Modeling the Dynamics of an Urban Emergency Medical Services System

Abstract

Emergency medical services (EMS) are essential to the delivery of on-demand urgent medical care to patients. The major challenge commonly encountered by EMS agencies is the effective allocation of resources, specifically ambulances and paramedics, such that coverage is sufficient and response times are minimized. Compounding the complexity is the growth and shift of populations which impact EMS demand. In this study, the system dynamics methodology was applied in order to develop a model, using industry data, that represents an urban EMS system. Utilizing this model, the relationship between call request demand and resource capacity was analyzed. A two-scenario analysis was conducted to observe the behavior of the model to a changing number of ambulances and population. A tipping point where continued population growth causes the system to reach peak capacity was identified. The resulting model will support EMS managers and dispatchers with demand planning and policy development.

Keywords: emergency medical services, system dynamics, ambulances, healthcare management

Word Count: 3532

I. Introduction

Emergency medical services, commonly referred to as ambulance, paramedic or prehospital emergency services, are a critical component in the delivery of urgent medical care to communities. Emergency medical service agencies (EMS) are organizations charged with the responsibility of providing out- of-hospital acute medical care to the population of a defined geographic area. EMS agencies also provide transportation to local clinical care facilities, such as hospitals and emergency departments, for patients who are unable to transport themselves due to the nature of their condition or circumstances. By their very nature, EMS systems are extraordinarily complex. The demand for ambulances is dynamic and is known to fluctuate spatially and temporally based on the time of day and day of the week (Channouf et al., 2007). EMS managers and dispatchers are faced with the evolving task of deploying ambulances and the personnel required to provide adequate coverage to their respective service area. Dispatchers have the option of redeploying their fleet to compensate for spatiotemporal fluctuations, but the scope of these adjustments is restricted by the pre-determined staffing plan for a given period. Industry and academic researchers have conducted various studies focused on developing novel deployment strategies, and associated staffing plans, in an effort to combat call volume variability and maximize service coverage (Rajagopalan et al., 2011). These deployment models, developed based on historical data, are ultimately dependent on a comprehensive understanding of demand. Related EMS researchers have sought to identify more sophisticated approaches to forecasting demand in order to improve predictive models for demand planning (Channouf et al., 2007) (Chen et al., 2016) (Setzler et al., 2009) (Aringhieri et al., 2016). EMS managers and dispatchers often describe deployment planning and redeployment decisions as an art, based on the experiences and intuition of the individual (Penner et al., 2016). In the context of systems dynamics, these behaviors are attributed to the mental models of individuals based on their perceptions of demand and the various decision rules established within the organization (Sterman, 2000).

The primary justification for applying a system dynamics methodology to EMS (ambulance) systems is to leverage an approach that, based on the current literature, has not been explored in depth. The current related work in the system dynamics literature has focused primarily on the development of models addressing broader emergency healthcare care issues. From a policy perspective, Cooke et al. examined the overcrowding of emergency departments in healthcare facilities and developed associated models (Cooke et al., 2007). Wang et al. focused on modeling the coordination and allocation of emergency services at a strategic level under extreme circumstances such as natural disasters (Wang et al., 2012). Historically, models addressing common EMS problems such as resource allocation and deployment have been approached using traditional operations research methodologies and discrete event simulations. By applying a system dynamics methodology to this problem, the endogenous and exogenous factors that explain the non-linear interactions between various elements and the feedback processes within the broader system were highlighted. Specifically, the presented model concentrates on observing the impact to the supply of ambulances considering fluctuations in call request demand overtime. While the short-term objective of this study is to advance the demand and capacity planning capabilities of emergency medical service agencies, the ultimate long-term goal is to improve patient outcomes and service delivery by understanding system weaknesses and capacity bottlenecks. This includes testing scenarios such as (1) determining the effect of adjusting the number of available ambulances in the EMS systems fleet and (2) measuring the impact of population growth on the local EMS system capacity. Lastly, this project seeks to alter the strongly held mental models of the managers and dispatchers in the industry. To conduct this study a collection of multi-year call data has been provided by MEDIC, an emergency medical services agency serving Mecklenburg County, North Carolina. The scope of the data used includes all 911 emergency medical service calls received over the course of a one year period (2003-2004). Based on the available data, the time horizon for this study will be focused on a 10-year period spanning from 2003 – 2013 within the scope of Mecklenburg County. Subsequent models and findings can be used to conduct related studies of other emergency medical service agencies using comparable datasets.

II. Model Description

i. Variable Identification

To begin the modeling process, a boundary table (Table 1) was developed to establish scope and identify which variables are included in, and excluded from, the study. Those variables considered to be under the influence and control of local EMS agencies and emergency healthcare providers were considered endogenous, while those variables outside the control of EMS agencies and emergency healthcare providers were categorized as exogenous. Several variables were determined to be excluded from this study including the population and healthcare resources outside of a service area. In the context of this investigation a defined service area is determined by the geographic area that an individual EMS agency is responsible for providing service coverage to; such as a county, city, or municipality. Additional external variables were excluded such as weather conditions, the average income of the population and non-clinical emergency first responders. After completing the boundary selection, a Causal Loop Diagram (Figure 1) was developed to illustrate the broader causal relationships among variables.

