Forecasting Emergency Department Arrivals via Regression with ARIMA errors and Facebook Prophet: The Case of a Montreal Hospital

Gerald Bayiapodog Potkah

A Thesis

in

The Department

of

Mechanical, Industrial & Aerospace Engineering (MIAE)

Presented in Partial Fulfillment of the Requirements

for the Degree of

Master of Applied Science (Industrial Engineering) at

Concordia University

Montréal, Québec, Canada

November 2021

CONCORDIA UNIVERSITY

School of Graduate Studies

This is to certify that the thesis prepared

By: **Gerald Bayiapodog Potkah**

Entitled: **Forecasting Emergency Department Arrivals via Regression with ARIMA errors and Facebook Prophet: The Case of a Montreal Hospital**

and submitted in partial fulfillment of the requirements for the degree of

Master of Applied Science (Industrial Engineering)

complies with the regulations of this University and meets the accepted standards with respect to originality and quality.

Signed by the Final Examining Committee:

______________________________ Chair
Dr. Onur Kuzgunkaya

______________________________ External Examiner
Dr. Andrea Schiffauerova

______________________________ Examiner
Dr. Onur Kuzgunkaya

______________________________ Supervisor
Dr. Nadia Bhuiyan

______________________________ Co-supervisor
Dr. Daria Terekhov

Approved by ______________________________
Martin D. Pugh, Chair
Department of Mechanical, Industrial & Aerospace Engineering (MIAE)

______________ 2021 ______________________________
Mourad Debbabi, Dean
Faculty of Engineering and Computer Science

Abstract

Forecasting Emergency Department Arrivals via Regression with ARIMA errors and Facebook Prophet: The Case of a Montreal Hospital

Gerald Bayiapodog Potkah

This thesis is motivated by a practical problem in emergency department (ED) operations management. Prolonged waiting times and overcrowding are prevalent in EDs as a result of the mismatch between demand (i.e., patient arrivals) and supply of ED services (Morley, Unwin, Peterson, Stankovich, & Kinsman, 2018). As the gateway to modern healthcare systems, EDs are faced with the arrival of patients with urgent and complex care needs, increased arrivals of the elderly, and high volume of low-acuity patient arrivals. To improve the operational efficiency and healthcare delivery, ED administrators have to make informed decisions about efficient allocation of resources; demand forecasting is a first step towards informing such decisions.

Using a Montreal hospital ED as a basis for our investigation, we evaluate the effectiveness of the rarely used regression with autoregressive integrated moving average errors (regARIMA) model in forecasting future daily and hourly ED arrivals. We also experimentally evaluate the performance of Facebook Prophet (fbprophet) and demonstrate its competitiveness with established forecasting methods. This insight is particularly valuable given that in the ED arrival forecasting literature, it is viewed as a "Blackbox" or "off-the-shelf" method and has not been used for comparison with other established methods. Furthermore, we investigate the hypothesis that public sporting events, particularly hockey, lead to increased arrivals to the ED by using hockey games as a predictor within our forecasting models.

Acknowledgments

I would like to profoundly thank Dr. Nadia Bhuiyan and Dr. Daria Terekhov for their guidance and encouragement during my graduate studies. Their patience and insightful discussions on the research topic helped hone my ability to be an independent researcher. I am greatly indebted to Dr. Bhuiyan for offering me the opportunity to pursue my studies at Concordia University. This research would not have been accomplished without the incredible support of my supervisors. It was an honour and privilege to work under their supervision.

I would also like to thank all the members past and present of the “data-driven-operations-research” lab in the industrial engineering department at Concordia University for their support and encouragement.

I am especially grateful to my mother, Hannah Muaka for her love, advice, and endless sacrifice; and my wife, Blandine, who has been a source of inspiration and has supported me unconditionally. Thank you both for your love, support and confidence. A huge thanks go out to my siblings Robert, Franklin, Elvis, Perry, and their respective families.

This thesis is also dedicated to my loving father Bernard Ivo Potkah of blessed memory.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Emergency departments (EDs) are the gateway to many acute care hospitals around the world. EDs face significant challenges in delivering high-quality patient care due to overcrowding which occurs when ED demand, expressed in terms of patient arrivals, exceeds ED capacity (Anneveld, Van Der Linden, Grootendorst, and Galli-Leslie (2013); Di Somma et al. (2015)). Overcrowding not only affects patient satisfaction but also the quality of treatment and prognosis (Kam, Sung, & Park, 2010).

Morley et al. (2018) classified causes of emergency department overcrowding (EDOC) into three categories: input, throughput, and output causes. Under input causes, the authors listed the arrival of patients with more urgent and complex care needs, increase in arrivals by the elderly, and high volume of low-acuity arrivals. Therefore, efficient management of patient flow in EDs has become an urgent issue for many hospital administrators. The ability to accurately forecast patient arrival (demand) can influence planning and guide the allocation of human and physical resources to facilitate patient flow in EDs. The efficiency in patient flow has the potential to minimize patient care delays, and improve the overall quality of care (Wargon, Casalino, & Guidet, 2010). Long-term demand forecasts may be used to analyze infrastructure and personnel expansion plans (Batal, Tench, McMillan, Adams, & Mehler, 2001), while short-term forecasts may support the operational planning of available resources daily (Marcilio, Hajat, & Gouveia, 2013).

The motivation of the work presented in this thesis is the need to study and develop better forecasts for EDs, particularly in Quebec. Patients of the province of Quebec, Canada, are experiencing very long waiting times in the ED. As of 2016, Quebec had the worst ED waiting times in the western world with 35 percent of patients waiting five hours or more for care, compared to 15 percent in Ontario, five percent in Germany and the United States, and two percent in Switzerland (Authier, 2016), as can be seen in Figure 1.1.

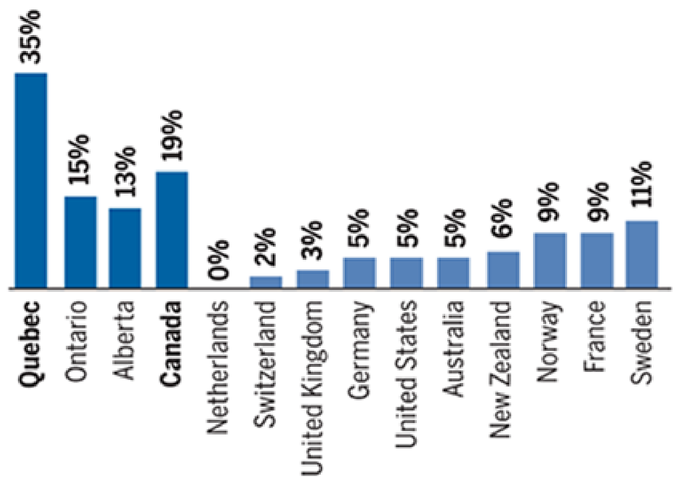


Figure 1.1: Emergency room wait times. Figure adapted from (Authier, 2016)

1.1 Demand Forecasting in EDs

ED demand forecasting can be classified into two categories. In the first category, different methods are compared and the method with the most accurate forecast is selected. However, in this setting, regression with autoregressive integrated moving average errors (regARIMA) has not been widely considered despite the method being an established forecasting model with multiple advantages over models that have been extensively studied. Also, Facebook Prophet (fbprophet), which is a fully automated statistical method (Taylor & Letham, 2018), has not been used because it is viewed as a "Blackbox" or "off-the-shelf" method and researchers want to develop more customized models. In the second category, a few researchers use an "off-the-shelf" method like fbprophet as a module within a bigger resource allocation framework with no real justification for why it was chosen over other established forecasting methods. These observations lead to questioning how fbprophet compares to a custom-designed ARIMA model, and whether fbprophet is then justified to be used as a module in larger studies.

To our knowledge, this thesis is the first work that highlights these nuances in the ED demand forecasting literature and provides a perspective for resource allocation planning problems that need to be addressed by the research community.

Another focus of this research is on the reasons for overcrowding in the ED. Physicians have hypothesized that public sporting events, particularly hockey, lead to increased arrivals to the emergency rooms. We, therefore, investigate this hypothesis by using hockey games, a popular sport in the city of Montreal, as a predictor within our forecasting models.

1.2 Contributions

In this thesis, we find that within the context of a Quebec hospital ED:

(1) a regression with ARIMA errors model, which exploits both exogenous variable(s) and internal dependency, is effective in forecasting both daily and hourly ED arrival counts.

(2) a Facebook Prophet model has a slightly better predictive power than regARIMA, for this particular data, and although we are a single study, this competitiveness provides some justification for its use as a module of a bigger framework focused on scheduling and resource allocation planning.

(3) there is inconclusive evidence of the influence of hockey games as a predictor of patient hourly arrival.

1.3 Organization of the Thesis

This thesis is organized as follows.

In Chapter 2, we provide a background of the ED process in Canada and a Montreal hospital in particular. Next, we present an exploratory analysis of both the current hospital ED physician schedule structure and historical patient arrival data to identify underlying patterns, relationships, and trends suggesting how to tackle the overcrowding and prolonged waiting time situation at a Montreal hospital ED.

In Chapter 3, the literature review is presented. First, we survey ED patient demand/arrival forecasting methods and discuss how these methods are categorized in the literature. Next, we survey predictor variables that are considered important for ED arrival prediction.

In Chapter 4, we investigate two rarely considered forecasting models for daily arrival prediction at a Montreal hospital based on five predictor variables. In addition, we investigate the importance of sporting events in forecasting future arrivals to a Montreal hospital ED based on physicians' observation of high arrival rates on days when there is a hockey game.

In Chapter 5, we extend the study in Chapter 4 by investigating the performance of regARIMA and fbprohet in predicting hourly arrivals.

In Chapter 6, we conclude and briefly describe some directions for future research.

Chapter 2

Canadian Emergency Department Setting and Initial Analysis

In this chapter, we provide a background of the typical emergency department (ED) process in Canada and a Montreal hospital in particular. We describe the results of exploratory data analysis on both the current hospital ED physician schedule structure and historical patient arrival data. This chapter shows that there is a mismatch between demand (i.e., number of patients) and the staff levels at the Montreal hospital ED, which suggests a need to forecast accurately in this particular setting.

2.1 Canadian ED setting

Canada has a publicly funded healthcare system (Evans (1988); Allin and Rudoler (2015)). The emergency care delivery setting can be explained by looking at the throughput of patients in the ED (Figure 2.1).

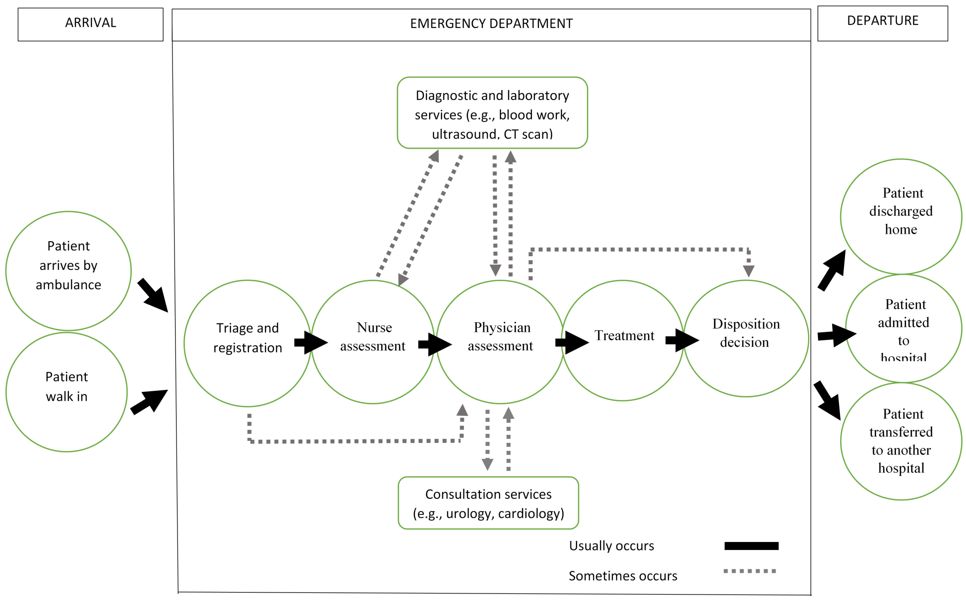


Figure 2.1: Patient Flow through an Emergency Department. Figure Adapted from the Office of the Auditor General of Ontario.

When patients arrive at the ED, they are assessed and placed into different priority classes based on the severity of their cases. In Canada, this is done with the guidance of a five-point Canadian Triage and Acuity Scale (CTAS), with level 1 being the most severe and level 5 the least severe. Each class is associated with a wait time target (Table 2.1). The purpose of the CTAS guideline is to help the EDs maintain an efficient flow of patients to meet the demand of new arrivals. However, the uncertainty in arrivals can lead to failure in performance benchmarks across the triage levels. The reader is referred to (J Murray (2003); Hinson et al. (2019)) for a review on CTAS performance in emergency medicine.

Table 2.1: CTAS Benchmark

Triage Level	**Description**	**Time to Nurse Reassessment**	**Time to Physician Assessment**
I	Resuscitation	Continuous	Immediate
II	Emergency	Every 15 min	$\leq$ 15 min
III	Urgent	Every 30 min	$\leq$ 30 min
IV	Less urgent	Every 60 min	$\leq$ 60 min

The less severe patients are directed to a waiting area while the severely ill are attended to immediately. For example, patients who arrive by ambulance usually go through an initial assessment immediately and a charge nurse decides their level of acuity. The waiting patients are usually assigned to an examination room in the order of their acuity level, to be assessed and treated by a physician. Some patients go through some sort of diagnostic procedure(s) and are reassessed by the attending physician. The physician evaluates the patient's condition and provides the required treatment and decides to either admit or discharge the patient or may have to wait for a consultant to get the disposition decision.

The Montreal hospital is an acute care institution. As part of the Integrated University Center (CIUSSS), its emergency department is one of the busiest for distress cases in Montreal with over 40,000 annual visits (Gouvernement du Québec, 2021). Physicians at this hospital are not paid by the hospital but are reimbursed directly by the Quebec government health insurance program on a fee-for-service basis. The hospital's ED operates 24 hours a day, 7 days a week.

To get a detailed picture of the time points for the analysis, we used a simplified version of the process flow presented in Figure 2.1, since the data obtained from the Montreal hospital's ED does not contain all the necessary patient routes in the ED; this simplified process is shown in Figure 2.2.

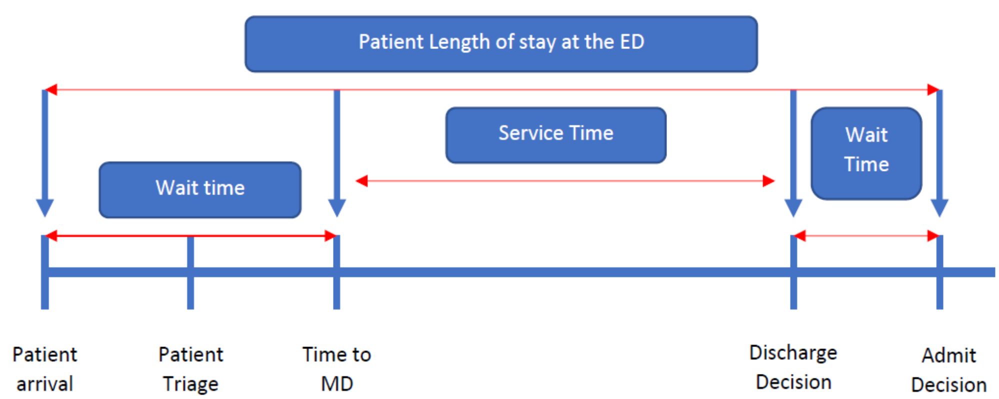


Figure 2.2: The Montreal hospital ED patient flow process time points

2.2 Data Description and Definitions

Two sources of data are used to understand the situation at the Montreal hospital ED: patient arrival data and past physician schedule data.

2.2.1 Patient Arrival Data

The data provided by the CIUSSS authority consisted of 39,950 individual patients who were registered and triaged in the hospital ED between April 1, 2017, and March 31, 2018. The patient-level information included

- time of arrival, i.e., the date and time the patient was registered and triaged,
- mode of arrival, i.e., the principal means by which a patient arrived at the ED,
- time to MD, i.e., the time a patient first sees a physician,
- time to admission, i.e., the time a physician decided to admit a patient,
- time to discharge, i.e., the time a physician decided to discharge a patient,
- patients' triage level, i.e., the appropriate level of care based on a patient's symptom(s) and medical history, and
- patient's classification based on their ability to walk unassisted, i.e. stretcher or ambulatory.

2.2.2 Physician Schedule Data

The data collected on the service side included the number, the type, and the target number of shifts assigned to each physician at the ED.

Based on information from the physicians, a physician is either assigned an ambulatory shift to cater to patients who can walk unassisted or a stretcher shift to cater to patients on stretchers. Physicians on-call are usually assigned to the stretcher task. Also, during the hours between 13:00 and 21:00, and overnight, a physician works in both treatment areas. The shift assignment is however based on the physician's skill set and seniority. Since the physicians work on a contract basis, and many are part-time (i.e., they have other medical practices), they dictate their availability, the type and the number of shifts they can handle in a planning horizon. Some physicians do not take night shifts anymore (due to seniority), others try to avoid doing weekends, while others prefer evenings etc. Hence, the distribution of shifts is highly variable.

However, the following set of schedule rules govern the distribution of shifts to physicians at the Montreal hospital ED:

- ***Non-availabilities***, i.e., each physician MUST give **1.5x** days of availabilities per shift requested per period. If only part of the day is left available, it will be considered a day of **0.5** availability.

- ***Undesirable shifts*** (i.e., all weekend shifts from 13:00 on Friday to 7:30 on Monday, and overnights). Based on staffing at the time, 35-40% of shifts per period for a physician are expected to be undesirable. His/her availabilities given must reflect this ratio.

- ***Overnight shifts***, i.e., preference for overnights will be a priority in assigning preferred shifts. If overnight shifts remain unfilled after all preferences accommodated, overnight shifts will be distributed equitably among remaining active/associate members of ED

- ***Christmas break***, i.e., each year, there will be a quota for shifts to work during the 2-week Christmas period. The quota will be determined by the number of shifts worked in the preceding year, not to exceed 7 shifts per physician.

- ***Summer***, i.e., each year, there will be a quota for shifts to work during the 3 period Summer block. The quota will be determined by the number of shifts worked in the preceding year.

- ***Spring break***, i.e., preference for time off will be given to those who worked the preceding Spring Break first.

- ***Holiday weekends***, i.e., each physician must make themselves available for a minimum of 1 of (Easter weekend or Thanksgiving weekend) **AND** a minimum of 1 of (Victoria Day weekend or St Jean Baptiste weekend). Christmas and New Year are covered under the Christmas schedule. Canada Day and Labour Day weekends are covered under the Summer period.

- ***vacations/extended non-availability***, i.e., any vacation or non-availability greater than 14 days is considered extended.

The most important of these rules is the rule about undesirable shifts; there is usually not enough availability for weekends.

Through our discussions, it seemed that in addition to the national CTAS benchmark (Table 1.1), the Montreal hospital ED has some internal targets for the number of patients that are to be seen based on triage code: CTAS 1 (1 patient/hr), CTAS2 (1.5 patients/hr), CTAS3 (2 patients/hr), CTAS4 (3 patients/hr), and CTAS5 (4 patients/hr).

