (3) transferred to another hospital or (4) may even die. In reality, there are some patients who need an ambulance to leave ED, especially those patients in area A who will go home or be transferred to another hospital. The ambulance is provided by a service center, it is common that there is a delay from requesting until it becomes available. So, the problem is the response time of the ambulance service because the patients keep occupying the physical resource of ED during the waiting time (also known as discharge pending). This may cause or worsen overcrowding because the extra waiting time was caused by other factors (neither ED nor patient). Therefore, not only the service efficiency in ED, but also the way of departure can affect the ED throughput (defined as the rate of patients leaving ED after all the care service). Having said that, it is necessary to include the model of ambulance response time as part of ED model to analyze the degree of the impact on an ED.

Considering that the ambulance response time depends on lots of uncertainties and it is possible to obtain enough samples from the ED information system. We used a data-driven model to fit the behavior of ambulance's response. After analyzing the response time recorded from Hospital Universitari Parc Taulí, we find that the response time is not a constant time and it has significant distribution shape. One of the commonly used methods for modeling a source of randomness when relative data are available is to fit a theoretical probability distribution to the data [102]. In order to choose a proper empirical distribution, a histogram of the data is one of the most useful tools for determining the shape of the underlying density function. Because the histogram is an essential graphical estimate of the density. After a histogram analysis, some estimation methods (e.g., maximum likelihood estimation, method of moments) can be used to estimate the parameters of the selected distribution.

More specifically, we analyzed the real data of 2011 from the Hospital Universitari Parc Taulí, following the tutorial of Law, A. in Ref. [102], the summary statistics information was shown in Table 5.5. According to this information, we chose gamma distribution $X \sim \Gamma(k, \theta)$ to fit the real data, and Maximum Likelihood Estimation (MLE) was used to estimate the shape parameter k and scale parameter θ. The estimated value of k was 1.6740, and θ was 37.9326. Density-histogram plot of the fitted gamma distribution and the real data was shown in Figure 5-2. Similar to the patient arrival model, the simulation users could define their own model or modify the parameter of the existing model to satisfy their own requirement.

Table 5.5: Summary statistics of the real ambulance's response time.

Summary statistic	**Value**
Mean	63.5
Median	49
Variance	2629
Coefficient of Variation	0.81
Skewness	1.56

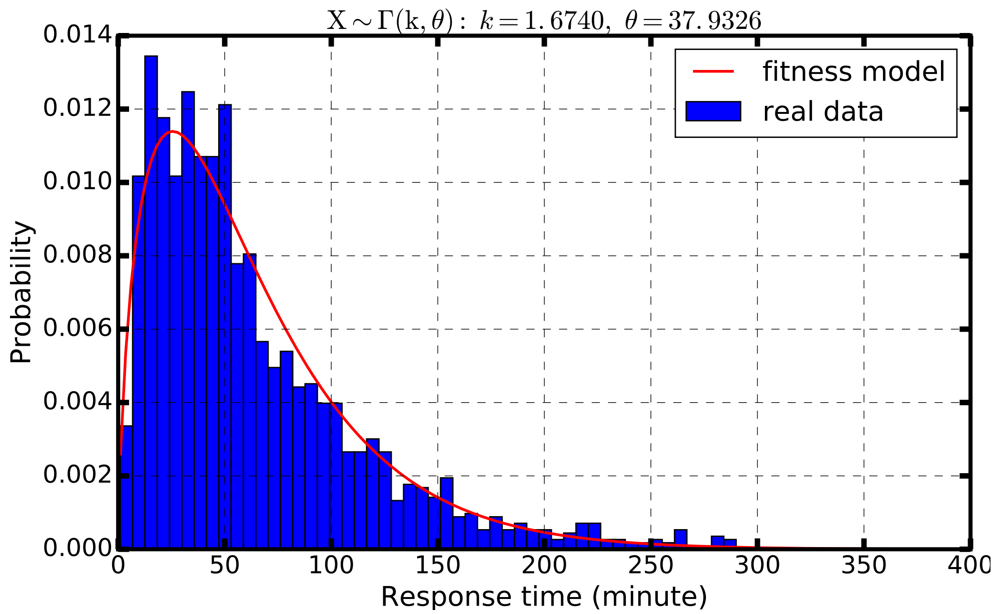


Figure 5-2: Density-histogram plot for the fitted gamma distribution and the real data.

5.5.2 Simulation

As shown in subsection 5.5.1, when a patient asked the ambulance service for departure, the patient continue to use the treatment place during the ambulance's response time period. Moving on from here, the ambulance's response time has some influence on the behavior of ED, for example, causing overcrowding or worsting overcrowding. The LoS of their following patients (patient who is awaiting a free bed) will also be increased because they have to wait longer for a free treatment place. To quantitatively analyze the influence, we simulate two more scenarios based on the saturated scenario in Table 5.3. In this saturated scenario, when daily arrival increased to 496 (150% of normal value), the ED meets serious overcrowding even with two more technicians have been added to the laboratory test room. Here, we simulate and consider the influence of ambulance response time for departure. In the first scenario, we use half of the actual ambulance response time described in subsection 5.5.1. In the second scenario, we assume the response time is zero (the ambulance service always becomes available immediately after requesting). Then, we simulated to see if the response time of ambulance can alleviate the overcrowding problem. The simulation results, i.e., patients' LoS under these two scenarios are shown in Table 5.6. It it clear that the overcrowding problem has been lessened with decrease of the ambulances' response time.

Table 5.6: Influence of ambulance response time to LoS.

Ambulance response time model	**Average LoS by acuity level(hour)**				
	1	2	3	4	5
current actual delay(*mean*=63 minutes)	13.23	12.90	**33.93**	4.02	4.16
50% of actual delay(*mean*=31 minutes)	12.70	12.60	**17.96**	3.94	4.03
without delay	12.04	12.51	**15.53**	3.86	3.86

In Table 5.6, we can see how the response time of ambulance affects the ED behavior (LoS as indicator). The effects are due to the fact that the patient continue to occupy the treatment place (carebox in area A or chair in area B) during the waiting period. The overcrowding problem, especially for patients with level 3, has been lessened when ambulance response time decreases to half of normal. Compare with totally no waiting (i.e., the ambulance service are always available to use), the improvement of cutting response time to the half of current situation is much more significant. Therefore, if constrained by budget, partially improve the ambulance response time still has significant benefit on solving ED's overcrowding problem. From the point of view of ED operation managers, this may provide alternative ways to optimize changes to the system with budget constraints.

5.6 Discussion

This chapter presents two case studies to show two of the possible uses of an ED simulator. The first case study is about dealing with the increasing patient arrival caused overcrowding problem, and the second is a quantitative analysis of the influence of ambulance (for departure) response time over the ED behavior.

