

Forecast emergency room visits – a major diagnostic categories based approach

Abdeljelil Aroua and Georges Abdul-Nour*

Industrial engineering dept. Université du Québec à Trois-Rivières, Canada

Received: 20 May 2014 / Accepted: 20 May 2015

Abstract. This work is a case study intended to explore the capability of three forecasting techniques to predict emergency department (ED) visits based on Major Diagnostic Categories. It is a part of a larger work aimed to improve ED patients' throughput time. The ED in this case is considered as a part of the health chain and the process of arrival and departure of patients are included. The prediction models presented in this work are initially established and validated from the historical 3-year emergency room visits at Sherbrook University Hospitals and uses the week as the period unit. Given that resources are consumed differently for each disease, a group of patients has been considered according to the major diagnostic categories (MDC). Three predictive models of the number of visits are considered and compared: linear regression model, SARIMA and multivariate SARIMA. The accuracy of the prediction models is evaluated by calculating the mean percentage error (MAPE) and the mean absolute error (MAE) between forecast and observed data. The medium term forecasting model for the number of admissions is determined according to the estimated admission ratio for each patient group, while the short term model is established according to a regression model based on age groups. SARIMAX offers the most accurate model with a MAPE ranging from 6% to 49% (group of a small number of visits). Twelve of the twenty-seven groups of patients account for nearly 90% of the total of emergency room visits and the weighted mean average percentage error (WMAPE) stands at 8%. The admission rates for each group of patients is based on Gauss' distribution and is different from one group to another. For many MDCs, strong correlations can be demonstrated between the admission rates and the patient age groups by using a quadratic regression. The prediction models explored in this paper aims to help managers to plan more efficiently the emergency department resources. The models can also be used to plan resources of other hospital departments since they give information about the number of admitted patients for each MDC.

Keywords: Emergency department visits, forecast, major diagnostic categories, ARIMA, ARIMAX

1 Introduction

The resources of the emergency department (ED) are some of the most difficult of all departments to schedule because of the complexity of its processes, the exposure to diversity and the nature of care requests. For several decades, we have observed phenomena of overcrowded ED and extension of waiting time for patients. These phenomena make it more difficult for governments and hospitals to provide good care services that are effective, safe and equitable to all citizens.

Three factors can be considered as the source of the overcrowding in the ED: input factors, throughput factors and output factors. Input factors consist in the number of patient visits and the distribution of those visits over time. Throughput factors are often associated with inefficiencies in resource management such as personal care and beds. Finally, output factors refer to the admission process and the hospital capacity to hospitalize new patients.

The biggest challenge for hospital managers resides in providing the necessary resources to meet the demand for care and in maximizing their efficiency. Thus, good models created to predict the number of ED visits and their dispositions are important tools that can help managers to make the right decisions.

This work aims to compare three prediction techniques for ED visits that will be useful to managers, by providing a better understanding of the number of visits and the number of patient admissions, and by providing input data for scheduling, simulating and queuing theory works in order to optimize ED resources.

2 Literature review

The prediction of patient visits and overcrowding in the EDs has received considerable attention in recent year due to the common desire to achieve a more efficient planning of the resources and to improve patient flow. Several

* Correspondence: abdeljelil.aroua@uqtr.ca

authors have addressed the problem of time series analysis and others have forged links between different related information sources.

Moreover, the prediction of patient visits has been treated in several ways; some authors have studied different techniques for predicting patient visits (all health problems combined), while others have focused on predicting the number of patients based on certain types of health problems (respiratory problems – Alexander G. Perry [1], abdominal problems – Sadeghi et al. [2], pediatric patients – Walsh et al. [3], etc).

Under the assumption that the bottleneck of the whole patient treatment system is the admission process [4], authors limited their researches by forecasting the number of patients admitted and not the number of patient visits to the ED [4,5].

2.1 Prediction of ED visits: analysis of time series

Several approaches and mathematical models have been proposed in the literature to forecast the demand for emergency care based on time series analysis and linear models, using variables of different types.

Abraham et al. [6] have established different forecasting models: moving average forecasting model, simple exponential smoothing forecasting model, and models of auto regressive integrated moving average (ARIMA) or seasonal auto regressive integrated moving average (SARIMA) models. In their study, the time variables were used as explanatory variables.

Wargen et al. [7] proposed a general linear model for predicting ED traffic in the region of Paris, using explanatory variables such as time (days, months, holidays).

Chen et al. [8] have used an ARIMA involving variables such as climate (temperature, humidity, level of rain) to estimate the number of patient visits classified in three different groups; traumatic, non-traumatic and pediatric visits.

The simultaneous use of temporal variables (day, month, season) and other variables, such as climate (temperature, humidity, precipitation, etc.), in forecasting techniques has aroused the interest of several researchers. Hye Jin Kam et al. [9] and Sun et al. [10] even improved the model by dividing patients in three acuity levels.