Table 1: Emergency Medical Services Systems Model Boundary Table

Endogenous Variables	Exogenous Variables	Excluded Variables
Call Response Time Call Service Time # of Ambulances # of Paramedics & EMTs Distance Between Request Location & Next Available Ambulance (Dispatch Coverage)	Population Incoming EMS Call Requests Size of Service Area (Square Miles) Populations Access to Healthcare % of Population without Healthcare Insurance	Weather Conditions Population outside Service Area Average Earnings/Income of Population Service Area Healthcare Resources outside of Service Area Non-Medical/Clinical Emergency First Responders

ii. Causal Loop Diagram

Many of the variables in the Causal loop diagram directly or indirectly impact two key variables, the number of EMS call requests received and the call response times. The primary objective of EMS systems is to reduce individual call response times for high priority calls by as much as possible. Chen et al. define call response time as the elapsed time from when an operator receives the EMS call until the arrival of the ambulance at the scene (Chen et al., 2016). Intuitively, the survival rate of patients with critical conditions is directly impacted by the response time of the dispatched ambulance. The "# of EMS Call Requests" variable is an accumulator representing the total number of call requests received for a given period (hour, day, week, etc.). The volume of calls received impacts the system's ability to respond to requests, thus increasing the call response time for individual request instances. The additional variables impacting call response times are related to logistics, i.e. the number of available resources, current traffic conditions and the distance between a call request and the responding ambulance. Other variables that impact the number of call requests in a period include, total population and variables related to the populations access to healthcare. This includes the percent of the population without health insurance whose access to healthcare services would be limited based on their circumstance. Ragin et al. conducted an observational study to identify the primary reasons for patients pursuing care from hospital emergency departments. They concluded that a patient's decision to seek care from an emergency department is "often a choice driven by lack of access to, or dissatisfaction with, other sources of care" (Ragin et al., 2005). These findings support a prominent hypothesis in healthcare research that as access to healthcare, and health insurance, increases the need for emergency medical services decreases.

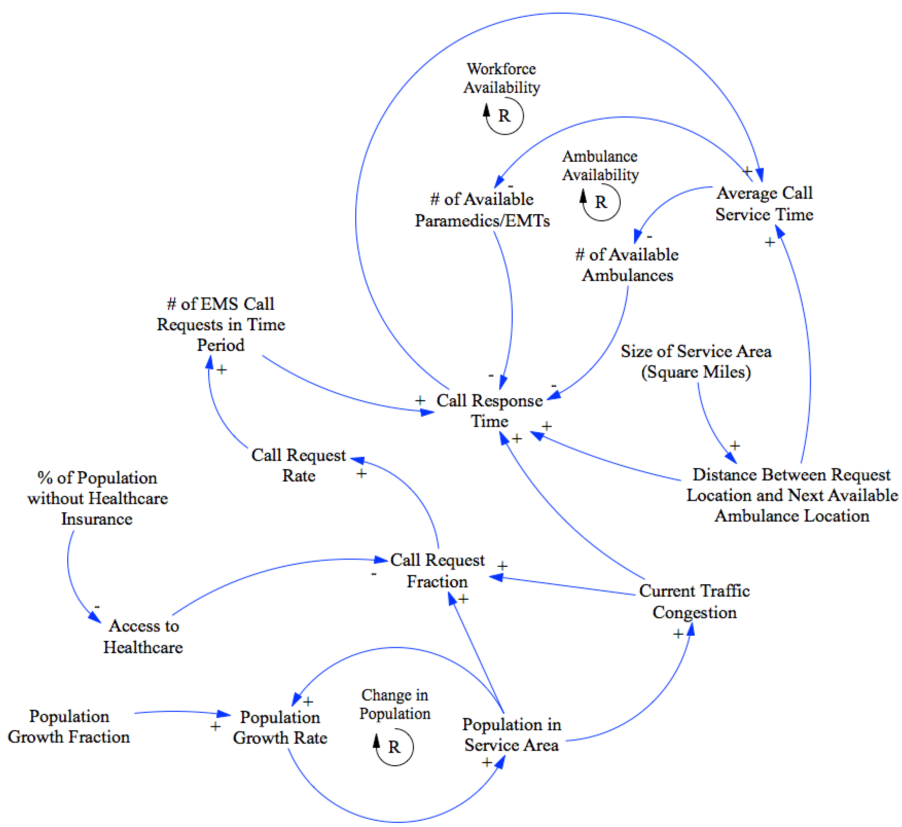

Figure 1: Emergency Medical Services Systems Causal Loop Diagram

The logic behind this hypothesis is that as more people have access to, and rely on, routine and preventative medical care, the less they rely on on-demand acute medical care services such as those provided by EMS agencies and emergency departments. The relationship that exists between emergency medical services and hospitals is created from emergency calls that require patient transfer to emergency department facilities. These types of calls account for most of the EMS calls received and responded to. In a more recent study conducted by MEDIC, the Charlotte Mecklenburg County EMS Agency, it was found that approximately 70-75% of all calls received required transport to a local hospital, i.e. emergency departments (Studnek et al., 2013). Reviewing the causal loop diagram shown in Figure 1, highlights the underlying complexity of EMS systems. To establish a base model, select variables shown in the broader causal loop diagram such as the request distance, traffic congestion and the population access to healthcare were excluded. A simplified causal loop diagram is shown below in Figure 2.

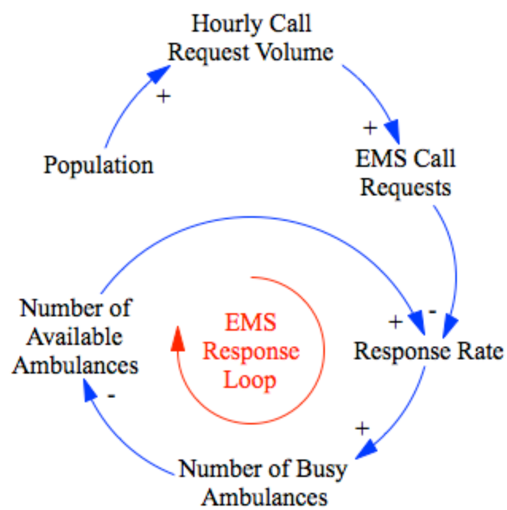

Figure 2: Simplified Emergency Medical Services Systems Causal Loop Diagram

iii. Stock & Flow Diagram

Continuing with the modeling process, a stock and flow diagram (Figure 3) was developed to depict which variables serve as accumulators, rates of change, and ancillary variables. For the time being, the following variables were omitted from the stock and flow analysis due to limited data availability and difficult associated with incorporation; Size of Service Area, Distance Between Request Location and Next Available Ambulance Location, Percent of Population without Healthcare Insurance, and Access to Healthcare. Furthermore, the Number of Available Paramedics/EMTs and the Number of Available Ambulances variables were combined into a stock variable labeled "Idle Available Ambulances". The underlying assumption here is that each available ambulance is staffed with the personnel required to respond to an emergency call request.