2.3 Diagnostic study

The Montreal hospital ED has seven daily shifts with start times at 7:00 AM, 8:00 AM, 1:00 PM, 3:00 PM, 4:00 PM, 6:00 PM, and 11:00 PM. To verify the half-hourly aggregate treatment capacity, we looked at the half-hourly distribution of the average number of patients present in the ED and the number of assigned physicians (i.e., staff level), between the period of September 11, 2017, to October 8, 2017 (Figure 2.3). The figure clearly shows a mismatch of patients flow and personnel schedules.

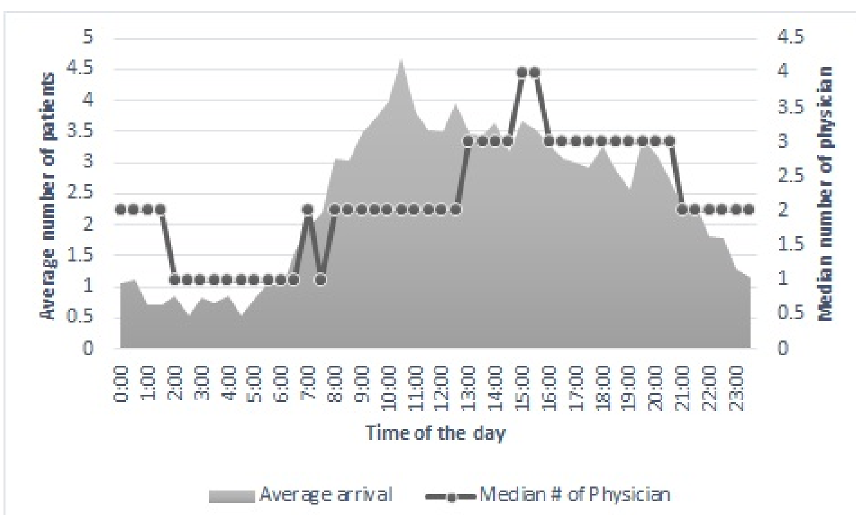


Figure 2.3: Demand volume pattern vs. Physician schedules

The above analysis was done using data covering a one-month planning horizon. To ensure that the analysis holds for all the planning horizons spanning the available data (i.e., 12 months worth of patient arrival data), the distributions of arrival over time of the day for all the planning horizons spanning the available data were checked for similarity. At the Montreal hospital ED, the planning horizon is a month. The graphs plotted over the same axes (Figure 2.4) show that the distributions are "alike", which suggests that the same schedule can be used for every day in the planning horizons under investigation.

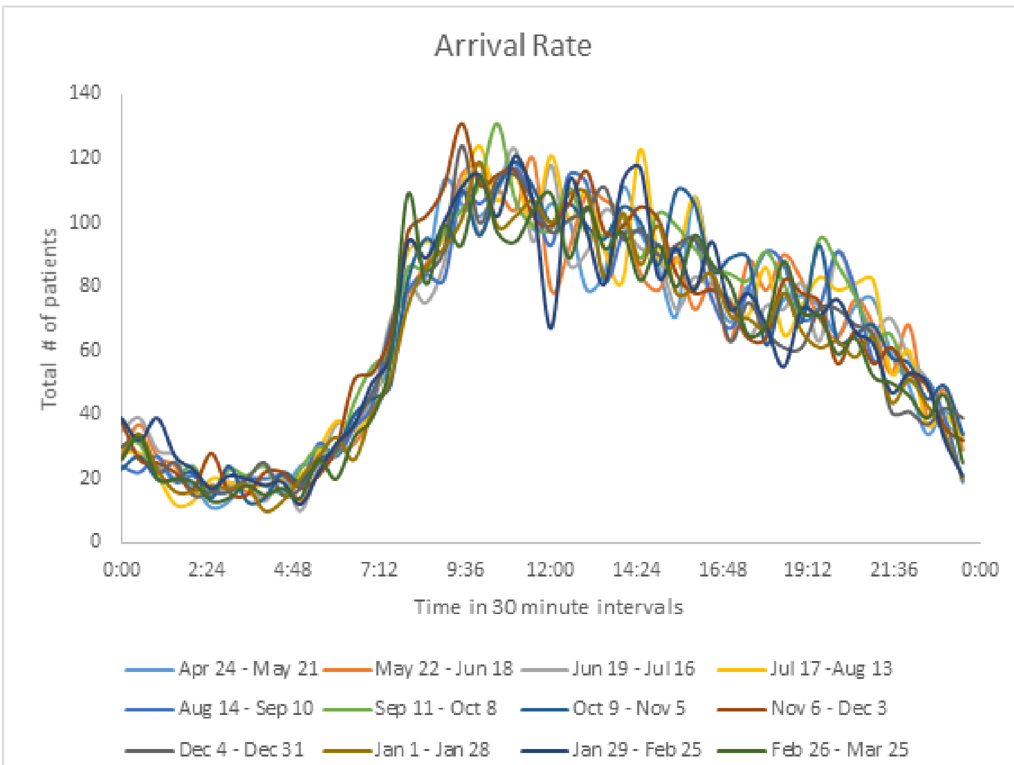


Figure 2.4: Arrivals vs. Time of Day: Different planning horizon

The equality of the distributions was further ascertained by applying the pairwise Kolmogorov-Smirnov (KS) test (Table 2.2). The null hypothesis of the KS test states that there is no difference between the two distributions. The KS test failed to reject the null hypotheses for all pairwise comparisons, indicating there is sufficient evidence to support the fact that there is no difference between the distributions of arrival over time of the day for all the planning horizons. In the table, 1^{st} corresponds to the first planning horizon (i.e., Apr 24 – May 21), 2^{nd} corresponds to second planning horizon (i.e., May 22 – June 18), etc.

Table 2.2: P-values for pairwise Kolmogorov-Smirnov (KS) test

	2nd	3rd	4th	5th	6th	7th	8th	9th	10th	11th	12th
1st	1.00	1.00	0.69	0.96	0.37	0.52	1.00	0.96	1.00	1.00	0.96
2nd		1.00	0.96	0.96	0.69	0.85	0.69	0.96	0.96	1.00	1.00
3rd			0.85	1.00	0.52	0.69	0.52	0.85	0.96	1.00	0.96
4th				1.00	0.96	1.00	0.96	0.85	0.69	0.85	0.96
5th					0.96	1.00	0.69	0.96	0.85	0.96	1.00
6th						1.00	0.69	0.69	0.37	0.52	0.85
7th							0.96	1.00	0.69	0.85	0.96
8th								0.96	0.96	0.96	0.69
9th									0.96	1.00	1.00
10th										0.96	1.00
11th											0.96

Also, the distributions of half-hourly arrival rate by the day of the week with a 95% confidence band (Figure 2.5) show where demand for care is highest and lowest within and between days. The peak times were 9:30 am, 9:30 am, 10:30 am, 11:00 am, 10:00 am, 1:00 pm, and 2:30 pm from Monday to Sunday respectively.

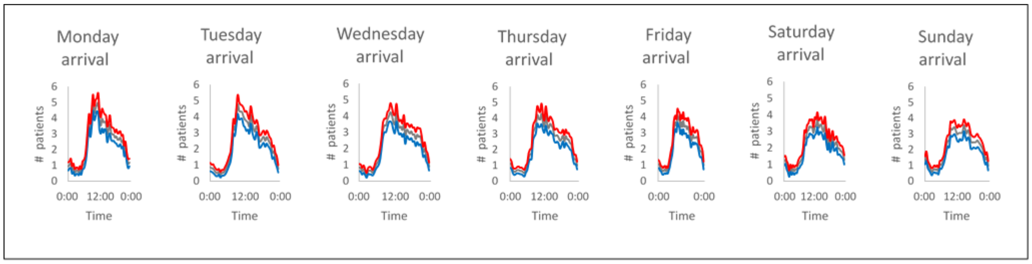


Figure 2.5: Arrivals vs. Time of Day: Different days of the week

The door-to-doctor time, the total time (in hours) from arrival to initial physician assessment (i.e., the time to MD), can be considered a bottleneck in ED operations that leads to overcrowding (Zhao et al., 2015). This period corresponds to the waiting time in the treatment area and it is a key indicator of overcrowding in Canadian EDs as determined by a Delphi study among experts (Ospina et al., 2007). The analysis of the door-to-doctor time revealed that the overcrowding in the Montreal hospital ED is a result of congestion caused by acuity levels 3 and 4 patients waiting for too long to get treatment (Figure 2.6). The number of patients waiting is dependent on the time of the day and the number of patients being treated is an indicator that reflects the efficiency of the ED process.

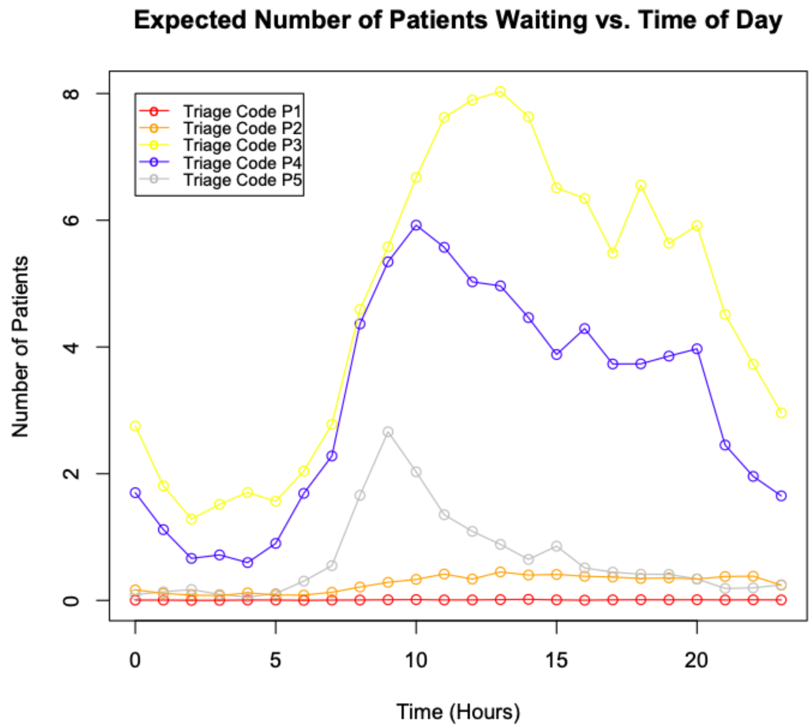


Figure 2.6: Congestion During the Day

The diagnostic study indicated the following about the patient demand:

(1) The number of patient arrivals varies over time of the day, but not over days of the week, suggesting that a daily schedule should be used for every day of the week.

(2) The proportion of patients served in successive half-hours is time-dependent and thus can be exploited in modelling.

(3) The demand distributions over time, of all planning horizons (28 days/planning horizon) for the 2017 – 2018 fiscal year, are statistically similar suggesting that a single physician schedule be used for all days of the year with no variation by days of the week or season.

(4) Congestion in the ED is time-dependent and hence the efficiency of the ED process is reflected in the number of patients being treated.

On the service side, the schedule rules were not strictly followed. For example, undesired shifts (i.e., weekend and overnight shifts) were not equally distributed and the vacation/non-availability rules were not respected. Also, the distribution of shifts was highly variable.

After discussing the diagnostic results with the ED physicians, they were interested in improving the waiting times. Therefore, our next research goal was to match the demand and supply of care. We proposed two possible recommendations to achieve this goal: (1) add extra shift(s) to the current schedule or (2) redesign the current shift schedule.

The ED administration believed adding an extra shift is a good first step. However, the follow-up question was where should the extra shift be added? Intuition will suggest that shifts should be added during periods of increased patient arrival. However, the likelihood of future patient demand is not well known and must be modelled using forecasting techniques(Ernst, Jiang, Krishnamoorthy, & Sier, 2004). Hence, the remainder of this thesis focuses on the evaluation of techniques for forecasting arrivals.

2.4 Conclusion

In this chapter, we provided a background of the ED process in Canada and a Montreal hospital in particular. We carried out an exploratory analysis on both the Montreal hospital ED physician schedule structure and historical patient arrival data to identify the underlying cause(s) of the mismatch between demand (i.e., number of patients) and the staff levels at the hospital ED. Our results show, in summary, that the physician schedule rules were not strictly followed and that congestion in the ED was caused by acuity levels 3 and 4 patients waiting for too long to get treatment. The results and the physicians' observations highlighted the need to forecast accurately in this particular setting, motivating the rest of this thesis.

Chapter 3

Literature Review

In EDs, management has to make resource allocation decisions at operational (short-term), tactical (medium-term) and strategic (long-term) levels. In all three levels of decision-making, forecasting methods play a key role. Forecasts based on long-term (i.e., monthly and yearly) forecasting horizons are important for strategic and tactical planning decisions, for example, staffing, while forecasts based on short-term (i.e., hourly and daily) horizons help with operational planning decisions such as rostering (Cote, Smith, Eitel, and Akçali (2013)).

The problem of forecasting ED arrivals has been extensively studied. An exhaustive review and analysis on applications of statistical forecasting in hospital emergency departments were conducted by Gul and Celik (2020). They identified forty-three studies and observed that four categories of models, namely, time series, regression-based, data mining-based, and hybrid were prevalent in those studies. Our review of the literature is based on these four modelling strategies because it is the current research trend, and it will provide the reader with the rationale as to why the forecasting methods used in this thesis were chosen.

This chapter is organized as follows. Section 3.1 provides a classification of the type of forecasting methods used in studies on patient arrival forecasting. Details of the type of predictors usually considered important for patient demand/arrival are given in Section 3.2 and concluding remarks are provided in Section 3.3.

3.1 Methods

In an extensive part of the literature, studies compare forecasting models from multiple categories on the same dataset and select the method with the most accurate forecasts. This comparison excludes models labelled as “Blackbox” or “off-the-shelf”. In this context, the superiority of one model over another depends on the data under analysis (De Gooijer & Hyndman, 2006).

Most studies focus on analyzing total patient arrivals (e.g., Whitt and Zhang (2019), Rostami-Tabar and Ziel (2020)) while very few propose separated forecasting models for different types of patients (e.g., Sun, Heng, Seow, and Seow (2009), Chen et al. (2011)) since the arrival pattern as well as the availability of resources to serve them vary. Some other studies analyze patients in different severity levels to reveal that they could be modelled by the same set of predictors (e.g., M. Xu, Wong, and Chin (2013). Therefore, we developed separate forecasting models for different priority types. We considered low (CTAS 3,4, and 5) and high (CTAS 1 and 2) acuity patients because the treatment area of most, if not all, hospitals in Canada is divided based on this classification. We also considered walk-in patients (i.e., based on the mode of arrival) and WILAAB patients (i.e., walk-in patients who are low acuity and ambulatory) who are more likely to be deterred from visiting the ED by bad weather (e.g., snowstorm) and for whom sporting events could be a factor in deciding or planning a visit to the ED.

Below, Figure 3.2 presents a table that lists most of the methods surveyed by Gul and Celik (2020) while Figure 3.1 presents a table that defines all the acronyms used to describe the methods.

Methods		Predictors		Category	
Acronyms	**Description**	**Acronyms**	**Description**	**Acronyms**	**Description**
ANN	Artificial Neural Network	**AMaxT**	Mean maximum temperature	**DM**	Data mining models
ARIMA	Autoregressive Integrated Moving Average	**AMinT**	Mean minimum temperature	**Hy**	Hybrid
ARIMAX	Autoregressive Integrated Moving Average with exogenous variable(s)	**AQ**	Air Quality	**Reg**	Regression models
ES	Exponential Smoothing	**AT**	Average (mean) temperature	**TS**	Time series models
fbprophet	Facebook Prophet	**BHL/ AHL**	Days before holiday/ Days after holiday		
GA	Genetic Algorithm	**CUR**	Care utilization ratio		
GEE	Generalized Estimating Equation	**DOW**	Days of the week		
GLM	Generalized Linear Model	**HL**	Holidays		
HW	Holt-Winters	**HZR**	Health zone or residence		
LR	Linear Regression	**IF**	Influenza		
LR1	Logistic Regression	**MaxT**	Maximum Temperature		
MA	Moving Average	**MinT**	Minimum Temperature		
MHW	Multiplicative Holt-Winters	**MOY**	Months of the year		
MLP	Multilayer Perceptron	**PH**	Public holidays		
MSARIMA	Multivariate Seasonal Autoregressive Integrated Moving Average	**QOY**	Quarter of the year		
MVARIMA	Multivariate Vector Autoregressive Integrated Moving Average	**RF**	Rainfall		
SARIMA	Seasonal Autoregressive Integrated Moving Average	**RH**	Relative Humidity		
SARIMAX	Seasonal Autoregressive Integrated Moving Average with exogenous variable(s)	**SB**	School breaks		
SMHW	Seasonal Multiplicative Holt-Winters	**SF**	Snowfall		
SR	Systematic Review	**SMI**	Stock Market Index		
SS	Simple seasonal exponential smoothing	**TD/TG**	Temperature difference (gap)		
SVM	Support Vector Machine	**TOD**	Time of the day		
SVR	Support Vector Regression	**WOY**	Week of the year		
TBATS	Trigonometric Exponential Smoothing State Space model	**WS**	Wind speed (air velocity)		
VAR	Vector Autoregression	**YD**	Yellow dust		

Figure 3.1: Acronyms of methods, category of methods and predictors in Figure 3.2