We can see from the two case studies that the flexibility and adaptability features of this model provide a way for ED simulation users (e.g., operations managers) to accommodate different scenarios without significant modification of the underlying model. It enables the simulation researchers to focus their effort on the understanding of ED behavior rather than developing a theoretical model each time. Due to the massive computational resources and big-data processing capacity provided by high performance computing techniques, the simulator could execute a large number of simulations in aa acceptable period of time. Then, applying data mining techniques on such model output data allows the decision makers to gain new insights into the complexity of the interrelated variables and, quantify the effect of changes on the overall performance of the ED.

Chapter 6

Case Study: Discover Macro-Level Features From Micro-Level Behaviors

Precise knowledge of self and precise knowledge of the threat leads to victory.

\- Sunzi's Art of War

6.1 Knowledge discovery

Means and methods to obtain knowledge about the inherent uncertainties and complexities of a system to support learning, problem solving, decision making, and policy formulation have attracted a great deal of research attention. In the analysis of complex system, the singularity is defined as a point near which the system exhibits extreme behavior. As demonstrated in Figure 6-1, the system behaves normally (linearly) to variables before some values, after that it behaves quite differently. That is to say, according considerable amount of actual system behavior records, it is reasonable to make some extrapolation. However, extrapolation may fail after a point (i.e., singularity). The singularity is normally difficult to predict, and the system behavior after the singularity also is very difficult to quantify with an analytical model or data-driven model. For a critical system like emergency department (ED), the estimation error may lead to wrong decision and further results in serious operations management problem.

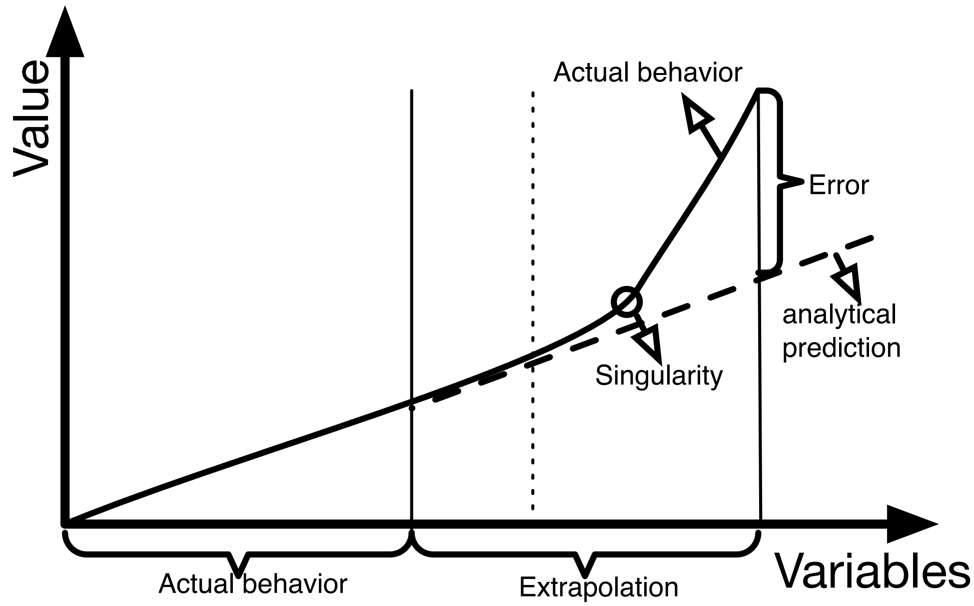


Figure 6-1: A demonstration of singularity.

Like a first-principle kinematic model of a physical system, one benefit of simulating the micro-level behavior (i.e., interactions between agents in this study) is that it provides a potential to know the root cause of macro-level features. That is, for an existing macroscopic phenomenon, it is possible to trace back to the micro-level behavior that accounts for it. Since an agent-based ED model simulates the individual behavior of system components and their interactions directly at individual level, if fine tuned, it is capable to quantify the system behavior after the singularity (extrapolation).

For a complex system like an ED, the macro-level features of the system emerge from the corresponding micro-level interactions. However, the interaction data generated from the agent simulator is massive and unreadable without being analyzed. To meet the massive simulation data generated by micro-level simulator as well as constantly changing requirement, we designed a layer architecture to simulate, monitor and discover knowledge for full insight into the complex system (see Figure 6-2).

As shown in Figure 6-2, the core of the knowledge discovery system is an agent-

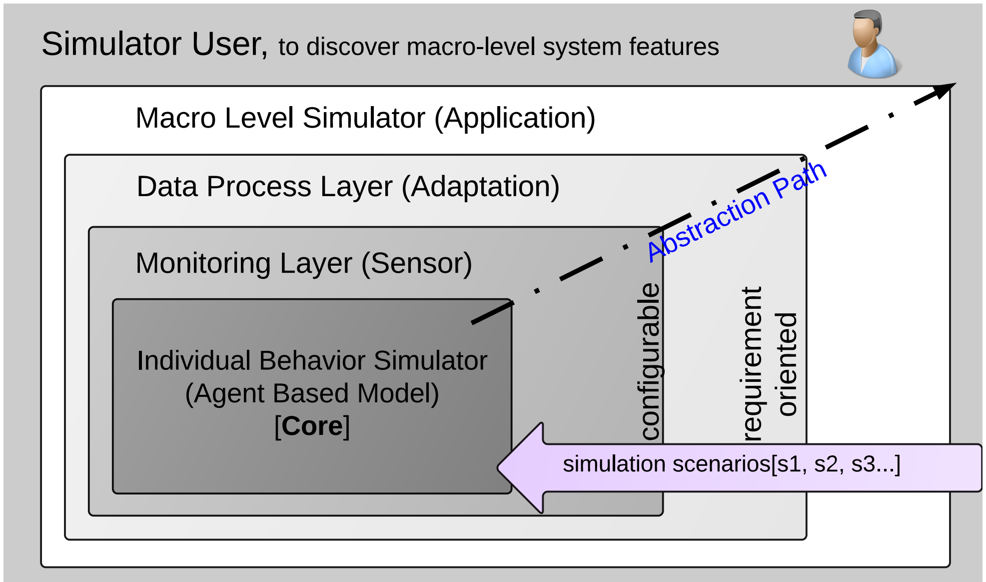


Figure 6-2: Layer architecture of application framework for knowledge discovering from micro-level behavior simulator. The micro-level simulator generates interaction information among the system components, configurable monitoring layer records all the needed interaction information and state information in a given format for the upper processing layer.

based micro-level behavior simulator. It can provide detailed interaction information among the smallest components of the ED as well as state information of the simulated environment. This information is the source of knowledge to understand behavior of the entire system. However, not all the data is required for specific analysis, the monitoring layer is designed to provide the flexibility on micro-level data collecting and processing (detailed in subsection 3.4.2).

Moreover, the simulation scenario is defined as a set of parameters for characterizing the agent-based model and environment (as shown in section 3.5). Therefore, from the perspective of the simulator users, the whole system is a macro-level features simulator because what the users will get is the macro-level information extracted from the micro-level data. However, different as a macro-level simulator, the simulation scenarios are directly designed in micro level without abstraction. This feature could simplify the use of simulation (straightforward to users). Moreover, it also provides better adaptivity, and capable with identify the root-cause (i.e., explain why the system behaves like predictions), i.e., transparent prediction.