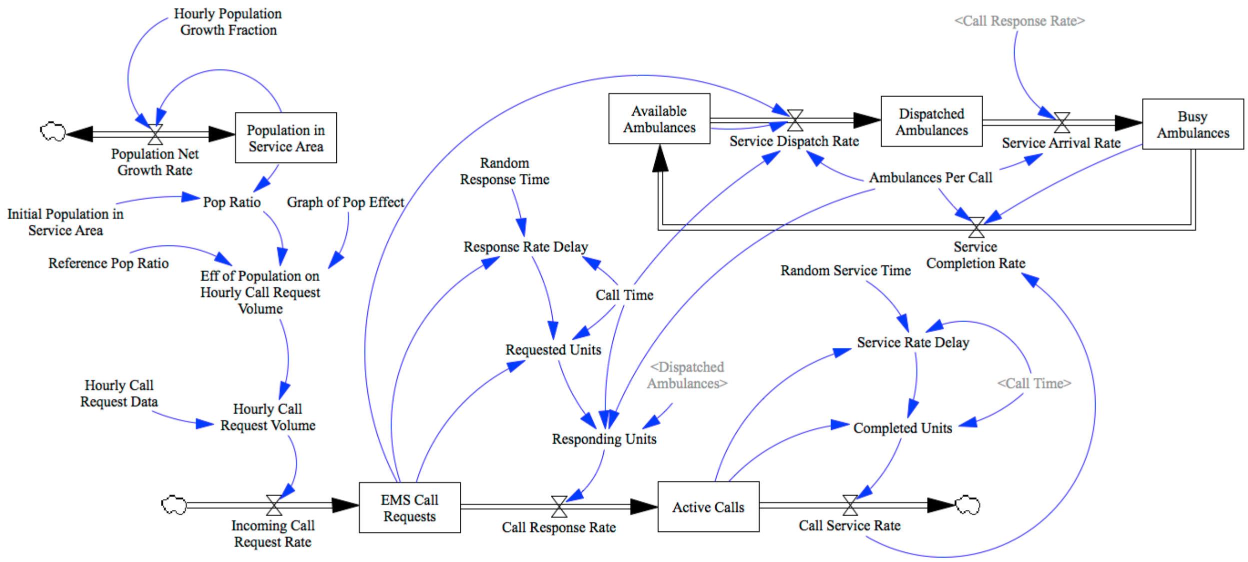

Figure 3: Emergency Medical Services Systems Stock and Flow Diagram

The stock and flow diagram includes six key stock variables representing the Population in the Service Area, Number of EMS Call Requests, The Number of Active Calls, Idle Available Ambulances, Dispatched Ambulances and Busy Ambulances. Typically, ambulance call response and service times in the industry are reported in minutes (Setzler et al., 2009). However, to ensure uniformity throughout the entire model, and a reasonable level of granularity for other factors such as population growth, the models time units were set to hours. At a macro level, the model consists of three primary flow structures representing; 1. Population (people), 2. Calls and, 3. Ambulances. The flows for calls and ambulances are co-flows with aging, because the flows directly interact with one-another as calls are received, responded to and completed.

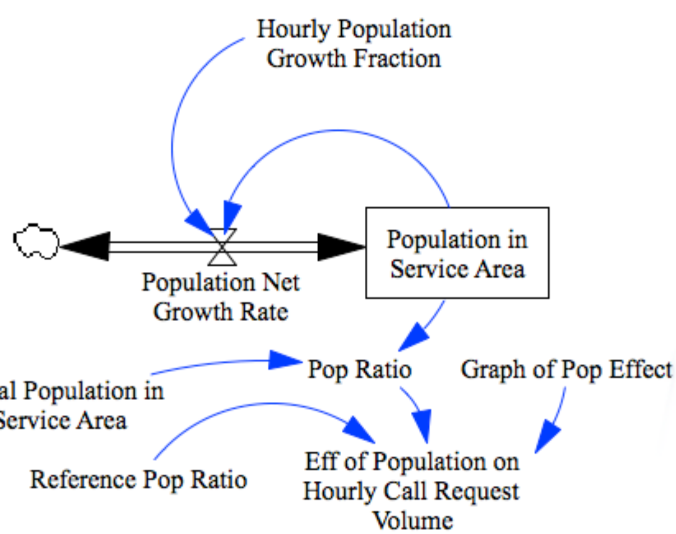

Figure 4: Stock and Flow Diagram – Population Flow Structure

Focusing specifically on the population flow structure (Figure 4), the initial population in service area variable is set to 755,000, based on 2003 Census population estimates for Mecklenburg County, NC (Bureau, 2016). Averaging population growth rates over a 10-year period, from 2003-2013, the annual population growth fraction was calculated to be approximately 2.74%, and then scaled down to percent/hour serving as the hourly population growth fraction value. As the population grows overtime, the associated call request volume is expected to increase in the service area. This relationship is illustrated by the population effect function. Turning now to the call flow structure (Figure 5), the Hourly Call Request Volume, calculated by multiplying the Hourly Call Request Data value by the Population Effect Function, serves as the primary input to the flow structure and determines the value of the Incoming Call Request Rate for each iteration. The Hourly Call Request Data value is generated based on one of two possible model configurations. The first configuration uses historical hourly call volumes extracted from the MEDIC dataset for one randomly selected week, 24 hours for 7 days, providing data sufficient for a 168-iteration simulation run. The second model configuration allows for considerably longer simulations, where the value of the Hourly Call Request Data is determined using a function built into Vensim PLE. The function outputs a normally distributed random integer value given the mean, standard deviation, maximum and minimum hourly call volume values calculated from the MEDIC dataset.