Authors (year)	Region	Forecast Horizon	Methods	Predictors	Category
Sun et al. (2009)	Singapor	Daily	ARIMA	PH, AQ, AT, RH	TS
Jones et al. (2009)	U.S.A	Hourly	VAR	TOD, DOW	TS
Wargon et al. (2009)	Multiple	All	SR	DOW, MOY, HL, AT, RF	TS , Reg
Kam et al. (2010)	Korea	Daily	MA, SARIMA, MSARIMA	MOY, DOW, QOY, HL, AT, MinT, MaxT, TG, RF, SF, AV, RH, YD	TS
Chen et al. (2011)	Taiwan	Monthly	ARIMA	AMaxT, AMinT, RH, RF, SMI	TS
Chase et al.(2012)	U.S.A	Hourly	LoR	CUR	Reg
Marcilio et al. (2013)	Brazil	Daily	GLM, GEE, SARIMA	DOW, HL, BHL, AHL, IF, AT	TS, Reg
Cote et al. (2013)	U.S.A	All	LR	Yr, TOD, DOW, MOY	Reg
Xu et al. (2013)	Hong Ko	Daily	ANN, LR	DOW, IF, HL, RF, AT, WS	DM, Reg
Kadri et al. (2014)	France	Daily	ARIMA	DOW	TS
Bergs et al.(2014)	Belgium	Monthly	ES	DOW	TS
Mai et al. (2015)	Australia	Monthly	ARIMA, MVARIMA	DOW	TS
Aroua & Abdul-Nour (2015)	Canada	Weekly	LR, SARIMA, SARIMAX	WOY, AT, MaxT, MinT, TD, RF, SF	Reg, TS
Zlotnik et al. (2015)	Spain	Daily	SVR, M5P, GEE	DOW	Reg
Ekstrom et al. (2015)	Sweden	Hourly	LR	DOW, TOD	Reg
Calegari et al. (2016)	Brazil	Daily	SS, SMHW, SARIMA, MSARIMA	MOY, DOW, AT, MinT, MaxT,TG, RF, WS, RH	TS
Xu et al. (2016)	China	Daily	ARIMA–LR, GLM, ARIMA, ARIMAX, ARIMA-ANN	MinT, MaxT, PH, DOW, MOY, SB	TS, Reg, Hy
Afilal et al. (2016)	France	Daily	ARIMA	DOW	TS
Rosychuck et al. (2016)	Canada	Monthly	SARIMA	Age, Sex, HZR	TS
Juang et al. (2017)	Taiwan	Monthly	ARIMA	DOW	TS
Carvalho-Silva et al. (2018)	Portugal	Daily	ARIMA, SARIMA,MA,MHW, HW, ES	RF, MaxT	TS
Jiang et al. (2018)	Hong Ko	Daily	ANN, GA, SARIMA, ARIMAX, LR, SVM	HOD, DOW, MOY, PH, BHL, AHL, MinT, MaxT, MeanT, MDP, MRH, RF, WS	DM, Reg, TS,Hy
Yucesan et al. (2018)	Turkey	Daily	LR, ARIMA, ANN, ES, ARIMA–ANN, ARIMA-LR.	MOY, DOW, TOD	Reg, TS, DM, Hy
Camiat et al. (2019) *	Canada	Daily	fbprophet	TOD, DOW	Reg
Whitt & Zhang (2019)	Israel	Daily	LR, SARIMA,SARIMAX, MLP	DOW, MOY, HL, MaxT, MinT, RF, BHL, AHL	Reg, TS, DM, Hy
Rostami-Tabar & Ziel (2020)	U.K	Daily	Naïve, AR, ES, LR	DOW, HL	TS, Reg
Choudhury & Urena (2020)	U.S.A	Hourly	ARIMA, HW, TBATS, ANN	TOD, DOW	TS, DM
Tuominen et al. (2021)*	Finland	Daily	SARIMA, fbprophet, GLM	MOY, DOW	TS, Reg

Figure 3.2: Summary of reviewed literature. A table adapted from Gul and Celik (2020). Red star (*) indicates studies that used Facebook Prophet model

3.1.1 Times series models (TS)

Among the reviewed papers, the vast majority applied time series models to forecast ED daily arrivals (Sun et al. (2009), Jones et al. (2009), Kam et al. (2010), Chen et al. (2011), Marcilio et al. (2013), Kadri, Harrou, Chaabane, and Tahon (2014), Kim et al. (2014), Mai, Aboagye-Sarfo, Sanfilippo, Preen, and Fatovich (2015), Aroua and Abdul-Nour (2015), Ekström, Kurland, Farrokhnia, Castrén, and Nordberg (2015), Calegari et al. (2016), Q. Xu, Tsui, Jiang, and Guo (2016), Rosychuk, Youngson, and Rowe (2016), Carvalho-Silva, Monteiro, de Sá-Soares, and Dória-Nóbrega (2018), Yucesan, Gul, and Celik (2018), Whitt and Zhang (2019), Choudhury and Urena (2020)). The simplicity and effectiveness of this category of models make it attractive for practical applications (Kadri et al., 2014). The models focus on three components (i) long-term trends, (ii) cyclical changes, and (iii) the effect of unexpected events; these models predict future arrivals based on past arrivals (Wargon, Guidet, Hoang, & Hejblum, 2009). Among the time series models, the autoregressive integrated moving average (ARIMA) and its different extensions (e.g., Seasonal ARIMA, Multivariate ARIMA) were the most widely used.

One major disadvantage of ARIMA forecasting is the process of identifying the correct model. If done without using an automated statistical package, the process is usually computationally expensive and subjective. In such a situation, experience and skill are key to choosing a reliable model. This disadvantage can be overcome, however, using auto.arima function in R, which utilizes the Akaike Information Criterion (AIC) to identify the best ARIMA model by selecting the right combination of parameters within the arguments provided.

Although ARIMA models can be effective in capturing seasonality and the overall trend, they however are generally poor at predicting series with turning points (i.e., values that fall significantly outside the linear trend) which must be included in the analysis.

3.1.2 Regression-based models (Reg)

Cote et al. (2013) wrote a tutorial for ED directors advocating the use of regression analysis to forecast ED arrivals. The authors explained that regression models are capable of handling a wide variety of data patterns typically found in the ED arrival time series and that their results are easy to understand and interpret. In this category of models, the arrivals are considered to be a function of independent variables (e.g., day-of-week, month, average daily temperature etc.), with the relationship defined by a straight line. Cote et al. (2013) demonstrated how to use a simple linear regression with calendar variables to forecast yearly, monthly, daily, and hourly arrivals to an ED. Marcilio et al. (2013) applied two variants of the linear model, namely a generalized linear model (GLM) and generalized estimating equations (GEE), to forecast the total daily patient arrivals to an ED in Sao Paulo, Brazil. The models were explored with the mean daily temperature as the predictor variable. Ekström et al. (2015) used linear regression to predict daily arrivals using visits to health care guide websites between 6 pm and midnight and day of the week as independent variables. The authors believe that using internet data allows for a model that reflects a behavioural trend. Whitt and Zhang (2019) investigated a linear regression model based on calendar and weather variables, while Rostami-Tabar and Ziel (2020) proposed a penalized linear regression model to generate both point and probabilistic daily forecast of ED attendance, considering fixed-date, flexible-date, and long-date holiday-based events in addition to existing effects in the historical data.

Unlike classical linear regression which depends on the sum of the effects of individual predictors, generalized additive models (GAMs) use the sum of smooth functions to predict outcomes. Facebook Prophet (fbprophet) uses an additive regression model with four components: (1) a piecewise linear or logistic growth curve to detect changes in trends by selecting change points from the historical data, (2) a seasonal component modelled using Fourier series, (3) a user-provided list of relevant holidays, and (4) regressor effect in the prediction of future ED arrivals (Taylor & Letham, 2018).

According to Taylor and Letham (2018), the fbprophet modelling approach blends the advantages of a fully automated statistical method and judgmental forecasting. The integration of both the statistical and judgemental methods have been discussed, in the forecasting literature, as the best way to obtain a more reliable forecast (Soyiri and Reidpath (2013), Hyndman and Athanasopoulos (2018)). Judgemental methods are used to adjust statistical forecasts, taking account of information that was not able to be incorporated into the statistical model (Hyndman, 2011). This characteristic of fbprophet is particularly useful for ED administrators who are non-statisticians/non-data-scientists but have extensive ED settings knowledge which is often needed to make reliable time-series forecasts.

However, none of the papers considered in Gul and Celik's survey (Figure 3.2) used Facebook Prophet (fbprophet), because it is viewed as a "Blackbox" or "off-the-shelf" method. On the other hand, some researchers have used fbprophet as a module towards some bigger goal, such as scheduling with no real justification for why it was chosen over other approaches. For example, Camiat, Restrepo, Chauny, Lahrichi, and Rousseau (2021) estimated the future demand (patient arrival) as a module in a scheduling framework using fbprophet.

3.1.3 Data mining-based models (DM)

Given the growing availability of data and computing power, data-mining methods like artificial neural networks (ANNs), convolutional neural networks (CNNs), recurrent neural networks (RNNs), etc. have become an important forecasting method considered by researchers. According to Gul and Celik (2020), the most frequently applied data mining-based models to forecast ED patient arrivals are ANNs. These models are data-driven, self-adaptive, and can model both linear and non-linear processes, which is common in the real world (Khashei & Bijari, 2010).

M. Xu et al. (2013) proposed a three-stage methodology to evaluate and forecast ED daily arrivals. The ANN model was constructed to represent the association between key variables identified in the first stage and patient arrivals and was trained and validated before being compared with the two benchmarking methods. Their result showed the ANN method outperformed the regression models and they recommended the use of the ANN model in practice. Whitt and Zhang (2019) investigated alternative models, including the multilayer perceptron (MLP) model, which is an ANN machine learning method, to forecast daily arrival totals using publicly available data of the Rambam Hospital ED in Haifa, Israel. Their results showed the MLP was outperformed by a SARIMA model. Yucesan et al. (2018) compared ANN to other models in forecasting the number of daily patient arrivals, showing that the ANN model outperformed the linear regression model using the same input variables.

A major detriment of using ANN is that it is computationally very expensive, time-consuming to train with traditional CPUs, and requires a lot of data (Donges, 2021). Furthermore, ANN modelling cannot be easily reproduced by hospital managers if the model is to be updated with new data.

3.1.4 Hybrid models (Hy)

Hybrid models are integrations of two or more methods. These models provide higher accuracy performance in many applications (Yucesan et al., 2018). Common combinations include regression + data mining, regression + time series, and regression + time series + data mining. The hybrid model combines the advantages of the individual models.

Using two real-world data sets collected from hospitals in DaLian, LiaoNing Province, China, Q. Xu et al. (2016) compared the forecasting performance of the regression + time series (ARIMA–LR) hybrid approach with several widely used models namely GLM, ARIMA, ARIMA with explanatory variables (ARIMAX), and ARIMA–ANN hybrid model. Their result showed the hybrid ARIMA–LR model outperformed existing models in terms of forecasting accuracy. M. Xu et al. (2013) used ARIMA-ANN to forecast the ED arrivals. Whitt and Zhang (2019) investigated alternative models to forecast daily arrival totals using publicly available data of the Rambam Hospital ED in Haifa, Israel. Their results showed the seasonal autoregressive integrated moving average with exogenous regressors (SARIMAX) model outperformed the others.

None of the forty-three papers surveyed by Gul and Celik from 2009 to 2018 explicitly mentioned regression with autoregressive integrated moving average errors (regARIMA) despite it being an established forecasting method. In contrast, another model, ARIMAX, of the same family of dynamic regression models as regARIMA, has been widely used. In regARIMA, the effects of the exogenous variables are not mixed up with the effects of the autoregression part, as is the case with ARIMAX, and hence the regression coefficients are interpreted the usual way (Hyndman, 2010). This characteristic makes regARIMA very attractive as it avoids interpretation errors which can be a serious issue in real-world applications.

3.2 Important predictors

Some factors have been shown to influence patient arrival rates in the literature of ED arrival forecasting (see Figure 3.2).

Snowfall, for example, has been shown to reduce ED volume in a study carried out in Toronto, Canada by Shah, Murray, Mamdani, and Vaillancourt (2019). Meanwhile, the predictive impact of temperature on daily ED patient arrival continues to be uncertain as the effect of the temperature variable depends on the geographical location and characteristics of the ED (Wargon et al. (2009), Marcilio et al. (2013)).

Also, the inclusion of public holidays as a factor of ED patient arrival has yielded contradictory results (Rostami-Tabar & Ziel, 2020). Rostami-Tabar and Ziel (2020) highlighted that some studies reported an increased ED patient volume on holidays, while others reported a decrease.

However, the impact of calendar variables is highly significant in ED arrivals forecasting. For example, Whitt and Zhang (2019) found that the day-of-week factor explained most of the variance in the daily arrival totals. They observed that being the first day of the week in Israel, Sunday, had the highest number of patient visits while the weekend days (i.e., Friday and Saturday) had fewer patient visits. In another study carried out in Singapore, Sun et al. (2009) reported that the highest arrival rates happened on Mondays. The authors attributed the high demand on Mondays to the closure of public primary care facilities on Sundays.

Wargon et al. (2009) named local events, such as fairs and exhibits, and sporting events as variables that may affect patient arrivals. Kadri et al. (2014) suggested that the variation of arrivals according to day-of-the-week in their study could be as a result of special events such as festivals and sports in the region. Hughes et al. (2018) highlighted that soccer games have a relatively minor influence on ED arrivals, detectable at the hourly level. They explained that a few hours before and during the period of a soccer game, the numbers of arrivals at the EDs where the local population has an interest in the game may be lower than usually expected, while the arrival may increase immediately following the end of the game. The authors also pointed out that studies that have investigated the impact of sporting events (both live and televised) on daily ED arrivals have yielded contrasting results: from no impact observed to increased arrivals due to assault. Ranse et al. (2019) demonstrated that there is an impact on ED services from mass gathering events while Weiss and Elixhauser (2016) demonstrated that some ED arrivals are sports-injury related.

3.3 Conclusion

In this chapter, we reviewed forecasting models for ED patient demand, which were grouped into four categories based on Gul and Celik's classification, namely time series, regression-based, data mining-based and hybrid models. There are advantages and disadvantages of utilizing models within and between these different classes of models. However, the literature does not portray a particular model or class of models as the best. Instead, it reinforces the idea that the superiority of one model over another depends on the data under analysis. Hence researchers try different models for a particular dataset.

We observed that regARIMA model, an established forecasting method, has not been mentioned in the literature, but ARIMAX has been widely used. Furthermore, the performance of fbprophet has not been compared to other established forecasting methods but has been used as a module in bigger studies.

We also reviewed independent variables considered important for predicting patient arrival to the ED. Weather variables, holidays and sports events were reported to have yielded contradictory results in terms of impacting patient arrival to the ED. That is, some studies reported an increase in ED patient arrival, while others reported a decrease. However, the impact of calendar variables, particularly day-of-week (DOW) was reported to be highly significant in ED arrivals forecasting.

In the subsequent chapter, we will present our study on forecasting the daily ED patient demand of an acute care institution located in Montreal, Quebec.

Chapter 4

Forecasting Daily Number of Patient Arrivals to An Emergency Department

4.1 Introduction

The arrival of patients with more urgent and complex care needs, the increase in the arrival of the elderly, and the high volume of low-acuity patient arrivals are some of the key causes of a supply-demand mismatch in emergency departments (EDs) resulting in prolonged ED waiting times and eventually problems with overcrowding (Morley et al., 2018). A reliable prediction of future arrivals (demand) provides ED administrators with insights necessary to make resource allocation decisions to eliminate the supply-demand mismatch and improve the operational efficiency and healthcare delivery of the ED.

There is extensive literature on comparing forecasting methods for ED arrival forecasting (Wargon et al. (2009), Gul and Celik (2020)). However, of the forty-three papers surveyed by Gul and Celik from 2009 to 2018, none explicitly mentioned regression with autoregressive integrated moving average errors (regARIMA) despite it being an established forecasting method. In contrast, another model, ARIMAX, of the same family of dynamic regression models as regARIMA, has been widely used. In regARIMA, the effects of the exogenous variables are not mixed up with the effects of the autoregression part, as is the case with ARIMAX, and hence the regression coefficients are interpreted the usual way (Hyndman, 2010). This characteristic makes regARIMA very attractive as it avoids interpretation errors which can be a serious issue in real-world applications.

Also, none of the papers considered in Gul and Celik's survey (Table 1, Chapter 3) used Facebook Prophet (fbprophet), because it is viewed as a "Blackbox" or "off-the-shelf" method and researchers want to develop more custom models. Yet, on the other hand, some researchers have used fbprophet as a module towards some bigger goal, such as scheduling (e.g., Camiat et al. (2021)) with no real justification for why it was chosen over other approaches, e.g., an ARIMA model.

According to Taylor and Letham (2018), the fbprophet modelling approach blends the advantages of fully automated statistical method and judgmental forecasting method. The integration of both the statistical and judgemental methods has been discussed, in the forecasting literature, as the best way to obtain a more reliable forecast (Soyiri and Reidpath, 2013, Hyndman and Athanasopoulos, 2018). Judgemental methods are used to adjust statistical forecasts, taking account of information that was not able to be incorporated into the statistical model (Hyndman, 2011). This characteristic of fbprophet is particularly useful for ED administrators who are non-statisticians/non-data-scientists but have extensive ED settings knowledge which is oftentimes needed to make reliable time-series forecasts.

Given this characteristic of fbprophet and the fact that there is literature evaluating forecasting methods without considering fbprophet, and literature that just uses fbprophet without much justification, in this chapter we investigate the naturally-arising questions: how does fbprophet compare to a custom-designed ARIMA model, and is fbprophet then justified to be used as a module in bigger studies?

In this chapter, we contribute to the existing literature by evaluating the effectiveness of the rarely used regARIMA model in forecasting future daily ED arrivals. Also, we experimentally evaluate the performance of fbprophet, a model that has not been thoroughly studied in the literature, and demonstrate its competitiveness with established forecasting methods. This insight is particularly valuable given that several papers have been using fbprophet as a step towards a bigger framework without evaluating its actual predictive performance. Although this is just one study, our results provide some justification for the use of fbprophet in studies that may use it as a part of a bigger framework. These contributions are demonstrated within the context of a Montreal hospital ED.

The remainder of this chapter is organized as follows. Section 4.2 presents the methodology. Detailed results and interpretations are provided in Section 4.3. Concluding remarks are provided in Section 4.4. Limitations and future research directions are described in Section 4.5.

4.2 Methodology

Our goal is to develop and evaluate models for predicting the number of patients arriving at the emergency department of a Montreal hospital. We first develop a custom regARIMA model and then compare it with the out-of-the-box fbprophet method.

4.2.1 Data

The data used consists of five emergency room patient arrival categories: low acuity patients (i.e., with CTAS 4 and 5); high acuity patients (i.e., with CTAS 1, 2, and 3); walk-in patients (i.e., based on the mode of arrival); WILAAB patients (i.e., walk-in patients who are low acuity and ambulatory); and total patients (i.e., a summation over all patient categories).[1]

4.2.2 Predictors

The choice of predictors is based on both background knowledge from the literature review and discussion with ED physicians.

In the literature of ED arrival forecasting, prior arrivals are considered the most important predictors of future arrivals, which is the basic assumption of time series models (Wiler, Griffey, & Olsen, 2011). Calendar variables such as day-of-week are highly significant. Some authors (e.g., Whitt and Zhang (2019), Sun et al. (2009)) found that the day-of-week factor explained most of the variance in daily ED arrival totals. Meanwhile, Rostami-Tabar and Ziel (2020) highlighted that some studies report an increase in ED patient volume on holidays, while others report a decrease. The importance of weather variables on ED arrivals depends on the geographical location and characteristics of the ED (Wargon et al. (2009), Marcilio et al. (2013)). Snowfall, for example, has been shown to reduce ED volume in a study carried out in Toronto, Canada by Shah et al. (2019). Some

[1] For a description of the CTAS classifications, see Chapter 2.

studies have highlighted that including temperature adds uncertainty to the forecasting model in exchange for little improvement on forecasting accuracy (Marcilio et al., 2013). At the Montreal hospital ED, the physicians had observed higher patient arrival rates on the days when there is a hockey game. Mass gathering events, such as fairs and exhibits, and sports have been suggested as variables that may affect patient arrivals (Wargon et al. (2009), Kadri et al. (2014)). Correia, Braillard, Combescure, Gerstel, and Spechbach (2018) concluded that their results support the hypothesis that the broadcasting of large-scale sporting events such as tennis matches decreases admission rates in emergency units. However, Hughes et al. (2018) pointed out that studies that have investigated the impact of sporting events (both live and televised) on daily ED arrivals have yielded conflicting results: from no impact to increased arrivals due to assault.