6.2 Case studies

Decision making in the field of healthcare service management assessment is not a simple task and it is important for different stakeholders. For example, patients are expecting efficient services, insurers are aiming for cost-effectiveness and the health industry is primarily interested in yield maximization. Understanding the complexity of such a system requires more than experience and intuition alone. In this chapter, we will demonstrate two case studies. The first one (subsection 6.2.1) is about ED resource configuration. The second one (subsection 6.2.2) is about the influence of micro-level behavior on macro-level functionality. The base configuration of the simulated ED, such as number of doctors, nurses as well as average attention time are specified in Table 6.1.

Table 6.1: Configuration of the emergency department (environment) and individual behavior model.

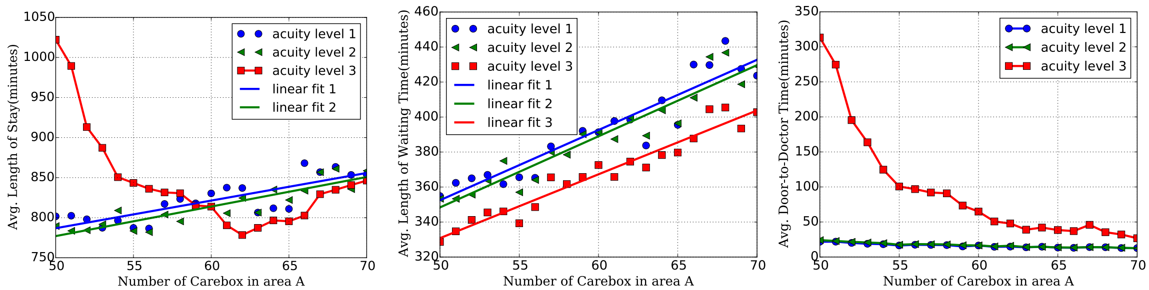
Resource	**Number**		**Avg. Service Time (ST, minute)**		**ST Dist.**
	day	night	first-interaction	follow-up	
jA	3	2	5	-	Gamma
sA	2	0	3	-	Gamma
jTN	3	1	8	-	Gamma
sTN	2	1	6	-	Gamma
jD_A	2		20	15	exponential
sD_A	4		15	13	exponential
jN_A	5		25	18	exponential
sN_A	5		20	14	exponential
jD_B	2		8	7	exponential
sD_B	5		6	5	exponential
jN_B	4		11	7	exponential
sN_B	4		7	5	exponential
Tr_{img}	5	2	45	-	Beta
Tr_{lab}	4	2	30	-	Beta
A_{CB}	50		-		-
B_{chair}	60		-		-
$Auxi$	3		15		exponential

In these case studies, as the staff work shifts, we consider that the medical test technicians, admission staff and triage nurse group run on two shifts and the number of staff is different during the day (6:30 - 18:30) and night (18:30 - 6:30) because the patients arrival rates are quite different. The rest of the staff work on one shift by turns. We simulate one scenario for 1608 hours, to avoid initialization bias, the simulation allowed a warm-up period of 168 hours, then monitoring was carried out during the following 1440 hours.

6.2.1 Influence of Capacity in Area A

ED overcrowding is defined as a situation where the demand for emergency services exceeds the ability of an ED to provide quality care within appropriate time frames. By observing more than 20 million patient visits to EDs over five years, Ref. [103] determined that the risk of death and hospital readmission increases with the degree of crowding in the ED. When an ED meets an overcrowding problem, usually there are many patients waiting in the waiting room or even receiving attention in a corridor. From intuition, when there are many patients waiting to enter the treatment zone, one of the most straightforward way to solve this problem is by expanding the capacity. In this case study, a cross-scenario analysis was used to discover the influence of additional careboxes in area A with the goal of solving the overcrowding problem. Regarding this requirement, "sensor" to monitor patient's behavior and carebox utilization are enabled via the user interactive application shown in Figure 3-6.

As one of an important KPIs of an ED, the patient's length of stay (LoS) is the time when a patient arrives at the ED to the time they depart from the ED. From the point view of the patient's state, the LoS consists of two parts: the total length of waiting time (LoW, total length of time on waiting for services) and the length of attention time (LoAt, total length of time on interacting), i.e., $LoS = LoW + LoAt$. Since the arrival patients keep the same in this case study and we assumed that the service a patient needs is determined entirely by properties of patient, i.e., the average $LoAt$ will keep no change among scenarios, the LoS differences among different scenarios are the length of waiting (LoW) time. The influence of area A capacity (carebox number) from the point of view of LoS is visualized in Figure 6-3. Due to the randomness, slightly small changes may somewhat result in inconsistent change, linear fit was used to demonstrate the trend in Figure 6-3a and Figure 6-3b.



(a) Length of stay. (b) Length of waiting time. (c) Door-to-doctor time.

Figure 6-3: The influence of additional carebox on patients' behavior. (note: the scale of vertical coordinates are different.)

In this case study, the start point is an overcrowding scenario, i.e., when the number of careboxes is 50, the studied ED is facing with overcrowding. Since patients with acuity 3 have the lowest priority to be assigned a free carebox in area A, they will be delayed first. From Figure 6-3a, it is clear to see that additional careboxes provides good results in patients with acuity level 3, the overcrowding problem is solved. However, different from what we would expect, patients with higher acuity meet bad influence because their LoS increased. As shown in Figure 6-3b, all the patients met increasingly longer waiting times for service with additional carebox. Therefore, the root cause is: after adding more careboxes, as the number of corresponding nurses and doctors did not increase accordingly, they cannot provide service to patient as instantly as before, so the patients need to wait more for their doctor and nurse (Figure 6-3b), which finally results in the increased LoS.

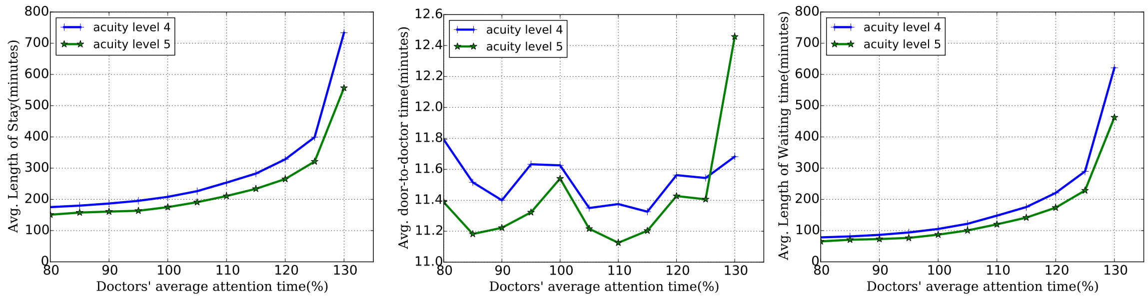
Additionally, a resource with a high occupation rate will be more sensitive to fluctuations in its arrival process than a resource with a lower occupation rate. Tracing back to the micro-level indicator by single scenario analysis, the average occupancy of doctors in area A (percentage of scheduled time spent on patient related activities) is 89.9%, and average occupancy of nurses is 92.3%. However, as shown in Figure 6-3c, one benefit of additional careboxes is the reduced length of "door-to-doctor" time (i.e., the number of minutes from patient arrival until seeing a doctor). That is to say, the patients can enter the treatment zone earlier and they may feel happier than waiting helplessly in the waiting room. Furthermore, as shown in Figure 6-3a, an additional 12 careboxes (i.e., 62 in total) in the ED may be a good choice for current staff configuration if there is no cost constraint, because the patients with acuity level 3 (about 30% of arrival) meet with the shortest LoS. Further studies could be done to find the tradeoff between patient satisfaction and cost constraint.