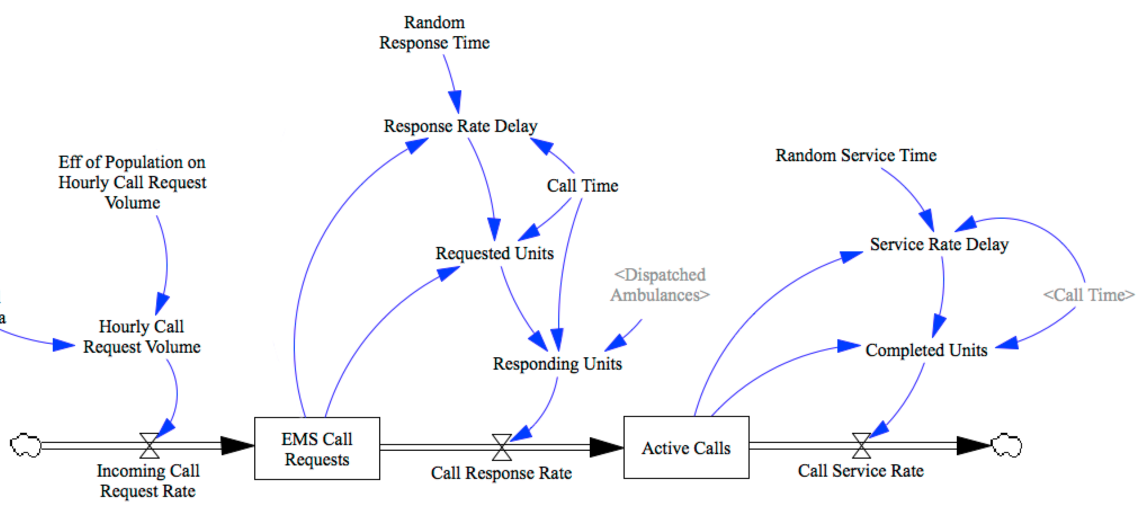

Figure 5: Stock and Flow Diagram – Call Flow Structure

Response and service time data values were also aggregated at an hourly level from the original dataset to calculate statistics and produce normally distributed random values using Vensim. Incoming calls are collected by the EMS Call Requests stock, and migrate to the Active Calls stock based on the Call Response Rate. As discussed earlier, response rate is the elapsed time between when an emergency call is received and when the dispatched ambulance arrives on scene. Therefore, the variable Response Rate Delay is responsible for moving calls to the active stock based on the randomly generated response time. A similar delay structure is implemented for Call Service Rate using randomly generated average service times. To ensure the proper number of calls are made active for a given period, the number of requested calls are compared against the number of ambulances currently dispatched. One key assumption of the model is that there is a one to one relationship between calls and ambulances. This is typically the case in practice, however, in some cases multiple ambulances are dispatched to the same call. The current model does not accommodate instances of calls requiring multiple emergency medical first responders.

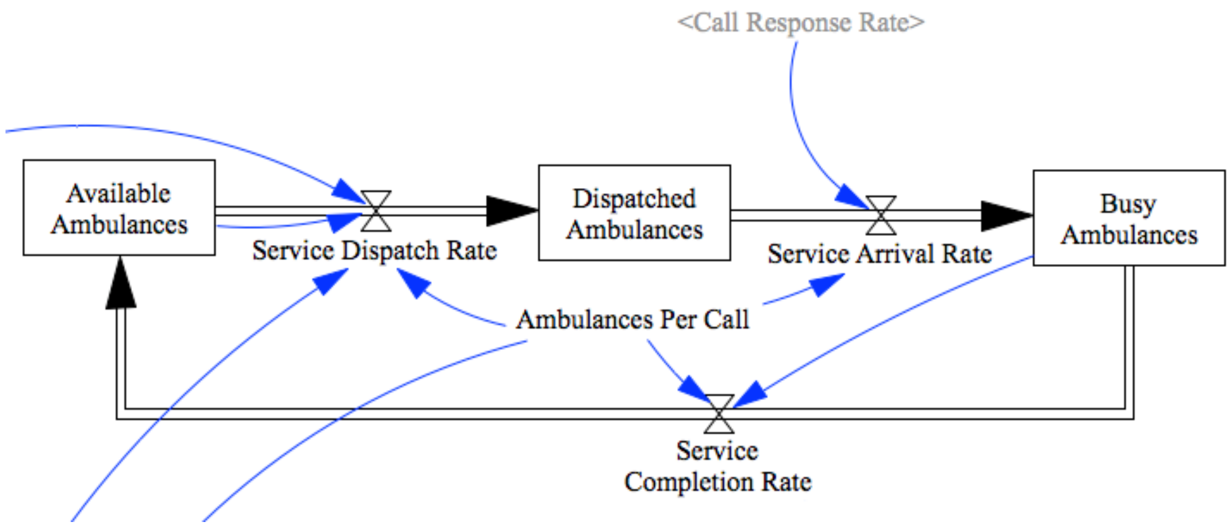

Figure 6: Stock and Flow Diagram – Ambulance Flow Structure

Serving as a co-flow structure to calls, ambulances travel through an aging loop, moving forward from one stock to the other based on status (Figure 6). The initial value of the available ambulances stock is set to 45, based on the estimated number of ambulances in operation during the beginning of the data collection period (Penner et al., 2016). Ambulances are moved from the available stock to the dispatched stock based on incoming call requests. If the number of call requests for a period is greater than the number of ambulances currently available, only the number of available ambulances are dispatched, creating a virtual queue in the call flow structure. Ambulances transition from dispatched to busy is based on the Call Response Rate. Likewise, ambulances return to the available stock is based on the Service Complete Rate.

III. Base Run & Model Credibility

To evaluate model performance and credibility, the call request volumes produced by the base model were compared against data used in a 2007 investigation conducted by researchers in Alberta, Canada. Channouf et al. performed a study concentrated on generating daily and hourly EMS call volume forecasts (Channouf et al., 2007). The data used for their investigation was provided by the Calgary EMS System in Alberta, Canada and spanned a period of 50 months from 2000-2004. In preparing the data for analysis, they aggregated call records based on the number of calls occurring during each hour of the day. They then plotted the hourly call volume counts over a one-year (Figure 7) and one-week (Figure 8) period to identify any clearly observable trends and seasonal components present in the data.