For possible predictors, we consider

- daily snowfall data obtained from the Government of Canada (Government of Canada, 2021),
- binary data representing daily sport event dates of local popular sports (hockey, soccer, tennis, and Formula 1 Grand Prix) obtained from online sources (i.e., Sport reference (2021), Major League Soccer (2021), National Bank Open (2021), and Grand Prix Canada (2021)),
- binary data representing daily festival event dates (Ville de Montréal, 2021),
- binary data representing dates of Quebec public holidays and school/university breaks obtained from online sources (i.e., Montreal Public Holidays (2021) and English Montreal School Board (2021)),
- day-of-week (DOW) obtained from the calendar.

4.2.3 Modelling

The modelling approach used in this thesis is based on considering a set of all possible candidate models for each arrival category, created by all possible combinations of the potential predictors. That is, given p predictors, 2^p models are fitted and the model with the lowest AICc is chosen as the "best" predictive model. The models are fitted using an automated algorithm, auto.arima() from the forecast package (Hyndman & Khandakar, 2008) in RStudio, version 1.3.1093.

We began by creating a design (or model) matrix, whereby each row represents an individual model and the successive columns correspond to the predictor variables to be included in the model. In the matrix, "1" indicates predictors to be considered in each model during the fitting process, and "0" indicates predictors to be left out. Since we are investigating five potential predictors, a model matrix of $2^5 = 32$ rows and 5 columns are created for each category of arrivals (Table 4.1).

Table 4.1: Model matrix based on potential predictors

Model	Snowfall	Sports	Festivals	Holidays	DOW
1	0	0	0	0	0
2	0	0	0	0	1
⋮	⋮	⋮	⋮	⋮	⋮
32	1	1	1	1	1

Next, we divide the dataset into training and test sets. The first 337 days of the dataset, covering the period from April 1, 2017, to March 3, 2018, are used as the training set and the remaining 28 days, covering the period from March 4, 2018, to March 31, 2018, are used as the test set. The training set is tested for stationarity using both the Augmented Dickey-Fuller (ADF) and the Kwiatkowski-Phillips-Schmidt-Shin (KPSS) tests, to ensure that the series is truly stationary. The null hypothesis for the ADF test is that the series is not stationary while for the KPSS test it is that the series is stationary. Both tests may have consistent or contradictory results (Seabold & Perktold, 2010). Possible contradictory results can be summarized as follow:

(1) ADF indicates stationarity and KPSS indicates non-stationarity. In such a scenario, the series is difference stationary. That is, differencing is to be used to make the series stationary and the differenced series is checked for stationarity.

(2) ADF indicates non-stationarity and KPSS indicates stationarity. In such a scenario, the series is trend stationary. The trend needs to be removed to make the series strict stationary and the detrended series is checked for stationarity. The simplest method to detrend a time series is by differencing.

In practice, differencing of a time series means subtracting subsequent observations from one another, following the formula:

$$diff(t) = x(t) - x(t-1)$$

where $diff(t)$ is the differenced series at time t and $x(t)$ is an observation of the original series at time t.

Auto Correlation Function (ACF) and partial ACF plots are created to verify seasonal patterns in the training data. Then, considering the information in each row of the design matrix, auto.arima() searches by trying all the possible parameters (within the arguments provided) and returns the model that minimizes AIC for the time series data with the corresponding AICc value. For more information on how auto.arima() works, see Section 8.7 of Hyndman and Athanasopoulos (2018).

The list of candidate models is sorted in ascending order based on the AICc values, and the AICc values are rescaled to

$$\Delta_i = AICc_i - AICc_{min} \tag{1}$$

where i represents the model number ($i = 1, \ldots, 32$) and $AICc_{min}$ is the minimum $AICc_i$ value. This transformation forces the "best" model to have $\Delta_i = 0$, while the rest of the models have positive values. According to Fabozzi, Focardi, Rachev, and Arshanapalli (2014), $\Delta_i \leq 2$ indicates there is substantial evidence to support that candidate model i is almost as good as the best model. In another study by Burnham, Anderson, and Huyvaert (2011), the authors stated that the cut-off should be at $\Delta_i \leq 7$. For this study, we use the Fabozzi et al. (2014) cut-off. We refer to models with $\Delta_i \leq 2$ as *Best subset models*. Residual analysis is performed to evaluate the goodness of fit of the model with the lowest AICc value.

The fbprophet procedure does not provide goodness-of-fit tests; hence, the design matrix was not used in fitting all possible fbprophet models. Instead, fbprophet models are built with the predictors included in the best ARIMA-based models. The fbprophet algorithm automatically detects changes in trends by selecting changepoints from the data, uses Fourier series to model seasonality, and fits nonlinear trends while adding predictors as exogenous variables (Taylor & Letham, 2018).

Both the fbprophet and the best regARIMA models are then used to create forecasts.

4.2.4 Forecast evaluation

To validate forecasting models, we adopt the "rolling-origin-update evaluation" technique explained by Tashman (2000). The basis of the rolling-origin-update approach is to create a more robust evaluation of the models' forecast ability within a context of a real-world application, where a model is built once, and used later with updated information as new values become available. That is, forecasts for a fixed horizon are performed by sequentially using past values from the test set to update the input information of the model, and changing the forecast origin accordingly. This iterative process continues until the end of the dataset. Below is a step-by-step description of the process:

(1) The models are fitted to the training dataset covering the period from April 1, 2017, to March 3, 2018 (337 days).

(2) The forecast values are calculated for week 1 of the test set (March 4, 2018, to March 10, 2018).

(3) The week 1 forecast values are then compared with the actual values by calculating Root Mean Square Error (RMSE), Mean Absolute Error (MAE), maximum of the maximum between zero and the difference in actual and forecast arrivals (Max. UNForcast), and the percentage of the actual values that fall within the 95% prediction interval.

(4) The forecasting origin is moved forward by 7 days with the additional data point used to update the input to the model. The forecasts for week 2 of the test set are generated, followed by a calculation of accuracy measures based on the forecasted values and the actual values.

(5) Steps II to IV are repeated until the end of the data set. A total of four weekly forecasts are generated.

4.3 Results and discussion

Preliminary data analysis of the 39,950 patients who visited the ED between April 1, 2017, and March 31, 2018, 77.3% of the arrivals were walk-ins, 22% were ambulance arrivals, 0.15% were

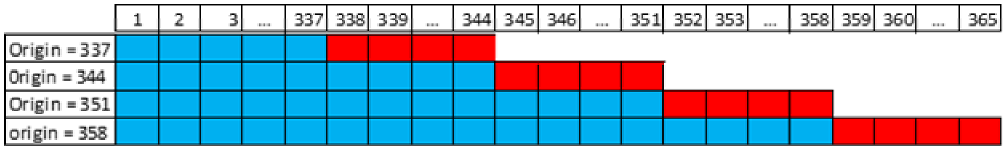


Figure 4.1: Rolling origin with constant 7 days holdout. Figure adapted from Svetunkov and Petropoulos (2018)

adapted transport arrivals, 0.04% were prison transport arrivals, and 0.45% were accompanied by police. For the patients who arrived by ambulance, 83.2% were classified as stretcher patients while 32% of walk-in arrivals were classified as stretcher patients.

At the Montreal ED, the resources needed for treating stretcher and ambulatory patients are different. Given that stretcher patients tend to be at higher acuity and consume more resources, it is crucial to ensure there are enough resources to meet this category of the demand quickly.

On average, 109 patients visit this ED daily. The majority of the daily demand is from the high acuity category and most of the arrivals are through walk-ins (Table 4.2). Also, the daily arrival variation within the week (Figure 4.2) and the year (Figure 4.3), in the form of boxplots, show that the greatest demand (i.e., the median number of arrivals) occurred on Monday and in September.

The greatest demand occurring on Monday is consistent with empirical knowledge and real observations. The same observation has been reported in previous studies (M. Xu et al., 2013). A possible explanation is based on the fact that most people visit EDs on weekdays because doing so allows them to take a day off especially given that Monday is the first day of the week. Also, September is full of events and activities in Montreal. It is the season for popular music, arts, design, cultural festivals, and Labor Day weekend activities (every first Monday in September is Labor Day under Quebec law). Schools and universities also reopen in September after the summer holidays.

Table 4.2: Descriptive statistics of the Montreal hospital daily ED arrivals by category

Horizon	**Arrival categories**	**Mean**	**Median**	**Max.**	**Min.**	**St. Dev**
Day	Total	109.5	109	137	76	11.1
	High Acuity	78.2	79	110	50	10.5
	Low Acuity	31.3	31	59	12	7.7
	Walk-in	84.6	84	112	38	11
	WILAAB	26	26	48	9	6.5

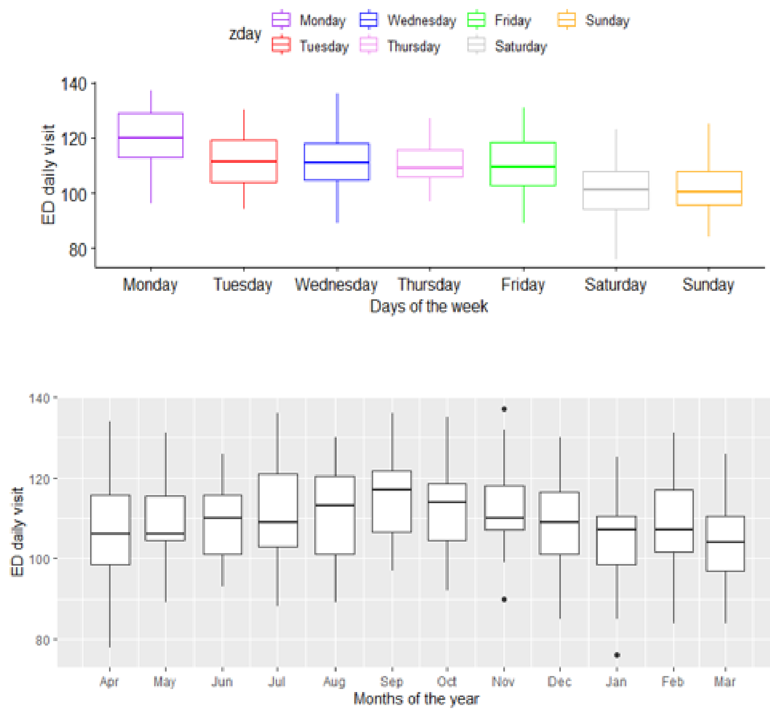


Figure 4.3: The number of daily arrival totals in a month view.

A time series plot of the patient arrival data (Figure 4.4) shows no obvious pattern in any of the arrival categories, but a slight decrease in the trend towards the end of the dataset. This suggests that the arrivals in each category are non-stationary. However, the ADF test rejects the null hypothesis of non-stationarity for all arrival categories, indicating there is sufficient evidence to support the stationarity of the daily arrival time series.

A time-series data is stationary when its properties such as mean, variance, autocorrelation, etc are not time-dependent but remain constant over time. This is important in modelling because, without a stationary series, the model describing the data will vary in accuracy at different time points.

4.3.1 Fitted models

The ACF plots (Figure 4.5) of the daily arrivals for the different patient categories show positive significant spikes on lags 7, 14, 21, 28, 35, and 42 for three of the arrival categories, indicating weekly seasonality. However, for the low acuity and WILAAB arrival categories, the significant spikes do not follow this pattern, they seem to occur on lags 14 and 28 (Figure 4.5 (c) and (d)). This

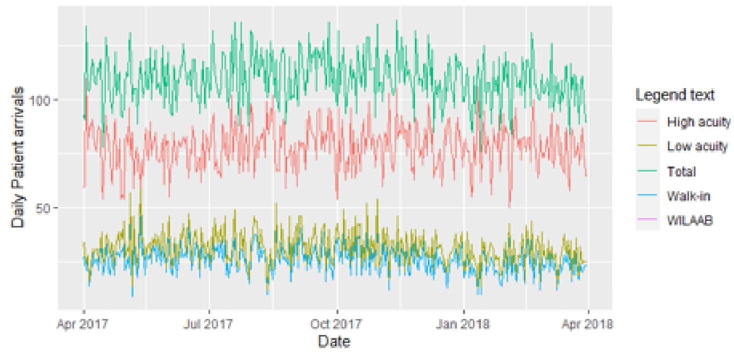


Figure 4.4: Time series plot of daily arrivals by category.

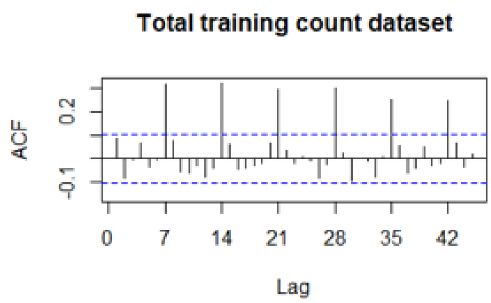
result suggests that for the total, walk-in and high acuity categories, there is a regular and predictable weekly pattern in the arrivals. For example, if the number of arrivals for Monday is known for a given week, we can have some degree of confidence that about the same number of patients will be arriving the Monday of the following week.

For the row with all zeros in the design matrix (Table 4.1), auto.arima() fits a Seasonal Autoregressive Integrated Moving Average (SARIMA), which is the ARIMA model we further consider as the base model in each arrival category (Table 4.3).

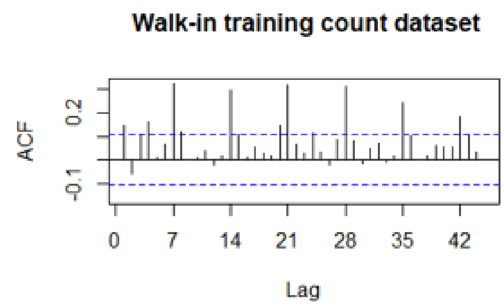
SARIMA is an extension of the ARIMA model that accounts for seasonal components. The model is described by seven parameters: (p, d, q) $(P, D, Q)[s]$. Where

- p represents the order of the autoregressive factor (AR),
- d represents the order of differentiation required to reduce nonstationarity in the data (I),
- q represents the order of the moving-average model (MA),
- P corresponding to p, represents the seasonal order of the autoregressive factor (SAR),
- D corresponding to d, represents the seasonal order of differentiation required to reduce nonstationarity in the data (I),
- Q corresponding to q, represents the seasonal order of the moving-average model (SMA), and
- s represents the seasonal lag which is 7 in our case since we have weekly seasonality.

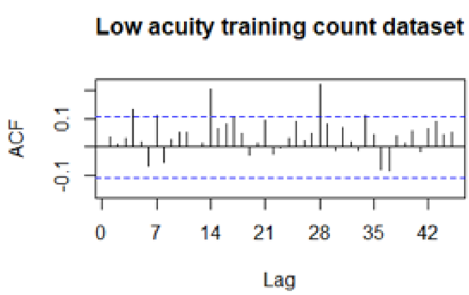
For example, if $p = 2$, then there are two AR parameters (i.e., ar1 and ar2).



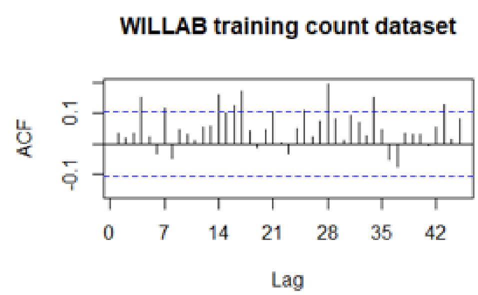
(a) Total training set



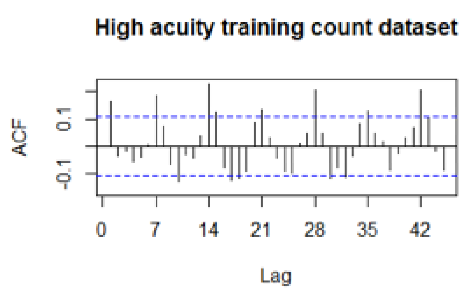
(b) Walk-in training set



(c) Low acuity training set



(d) WILAAB training set



(e) High acuity training set

Figure 4.5: The ACF of the Daily arrival for the different patient categories.

Table 4.3: Estimated coefficients of the base models for each arrival category.

Arrival Category	Base Model	Coefficients	Estimates.	Std. Error
Total	ARIMA(0,0,1)(2,0,0)[7]	ma1	0.11	0.06
		sar1	0.23	0.05
		sar2	0.26	0.05
		intercept	109.86	1.16
High acuity	ARIMA(0,0,1)(2,0,0)[7]	ma1	0.15	0.06
		sar1	0.14	0.05
		sar2	0.20	0.06
		intercept	78.38	0.94
Low acuity	ARIMA(0,1,1)(2,0,2)[7]	ma1	-0.98	0.02
		sar1	0.09	0.19
		sar2	0.72	0.19
		sma1	-0.04	0.21
		sma2	-0.56	0.22
Walk-in	ARIMA(2,1,1)(2,0,0)[7]	ar1	0.10	0.06
		ar2	-0.13	0.06
		ma1	-0.97	0.01
		sar1	0.22	0.06
		sar2	0.19	0.06
WILAAB	ARIMA(0,1,1)(1,0,2)[7]	ma1	-0.98	0.02
		sar1	0.92	0.09
		sma1	-0.88	0.11
		sma2	0.04	0.06

Given that the best subset of models (i.e., models with $\Delta_i \leq 2$ from all possible combinations in Table 4.1) have similar predictive power, and knowing that each predictor contributes to a model's predictive power, we examine the pattern of the predictors in this subset of models to identify the most important predictor(s). If, for example, the best model in the subset (i.e., model with $\Delta_i = 0$) contains both holiday and sport as predictors and all the other models contain sports but not all contain holiday, then we consider that sport is more important than holidays in forecasting future arrivals, given the data. Tables 4.4 to 4.8 present the best subsets of models (out of 32 fitted), identified as those models with $\Delta_i \leq 2$ (where Δ_i is calculated according to (1)), for each arrival category. Note that the models with regressors are all regARIMA models.

We found that for the best model in the total arrival category, holidays and DOW are important predictors (Tables 4.4). For the other arrival categories, DOW is the important predictor (Tables 4.5

– 4.8).

This result was consistent with the literature. Holidays have been shown in the literature to be an important predictor of overall patient arrival (i.e., total arrivals increase or decrease during holidays). Calendar variables (e.g., DOW) has also been shown to be highly significant in ED arrivals forecasting (Sun et al. (2009); Whitt and Zhang (2019); Rostami-Tabar and Ziel (2020)). Also, when patient arrivals have been categorized, the holidays have not been important in predicting future arrivals (see M. Xu et al. (2013))

In Table 4.6, Festivals does not conform to the pattern of "the best predictor is the one that is present in all models in the best subset", but DOW does. Hence, DOW is kept as the most important predictor for the walk-in arrival category.

Given that the best model contains Festivals, which has been ruled out as an important predictor, we selected the next best model (with $\Delta_i = 0.4$) which contains the DOW predictor.

In Table 4.8, the base model is included as the best subset model which causes DOW to not completely conform to the pattern of "the best predictor is the one that is present in all models in the best subset". However, we kept DOW as an important predictor for the low acuity arrival category because the base model is the only model which does not include DOW, and it is not the next best model in the list, based on AICc value.

Table 4.4: Best subset models for Total arrivals. Chosen model is highlighted in blue.