6.2.2 Behavior of Doctor in Area B

Identifying the primary causes of overcrowding in an ED is a critical step in knowing how to increase throughput. In this case study, the quantitive association between a doctors' micro-level behavior and macro-level patients' average LoS was discovered via cross-scenario analysis.

Attention time, also known as service time, is the length of time for one interaction. The length is determined by service provider drawn from an exponential distribution. Taking the interaction among doctor and patient as an example, it is different in terms of doctors' expertise, the patients' condition (patient's acuity level, age) and interaction times (fist interaction or follow-up). The initial average value for the exponential distribution was shown in Table 5.1. As we assumed that the relationship between the patients and doctors is always static, only when doctors change shifts do they detach their patients to other doctors on the next shift. In addition, according to empirical data, the first interaction with a patient always takes longer. Thus, there are two values for the attention time model of a doctor and this will result in four for the group of doctors (consisting of junior and senior) in area B. Therefore, to study the overall influence of doctor group in area B, in scenarios' design step, we change the average length of attention time of doctor in area B by percentage independently. As micro-level data monitor configuration, only a "sensor" for recording patients' behavior was enabled through the separate application shown in Figure 3-6. The effects of doctors' behavior on LoS and "door-to-doctor" time are illustrated in Figure 6-4.

Figure 6-4a clearly shows the significant impact of doctors' behavior on systemic functionality. With the increasing length of doctors' attention time (e.g., working with lower efficiency or more carefully diagnose), average patient LoS increased dra-



(a) Length of Stay. (b) Door-to-doctor time. (c) Length of waiting time.

Figure 6-4: The effect of length of doctors' attention time on macro-level LoS and the root cause identification. The horizontal axis is the percentage against the normal configuration show in Table 5.1. (note: the vertical coordinate scale of (b) is quite different as (a) and (c).)

matically in area B. After analyzing the "door-to-dctor" time (Figure 6-4b), we find that the patients enter the treatment area after a very short waiting time, which means that the increased LoS is not because of "door-to-doctor" time. Thus, the patient spent most of their LoS in the treatment area. Moving on from here, we analyzed the length of waiting time in treatment phrase (Figure 6-4c). It is clear that waiting for doctors' attention is the root cause of the increasing LoS. Furthermore, Figure 6-4a also provides the singularity of this micro-to-macro association, that is to say, if doctor's attention increases more than 125%, patients' LoS will increase very fast. This information is useful for managers to avoid mistake in intuitive thinking. In summary, this case study quantifies the effects of micro-level behavior on macro-level LoS, further study can be done to balance the quality of service and efficiency of ED system under specific situations.

6.3 Discussion

Due to the massive computational resources and big-data processing capacity provided by high performance computing techniques, the simulator could execute a large number of simulation scenarios in an acceptable time period. Then, applying data mining techniques on such model output data allows decision makers to gain new insights into the complexity of the interrelated variables and the effect of changes on the systematic performance of the ED.

This chapter presents an approach to discover knowledge of emergency department through simulating individual behavior of its components (discover knowledge via "playing" with the simulator). The behavior simulation model (described in chapter 3) can generate interaction information under various configuration scenarios. Analyzing this interaction information thoroughly enables knowledge discovery towards a better understanding of the complex systemic behavior. This makes it possible to explore association between micro-level behaviors of individuals and macro-level patterns that emerge from their interactions, thus assisting users to better understand a system's behavior under various conditions. Additionally, a layer-based architecture was used to achieve flexibility and configurability. This proposed framework can be used to promote learning, hypothesis testing, decision making support, and policy formulation after being properly validated, offering the user and organization the ability to understand the complexity of healthcare systems and to facilitate the redesign of optimal outcomes.

Chapter 7

Conclusion and Future Work

It seems that perfection is attained not when there is nothing more to add, but when there is nothing more to remove.

- Antoine de Saint Exupéry

7.1 Conclusion

Hospital based emergency departments (EDs) are highly integrated service units to primarily handle the needs of the patients arriving without prior appointment, and with uncertain conditions. Due to the complexity of the ED system and uncertainties to the ED system, efficient management becomes a big challenge. Incorrect decisions may lead to serious consequences on the quality of service and cause unnecessary deaths.

Prediction, explanation & optimization are challenging for a complex system like emergency department. A precise ED simulator enables managers to make better decisions by letting them see the impact of changes before implementing them. This article presented an agent-based model of EDs which could be used to settle problems such as prolonged waiting times, inefficient use of ED resources, and unbalanced staff scheduling. The model was built from bottom up, i.e., the systemic behaviors were emerged from the simulation of all system components. More specifically in this model, policies such as staffing, human factors such as sanitary staff behavior, new cases such as a flu outbreak could be set up and their effects on system performance such as waiting time and throughput could be quantified. With the amount of adjustable parameters, the simulator is customizable to simulate a variety of scenarios. The presented simulator is currently working as a platform to study Methicillin-resistant Staphylococcus Aureus (MRSA) transmission in EDs and as an experimental platform of EDs to provide data under various scenarios for knowledge discovery.

To achieve high fidelity and credibility in conducting prediction, explanation and exploration of the actual system with simulation models, a rigorous calibration and validation procedure should firstly be applied. However, one of the key issues in calibration is the acquisition of valid source information from the target system. This thesis also developed a simulation-and-optimization based systematic method to automatically calibrate a general emergency department model with incomplete data. The proposed calibration method enables simulation users to calibrate the general model for simulating their system without the involvement of model developers. We believe that it is promising for promoting the application of simulation in ED-related studies. In addition, the integration of the ED simulator and optimization techniques originally developed for model parameters calibration could also be used for systematic performance optimization, i.e., by changing the objective function and variable constraints. For example, with constraints (budget, place or quality of service guarantees) and design parameters, the proposed simulation-based optimization workflow could be used to find the optimal (and suboptimal) design parameters to achieve best system performance.

In summary, starting from simulating the EDs, our efforts proved the feasibility and ideality of using an agent based modeling & simulation techniques to study healthcare systems. The cross-validation results showed that the developed ED simulator can accurately represent the emergent behavior of the complex ED system. Some demo applications proved that the simulator is ready to work as part of decision support system.