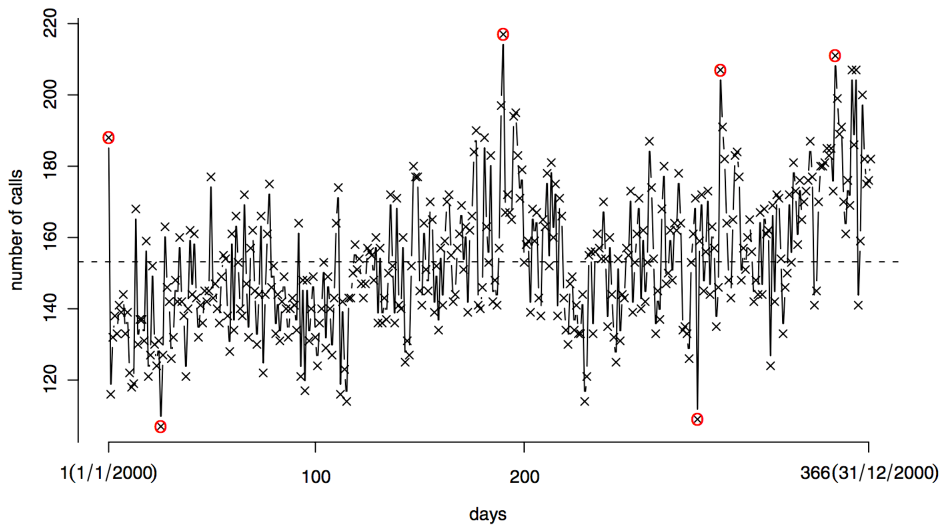

Figure 7: One-Year Daily Call Volume; Jan-Dec 2000; Alberta, Canada (Channouf et al., 2007)

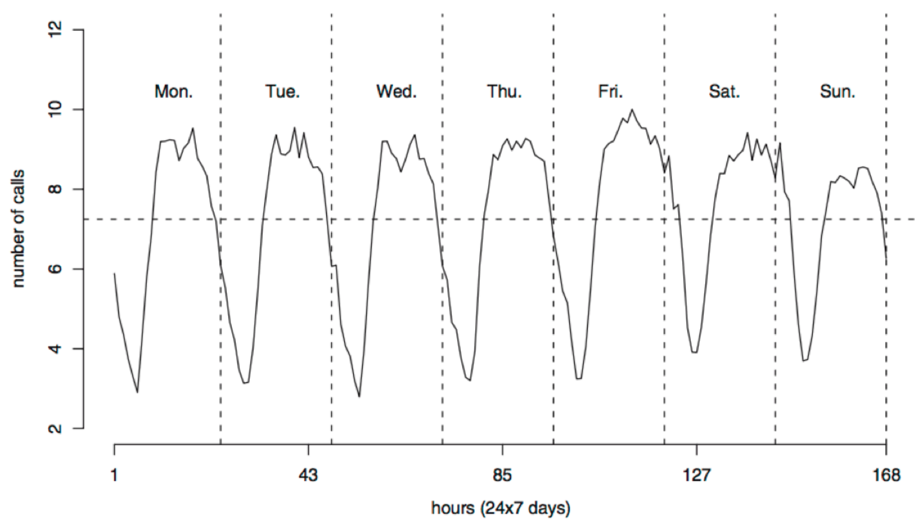

Figure 8: Hourly Call Volume; One Week Period; Alberta, Canada (Channouf et al., 2007)

From the one-year and one-month perspectives a clear positive trend and seasonal behavior are visible, with demand reaching peak values during the months of July and December. One would assume these seasonal peaks are related to increased holiday travel, events, and activities common during these months. Channouf et al. attributed the positive upward trend as likely being caused by urban population growth and, as further underlined by McConnel and Wilson, the advancement of the aging society (McConnel et al., 1998). Visualizing the hourly call volumes at the one-week perspective uncovered an oscillation demand pattern mode, with demand reaching its highest values between the hours of 10:00am and 8:00pm Sunday-Thursday and spanning into the late night/early morning on Friday and Saturday (Channouf et al., 2007). While the data used in Channouf et al.'s study was collected during a similar period (2000-2004), call volume behavior is expected to vary slightly in the model due to the differences in the populace between Alberta, Canada and Mecklenburg County, NC. Figure 9 shows the results for EMS call requests hour by hour after running the base model for one-week (168-hrs.) while Figure 10 shows the results for the same variable over a 10-year period with a 2.74% average annual population growth. It can be observed that as the population increases in the service area, the current number of available ambulances is sufficient to meet the growing call demand. The impact of adjusting the number of ambulances in the system are explored during in the scenario analysis.

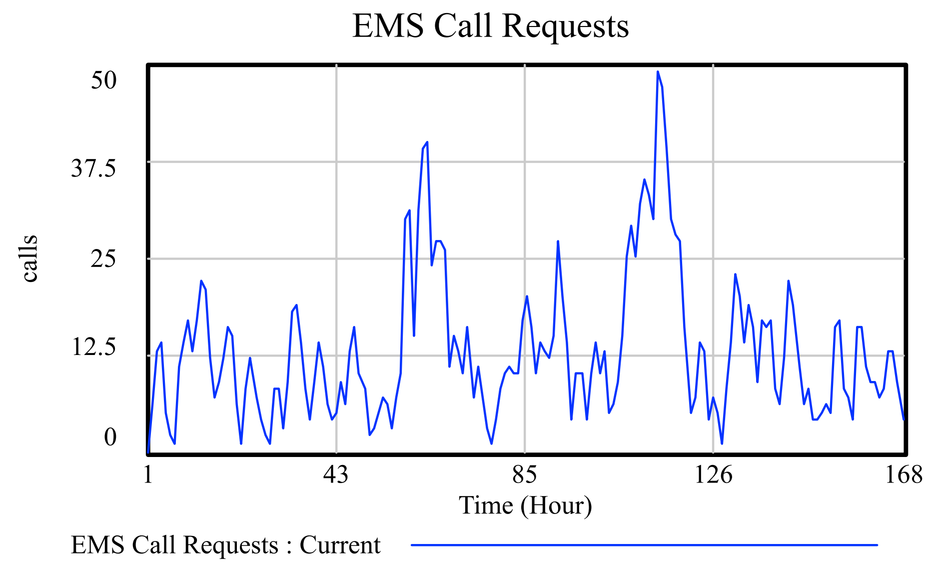

Figure 9: EMS Call Requests; Base Model; One Week Period; MEDIC Mecklenburg County, NC