Model	Sports	Festivals	Holidays	Snow	DOW	AICc	delta
M27	0	0	1	0	1	2467.74	0
M11	1	0	1	0	1	2469.74	2

Table 4.5: Best subset models for High acuity arrivals. Chosen model is highlighted in blue.

Model	Sports	Festivals	Holidays	Snow	DOW	AICc	delta
M31	0	0	0	0	1	2482.22	0
M23	0	1	0	0	1	2483.1	0.88
M27	0	0	1	0	1	2484.11	1.89

Table 4.6: Best subset models for Walk-in arrivals. Chosen model is highlighted in blue.

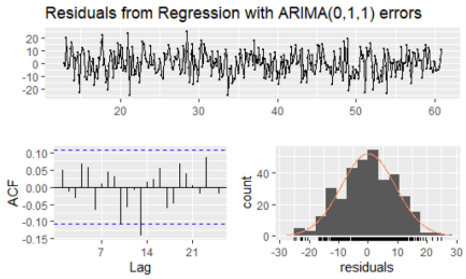
Model	Sports	Festivals	Holidays	Snow	DOW	AICc	delta
M23	0	1	0	0	1	2449.96	0
M31	0	0	0	0	1	2450.36	0.4
M19	0	1	1	0	1	2451.46	1.5
M27	0	0	1	0	1	2451.47	1.51
M7	1	1	0	0	1	2451.71	1.75

Table 4.7: Best subset models for WILAAB arrivals. Chosen model is highlighted in blue.

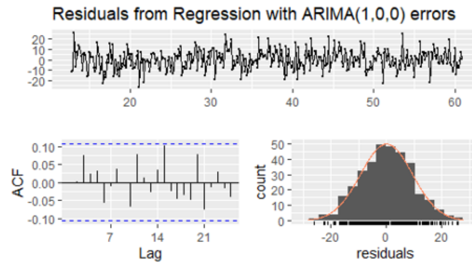
Model	Sports	Festivals	Holidays	Snow	DOW	AICc	delta
M31	0	0	0	0	1	2206.44	0
M15	1	0	0	0	1	2207.73	1.29
M23	0	1	0	0	1	2208.43	1.99
M29	0	0	0	1	1	2208.44	2

Table 4.8: Best subset models for Low acuity arrivals. Chosen model is highlighted in blue.

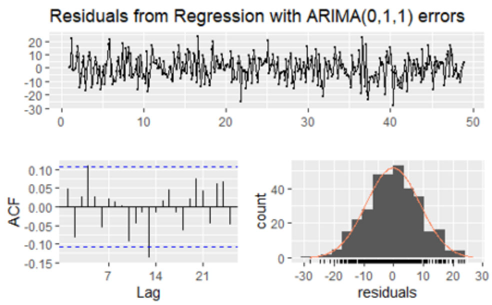
Model	Sports	Festivals	Holidays	Snow	DOW	AICc	delta
M31	0	0	0	0	1	2313.43	0
M29	0	0	0	1	1	2313.98	0.55
M23	0	1	0	0	1	2314.14	0.71
M15	1	0	0	0	1	2314.71	1.28
M27	0	0	1	0	1	2315	1.57
M21	0	1	0	1	1	2315.03	1.6
M32	0	0	0	0	0	2315.07	1.64
M19	0	1	1	0	1	2315.3	1.87
M7	1	1	0	0	1	2315.43	2



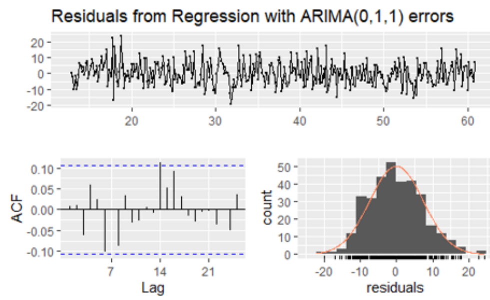
(a) Total arrival



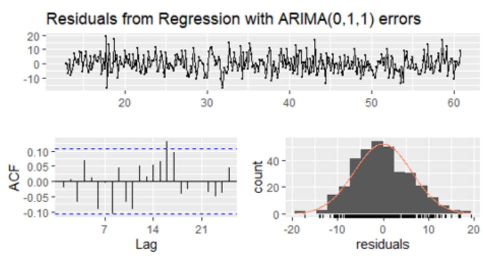
(b) High acuity arrival



(c) Walk-in arrival



(d) Low acuity arrival



(e) WILAAB arrival

Figure 4.6: Residual analysis of the regARIMA forecasting models.

Figure 4.6 shows the results of residual analysis for the best model in each arrival category. Based on the diagnostic measures of model residuals, the assumptions of the regARIMA model fit appear satisfied for all arrival categories. The top subplot is the plot of residuals over time, which shows that the variation of the residuals stays much the same across the historical data, and therefore the residual variance can be treated as constant. The bottom-left subplot is the ACF plot of the residuals, which indicates that the serial correlation present in the data has adequately been dealt with; all but a few spikes of the ACF plots of some of the models exceed the upper and lower limit and hence provide a good fit. The Histogram (bottom-right) shows reasonable conformity with normality for all categories.

After choosing the best model from the best subsets presented in Tables 4.4 to 4.8 (highlighted in blue) and confirming that those models satisfy the appropriate characteristics of residual analysis, the order and regressor(s) of the best ones for each category are presented in Table 4.9.

Table 4.9: The order and regressor(s) of regARIMA models by arrival category

Arrival category	**ARIMA error order**	**Regressors**
Total	(0,1,1)	Holidays and Day-of-week
High Acuity	(1,0,0)	Day-of-week
Low Acuity	(0,1,1)	Day-of-week
Walk-in	(0,1,1)	Day-of-week
WILAAB	(0,1,1)	Day-of-week

From Table 4.9, we see that, of all the arrival categories, only the high acuity category was not differenced (implying the second stationarity check via the KPSS test failed to reject).

The low acuity, walk-in and WILAAB arrival categories resulted in similar forecasting models being chosen and using the same predictor, i.e., DOW. This was not the result we expected. Our reason for including the WILAAB arrival stream, in particular, was centred on finding out the importance or influence of the sports predictor to the ED arrivals. By analyzing the WILAAB arrival category, which is made up of walk-in patients who are low acuity and ambulatory, we expected sports to be an important factor for the WILAAB arrivals.

On the other hand, it was not surprising that all three arrival streams resulted in similar forecasting models with the DOW as important predictors. In our data, over 70% of the walk-in patients are ambulatory of which a vast majority are classified as low acuity. For low acuity patients, DOW is a factor in deciding or planning a visit to the ED. This is in contrast to high acuity patients whose decision to go to the ED is mainly based on their severity irrespective of what day of the week it is.

We, therefore, continued the analysis with three arrival categories: total, low acuity and high acuity. Our choice is based on the fact that the arrival pattern as well as the resources to serve them vary between low acuity and high acuity categories at the Montreal hospital ED and most, if not all, hospitals in Canada. In our data, 77.3% of the total arrivals were walk-ins of which over 70% are low acuity. Also, 22% of the total arrivals were ambulatory and over 94% of ambulatory patients are registered with CTAS scores 3, 4, and 5.

Table 4.10: Estimated coefficients of regARIMA models by arrival category.

Arrival Category	Best model	Coefficients	Estimates	Std. Error
Total	regARIMA (0,1,1)	ma1	-0.95	0.02
		Mon	17.79	1.84
		Tue	9.29	1.85
		Wed	8.66	1.85
		Thu	8.10	1.84
		Fri	8.50	1.84
		Sat	-1.21	1.83
		Holidays	-2.05	1.39
High acuity	regARIMA (1,0,0)	ar1	0.15	0.05
		intercept	73.34	1.37
		Mon	12.54	1.78
		Tue	7.08	1.91
		Wed	8.06	1.93
		Thu	7.07	1.93
		Fri	2.90	1.91
		Sat	-2.32	1.78
Low acuity	regARIMA (0,1,1)	ma1	-0.96	0.02
		Mon	5.21	1.47
		Tue	2.39	1.47
		Wed	0.79	1.48
		Thu	1.17	1.48
		Fri	5.63	1.48
		Sat	1.20	1.47
Walk-in	regARIMA (0,1,1)	ma1	-0.94	0.02
		Mon	15.05	1.79
		Tue	7.78	1.79
		Wed	6.68	1.79
		Thu	7.03	1.79
		Fri	6.71	1.79
		Sat	-1.66	1.78
WILAAB	regARIMA (0,1,1)	ma1	-0.96	0.01
		Mon	3.65	1.26
		Tue	1.24	1.26
		Wed	-0.65	1.26
		Thu	1.15	1.26
		Fri	4.45	1.26
		Sat	1.11	1.25

Table 4.10 shows the estimated regression coefficients and the standard error of the regARIMA models for the different arrival categories which are interpreted the usual way. For example, in the total arrival category, about 17.8 more arrivals should be expected on a typical Monday than a typical Sunday (Mon's coefficient) and about 1.2 fewer arrivals should be expected on a typical Saturday than a typical Sunday (Sat's coefficient). The other weekday coefficients are interpreted similarly. Further, the coefficient for holidays indicates that irrespective of the day, daily ED arrivals are decreasing at a rate of about 2 on holidays.

In table 4.10, the best model for the high acuity arrival category has an intercept while the best models in the other categories do not. The reason for this is that auto.arima() provides an estimate of the intercept of an ARIMA-based model by default when the time series was not differenced (i.e., stationary time series, d=0) if it improves the AIC (Hyndman, 2012). In the context of our problem, the intercept can be interpreted as the overall expected number of daily high acuity arrivals.

In addition to the RegArima models, a fbprophet model was fitted for each of these three arrival categories, taking into consideration the important predictors of the respective categories. The goal is to compare the predictive power of the best models from each category (identified above) to that of a fbprophet model.

4.3.2 Forecast Evaluation

Tables 4.11 to 4.13 summarize the accuracy of regARIMA and fbprophet models in forecasting future patient daily arrivals from March 4, 2018, to March 31, 2018.

In Table 4.11, the performance of both the regARIMA and fbprophet models in forecasting future daily arrivals is shown to be identical for the % within 95% P.I. metric and different for the other metrics. The regARIMA model slightly outperforms the fbprophet model based on the max. UNForcast metric, while the fbprophet model slightly outperforms the regARIMA model when considering RMSE and MAE metrics.

Table 4.11: Performance comparison of regARIMA and fbprophet models for Total arrivals.

	regARIMA (0,1,1)				Facebook Prophet			
Day	**RMSE**	**MAE**	**% within P.I.**	**Max. UF**	**RMSE**	**MAE**	**% within P.I.**	**Max. UF**
Sun	10.17	8.5	100	0	9.15	7.75	100	0
Mon	8.56	7.75	100	9	7.68	6.5	100	10
Tue	7.6	6.75	100	2	6.52	6	100	4
Wed	8.59	7.75	100	3	7.52	7	100	5
Thu	8.9	8.25	100	6	8.22	8	100	8
Fri	9.86	7.75	75	6	9.72	8	75	8
Sat	6.76	5.25	100	10	6.71	5.5	100	11

In Table 4.12, the performance of both the regARIMA and fbprophet models is shown to be identical for both the % within 95% P.I and max. UNForcast metrics. For the RMSE and MAE metrics, the cells highlighted in blue represent the days of the week when one model slightly outperformed the other.

Table 4.12: Performance comparison of regARIMA and fbprophet models for High acuity arrivals.

	regARIMA (1,0,0)				Facebook Prophet			
Day	**RMSE**	**MAE**	**% within P.I.**	**Max. UF**	**RMSE**	**MAE**	**% within P.I.**	**Max. UF**
Sun	8.34	6.5	100	0	8.57	6.5	100	0
Mon	8.06	7.5	100	8	8.2	7.75	100	8
Tue	8.41	8.25	100	7	8.56	8.25	100	7
Wed	10.23	8.75	100	1	10.99	9.25	100	1
Thu	6.18	5.25	100	2	6.12	5	100	1
Fri	10.21	9.25	100	13	10.21	9.25	100	13
Sat	10.67	9.75	100	17	10.46	9.5	100	17

In Table 4.13, the performance of both the regARIMA and fbprophet models is identical for the % within 95% P.I. and max. UNForcast metrics, while for the RMSE and MAE metrics, one model outperformed the other for certain days of the week.

Table 4.13: Performance comparison of regARIMA and fbprophet models for Low acuity arrivals.

	regARIMA (0,1,1)				Facebook Prophet			
Day	**RMSE**	**MAE**	**% within P.I.**	**Max. UF**	**RMSE**	**MAE**	**% within P.I.**	**Max. UF**
Sun	4.15	2.75	100	0	3.67	2.5	100	0
Mon	1.8	1.25	100	3	1.94	1.75	100	3
Tue	6.3	5.25	100	2	6.02	4.75	100	2
Wed	5.81	4.25	100	11	5.81	4.25	100	11
Thu	3.24	3	100	4	3.5	3.25	100	5
Fri	9.97	8	75	4	9.75	7.5	75	4
Sat	7.12	6.75	100	7	7.04	6.5	100	8

The max. UNForcast metric, which counts the number of under-forecasts (i.e., forecast less than the actual), is a good indicator of forecast bias. The result of this particular metric will be useful in practice by providing the ED administrators to avoid excess staffing situations.

For the total and low acuity arrivals, one out of the four actual values on Friday fell outside the 95% prediction interval. This may seem contradictory to the proposal that 95% of the actual values should be contained in the 95% prediction interval. However, prediction intervals are defined under a probability construct which remains true only if the rolling-origin weekly forecasting is continued for a longer period. That is, if the forecasting continues for a longer period, approximately 95% of the actual values will be contained in the proposed 95% prediction interval.

The regARIMA models were chosen over the base (SARIMA) models based on AICc values. Residual analysis suggests that the regARIMA model, which exploits both exogenous variable(s) and internal dependency, is appropriate to forecast daily ED arrival counts. The fbprophet model shows similar predictive power to regARIMA for this particular data and this competitiveness provides some justification for its use as a module of a bigger framework.

The plot of the predicted daily arrival counts and the observed values over the period from March 4, 2018 to March 31, 2018 (the test set) is shown in Figures 4.7, 4.8, and 4.9. The models can pick up the general pattern in the daily ED arrival data for the total and high acuity arrivals with some notable differences (i.e., some few peaks and troughs were not picked up). However, for the low acuity arrivals, both models did not quite pick up the general pattern.

In the preliminary analysis, the ACF plot revealed that the low acuity category of arrivals did not follow a defined weekly seasonal pattern, but rather a weak biweekly pattern. The low acuity model should therefore be used with caution when making a weekly prediction, i.e., forecasting next Monday's arrival based on the current Monday's arrival.

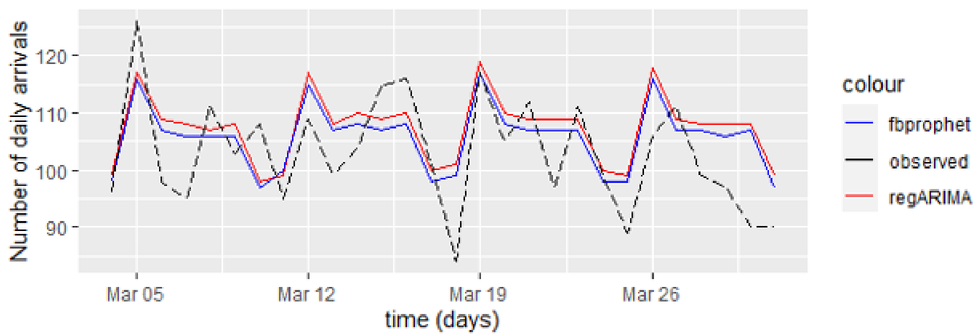


Figure 4.7: Forecasted ED daily patient arrival by regARIMA and fbprophet models for Total arrival and out-of-sample (test) set (March 4, 2018–March 31, 2018).

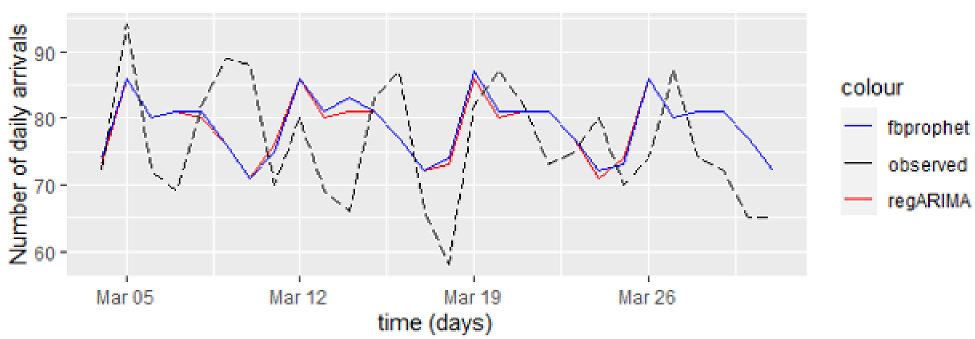


Figure 4.8: Forecasted ED daily patient arrival by regARIMA and fbprophet models for High acuity arrival and out-of-sample (test) set (March 4, 2018–March 31, 2018).

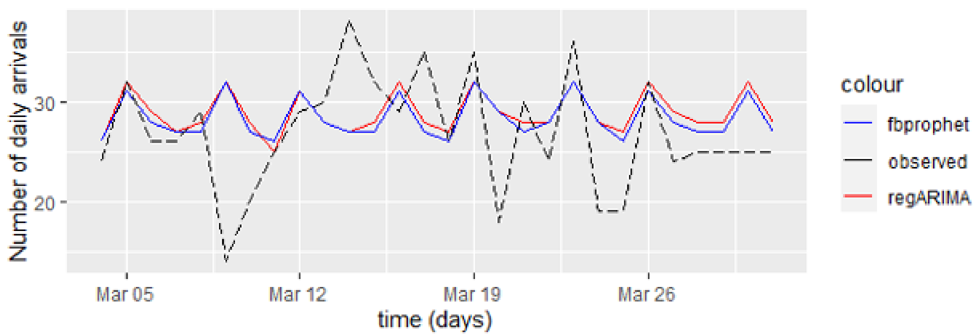


Figure 4.9: Forecasted ED daily patient arrival by regARIMA and fbprophet models for Low acuity arrival and out-of-sample (test) set (March 4, 2018–March 31, 2018).

4.4 Concluding remarks

In this chapter, we investigated regARIMA and fbprophet models for daily arrival prediction at a Montreal hospital based on five predictor variables (holidays, sports, festivals, day-of-week, and snowfall). We used the all possible regression model approach to select the best model of each arrival category based on the AICc value. The selected model for each arrival category contained the important predictor(s). Five different arrival categories and five predictor variables were investigated. For each category of arrival, thirty-two models were fitted using auto.arima() function from the forecast package in R and RegARIMA was the best fit model in each category. Holiday and DOW were the important predictors for the total arrival category while DOW was the most important predictor for the other four arrival categories namely low acuity (i.e., with CTAS 4 and 5), high acuity (i.e., with CTAS 1, 2, and 3), walk-in (i.e., based on the mode of arrival), and WILAAB (i.e., walk-in patients who are low acuity and ambulatory).