Although this work was focused on the ED, the model methods and framework developed in this thesis could most likely be applied to many other healthcare entities, such as an intensive care unit, a comprehensive full service hospital, a managed care organization, or a vertically integrated system of primary care providers, and outpatient services.

7.2 Future Research Directions

- Do statistical sensitivity analysis of the variables of the emergency department simulator. The variable sensitivity information could be used to build a knowledge base or a metamodel (model of a model) of EDs.

- Connect the emergency department simulator with the hospital to study disordered system behavior based on the integration of first-principles model and data-driven model (with real operation data).

- Calibrate the general model for all emergency departments in a regional area and, connect these simulators together for short-term (several hours) occupancy prediction. Then, a load balancing scheme of the incoming patients could be designed based on these future occupancy predictions.

- The framework developed in our work could be used to build a full model of integrated care system. A full model of integrated care system will be able to represent a comprehensive tool to quantitatively evaluate prospective planned changes to the integrated care system for decision making, and open a wide field of possible simulation scenarios for a better understanding of the integrated care complex system.

7.3 List of Publications

The research presented in this thesis has been published in the following papers:

1. **Zhengchun Liu**, Francisco Epelde, Dolores Rexachs and Emilio Luque. A Bottom-up Simulation Method to Quantitatively Predict Integrated Care System Performance. Proceedings of the the 16th International Conference for Integrated Care [104].

2. **Zhengchun Liu**, Eduardo Cabrera, Dolores Rexachs, Francisco Epelde, and Emilio Luque. Simulating the Micro-level Behavior of Emergency Department for Macro-level Features Prediction. Proceedings of the 2015 Winter Simulation Conference [37].

3. **Zhengchun Liu**, Eduardo Cabrera, Manel Taboada, Francisco Epelde, Dolores Rexachs and Emilio Luque. Quantitative Evaluation of Decision Effects in the Management of Emergency Department Problems. Proceedings of the 2015 International Conference on Computational Science [36].

4. **Zhengchun Liu**, Eduardo Cabrera, Dolores Rexachs and Emilio Luque. A Generalized Agent-Based Model to Simulate Emergency Departments. Proceedings of the 6th International Conference on Advances in System Simulation [38].

5. **Zhengchun Liu**, Eduardo Cabrera, Dolores Rexachs and Emilio Luque. Study of Emergency Department by Using High Performance Computing. XXV Jornadas de Paralelismo. September 16-18, 2014. (*national conference*)

6. **Zhengchun Liu**, Emilio Luque, Dolores Rexachs, Francisco Epelde, Eduardo Cabrera, Manel Taboada. A Simulator of Emergency Departments for Decision Support and QoS Improving. V CONFERENCE: R+D+I Research and Development in ICT and Health. Girona, Spain, June, 2014. (*poster*)

Bibliography

[1] J. Hurwitz, J. Lee, K. Lopiano, S. McKinley, J. Keesling, J. Tyndall, A flexible simulation platform to quantify and manage emergency department crowding, BMC Medical Informatics and Decision Making 14 (1) (2014) 1–11. doi:10.1186/1472-6947-14-50.

[2] J. McCusker, O. Ardman, F. Bellavance, E. Belzile, S. Cardin, J. Verdon, Use of community services by seniors before and after an emergency visit, Canadian Journal on Aging 20 (2001) 193–210. doi:10.1017/S0714980800012976.

[3] N. R. Hoot, D. Aronsky, Systematic review of emergency department crowding: Causes, effects, and solutions, Annals of Emergency Medicine 52 (2) (2008) 126–136. doi:10.1016/j.annemergmed.2008.03.014.

[4] F. Zeinali, M. Mahootchi, M. M. Sepehri, Resource planning in the emergency departments: A simulation-based metamodeling approach, Journal of Simulation Modelling Practice and Theory 53 (2015) 123–138. doi:10.1016/j.simpat.2015.02.002.

[5] S. Brailsford, J. Vissers, OR in healthcare: A european perspective, European Journal of Operational Research 212 (2) (2011) 223–234. doi:10.1016/j.ejor.2010.10.026.

[6] F. Kadri, S. Chaabane, C. Tahon, A simulation-based decision support system to prevent and predict strain situations in emergency department systems, Journal of Simulation Modelling Practice and Theory 42 (2014) 32 – 52. doi:10.1016/j.simpat.2013.12.004.

[7] P. Escudero-Marin, M. Pidd, Using abms to simulate emergency departments, in: Proceedings of the 2011 Winter Simulation Conference, 2011, pp. 1239–1250. doi:10.1109/WSC.2011.6147773.

[8] S. Robinson, Simulation: the practice of model development and use, Palgrave Macmillan, 2004.

[9] T. Eldabi, R. Paul, A proposed approach for modeling healthcare systems for understanding, in: Proceedings of the 2001 Winter Simulation Conference, Vol. 2, 2001, pp. 1412–1420. doi:10.1109/WSC.2001.977464.

[10] M. Gunal, M. Pidd, Understanding Accident and Emergency Department Performance using Simulation, in: Proceedings of the 2006 Winter Simulation Conference, 2006, pp. 446–452. doi:10.1109/WSC.2006.323114.

[11] J. K. Cochran, K. Roche, A queuing-based decision support methodology to estimate hospital inpatient bed demand, Journal of the Operational Research Society 59 (11) (2008) 1471–1482. doi:10.1057/palgrave.jors.2602499.

[12] M. A. Ahmed, T. M. Alkhamis, Simulation optimization for an emergency department healthcare unit in Kuwait, European Journal of Operational Research 198 (3) (2009) 936–942. doi:10.1016/j.ejor.2008.10.025.

[13] J. K. Cochran, K. T. Roche, A multi-class queuing network analysis methodology for improving hospital emergency department performance, Computers & Operations Research 36 (5) (2009) 1497–1512. doi:10.1016/j.cor.2008.02.004.

[14] L. Holland, L. Smith, K. Blick, Reducing laboratory turnaround time outliers can reduce emergency department patient length of stay: An 11-hosiptal study, American Journal of Clinical Pathology 124 (5) (2005) 672–674. doi:10.1309/E9QPV6G2FBMJ.B.

[15] M. A. Centeno, H. R. Dodd, M. Aranda, Y. Sanchez, A simulation study to increase throughput in an endoscopy center, in: Proceedings of the 2010

Winter Simulation Conference, 2010, pp. 2462–2473. doi:10.1109/WSC.2010.5678942.

[16] P. Bhattacharjee, P. K. Ray, Patient flow modelling and performance analysis of healthcare delivery processes in hospitals: A review and reflections, Computers & Industrial Engineering 78 (2014) 299 – 312. doi:10.1016/j.cie.2014.04.016.

[17] S. Barnes, B. Golden, S. Price, Applications of Agent-Based Modeling and Simulation to Healthcare Operations Management, in: B. T. Denton (Ed.), Handbook of Healthcare Operations Management, Vol. 184, Springer New York, 2013, pp. 45–74. doi:10.1007/978-1-4614-5885-2.