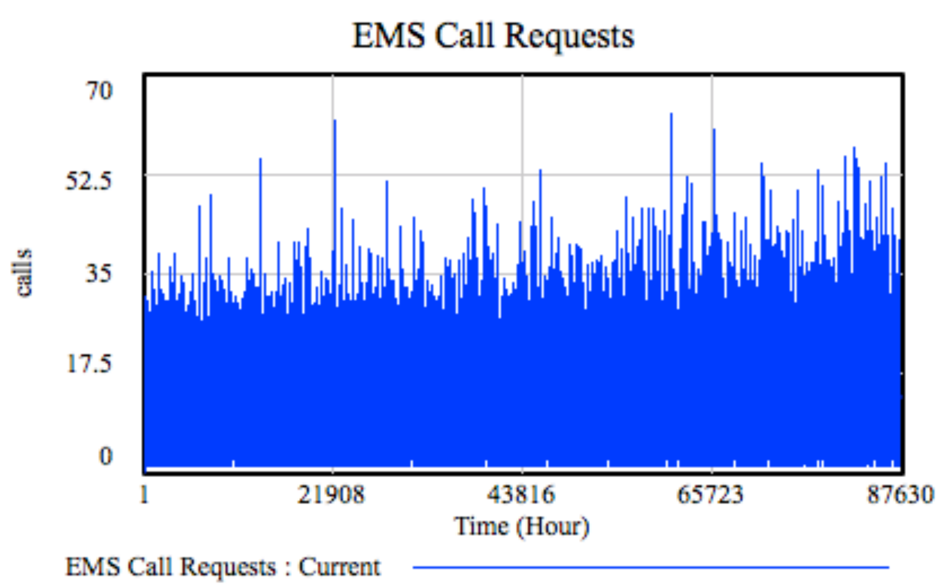

Figure 10: EMS Call Requests; Base Model; 10 Year Period; MEDIC Mecklenburg County, NC

Consistent with the data reviewed by Channouf et al., hourly call volume followed an oscillation demand pattern mode and annual hourly call volume exhibits a subtle positive trend over time. To further appraise the credibility, the results produced by the ambulance stocks were reviewed (Figure 11). As expected, the oscillatory behavior of call requests carries over into the behaviors of ambulances, with available and dispatched ambulances having an inverse relationship. The delay inherit to the call response and service times can also be seen.

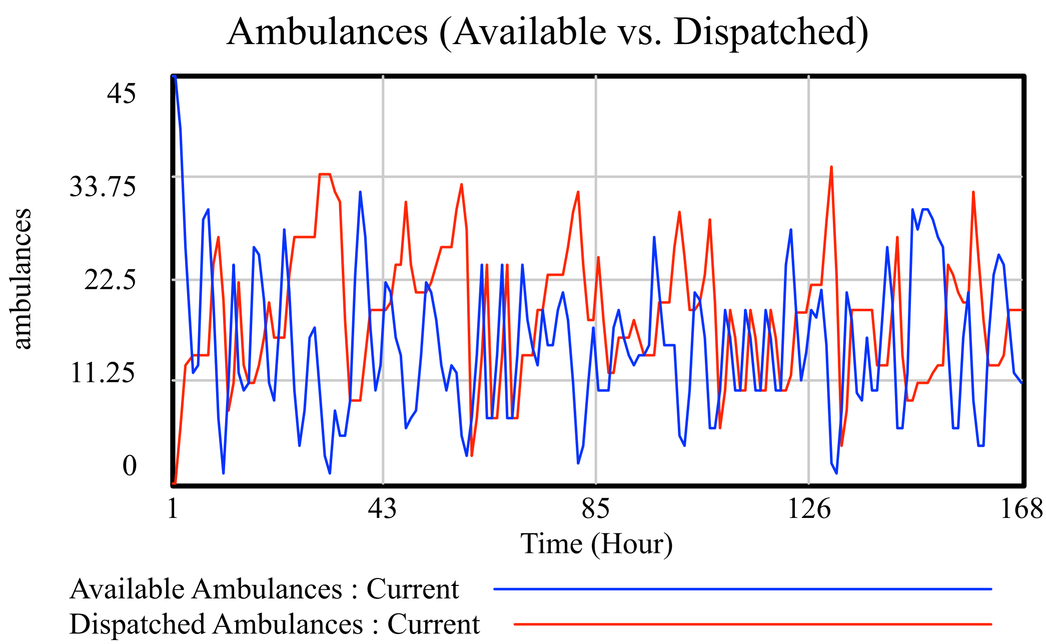

Figure 11: Available vs. Dispatched Ambulances; One-Week Simulation Results

IV. Model Analysis

i. Scenario #1 – Adjusting Fleet Capacity

To assess model durability, two scenarios were crafted to stress test the dynamics of the system, specifically focusing on the relationship between supply and demand. The model analysis starts by evaluating the impact of adjusting the initial value of available ambulances, which represents the number of vehicles in an EMS agencies fleet for a given period. The base model was run with an initial value of 45 as illustrated in the model credibility section. Scaling this up and down the influences of having arbitrarily extreme values of 80 and 20 total ambulances were observed. Viewing the results for Busy Ambulances over a one-week period, given 80 initial available ambulances (Figure 12), the number of busy ambulances rarely crests over 30. This represents a significant excess capacity that would equate to higher agency operating costs related to overstaffing and unnecessary ambulance deployments. Conversely, in the case of only having 20 available ambulances in the fleet (Figure 13), the system is in a constant state of response with demand exceeding supply. As such, calls go unanswered and backlog as displayed in the EMS Call Requests stock (Figure 14). Operationally, this would result in ineffective service coverage, overworked staff, and an increased mortality rate (Erkut et al., 2008).