One objective of this study was to ascertain the importance of sports events in forecasting future arrivals to a Montreal hospital ED based on the physicians' observation of high arrival rates on days when there is a hockey game. We hypothesized that the increased arrival rate was because some people put off seeking medical care until after sporting events and some arrivals may be assault-related; hence increasing the ED patient arrival rates "post-game". The mode of arrival of the vast majority of these patients will be via walk-in. Hence forecasts of walk-in and WILAAB arrivals could be a valuable input for resource allocation planning and flow management. For example, if on a sports event day, the ED knows the average number of WILAAB arrivals, the ED might plan to redirect these arrivals to other primary care services or call in more staff. However, sports and festivals did not prove to be important predictors of these arrival categories. Three of the arrival categories, i.e., Walk-in, WILAAB, and low acuity, resulted in the same model containing the same predictor (DOW) so we opted to continue the analysis with the total, low acuity, and high acuity categories.

The results of the forecast evaluation show that fbprophet either slightly outperformed or was at par in performance with regARIMA. This competitiveness of fbprophet provides some justification for its use as a module of a bigger framework.

4.5 Limitations and future work

There are several limitations of our study. The most obvious is the degree to which it can be generalized based on the fact that we studied a single hospital with specific characteristics. However, for arrival counts which are always autocorrelated and influenced by external variables, RegARIMA and fbprophet are appropriate models to use because they are both easier to fit and interpret in real-life applications.

We had a one-year worth of data spanning two consecutive years. The data was not enough to capture the long-term relationship between the arrivals and the predictor variables. A high percentage of missing values for the snowfall data were observed, and mean imputation was used to calculate missing values. Also, we did not have access to socio-demographic variables which may affect ED arrivals (e.g., age, sex, education, marital status, household, employment, and income). The inclusion of these types of predictors would further improve the predictive power of the models and add value to the literature.

Future work can focus on four aspects to improve the results of this study. The first is to collect more arrival data such that the model is trained over multiple full calendar years. Secondly, we will investigate the impact of socio-demographic factors on the number of daily patient arrivals to improve the precision of our forecasting model. The third aspect is to integrate this study in a bigger framework of descriptive, predictive and prescriptive analytics. The fourth future work direction is to use our approach in a comparative study of arrival forecasting between multiple Montreal hospitals.

Chapter 5

Forecasting Hourly Number of Patient Arrivals to An Emergency Department

5.1 Introduction

In Chapter 4, we focused on the question of forecasting daily arrivals to an emergency room via regression with ARIMA errors models and Facebook Prophet models. Local popular sports (i.e., hockey, soccer, tennis, and Formula 1 Grand Prix) were combined and designated as a single binary variable called Sports, which takes the value of 1 if a given day is a sports event date and 0 otherwise. The sports variable was hypothesized to be an important predictor of daily patient arrivals. The result showed that Sports was not included as a predictor in the best models (i.e., models with $\Delta_i = 0$). However, some of the best subset ($\Delta_i \leq 2$) models contained Sports as a predictor (see Tables 4.4 – 4.8), providing a hint that the sports predictor could be important for this data. Given the short duration of the sporting events (e.g., approximately 3 hours for hockey and 2 hours for soccer), we may have missed important hour-by-hour impacts of the sports predictor by investigating the daily ED arrivals (Hughes et al., 2018). Also, the Montreal hospital ED physicians made their observation about increased arrival rate only on days when there is a hockey game. Hence, this chapter aims to extend the study in Chapter 4 by investigating hourly arrival forecasting taking into consideration the hours during and after both home and away hockey games as predictors.

If there is any group of patients for whom hockey games is a factor in deciding or planning a visit to the ED, it's the low acuity patients (see Hughes et al. (2018)); in contrast, for high acuity patients, the decision to go to the ED is mainly due to their severity. This observation influenced our decision to focus on the walk-in and WILAAB arrival categories in this chapter. In our data, 77.3% of the arrivals are walk-ins of which the vast majority are low acuity patients (i.e., over 70%). Ambulance arrivals, which are mostly high acuity patients, make up 22% of the arrivals (see section 3 Chapter 4).

We, therefore, hypothesize two scenarios that can cause an intra-day increase in the number and type of arrivals to EDs based on the influence of sports on human behaviour irrespective of whether the events are being followed remotely (e.g., television) or attended locally. First, some people put off seeking medical care until after sporting events, thereby increasing the ED patient arrival rates in the periods immediately following the events. Secondly, some of the arrivals during the periods of the events and after the events are assault-related(Hughes et al., 2018) and sports-injury related (Weiss & Elixhauser, 2016).

In this chapter, we evaluate the effectiveness of the Dynamic Harmonic regression model with an ARIMA error structure (DHregARIMA) for predicting future hourly ED arrivals and experimentally evaluate the performance of Facebook Prophet (fbprophet) to demonstrate its competitiveness with established methods. In addition, we show, in the context of this specific hospital ED, that there is inconclusive evidence of the influence of hockey games as a predictor of patient hourly arrivals.

The remainder of this chapter is organized as follows. The methods used and the data are described in Section 5.2. Detailed results and interpretations are provided in Section 5.3. Concluding remarks are provided in Section 5.4. Limitations and future research directions are described in Section 5.5.

5.2 Methodology

Our goal is to develop and evaluate models for predicting the hour-by-hour number of patients arriving at the emergency department of a Montreal hospital, exploring the impact of hockey games on ED arrivals. We first develop a custom DHregARIMA model and then compare it with the out-of-the-box Facebook Prophet (fbprophet) method.

5.2.1 Data

Two key categories of patient arrival are considered: Walk-in patients (i.e., based on the mode of arrival) and WILAAB patients (i.e., walk-in patients who are low acuity and ambulatory). The data comprised counts of patients' arrival times at a Montreal hospital ED between April 1, 2017, to March 30, 2018, aggregated to obtain hourly arrivals.

5.2.2 Predictors

(Hughes et al., 2018) highlighted that the potential intra-day impact of sporting events on ED arrivals has been reported to decrease before and during the events, while it increases following some events. For this study, we considered four possible hockey-game-day predictors:

- hours during home games (HHG),
- hours following home games (HFHG),
- hours during away games (HAG), and
- hours following away games (HFAG).

The hockey game dates were obtained from the online source Sport reference (2021). Within the timeframe of our study (i.e., April 1, 2017, to March 30, 2018), forty-three games were played in Montréal and forty-five games were played way from Montréal, bringing the total to eighty-eight hockey games played.

5.2.3 Modelling

We followed the same approach we used in Chapter 4 to develop and choose the best models in this study. RStudio, version 1.3.1093, was used for the modelling and model evaluation process. Results of the modelling process, including results of statistical tests, are provided in the next section (Section 5.3).

We began by creating, for each arrival category, a design (or model) matrix which is made up of all possible candidate models, created by all possible combinations of the potential predictors. Each row of the matrix represents an individual model and the successive columns with "1" correspond to the predictor variables to be included in the model. Since we are investigating four potential predictors, a model matrix of 2^4=16 rows and 4 columns are created for each category of arrivals (Table 5.1).

Table 5.1: Model matrix based on potential predictors.

Model	**HHG**	**HFHG**	**HAG**	**HFAG**
1	1	1	1	1
2	1	1	1	0
⋮	⋮	⋮	⋮	⋮
16	0	0	0	0

Next, we divided the dataset into training and test sets. The first 8064 hours of the dataset, covering the period from April 1, 2017, to March 2, 2018, were used as the training set and the remaining 672 hours, covering the period from March 3, 2018, to March 30, 2018, were used as the test set. The last day was discarded because it was not a complete 24-hour day. Before applying an ARIMA-based model to a time series data, the data must be stationary. A stationary time series data is one whose mean and variance are not time-dependent. To objectively determine if a time series is stationary, a statistical hypothesis test of stationarity (also known as a unit root test) was performed. In our analysis, we used both the Augmented Dickey-Fuller (ADF) and the Kwiatkowski-Phillips-Schmidt-Shin (KPSS) tests.

Auto Correlation Function (ACF) and partial ACF plots are created to verify seasonal patterns in the training data. Hourly data usually has three types of seasonality: a daily pattern, a weekly pattern, and an annual pattern. To specify all of the frequencies (or periods) that might be relevant for our data, a periodogram was used. A periodogram is a Fast Fourier Transform (FFT) based non-parametric method used in identifying the dominant frequencies (or periods) of a time series (Welch, 1967). Its plot shows a high peak at a frequency if the time series has a strong sinusoidal signal at that frequency. The dominant periods were then used to find the optimal number of Fourier terms needed to model the seasonal pattern. The optimal number of Fourier terms was chosen to minimize the AICc (Hyndman & Athanasopoulos, 2018).

Then, considering the information in each row of the design matrix, and the Fourier terms needed for the seasonal pattern included as predictors in each candidate model, auto.arima() searched and returned a DHregARIMA model with a corresponding AICc value. The list of candidate models was sorted in ascending order based on the AICc values, and the AICc values were rescaled to

$$\Delta_i = AICc_i - AICc_{min},$$

where $AICc_{min}$ is the minimum $AICc_i$ value. The "best" model, in each arrival category, has $\Delta_i = 0$. For each best DHregARIMA model, a corresponding fbprophet model was fitted using prophet(). Both the fbprophet and the DHregARIMA models were then used to create 24 hours ahead forecasts from March 3, 2018, to March 30, 2018.

5.2.4 Forecast evaluation

The "rolling-origin-update evaluation" technique described in Chapter 4 was used to validate the forecasting models for each of the forecasting horizons. That is, forecasts for a fixed horizon are performed by sequentially using past values from the test set to update the input information of the model, and changing the forecast origin accordingly.

(1) The models are fitted to the training dataset covering the period from April 1, 2017, to March 2, 2018 (8,064 hours).

(2) The forecast values are calculated for 24 hours of the test set (March 3, 2018).

(3) The 24 hours forecast values are then compared with the actual values by calculating Root Mean Square Error (RMSE), Mean Absolute Error (MAE), maximum of the maximum between zero and the difference in actual and forecast arrivals (Max. UNForcast), and the percentage of the actual values that fall within the 95% prediction interval.

(4) The forecasting origin is moved forward by 24 hours with the additional data point used to update the input to the model. The forecasts for the next 24 hours of the test set are generated

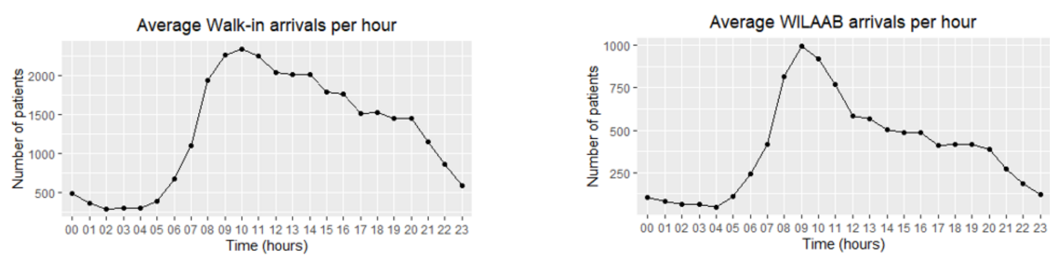
followed by a calculation of accuracy measures based on the forecasted values and the actual values.

(5) Steps 2 to 4 are repeated until the end of the data set. A total of four weekly forecasts are generated.

5.3 Results and discussion

5.3.1 Preliminary data analysis

The average hourly arrival by category (Figure 5.1) shows lower arrival rates from 00:00 AM to 06:00 AM, and peaked at 10:00 AM for the Walk-in category and 09:00 AM for the WILAAB category. 77.3% of all arrivals at the ED were through walk-ins, of which 30.7% were WILLAB arrivals. On average, 4 patients walk into this ED per hour (Table 5.2).



(a) Walk-in arrival

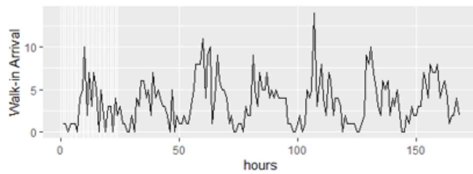
(b) WILAAB arrival

Figure 5.1: Average hourly arrival totals by category.

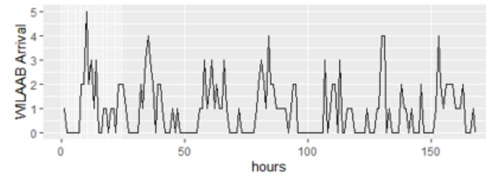
Table 5.2: Descriptive statistics of hourly arrivals by category

Horizon	**Arrival categories**	**Mean**	**Median**	**Max.**	**Min.**	**St. Dev**
Hour	Walk-in	3.5	3	17	0	2.7
	WILAAB	1.1	1	9	0	1.3

Time series plots of a full week of the arrivals, made up of 168 hourly observations in the dataset (i.e., April 1, 2017, to April 6, 2017) suggested the classic sinusoidal pattern associated with hourly arrivals (Figure 5.2).



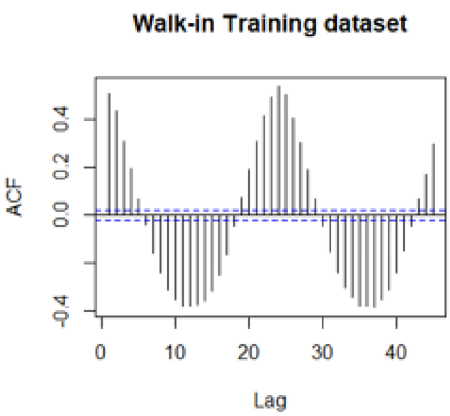
(a) Walk-in arrival



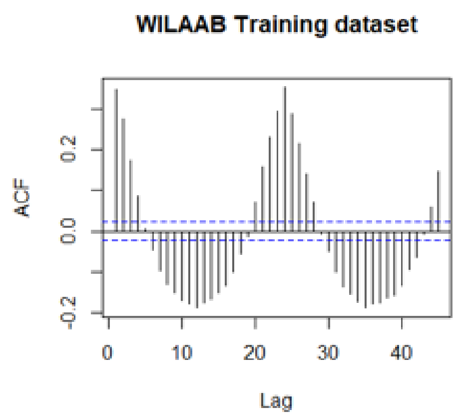
(b) WILAAB arrival

Figure 5.2: Time series plot of a Montreal hospital ED hourly arrivals (April 1, 2017, to April 6, 2017).

The Auto Correlation Function (ACF) plots of both arrival categories show a sinusoidal pattern with positive and negative significant spikes at lag intervals of 24, indicative of daily seasonality (Figure 5.3).



(a) Walk-in arrival



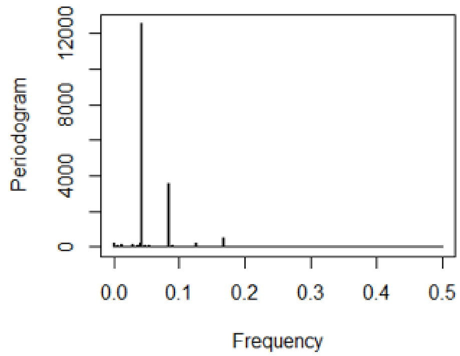
(b) WILAAB arrival

Figure 5.3: ACF plots of the training set by arrival category

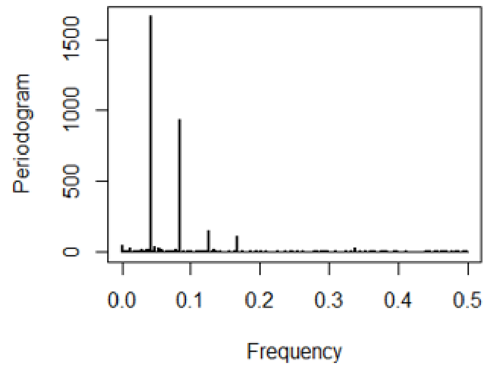
5.3.2 Fitted models

In this subsection, we present the results from fitting the DHregARIMA model to our time series data. Before fitting the models, the ADF test rejected the null hypothesis of non-stationarity for both arrival categories, indicating there is sufficient evidence to support the stationarity of the hourly arrival time series.

The periodogram of both time-series (Figure 5.4) detected the daily frequency (i.e., 24.03 and 23.96) to be the only dominant period. This implied that our data did not have a significant weekly pattern, and the wavelike patterns observed in both hourly ED arrival categories are due to the daily frequency.



(a) Walk-in arrival



(b) WILAAB arrival

Figure 5.4: Periodogram plots showing dominant spike at low frequency.

To find the number of Fourier terms (K) required to model the daily seasonality (i.e., the period of 24), the Fourier function, which takes the training set and K as arguments, was used as an external regressor in auto.arrima(). Then in a for loop that varied k from K =1 to K=12,[1] auto.arima() searched for the best model with the lowest AICc value. The number of Fourier terms was selected by minimizing the AICc. The order of the ARIMA model (i.e., p, d, and q) is also selected by minimizing the AICc, although that is done within the auto.arima() function. Based on the dominant period of 24, the number of Fourier terms that minimize AICc was determined to be nine for the walk-in arrival category and ten for the WILAAB arrival category. This implied fitting a pair of nine and ten sine and cosine functions as covariates to capture the seasonal variation in the walk-in and WILAAB time series data, respectively.

For the Walk-in arrival category, regression with ARIMA errors of order (2,0,2) provided adequate model fit. None of the hockey-game-day predictors proved to be important in forecasting future walk-in arrivals (Table 5.3) since none of them conforms to the pattern of “the best predictor is the one that is present in all models in the best subset”.

[1]K must not be greater than period/2 (see https://rdrr.io/github/robjhyndman/forecast/src/R/season.R)

Table 5.3: Fitted DHregARIMA models for the Walk-in arrival category.

Model	HHG	HFHG	HAG	HFAG	AICc	delta
M16	0	0	0	0	33000.51	0
M8	1	0	0	0	33000.76	0.25
M12	0	1	0	0	33001.89	1.38
M4	1	1	0	0	33002.29	1.78
M14	0	0	1	0	33002.31	1.8
M15	0	0	0	1	33002.4	1.89

For the WILAAB arrival category, regression with ARIMA errors of order (4,0,0) provided adequate model fit. Here also, none of the hockey-game-day predictors proved to be important in forecasting future WILAAB arrivals (Table 5.4) since none of them conforms to the pattern of "the best predictor is the one that is present in all models in the best subset".

Table 5.4: Add caption

Model	HHG	HFHG	HAG	HFAG	AICc	delta
M8	1	0	0	0	24010.32	0
M16	0	0	0	0	24010.74	0.42
M7	1	0	0	1	24010.97	0.65
M15	0	0	0	1	24011.39	1.07
M4	1	1	0	0	24012.11	1.79
M6	1	0	1	0	24012.32	2

In Tables 5.3 and 5.4, models with $\Delta_i \leq 2$, which we referred to as best subset models, are considered to have predictive power similar to the best models with $\Delta_i = 0$ (Fabozzi et al. (2014)). For walk-in arrivals, the best model is a DHregARIMA model which does not contain any hockey-game-day predictor. However, the next best model (with $\Delta_i = 0.25$) contains the hockey-game-day predictor HHG. Meanwhile, for WILAAB arrivals, the best model is a DHregARIMA model which contains the hockey-game-day predictor HHG and the next best model (with $\Delta_i = 0.42$) does not contain any hockey-game-day predictor.