[18] P. Koelling, M. J. Schwandt, Health systems: a dynamic system - benefits from system dynamics, in: Proceedings of the Winter Simulation Conference, 2005., 2005, pp. 1321–1327. doi:10.1109/WSC.2005.1574393.

[19] J. B. Jun, S. H. Jacobson, J. R. Swisher, Application of discrete-event simulation in health care clinics: A survey, The Journal of the Operational Research Society 50 (2) (1999) 109–123. doi:10.2307/3010560.

[20] A. A. Tako, K. Kotiadis, PartiSim: A multi-methodology framework to support facilitated simulation modelling in healthcare, European Journal of Operational Research 244 (2) (2015) 555–564. doi:10.1016/j.ejor.2015.01.046.

[21] D. Fone, S. Hollinghurst, M. Temple, A. Round, N. Lester, A. Weightman, K. Roberts, E. Coyle, G. Bevan, S. Palmer, Systematic review of the use and value of computer simulation modelling in population health and health care delivery, Journal of Public Health 25 (4) (2003) 325–335. doi:10.1093/pubmed/fdg075.

[22] M. J. North, C. M. Macal, Managing business complexity: discovering strategic solutions with agent-based modeling and simulation, Oxford University Press, 2007.

[23] D. Heard, G. Dent, T. Schifeling, D. Banks, Agent-based models and microsimulation, Annual Review of Statistics and Its Application 2 (2015) 259–272. doi:10.1146/annurev-statistics-010814-020218.

[24] A. Borshchev, A. Filippov, From system dynamics and discrete event to practical agent based modeling: reasons, techniques, tools, in: Proceedings of The 22nd International Conference of the System Dynamics Society, 2004, pp. 959–966.

[25] M. Gul, A. F. Guneri, A comprehensive review of emergency department simulation applications for normal and disaster conditions, Computers & Industrial Engineering 83 (2015) 327–344. doi:10.1016/j.cie.2015.02.018.

[26] C. Rudin, Algorithms for interpretable machine learning, in: Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining, ACM, 2014, p. 1519.

[27] C. M. Macal, M. J. North, Agent-based modeling and simulation, in: Proceedings of the 2009 Winter Simulation Conference, 2009, pp. 86 – 98. doi:10.1109/WSC.2009.5429318.

[28] D. Dorner, The logic of failure: Why things go wrong and what we can do to make them right.

[29] J. D. Sterman, Modeling Managerial Behavior: Misperceptions of Feedback in a Dynamic Decision Making Experiment, Management Science 35 (3) (1989) 321–339. doi:10.1287/mnsc.35.3.321.

[30] E. Cabrera, M. Taboada, M. L. Iglesias, F. Epelde, E. Luque, Simulation optimization for healthcare emergency departments, Procedia Computer Science 9 (2012) 1464–1473. doi:10.1016/j.ejor.2008.10.025.

[31] M. Taboada, E. Cabrera, F. Epelde, M. L. Iglesias, E. Luque, Using an agent-based simulation for predicting the effects of patients derivation poli-

cies in emergency departments, Procedia Computer Science 18 (2013) 641–650. doi:10.1016/j.procs.2013.05.228.

[32] M. Taboada, E. Cabrera, M. L. Iglesias, F. Epelde, E. Luque, An agent-based decision support system for hospitals emergency departments, Procedia Computer Science 4 (2011) 1870–1879. doi:10.1016/j.procs.2011.04.203.

[33] C. Jaramillo, M. Taboada, F. Epelde, D. Rexachs, E. Luque, Agent based model and simulation of MRSA transmission in emergency departments, in: 2015 International Conference On Computational Science, Vol. 51, 2015, pp. 443 – 452. doi:10.1016/j.procs.2015.05.267.

[34] C. Jaramillo, D. Rexachs, E. Luque, F. Epelde, M. Taboada, Modeling the contact propagation of nosocomial infection in hospital emergency departments, in: The Sixth International Conference on Advances in System Simulation, 2014, pp. 84–89.

[35] E. Bruballa, M. Taboada, E. Cabrera, D. Rexachs, E. Luque, Simulation and big data: A way to discover unusual knowledge in emergency departments: Work-in-progress paper, in: Future Internet of Things and Cloud (FiCloud), 2014 International Conference on, 2014, pp. 367–372. doi:10.1109/FiCloud.2014.65.

[36] Z. Liu, E. Cabrera, M. Taboada, F. Epelde, D. Rexachs, E. Luque, Quantitative evaluation of decision effects in the management of emergency department problems, in: 2015 International Conference On Computational Science, Vol. 51, 2015, pp. 433 – 442. doi:10.1016/j.procs.2015.05.265.

[37] Z. Liu, D. Rexachs, E. Luque, F. Epelde, E. Cabrera, Simulating the micro-level behavior of emergency department for macro-level features prediction, in: 2015 Winter Simulation Conference (WSC), 2015, pp. 171–182. doi:10.1109/WSC.2015.7408162.

[38] Z. Liu, E. Cabrera, D. Rexachs, E. Luque, A Generalized Agent-Based Model to Simulate Emergency Departments, in: The Sixth International Conference on Advances in System Simulation, IARIA, 2014, Nice, France, 2014, pp. 65–70.

[39] G. A.Gray, T. G.Kolda, Algorithm 856: Appspack 4.0: Asynchronous parallel pattern search for derivative-free optimization, ACM Transactions on Mathematical Software 32(3) (2006) 485–507.

[40] T. G. Kolda, Revisiting asynchronous parallel pattern search for nonlinear optimization, SIAM Journal on Optimization 16(2) (2006) 563–586. doi:10.1137/040603589.

[41] J. D. Griffin, T. G. Kolda, Asynchronous parallel generating set search for linearly-constrained optimization, Technical Report, Sandia National Laboratories, Livermore, CA July 2006.

[42] W. Rashwan, A. Arisha, Modeling behavior of nurses in clinical medical unit in university hospital: Burnout implications, in: 2015 Winter Simulation Conference (WSC), 2015, pp. 3880–3891. doi:10.1109/WSC.2015.7408544.

[43] R. Blasak, W. Armel, D. Starks, M. Hayduk, The use of simulation to evaluate hospital operations between the emergency department and a medical telemetry unit, in: Proceedings of the 2003 Winter Simulation Conference, Vol. 2, 2003, pp. 1887–1893. doi:10.1109/WSC.2003.1261649.

[44] M. J. Cote, Patient flow and resource utilization in an outpatient clinic, Socio-Economic Planning Sciences 33 (3) (1999) 231–245. doi:10.1016/j.procs.2011.04.203.

[45] D. Sinreich, Y. Marmor, Emergency department operations: The basis for developing a simulation tool, IIE Transactions 37 (3) (2005) 233–245. doi:10.1080/07408170590899625.

[46] S. F. Railsback, V. Grimm, Agent-Based and Individual-Based Modeling: A Practical Introduction, Princeton University Press, 2011.