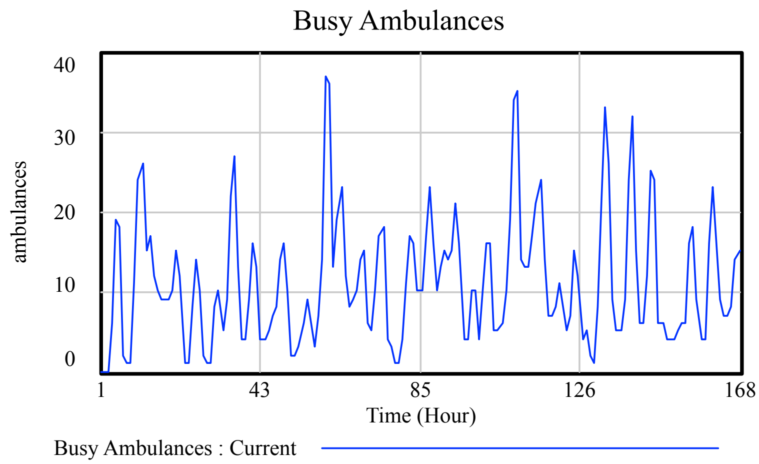

Figure 12: Busy Ambulances; 80 Initial Available Ambulances; One-Week Simulation Results

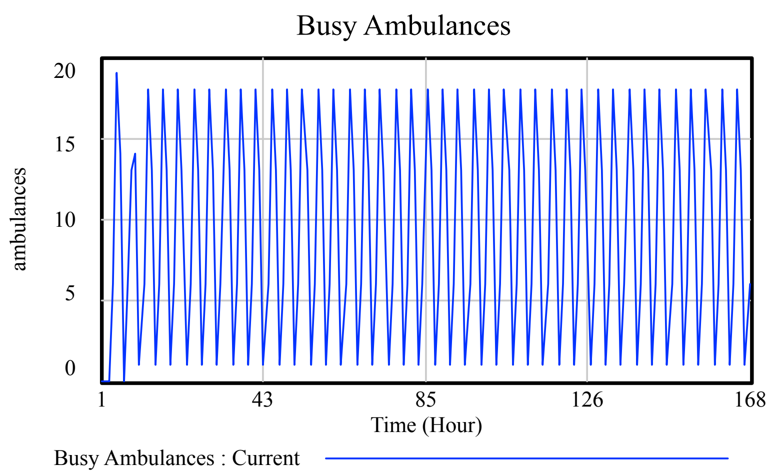

Figure 13: Busy Ambulances; 20 Initial Available Ambulances; One-Week Simulation Results

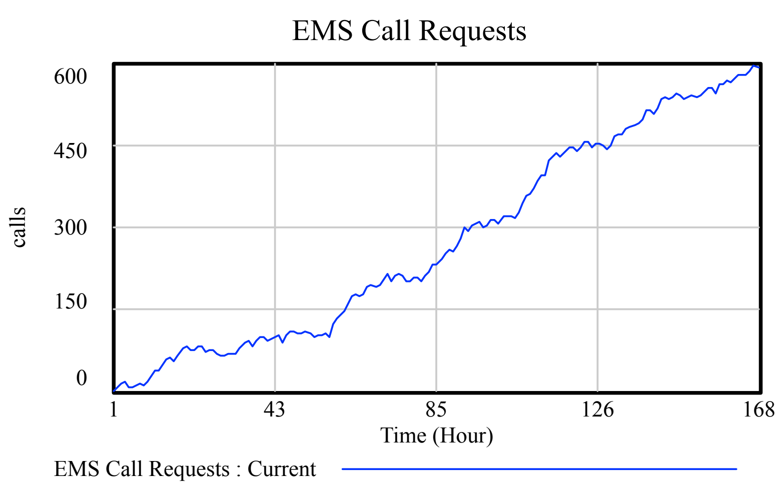

Figure 14: EMS Call Requests; 20 Initial Available Ambulances; One-Week Simulation Results

ii. Scenario #2 – Impact of Population Growth

Exploring yet another scenario, the impact of call volume demand and ambulance response capacity given an abnormal increase in the population were measured over time. For the base model the annual population growth fraction was set to 2.74%, based on the average population change in Mecklenburg County over a 10-year period (2003-2013) (Bureau, 2016). To mimic an extreme scenario this value was changed to 10.0% and another simulation ran for 87,360 hour iterations (10-years). At this extreme growth rate, population grows exponentially to approximately 2.25 million people (Figure 15). Around the 3.5-year mark, approximately 1.1 million people, the peak system capacity is reached, with 45 ambulances, and the number of EMS Call Requests starts growing exponentially as calls go unanswered (Figure 16).

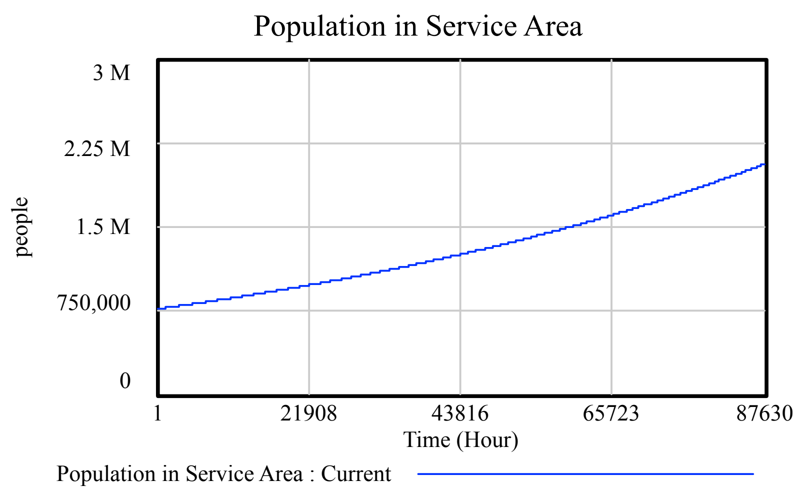

Figure 15: Population in Service Area; 10-percent population growth; 87,630-hour Simulation