Given that the DHregARIMA model, which does not contain any predictor, is either the best (Walk-in category) or belongs to the best subset of models (WILAAB category), we cannot assuredly conclude that HHG is an important predictor of WILAAB arrival without a tie-breaker. We, therefore, conclude that there is inconclusive evidence of the influence of hockey games as a predictor of walk-in and WILAAB patient hourly arrival.

The best model for the Walk-in category has 9 pairs of Fourier terms and can be written as

$$y_t = \beta_0 + \sum_{k=1}^{9}[\alpha_k \sin(\frac{2\pi kt}{24}) + \gamma_k \cos(\frac{2\pi kt}{24})] + \eta_t$$

where η_t is an ARIMA (2,0,2) process. With the total number of degrees of freedom of 22 (eighteen coming from Fourier parameters, two from the 2 autoregression (AR) parameters and two from the 2 moving average (MA) parameters).

For the WILAAB category, the best model has 10 pairs of Fourier terms and can be written as

$$y_t = \beta_0 + \sum_{k=1}^{10}[\alpha_k \sin(\frac{2\pi kt}{24}) + \gamma_k \cos(\frac{2\pi kt}{24})] + \eta_t$$

where η_t is an ARIMA (4,0,0) process. With the total number of degrees of freedom of 24 (twenty coming from Fourier parameters and four from the 4 autoregression (AR) parameters). The parameter estimates and standard errors are provided in Tables 5.5 and 5.6.

In both Tables 5.5 and 5.6, auto.arima() provides an estimate of the intercept of the DHregARIMA models by default when the time series was not differenced (i.e., stationary time series, $d = 0$) if it improves the AIC (Hyndman, 2012). In the context of our problem, the intercept can be interpreted as the overall expected number of hourly patient arrival counts in each category. The result shows that, in an hour, three patients are expected to walk into the ED. One of the patients is expected to be low acuity and ambulatory (i.e., WILAAB). The sine and cosine estimates are used to estimate peak timing and the amplitude of variation. They serve the purpose of capturing the seasonal variability that is present in our data.

Table 5.5: Estimated coefficients of the best DHregARIMA model for the walk-in arrival category.

Arrival category	Best Model	Coefficients	Estimates	Std. Error
Walk-in	regARIMA (2,0,2)	ar1	-0.207	0.122
		ar2	0.552	0.124
		ma1	0.251	0.127
		ma2	-0.486	0.129
		intercept	3.553	0.024
		S1-24	-1.396	0.033
		C1-24	-2.125	0.033
		S2-24	-0.647	0.031
		C2-24	0.64	0.031
		S3-24	0.286	0.029
		C3-24	-0.165	0.029
		S4-24	-0.217	0.029
		C4-24	-0.269	0.029
		S5-24	0.107	0.028
		C5-24	0.074	0.028
		S6-24	0.048	0.028
		C6-24	-0.055	0.028
		S7-24	0.025	0.028
		C7-24	-0.006	0.028
		S8-24	0.018	0.028
		C8-24	0.07	0.028
		S9-24	0.003	0.028
		C9-24	-0.098	0.028

Table 5.6: Estimated coefficients of the best DHregARIMA model for the WILAAB arrival category.

Arrival Category	Best model	Coefficients	Estimates	Std. Error
WILAAB	regARIMA (4,0,0)	ar1	0.067	0.011
		ar2	0.061	0.011
		ar3	0.033	0.011
		ar4	0.028	0.011
		intercept	1.09	0.015
		S1-24	-0.257	0.02
		C1-24	-0.857	0.02
		S2-24	-0.386	0.018
		C2-24	0.269	0.018
		S3-24	0.252	0.017
		C3-24	-0.074	0.017
		S4-24	-0.127	0.016
		C4-24	-0.11	0.016
		S5-24	0.059	0.016
		C5-24	0.043	0.016
		S6-24	0.017	0.016
		C6-24	-0.035	0.016
		S7-24	-0.01	0.016
		C7-24	0.03	0.016
		S8-24	0.009	0.016
		C8-24	0.013	0.016
		S9-24	0.016	0.016
		C9-24	-0.044	0.016
		S10-24	-0.031	0.016
		C10-24	0.029	0.016

For each of the best DHregARIMA models, a corresponding fbprophet model was fitted.

5.3.3 Forecast Evaluation

Table 5.7 summarizes the accuracy of DHregARIMA and fbprophet models in forecasting future patient arrivals every 24 hours from March 3, 2018, to March 30, 2018.

The performance of both models in forecasting future hourly arrivals is shown to be comparable. The forecast accuracy metrics are near identical for the 24-hour rolling forecast horizons of both arrival categories except for the % within 95% P.I. metric, where fbprophet outperformed DHregARIMA in forecasting the WILAAB arrivals.

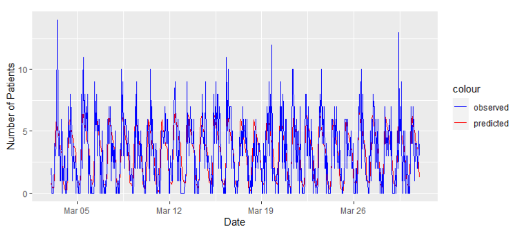
Table 5.7: Performance comparison of fbprophet and DHregARIMA models

Arrival category	**Model**	**RMSE**	**MAE**	**% within P.I.**	**Max UF**
Walk-in	Facebook prophet	1.8	1.43	95.1	9
	regARIMA (2,0,2)	1.83	1.44	95.2	8
WILAAB	Facebook prophet	1.03	0.78	94.2	5
	regARIMA (4,0,0)	1.05	0.78	86.3	5

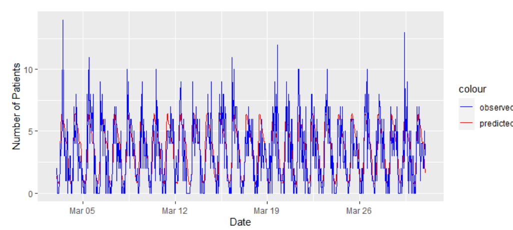
The plots of the predicted hourly arrivals and the observed values for both Walk-in and WILAAB arrivals, over the period from March 3, 2018, to March 30, 2018, are presented in Figure 5.5. Panels (a) and (b) are for the fbprophet and DHregARIMA of the Walk-in arrivals while panels (c) and (d) are for the fbprophet and DHregARIMA of the WILAAB arrivals. The blue lines represent the observed values while the red line represents the predicted values. There is a resemblance in each of the graphs produced by both modelling strategies for each arrival category. This is unsurprising since these models showed comparable performance in the prediction of future arrivals (Table 5.7). The models can capture the seasonal variability that is present in the out-of-sample (test) data set for both the walk-in and WILAAB categories using Fourier terms as predictors.

It is important to note that we believed the WILAAB group of patients are more likely to decide or plan a visit to the ED after both home and away hockey games, thereby supporting the doctors' hypothesis of increased arrivals on days when there is a hockey game. However, given that our finding indicates that the hockey game is not an important predictor for this arrival stream, the prediction of WILAAB arrivals might not be useful in our context. The walk-in category represents the largest arrival stream in our case. it is therefore important in practice to be able to produce accurate short-term (i.e., hourly) forecasts for this arrival stream.

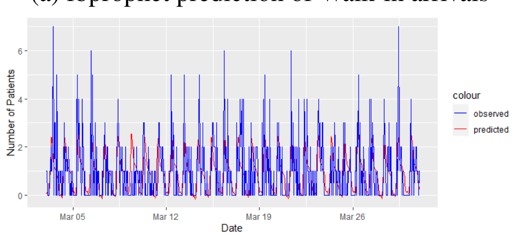
The fbprophet modelling approach blends the advantages of fully automated statistical method and judgmental forecasting method ((Taylor & Letham, 2018)); an integration that has been described as the best way to obtain a more reliable forecast (Soyiri and Reidpath (2013), Hyndman and Athanasopoulos (2018)). This characteristic of fbprophet is particularly useful for ED administrators who are non-statisticians/non-data-scientists but have extensive ED settings knowledge which is oftentimes needed to make reliable time-series forecasts. An ED administrator can predict how many arrivals are expected in 4, 6, 8, 12, etc. hours in future (corresponding to the beginning, midway, and/or the end of shifts), and then take steps to have more resources to meet the future demand.



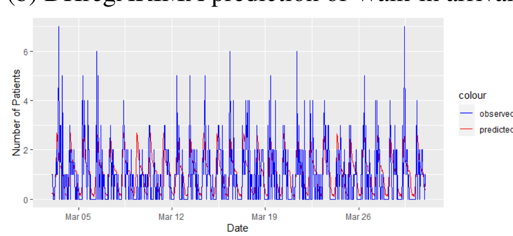
(a) fbprophet prediction of Walk-in arrivals



(b) DHregARIMA prediction of Walk-in arrivals



(c) fbprophet prediction of WILAAB arrivals



(d) DHregARIMA prediction of WILAAB arrivals

Figure 5.5: 24-hour rolling horizon prediction of Walk-in and WILAAB arrival test sets(March 3,2018-March 30,2018)

5.4 Concluding remarks

In this chapter, we investigated DHregARIMA and fbprophet models for hourly arrival prediction at a Montreal hospital based on four possible hockey-game-day predictors variables namely hours during home games (HHG), hours following home games (HFHG), hours during away games (HAG), and hours following away games (HFAG). Two arrival categories walk-in and WILAAB were investigated. We used the all possible regression model approach to select the best model of each arrival category based on the AICc value. The selected model for each arrival category contained Fourier terms used as predictors to capture the seasonal variability. For each category of arrival, sixteen models were fitted using auto.arima() function from the forecast package in R and DHregARIMA was the best fit model in each category.

The main objective of this study was to ascertain the importance of hockey-game-day predictors in forecasting future arrivals to a Montreal hospital ED based on the physicians' observation of high arrival rates on days when there is a hockey game. However, our results show, in the context of this specific hospital ED, that there is inconclusive evidence of the influence of hockey games as a predictor of patient hourly arrivals. Therefore, the selected models for the walk-in (i.e., based on the mode of arrival) and WILAAB (i.e., walk-in patients who are low acuity and ambulatory) arrival categories did not contain any of the four possible hockey-game-day predictors.

Both the fbprophet and DHregARIMA models showed comparable predictive power for this particular data. The competitiveness of fbprophet provides some justification for its use in studies that intend to use fbprophet as a part of a bigger framework without focusing on a particular forecasting aspect.

5.5 Limitations and future work

As previously stated, the most obvious limitation of this study is the degree to which it can be generalized based on the fact that we studied a single hospital with specific characteristics.

Also, we did not have access to socio-demographic variables which may affect ED arrivals (e.g., age and gender) which have been associated with violence-related ED arrivals (Hughes et al., 2018). The inclusion of these types of predictors would further improve the predictive power of the models.

Future work can focus on two aspects to improve the result of this study. The first is to investigate the impact of socio-demographic factors on the number of hourly patient arrivals to improve the precision of our forecasting model. The second future work direction is to use our approach in a comparative study of arrival forecasting between multiple Montreal hospitals.

Chapter 6

Conclusions and future research directions

In this final chapter, we summarize the work on ED demand forecasting presented in previous chapters, restate the major contributions of this thesis, and state some possible directions for future work.

6.1 Summary and contributions

As the gateway to hospitals, EDs are often plagued by congestion (overcrowding), which can be defined as a situation in which the need for services exceeds an ED's capacity to provide these services within a reasonable time (Sinclair, 2007). Overcrowding in ED affects patient satisfaction, the quality of treatment and prognosis (Kam et al., 2010). Many solutions have been proposed to solve this global problem. The underlying message in all the solutions is that ED overcrowding is a systemic problem and hence requires a systemic solution (Hwang et al. (2011), Morley et al. (2018)). In this dissertation, we focus on ED patient demand forecasting to mitigate the severity and duration of ED overcrowding in the Quebec setting.

The premise of the thesis is based on the observation that the literature in the emergency department (ED) demand forecasting can be classified into two categories. In the first category, different methods are compared and the method with the most accurate forecast is selected. However, in this setting, regression with ARIMA errors (regARIMA) has not been widely considered despite it being an established forecasting model with multiple advantages over models that have been extensively studied. Also, Facebook Prophet (fbprophet), which is a fully automated statistical method (Taylor & Letham, 2018), has not been used because it is viewed as a "Blackbox" or "off-the-shelf" method and researchers want to develop more custom models. In the second category, a few researchers use an "off-the-shelf" method like fbprophet as a module within a bigger resource allocation framework with no real justification for why it was chosen over other established forecasting methods.

Given that there is a literature evaluating forecasting methods without considering fbprophet, and literature that just uses fbprophet without much justification, led to the consideration of the naturally-arising questions: how does fbprophet compare to a custom-designed ARIMA model? and is fbprophet then justified to be used as a module in bigger studies?

We investigated the rarely considered forecasting model regARIMA and fbprophet, and evaluated their performance for daily and hourly arrival prediction. Using the all-possible-regression-model approach, the best models for different patient arrival categories were selected based on the Akaike information criterion (AICc). Each selected model contained the important predictor(s) for that arrival category.

In Chapter 4, five predictors, namely sports, holidays, day-of-week (DOW), snowfall, and festivals were investigated. DOW was found to be the most important predictor. On the other hand, the occurrence of sports events, which was considered a potentially important predictor based on the physicians' observation of high arrival rates on days when there is a hockey game, did not prove important for daily arrival forecasting. However, the inclusion of sports as a predictor in some of the best subset ($\Delta_i \leq 2$) models suggested that the sports predictor could be important for this data. Furthermore, considering that sporting events have a short duration (e.g., approximately 3 hrs for hockey and 2 hrs for soccer), we may have missed important hour-by-hour impacts of the sports predictor. Also, the Montreal hospital ED physicians made their observation about increased arrival rate only on days when there is a hockey game. Taking all of these points into consideration, we extended the study of Chapter 4 by investigating hourly arrival forecasting in Chapter 5. Two arrival streams were considered to analyze the importance of the hockey game as a predictor. Our choice of the arrival streams was influenced by the fact that the vast majority of the patients in these streams are low acuity patients for whom hockey games could be a factor in deciding or planning a visit to the ED; in contrast to high acuity patients whose decision to go to the ED is mainly due to their severity. The hours during and after both home and away hockey games were considered as predictors. The main objective was to ascertain the importance of hockey-game-day predictors in forecasting future arrivals to a Montreal hospital ED based on the physicians' observation of high arrival rates on days when there is a hockey game. However, our results showed, in the context of this specific hospital ED, inconclusive evidence of the influence of hockey games as a predictor of patient hourly arrivals.

The major contributions of this thesis are as follows. First, we evaluate the effectiveness of the rarely used regARIMA model in forecasting future daily and hourly ED arrivals. Second, we experimentally evaluate the performance of fbprophet, a model that has not been thoroughly studied in the literature and demonstrate its competitiveness with established forecasting methods. This insight is particularly valuable given that several papers have been using fbprophet as a step towards a bigger framework without evaluating its actual predictive performance. Although ours is a single study, our results provide some justification for the use of fbprophet in studies that intend to use fbprophet as part of a bigger framework without focusing on a particular forecasting aspect. Third, we investigate the hypothesis that public sporting events, particularly hockey, lead to increased arrivals to the ED by using hockey games as a predictor within our forecasting models.

6.2 Future research directions

Following the work presented in this thesis, we suggest future work in four directions. The first is to collect more arrival data such that the forecasting model is trained over multiple full calendar years. Given that the COVID-19 pandemic altered ED arrival patterns globally, and that the data used in training our models was obtained prior to the pandemic, it is very important to retrain, validate, and test the models with data that contains arrival from both COVID-19 and non-COVID-19 patients. Secondly, we will investigate the impact of socio-demographic factors on the number of daily patient arrivals. These two aspects are important for improving the precision of our forecasting model and increasing the probability of a successful implementation. The third future work direction is to use our approach in a comparative study of arrival forecasting between multiple Montreal hospitals.

Forecasting of future arrivals (i.e., demand) provides insights necessary to make resource allocation decisions in EDs, and healthcare in general. Another future work direction, therefore, is improving the scheduling of healthcare personnel to eliminate the supply-demand mismatch and improve the operational efficiency and healthcare delivery of the ED. Forecasting is a critical first step for developing robust personnel scheduling (or rostering) tools (Ernst et al., 2004). Some preliminary work was done in this direction after the diagnostic study findings (see Chapter 2). The ED administrator was interested in identifying where to add extra shifts to improve the mismatch between demand and supply of care (i.e., staff level). We were not convinced it was optimal to simply add shifts during periods of increased patient arrival (i.e., during peak demand periods). Hence, we formulated a prescriptive analytics tool, i.e., a deterministic shift scheduling Mixed Integer Programming (MIP) model. We present our preliminary work in this direction, including the MIP model, in Appendix A. The analysis presented in Appendix A was scenario-based, using the mean number and the maximum number of arrivals at each epoch. In the context of an ED, staff are needed to perform duties that arise from stochastic patient demand. The likelihood of future patient demand is not well known and must be modelled using forecasting techniques (Ernst et al., 2004). Now that we have created demand forecasting models in Chapters 4 and 5, further research would focus on finding ways to incorporate uncertainty into the optimization model using the forecasts. One idea is to use arrival prediction intervals to create an uncertainty set in a robust optimization model. Another idea is the creation of confidence interval capacity constraints in the model using the forecast interval limits.

6.3 Conclusion

Determining staffing requirements, i.e., determining the number of staff needed at different times over some planning period is critical for personnel scheduling in any service organization. In the context of an ED, where the staff are needed to perform duties that arise from stochastic patient demand, the likelihood of future patient demand is not well known and must be modelled using forecasting techniques (Ernst et al., 2004). In this thesis, we investigated the use of several forecasting techniques for forecasting arrivals, using a Montreal hospital ED as a basis for our investigation. We hope that forecasting studies such as ours will lead to the use of forecasting techniques as a precursor for optimizing staffing resources in emergency departments.

Appendix A

Future Direction on Prescriptive Modelling

A.1 Preliminary Optimization Model

Using the historical arrival data, we studied two demand profiles: the mean number and the maximum number of arrivals at each epoch (these numbers were obtained from the data set described in Chapter 2). The model was designed to minimize overall unmet demand and generate a daily schedule of optimal shift start times for ED physicians. The Montreal hospital ED administrator was looking to add an extra shift to the current seven daily shifts of varying length.

In the context of EDs and healthcare in general, a shift is assigned to a physician. Therefore, adding an extra shift could mean opening a completely new shift at a time interval different from those covered by the current shift structure or assigning two physicians to a current shift.

Basic notation

L: set of all possible start times, l = 1, 2, ..., 35 where $(l = 1) \equiv$ 7:00 AM and $(l = 35) \equiv$ 12:00 AM.

T: set of time period, t = 1, 2, ..., 48 where $(t = 1) \equiv$ 7:00 AM, $(t = 2) \equiv$ 7:30 AM, and $(t = 48) \equiv$ 6:30 AM

C: current shift start times (subset of L), i.e. l = $\{1, 3, 13, 17, 19, 23, 33\}$,

parameters

A_t: average number of patients arriving per half-hour *t*.