[47] D. Lin, J. Patrick, F. Labeau, Estimating the waiting time of multi-priority emergency patients with downstream blocking, Health Care Management Science 17 (1) (2014) 88–99. doi:10.1007/s10729-013-9241-3.

[48] J. L. Wiler, E. Bolandifar, R. T. Griffey, R. F. Poirier, T. Olsen, An Emergency Department Patient Flow Model Based on Queueing Theory Principles, Academic Emergency Medicine 20 (9) (2013) 939–946. doi:10.1111/acem.12215.

[49] E. Boudreaux, E. O'Hea, Patient satisfaction in the Emergency Department: A review of the literature and implications for practice, Journal of Emergency Medicine 26 (1) (2004) 13–26. doi:10.1016/j.jemermed.2003.04.003.

[50] E. J. Rising, R. Baron, B. Averill, A systems analysis of a university-health-service outpatient clinic, Operations Research 21 (5) (1973) 1030–1047.

[51] W. M. Hancock, P. F. Walter, The use of computer simulation to develop hospital systems, ACM SIGSIM Simulation Digest 10 (4) (1979) 28–32. doi:10.1145/1102815.1102819.

[52] T. O. Paulussen, A. Zöller, A. Heinzl, L. Braubach, A. Pokahr, W. Lamersdorf, Patient scheduling under uncertainty, in: Proceedings of the 2004 ACM Symposium on Applied Computing, SAC 04, 2004, pp. 309–310. doi:10.1145/967900.967966.

[53] M. A. Badri, J. Hollingsworth, A simulation model for scheduling in the emergency room, International Journal of Operations & Production Management 13 (3) (1993) 13–24. doi:10.1108/01443579310025989.

[54] D. Gove, D. Hewett, A hospital capacity planning model, OR Insight 8 (2) (1995) 12–15. doi:10.1057/ori.1995.8.

[55] E. Kozan, M. Diefenbach, Hospital emergency department simulation for resource analysis, Industrial Engineering & Management Systems 7 (2) (2008) 133–142.

[56] J. Kuljis, R. J. Paul, L. K. Stergioulas, Can health care benefit from modeling and simulation methods in the same way as business and manufacturing has?, in: 2007 Winter Simulation Conference, 2007, pp. 1449–1453. doi:10.1109/WSC.2007.4419755.

[57] S. Y. Shin, H. Balasubramanian, Y. Brun, P. L. Henneman, L. J. Osterweil, Resource scheduling through resource-aware simulation of emergency departments, in: Proceedings of the 5th International Workshop on Software Engineering in Health Care, SEHC 13, IEEE Press, Piscataway, NJ, USA, 2013, pp. 64–70. doi:10.1109/SEHC.2013.6602480.

[58] C. M. Macal, M. J. North, Tutorial on agent-based modeling and simulation, in: Proceedings of the Winter Simulation Conference, 2005., 2005, pp. 2–15. doi:10.1109/WSC.2005.1574234.

[59] P. Escudero-Marin, M. Pidd, Using ABMS to simulate Emergency Departments, in: Proceedings of the 2011 Winter Simulation Conference, 2011, pp. 1239–1250. doi:10.1109/WSC.2011.6147773.

[60] M. Laskowski, S. Mukhi, Electronic Healthcare: First International Conference, eHealth 2008, London, UK, September 8-9, 2008. Revised Selected Papers, Springer Berlin Heidelberg, Berlin, Heidelberg, 2009, Ch. Agent-Based Simulation of Emergency Departments with Patient Diversion, pp. 25–37. doi:10.1007/978-3-642-00413-1_4.

[61] S. S. Jones, R. S. Evans, An Agent Based Simulation Tool for Scheduling Emergency Department Physicians, AMIA Annual Symposium Proceedings 2008 (2008) 338–342.

[62] P. Barach, J. K. Johnson, Understanding the complexity of redesigning care around the clinical microsystem., Quality and Safety in Health Care 15 (suppl 1) (2006) 10–16. doi:10.1136/qshc.2005.015859.

[63] M. E. Lim, A. Worster, R. Goeree, J.-E. Tarride, Simulating an emergency department: the importance of modeling the interactions between physicians and delegates in a discrete event simulation., BMC medical informatics and decision making 13 (1) (2013) 59. doi:10.1186/1472-6947-13-59.

[64] R. Axtell, Why Agents? On the Varied Motivations for Agent Computing in the Social Sciences, Center on Social and Economics Dynamics - The Brookings Institution (17) (2000) 1–23.

[65] H. Bibi, A. Nutman, D. Shoseyov, M. Shalom, R. Peled, S. Kivity, J. Nutman, Prediction of Emergency Department Visits for Respiratory Symptoms Using an Artificial Neural Network, Chest 122 (5) (2002) 1627–1632. doi:10.1378/chest.122.5.1627.

[66] M. Thorwarth, A. Arisha, A simulation-based decision support system to model complex demand driven healthcare facilities, in: Proceedings of the 2012 Winter Simulation Conference, 2012, pp. 1–12. doi:10.1109/WSC.2012.6465019.

[67] P. O. Siebers, C. M. Macal, J. Garnett, D. Buxton, M. Pidd, Discrete-event simulation is dead, long live agent-based simulation!, Journal of Simulation 4 (3) (2010) 204–210. doi:10.1057/jos.2010.14.

[68] C. Macal, M. North, Introductory tutorial: Agent-based modeling and simulation, in: Proceedings of the Winter Simulation Conference 2014, 2014, pp. 6–20. doi:10.1109/WSC.2014.7019874.

[69] S. Brailsford, Discrete-event simulation is alive and kicking!, Journal of Simulation 8 (1) (2013) 1–8. doi:10.1057/jos.2013.13.

[70] A. Djanatliev, R. German, Prospective healthcare decision-making by combined system dynamics, discrete-event and agent-based simulation, in: Proceedings of the 2013 Winter Simulation Conference, 2013, pp. 270–281. doi:10.1109/WSC.2013.6721426.

[71] M. Hofmann, On the Complexity of Parameter Calibration in Simulation Models, The Journal of Defense Modeling and Simulation: Applications, Methodology, Technology 2 (4) (2005) 217–226. doi:10.1177/154851290500200405.

[72] M. Wagner, W. Cai, M. H. Lees, H. Aydt, Evolving agent-based models using self-adaptive complexification, Journal of Computational Science 10 (2015) 351–359. doi:10.1016/j.jocs.2015.03.005.

[73] J. Zhong, N. Hu, W. Cai, M. Lees, L. Luo, Density-based evolutionary framework for crowd model calibration, Journal of Computational Science 6 (2015) 11–22. doi:10.1016/j.jocs.2014.09.002.

[74] J. Zhong, W. Cai, Differential evolution with sensitivity analysis and the Powell's method for crowd model calibration, Journal of Computational Science 9 (2015) 26–32. doi:10.1016/j.jocs.2015.04.013.