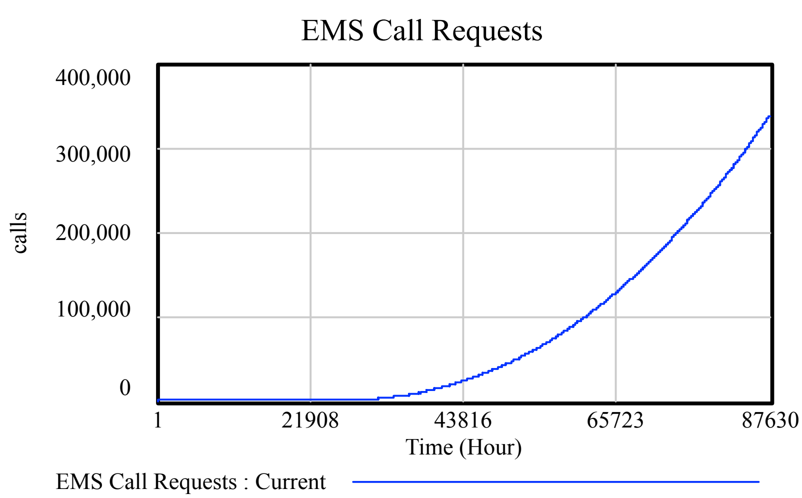

Figure 16: EMS Call Requests; 10-percent population growth; 87,630-hour Simulation

V. Conclusion & Future Work

In this investigation, the system dynamics methodology was applied to develop a model of an emergency medical services system in an urban setting. The primary objective was to apply the system dynamics approach, that have not been applied to the EMS field. Specifically, the impact of a growing population on the EMS system as well as the effect of adjusting capacity, i.e. the number of available ambulances, were analyzed. As expected, as the population grows overtime the associated demand increases and the system becomes more constrained requiring additional ambulances. Eventually, the system reaches a peak capacity and, without additional ambulance deployments, calls go unanswered, backlog and therefore service times increase. In the results of the base run, the initial value of 45 available ambulances was sufficient to meet demand. During the scenario analysis, the number of ambulances were adjusted to arbitrary values (approximately +/-50%) to observe the model's behavior. In its current state, the model can be used by EMS managers and dispatchers to identify the point of peak capacity, adjust for population growth rate, and determine the number of ambulances required. Several exogenous variables such as traffic conditions and the percent of the population without access to health insurance were excluded from the model. Future work could incorporate these variables, as well as explore additional demand factors such as the impact of the aging population on demand as discussed by McConnel & Wilson (McConnel et al., 1998). Furthermore, future models could include flow structures for different call priorities. For instance, high priority calls requiring shorter response times would take precedence over lower priority calls that can be delayed for longer periods of time.

VI. References

Aringhieri, R., Bruni, M. E., Khodaparasti, S., & van Essen, J. T. (2016). Emergency Medical Services and beyond: Addressing new challenges through a wide literature review. *Computers & Operations Research*. doi:http://dx.doi.org/10.1016/j.cor.2016.09.016

Bureau, U. S. C. (2016). United States Census Bureau Population Estimates. Retrieved from http://www.census.gov/popest/index.html

Channouf, N., L'Ecuyer, P., Ingolfsson, A., & Avramidis, A. N. (2007). The application of forecasting techniques to modeling emergency medical system calls in Calgary, Alberta. *Health Care Management Science, 10*(1), 25-45. doi:10.1007/s10729-006-9006-3

Chen, A. Y., Lu, T.-Y., Ma, M. H.-M., & Sun, W.-Z. (2016). Demand Forecast using Data Analytics for the Pre-allocation of Ambulances. *IEEE Journal of Biomedical and Health Informatics, 20*(4), 1178-1187.

Cooke, D. L., Yang, H., Curry, G., Rogers, P., Rohleder, T., Lee, R. C., & Strong, D. (2007). *Introducing system dynamics modeling to health care in Alberta*. Paper presented at the Proceedings of the 25th International Conference of the System Dynamics Society.

Erkut, E., Ingolfsson, A., & Erdoğan, G. (2008). Ambulance location for maximum survival. *Naval Research Logistics (NRL), 55*(1), 42-58.

McConnel, C. E., & Wilson, R. W. (1998). The demand for prehospital emergency services in an aging society. *Social Science & Medicine, 46*(8), 1027-1031. doi:http://dx.doi.org/10.1016/S0277-9536(97)10029-6

Penner, J., Studnek, J., & Infinger, A. (2016, June, 24 2016) *Interview: EMS Call Data for Research/Interviewer: C. Saydam, J. Martin, & V. Vasudev*.

Ragin, D. F., Hwang, U., Cydulka, R. K., Holson, D., Haley, L. L., Richards, C. F., . . . The Emergency Medicine Patients' Access To Healthcare Study, I. (2005). Reasons for Using the Emergency Department: Results of the EMPATH Study. *Academic Emergency Medicine, 12*(12), 1158-1166. doi:10.1197/j.aem.2005.06.030

Rajagopalan, H. K., Saydam, C., Setzler, H., & Sharer, E. (2011). Ambulance Deployment and Shift Scheduling: An Integrated Approach. *Journal of Service Science and Management, Vol.04No.01*, 13. doi:10.4236/jssm.2011.41010

Setzler, H., Saydam, C., & Park, S. (2009). EMS call volume predictions: A comparative study. *Computers & Operations Research, 36*(6), 1843-1851. doi:http://dx.doi.org/10.1016/j.cor.2008.05.010

Sterman, J. (2000). Business Dynamics - Systems Thinking and Modeling for A Complex World. *McGraw Hill: Boston*.

Studnek, J. R., Vandeventer, S., Infinger, A. E., Young, K., & Keith, J. (2013). Increasing Cardiac Arrest Survival by Using Data & Process Improvement Measures. *Journal of Emergency Medical Services, 38*(7), 68-70.

Wang, X., Wang, Q., & Chen, M. (2012). *System dynamics model of unconventional emergency resource coordination system*. Paper presented at the Computational Sciences and Optimization (CSO), 2012 Fifth International Joint Conference on.