P_t: median number of patients seen per half-hour *t*

M_{tl}: a shift coverage matrix with entry equal to 1 if the time period t is covered in shift starting at time l, and 0 otherwise

Decision variables

x_l = 1 if shift with start time *l* is assigned; 0 otherwise.

U_t = unmet demand in time period *t*.

D_t = total demand in time period *t*.

Y_t = total number of patients potentially treated per half-hour *t*.

Model

$$\text{Minimize} \quad \sum_{t \in T} U_t \tag{2}$$

$$\text{s.t.} \quad \sum_{l \in L} M_{tl} x_l \geq 1 \qquad \forall t \in T, \tag{3}$$

$$\sum_{l \in L} x_l M_{tl} P_t = Y_t \qquad \forall t \in T, \tag{4}$$

$$D_t = A_t \qquad t = 1, \tag{5}$$

$$U_t \geq D_t - Y_t \qquad t = 1, \tag{6}$$

$$D_t = A_t + U_{t-1} \qquad \forall t \in T/t \neq 1, \tag{7}$$

$$U_t \geq D_t - Y_t \qquad \forall t \in T/t \neq 1, \tag{8}$$

$$\sum_{l \in L} x_l = N \qquad \forall t \in T, \tag{9}$$

$$x_l = 1 \qquad l = \{1, 3, 4, 13, 17, 19, 23, 33\}, \tag{10}$$

The objective function (2) minimizes the unmet demand. Constraint (3) ensures at least one physician is assigned to time period t. Constraint (4) is an equation that defines the total number of patients treated every half-hour *t*. The group of constraints (5), (6), (7), and (8) together calculates the total demand from patients in time period *t*, and also ensures that the unmet demand at this time period does not exceed the difference between the total demand and the number of patients that can be seen. Constraint (9) ensures that the specified number of shifts are selected each time the model is run. For example, if $N = 8$, exactly 8 daily shifts will be selected. When activated (i.e., when $x_l = 1$), Constraint (10)) ensures that the current Montreal hospital ED shifts are selected as optimal. The model was solved using CPLEX Optimization Studio.

A.1.1 Computational Experiment

A computational experiment was carried out to find the best schedule structure to cover the demand (i.e., where to add extra shifts). The experiment was based on varying the shift pattern matrix M_{tl}, a given set of predefined shifts. Four different variants of matrix M_{tl} were developed for this study:

- Variant 1 contained forty-eight 8-hour shifts with the first shift starting at 7:00 AM and a new shift beginning in every half-hour interval.
- Variant 2 contained 7 shifts of different lengths (i.e., four 9-hour, a 9.5-hour, a 10-hour, and an 8-hour) with different start time sequences. These were the shift types used by the Montreal hospital ED.
- Variant 3 was formed by concatenating Variant-1 and Variant-2 matrices, hence it contained fifty-five shifts.
- Variant 4 was formed by concatenating two Variant-2 matrices, hence it contained fourteen shifts.

The model was run with each of these matrices; after each run, the objective value (i.e., the sum of unmet demand), optimal shift starting times, and the maximum unmet demand and the period during which maximum unmet demand occurred were recorded.

The first run was with variants 1, 2, and 3, selecting seven daily shifts to compare the performance of the current Montreal hospital ED shift schedule (variant-2) to the seven optimal shifts selected from the other matrices (i.e., variants 1 and 3). The goal of the first run was to find out if we only needed to change the current shift structure for the length of each of the shifts in the schedule, without adding an extra shift.

To find where to add an extra shift to the current schedule, the model was run the second time with each of the matrices (excluding variant 2 matrix), in succession, to select eight optimal shifts. For the run with variants 3 and 4 matrices, x_l is set to 1 for $l = \{1, 3, 4, 13, 17, 19, 23, 33\}$ so that the seven Montreal hospital ED shifts are selected as optimal and the eighth shift is selected to minimize the unmet demand. For the run with variant 4 matrix, the goal was to find out if the extra shift could be one of the current seven shifts (i.e., a duplicate) which will imply assigning two physicians to the selected eighth shift.

The model was run a third time with variant 3 matrix to select eight shifts without setting x_l to find out if shifts from the current Montreal hospital schedule will be selected and what level of improvement in terms of demand coverage will be achieved.

Results

The results showed a significant reduction in total unmet demand with the addition of an extra morning shift to the current Montreal hospital ED shift schedule. The demand coverage improved by 70% when the shift from 7:30 AM to 3:30 PM was added to the Montreal hospital ED current schedule; the unmet demand dropped significantly, from 1,805 to 541. The highlighted row in Table 6.1 represents the Montreal hospital ED schedule with the additional shift (i.e., keeping 7 of the current shifts and adding one additional one). However, the most noteworthy reduction of unmet demand (i.e., 159) came as a result of running the model with the variant 3 matrix and selecting eight shifts without constraining the shifts to be the ones currently in use. The shifts selected were different from the shifts in the current Montreal hospital ED schedule.

Table A.1: Result of computational experiments with various shift matrices showing the number of shifts selected, the objective value (i.e., the sum of unmet demand), the maximum unmet demand, and the period during which maximum unmet demand occurred.

Matrix (# shifts)	# shifts selected	objective value	max unmet D.	period
Variant 1 (48)	7	941	44	6:30 AM
	8	213	13	6:30 AM
Variant 2 (7)	7	1,805	51	6:00 PM
Variant 3 (55)	7	831	39	6:30 AM
	8	159	9	8:30 PM
	8	541	20	12:30 PM
Variant 4 (14)	8	431	19	12:30 PM

Changing the demand profile (i.e., from maximum to the mean number of arrivals per half-hour) did not change the location of where to add an extra shift (Table A.2).

Table A.2: Two possible choices of an extra shift.

Patient Demand	Added Shift
Mean #	7:00 AM - 3:00 PM
Max #	7:30 AM - 3:30 PM

Based on this result, the hospital responded by changing one of their current shifts to start earlier and to have more doctors at 10 AM.

References

Allin, S., & Rudoler, D. (2015). The Canadian health care system. 2017. *International profiles of health care systems*, *21*.

Anneveld, M., Van Der Linden, C., Grootendorst, D., & Galli-Leslie, M. (2013). Measuring emergency department crowding in an inner city hospital in the Netherlands. *International journal of emergency medicine*, *6*(1), 1–7.

Aroua, A., & Abdul-Nour, G. (2015). Forecast emergency room visits–a major diagnostic categories based approach. *International Journal of Metrology and Quality Engineering*, *6*(2), 204.

Authier, P. (2016). *Montreal Gazette: quebec has worst emergency room wait times.* https://montrealgazette.com/news/quebec/quebec-has-worst-emergency-room-wait-times-health-and-welfare-commissioner-says. (Accessed: 2016-06-02)

Batal, H., Tench, J., McMillan, S., Adams, J., & Mehler, P. S. (2001). Predicting patient visits to an urgent care clinic using calendar variables. *Academic Emergency Medicine*, *8*(1), 48–53.

Burnham, K. P., Anderson, D. R., & Huyvaert, K. P. (2011). Aic model selection and multimodel inference in behavioral ecology: some background, observations, and comparisons. *Behavioral ecology and sociobiology*, *65*(1), 23–35.

Calegari, R., Fogliatto, F. S., Lucini, F. R., Neyeloff, J., Kuchenbecker, R. S., & Schaan, B. D. (2016). Forecasting daily volume and acuity of patients in the emergency department. *Computational and mathematical methods in medicine*, *2016*.

Camiat, F., Restrepo, M. I., Chauny, J.-M., Lahrichi, N., & Rousseau, L.-M. (2021). Productivity-driven physician scheduling in emergency departments. *Health Systems*, *10*(2), 104–117.

Carvalho-Silva, M., Monteiro, M. T. T., de Sá-Soares, F., & Dória-Nóbrega, S. (2018). Assessment of forecasting models for patients arrival at emergency department. *Operations Research for Health Care*, *18*, 112–118.

Chen, C.-F., Ho, W.-H., Chou, H.-Y., Yang, S.-M., Chen, I.-T., & Shi, H.-Y. (2011). Long-term prediction of emergency department revenue and visitor volume using autoregressive integrated moving average model. *Computational and mathematical methods in medicine*, *2011*.

Choudhury, A., & Urena, E. (2020). Forecasting hourly emergency department arrival using time series analysis. *British Journal of Healthcare Management*, *26*(1), 34–43.

Correia, J. C., Braillard, O., Combescure, C., Gerstel, E., & Spechbach, H. (2018). Admission rates in emergency departments in geneva during tennis broadcasting: a retrospective study. *BMC emergency medicine*, *18*(1), 1–9.

Cote, M. J., Smith, M. A., Eitel, D. R., & Akçali, E. (2013). Forecasting emergency department arrivals: a tutorial for emergency department directors. *Hospital topics*, *91*(1), 9–19.

De Gooijer, J. G., & Hyndman, R. J. (2006). 25 years of time series forecasting. *International journal of forecasting*, *22*(3), 443–473.

Di Somma, S., Paladino, L., Vaughan, L., Lalle, I., Magrini, L., & Magnanti, M. (2015). Overcrowding in emergency department: an international issue. *Internal and emergency medicine*, *10*(2), 171–175.

Donges, N. (2021). *Online: 4 reasons why deep learning and neural networks aren't always the right choice.* https://builtin.com/data-science/disadvantages-neural-networks. (Accessed: 2021-11-02)

Ekström, A., Kurland, L., Farrokhnia, N., Castrén, M., & Nordberg, M. (2015). Forecasting emergency department visits using internet data. *Annals of emergency medicine*, *65*(4), 436–442.

English Montreal School Board, O. (2021). *Online: 2016-18 school holidays in quebec.* https://www.emsb.qc.ca/. (Accessed: 2021-07-21)

Ernst, A. T., Jiang, H., Krishnamoorthy, M., & Sier, D. (2004). Staff scheduling and rostering: A review of applications, methods and models. *European journal of operational research*, *153*(1), 3–27.

Evans, R. G. (1988). " we'll take care of it for you" health care in the Canadian community.

Daedalus, 155–189.

Fabozzi, F. J., Focardi, S. M., Rachev, S. T., & Arshanapalli, B. G. (2014). *The basics of financial econometrics: Tools, concepts, and asset management applications*. John Wiley & Sons.

Gouvernement du Québec, Q. (2021). *Integrated Health and Social Services Centres (CISSS) and Integrated University Health and Social Services Centres (CIUSSS).* https://www.quebec.ca/en/health/health-system-and-services/service-organization/cisss-and-ciusss. (Accessed: 2021-11-14)

Government of Canada, C. (2021). *Canada: historical climate data.* https://climate.weather.gc.ca/. (Accessed: 2021-04-06)

Grand Prix Canada, O. (2021). *Online: 2017 and 2018 formula 1 grand prix du canada schedule.* https://www.gpcanada.ca/en/. (Accessed: 2021-07-21)

Gul, M., & Celik, E. (2020). An exhaustive review and analysis on applications of statistical forecasting in hospital emergency departments. *Health Systems*, *9*(4), 263–284.

Hinson, J. S., Martinez, D. A., Cabral, S., George, K., Whalen, M., Hansoti, B., & Levin, S. (2019). Triage performance in emergency medicine: a systematic review. *Annals of emergency medicine*, *74*(1), 140–152.

Hughes, H. E., Colón-González, F. J., Fouillet, A., Elliot, A. J., Caserio-Schonemann, C., Hughes, T. C., … others (2018). The influence of a major sporting event upon emergency department attendances; a retrospective cross-national european study. *PloS one*, *13*(6), e0198665.

Hwang, U., McCarthy, M. L., Aronsky, D., Asplin, B., Crane, P. W., Craven, C. K., … others (2011). Measures of crowding in the emergency department: a systematic review. *Academic Emergency Medicine*, *18*(5), 527–538.

Hyndman, R. J. (2010). *Online: constants and arima models in r.* https://robjhyndman.com/hyndsight/arimax/. (Accessed: 2021-07-31)

Hyndman, R. J. (2011). *Forecasting: An overview.*

Hyndman, R. J. (2012). *Online: constants and arima models in R.* https://robjhyndman.com/hyndsight/arimaconstants/. (Accessed: 2021-07-31)

Hyndman, R. J., & Athanasopoulos, G. (2018). *Forecasting: principles and practice*. OTexts.

Hyndman, R. J., & Khandakar, Y. (2008). Automatic time series forecasting: the forecast package for r. *Journal of statistical software*, *27*(1), 1–22.

J Murray, M. (2003). The Canadian triage and acuity scale: A Canadian perspective on emergency department triage. *Emergency medicine*, *15*(1), 6–10.

Jones, S. S., Evans, R. S., Allen, T. L., Thomas, A., Haug, P. J., Welch, S. J., & Snow, G. L. (2009). A multivariate time series approach to modeling and forecasting demand in the emergency department. *Journal of biomedical informatics*, *42*(1), 123–139.

Kadri, F., Harrou, F., Chaabane, S., & Tahon, C. (2014). Time series modelling and forecasting of emergency department overcrowding. *Journal of medical systems*, *38*(9), 1–20.

Kam, H. J., Sung, J. O., & Park, R. W. (2010). Prediction of daily patient numbers for a regional emergency medical center using time series analysis. *Healthcare informatics research*, *16*(3), 158–165.

Khashei, M., & Bijari, M. (2010). An artificial neural network (p, d, q) model for timeseries forecasting. *Expert Systems with applications*, *37*(1), 479–489.

Kim, S. W., Li, J. Y., Hakendorf, P., Teubner, D. J., Ben-Tovim, D. I., & Thompson, C. H. (2014). Predicting admission of patients by their presentation to the emergency department. *Emergency Medicine Australasia*, *26*(4), 361–367.

Mai, Q., Aboagye-Sarfo, P., Sanfilippo, F. M., Preen, D. B., & Fatovich, D. M. (2015). Predicting the number of emergency department presentations in w estern a ustralia: A population-based time series analysis. *Emergency Medicine Australasia*, *27*(1), 16–21.

Major League Soccer, O. (2021). *Online: 2017-18 montreal impact schedule and statistics.* https://www.mlssoccer.com/. (Accessed: 2021-07-21)

Marcilio, I., Hajat, S., & Gouveia, N. (2013). Forecasting daily emergency department visits using calendar variables and ambient temperature readings. *Academic emergency medicine*, *20*(8), 769–777.

Montreal Public Holidays, O. (2021). *Online: 2017-18 statutory holidays in canada.* https://www.statutoryholidays.com/. (Accessed: 2021-07-21)

Morley, C., Unwin, M., Peterson, G. M., Stankovich, J., & Kinsman, L. (2018). Emergency department crowding: a systematic review of causes, consequences and solutions. *PloS one*, *13*(8), e0203316.

National Bank Open, O. (2021). *Online: 2017 and 2018 roger cup schedule.* https://omniumbanquenationale.com/en/. (Accessed: 2021-07-21)

Ospina, M. B., Bond, K., Schull, M., Innes, G., Blitz, S., & Rowe, B. H. (2007). Key indicators of overcrowding in canadian emergency departments: a delphi study. *Canadian Journal of Emergency Medicine*, *9*(5), 339–346.

Ranse, J., Lenson, S., Keene, T., Luther, M., Burke, B., Hutton, A., ... Crilly, J. (2019). Impacts on in-event, ambulance and emergency department services from patients presenting from a mass gathering event: A retrospective analysis. *Emergency Medicine Australasia*, *31*(3), 423–428.

Rostami-Tabar, B., & Ziel, F. (2020). Anticipating special events in emergency department forecasting. *International Journal of Forecasting*.

Rosychuk, R. J., Youngson, E., & Rowe, B. H. (2016). Presentations to emergency departments for copd: A time series analysis. *Canadian respiratory journal*, *2016*.

Seabold, S., & Perktold, J. (2010). Statsmodels: Econometric and statistical modeling with python. In *Proceedings of the 9th python in science conference* (Vol. 57, p. 61).

Shah, S., Murray, J., Mamdani, M., & Vaillancourt, S. (2019). Characterizing the impact of snowfall on patient attendance at an urban emergency department in toronto, canada. *The American journal of emergency medicine*, *37*(8), 1544–1546.

Sinclair, D. (2007). Emergency department overcrowding–implications for paediatric emergency medicine. *Paediatrics & child health*, *12*(6), 491–494.

Soyiri, I. N., & Reidpath, D. D. (2013). An overview of health forecasting. *Environmental health and preventive medicine*, *18*(1), 1–9.

Sport reference, O. (2021). *Online: 2017-18 montreal canadiens roster and statistics.* https://www.sports-reference.com/. (Accessed: 2021-07-21)

Sun, Y., Heng, B. H., Seow, Y. T., & Seow, E. (2009). Forecasting daily attendances at an emergency department to aid resource planning. *BMC emergency medicine*, *9*(1), 1–9.

Svetunkov, I., & Petropoulos, F. (2018). Old dog, new tricks: a modelling view of simple moving averages. *International Journal of Production Research*, *56*(18), 6034–6047.

Tashman, L. J. (2000). Out-of-sample tests of forecasting accuracy: an analysis and review. *International journal of forecasting*, *16*(4), 437–450.

Taylor, S. J., & Letham, B. (2018). Forecasting at scale. *The American Statistician*, *72*(1), 37–45.

Ville de Montréal, Q. (2021). *Quebec: calendar of festivals and cultural events.* http://ville.montreal.qc.ca/culture/en/calendar-festivals-and-cultural-events-2017. (Accessed: 2021-07-12)

Wargon, M., Casalino, E., & Guidet, B. (2010). From model to forecasting: a multicenter study in emergency departments. *Academic Emergency Medicine*, *17*(9), 970–978.

Wargon, M., Guidet, B., Hoang, T., & Hejblum, G. (2009). A systematic review of models for forecasting the number of emergency department visits. *Emergency Medicine Journal*, *26*(6), 395–399.

Weiss, A. J., & Elixhauser, A. (2016). Sports-related emergency department visits and hospital inpatient stays, 2013: Statistical brief# 207.

Welch, P. (1967). The use of fast fourier transform for the estimation of power spectra: a method based on time averaging over short, modified periodograms. *IEEE Transactions on audio and electroacoustics*, *15*(2), 70–73.

Whitt, W., & Zhang, X. (2019). Forecasting arrivals and occupancy levels in an emergency department. *Operations Research for Health Care*, *21*, 1–18.

Wiler, J. L., Griffey, R. T., & Olsen, T. (2011). Review of modeling approaches for emergency department patient flow and crowding research. *Academic Emergency Medicine*, *18*(12), 1371–1379.

Xu, M., Wong, T.-C., & Chin, K.-S. (2013). Modeling daily patient arrivals at emergency department and quantifying the relative importance of contributing variables using artificial neural network. *Decision Support Systems*, *54*(3), 1488–1498.

Xu, Q., Tsui, K.-L., Jiang, W., & Guo, H. (2016). A hybrid approach for forecasting patient visits in emergency department. *Quality and Reliability Engineering International*, *32*(8), 2751–2759.

Yucesan, M., Gul, M., & Celik, E. (2018). A multi-method patient arrival forecasting outline for hospital emergency departments. *International Journal of Healthcare Management*.

Zhao, Y., Peng, Q., Strome, T., Weldon, E., Zhang, M., & Chochinov, A. (2015). Bottleneck detection for improvement of emergency department efficiency. *Business Process Management Journal*.