[75] M. Fehler, F. Klügl, F. Puppe, Approaches for Resolving the Dilemma between Model Structure Refinement and Parameter Calibration in Agent-Based Simulations, in: Proceedings of the Fifth International Joint Conference on Autonomous Agents and Multiagent Systems, AAMAS, 2006, pp. 120–122. doi:10.1145/1160633.1160651.

[76] M. Fehler, F. Klügl, F. Puppe, Engineering Societies in the Agents World V: 5th International Workshop. Revised Selected and Invited Papers, Springer Berlin Heidelberg, 2005, Ch. Techniques for Analysis and Calibration of Multi-agent Simulations, pp. 305–321. doi:10.1007/11423355_22.

[77] D. A. Samuelson, C. M. Macal, Agent-based simulation comes of age, OR/MS Today 33 (4).

[78] J. L. Devore, Probability and Statistics for Engineering and the Sciences, Duxbury Press, 2011.

[79] R. S. Bermejo, C. C. Fadrique, B. R. Fraile, E. F. Centeno, S. P. Cueva, E. María, Triage in Spanish hospitals, Emergencias 25 (1) (2013) 66–70.

[80] M. J. Bullard, B. Unger, J. Spence, E. Grafstein, Revisions to the Canadian emergency department triage and acuity scale (CTAS) adult guidelines, Canadian journal of emergency medicine 10 (2) (2008) 136–151.

[81] A. J. Forster, I. Stiell, G. Wells, A. J. Lee, C. Van Walraven, The effect of hospital occupancy on emergency department length of stay and patient disposition, Academic Emergency Medicine 10 (2) (2003) 127–133. doi:10.1197/aemj.10.2.127.

[82] A.-M. Kelly, M. Bryant, L. Cox, D. Jolley, Improving emergency department efficiency by patient streaming to outcomes-based teams., Australian health review 31 (1) (2007) 16–21. doi:10.1071/AH070016.

[83] R. Ding, M. L. McCarthy, G. Li, T. D. Kirsch, J. J. Jung, G. D. Kelen, Patients Who Leave Without Being Seen: Their Characteristics and History of Emergency Department Use, Annals of Emergency Medicine 48 (6) (2006) 686–693. doi:10.1016/j.annemergmed.2006.05.022.

[84] M. Johnson, S. Myers, J. Wineholt, M. Pollack, A. L. Kusmiesz, Patients who leave the emergency department without being seen, Journal of Emergency Nursing 35 (2) (2009) 105 – 108. doi:10.1016/j.jen.2008.05.006.

[85] M. Kennedy, C. E. MacBean, C. Brand, V. Sundararajan, D. McD Taylor, Review article: Leaving the emergency department without being seen, Emergency medicine Australasia 20 (4) (2008) 306–313. doi:10.1111/j.1742-6723.2008.01103.x.

[86] P. M. Sloot, R. Quax, Information processing as a paradigm to model and simulate complex systems, Journal of Computational Science 3 (5) (2012) 247 – 249. doi:10.1016/j.jocs.2012.07.001.

[87] R. Carmen, M. Defraeye, I. Van Nieuwenhuyse, A Decision Support System for Capacity Planning in Emergency Departments, International Journal of Simulation Modelling 14 (2) (2015) 299–312. doi:10.2507/IJSIMM14(2)10.308.

[88] J. Holland, Studying complex adaptive systems, Journal of Systems Science and Complexity 19 (2006) 1–8. doi:10.1007/s11424-006-0001-z.

[89] J. L. Devore, Probability and Statistics for Engineering and the Sciences, Duxbury Press, 2011.

[90] L. Lamport, Time, clocks, and the ordering of events in a distributed system, Communications of the ACM 21 (7) (1978) 558–565.

[91] D. A. Marshall, L. Burgos-Liz, M. J. IJzerman, N. D. Osgood, W. V. Padula, M. K. Higashi, P. K. Wong, K. S. Pasupathy, W. Crown, Applying dynamic simulation modeling methods in health care delivery research-the SIMULATE checklist: report of the ISPOR simulation modeling emerging good practices task force., Value in Health 18 (1) (2015) 5–16. doi:10.1016/j.jval.2014.12.001.

[92] U. Wilensky, NetLogo. http://ccl.northwestern.edu/netlogo/., Center for Connected Learning and ComputerBased Modeling Northwestern University Evanston IL.

[93] C. A. Chung, Simulation Modeling Handbook: a Practical Approach, CRC Press, 2004.

[94] N. Hoot, S. Epstein, T. Allen, S. e. a. Jones, Forecasting Emergency Department Crowding: An External, Multicenter Evaluation, Annals of Emergency Medicine 54 (4) (2009) 514–522. doi:10.1016/j.annemergmed.2009.06.006.

[95] S. Welch, J. Augustine, C. A. Camargo, C. Reese, Emergency Department Performance Measures and Benchmarking Summit, Academic Emergency Medicine 13 (10) (2006) 1074–1080. doi:10.1197/j.aem.2006.05.026.

[96] S. J. Welch, B. R. Asplin, S. Stone-Griffith, S. J. Davidson, J. Augustine, J. Schuur, Emergency department operational metrics, measures and definitions: Results of the second performance measures and benchmarking sum-

mit, Annals of Emergency Medicine 58 (1) (2011) 33–40. doi:10.1016/j.annemergmed.2010.08.040.

[97] F. Osterreicher, I. Vajda, A new class of metric divergences on probability spaces and its applicability in statistics, Annals of the Institute of Statistical Mathematics 55 (3) (2003) 639–653. doi:10.1007/BF02517812.

[98] D. Endres, J. Schindelin, A new metric for probability distributions, IEEE Transactions on Information Theory 49 (7) (2003) 1858–1860. doi:10.1109/TIT.2003.813506.

[99] J. Lin, Divergence measures based on the shannon entropy, IEEE Transactions on Information Theory 37 (1) (1991) 145–151. doi:10.1109/18.61115.

[100] M. J. D. Powell, An efficient method for finding the minimum of a function of several variables without calculating derivatives, The Computer Journal 7 (2) (1964) 155–162. doi:10.1093/comjnl/7.2.155.

[101] C. Sutti, Local and global optimization by parallel algorithms for mimd systems, Annals of Operations Research 1 (2) (1984) 151–164. doi:10.1007/BF01876145.

[102] A. Law, A tutorial on how to select simulation input probability distributions, in: Proceedings of the 2012 Winter Simulation Conference (WSC), 2012, pp. 1–15. doi:10.1109/WSC.2012.6465281.

[103] A. Guttmann, M. J. Schull, M. J. Vermeulen, T. A. Stukel, Association between waiting times and short term mortality and hospital admission after departure from emergency department: population based cohort study from Ontario, Canada., BMJ 342 (2011) d2983. doi:10.1136/bmj.d2983.

[104] Z. Liu, F. Epelde, D. Rexachs, E. Luque, A bottom-up simulation method to quantitatively predict integrated care system performance, in: International Journal for Integrated Care, the 16th International Conference for Integrated Care, 2016.