Table 1 summarizes the main literature on forecasting techniques of ED visits.

2.2 Prediction of ED visits: using related information sources

The idea of exploiting related sources of information to be translated into meaningful information is not new. This technique is often used in the construction field in which the knowledge of the number of sites that had permissions can help creating a sales forecasting of products and services related to a given sector. Similarly, in the health sector, and specifically for emergency visits, Perry [1] showed the possibility of using the number of

calls received by “Telehealth Ontario” to draw a prediction of ED visits for patients suffering from respiratory problems.

2.3 Prediction of the number of admitted patients: time series analysis

Many authors believe that improving the flow of emergency patients is a direct consequence of reducing the patient waiting time in the admission process. They focused their researches on understanding and estimating the number of admitted patients. The techniques used and encountered during the literature review can be found below.

Several authors [5,11,12] have used the techniques of time series analysis in their work to predict the number of admitted patients. Boyle et al. [5] provided a model with a MAPE of 11% for daily admission. Abraham [6] noted an unpredictable fact for medium and long term admitted patient number and the techniques used are only appropriate for a horizon of less than one week.

2.4 Prediction of the admitted patient number: opinion of health professionals

Making predictions at the triage on whether a patient presents high probabilities to be admitted or not has been the subject of several studies that considered that, at this stage of the patient treatment, such information can help in making appropriate decisions that could limit emergency overpopulation. Peck et al. [13] demonstrated that such techniques can predict the number of patients admitted based on the number of ED visits. In their study, they compare different prediction methods, using different prediction factors such as the age, the arrival method, the acuity level and the primary complaint of the patients. Moreover, at the triage stage, the reactivity of the system is limited and the decisions are of an operational type and on very short terms. Table 2 summarizes the literature on the modeling of emergency admitted patients.

3 Methodology

Why trying to model the ED visit process? The answer is simple: “To make future forecasts for a better planning of the resources in order to meet the demand for emergency care”. Of course, estimating the total number of patient visits per time unit allows emergency managers to plan the deployment of the necessary resources to meet the demand. However, this process has its limits since the quantity of resources consumed varies from a patient to another; for example an admitted patient will consume more resources than a non-admitted one (Ministry of Health and Social Services of Quebec [14]).

Modeling the number of patient visits to the ED by grouping the patients into categories based on their health

Table 1. Summary of forecasting models for ED visits.

Autors	Input variables	Output variables	Models	Results	Limitations	Regions
Wargen et al. [6]	Holidays Weekdays Months Trend ED visits (Hist.)	ED visits (Hist.)	General linear model	MAPE = 5.3%	No info about the number of admissions waiting for beds.	France
Cheich-Fan Chen et al. [8]	Weather temperature (Average, min, max) Relative humidity, rain, Stock index fluctuation, ED visits (Hist.)	ED visits (Hist.)	ARIMA	MAPE = 5.73% ~ 21.18%.	Region of study: influence of some socio-economic factors not included. No info about the number of admissions waiting for beds.	Taiwan
Hye Jin Kam et al. [9]	Months Weekdays Holidays Chuseok Seasons Temperature (average max, min and diff) Rain, snow, wind speed, relative humidity and yellow dust ED visits (Hist.)	ED visits (Hist.)	General linear model SARIMA SARIMAX	MAPE (GLM) = 11.2% MAPE (SARIMA - UV) = 8.48% MAPE (SARIMA - MV) = 7.44%	No info about the number of admissions waiting for beds.	South Korea
Yan Sun et al. [10]	Months Weekdays Holidays Temperature Relative humidity PSI (Pollutant Standards Index) ED visits (Hist.)	ED visits (Hist.)	ARIMA	MAPE (total) = 4.8% MAPE (Acuity P1) = 16.8% MAPE (Acuity P2) = 6.7% MAPE (Acuity P3) = 8.6%	Prediction of ED visits for each acuity level but no info about the number of admissions waiting for beds.	Singapore
Spencer et al. [12]	Days, Holidays, Months Maximum temperature Interactions	ED visits (Hist.)	SARIMA Regression SC Regression SC with climatic variables Exponential smoothing Neural network	Good precisions for auto-regression models.	No info about the number of admissions waiting for beds. Analysis limited to one region	United States
Hoot et al. [24]	ED visits (Hist.) Visiting frequency	ED visits (Hist.)	Simulation tools for forecasting.	Good precision for 2 hours horizons.	Many simplifying assumptions were made in the process of Application of the Forecast	United States

problems instead of simply considering the number of patient visits (any health problems combined) offers several advantages:

- a- It explores the seasonal and evolutionary character of each health problem. In the Ancient Greece, Hippocrates observed that during fall, diseases are the most acute and the most deadly in all and that spring is the healthiest and the least deadly of all

seasons. In his Aphorisms, Hippocrates noted many correlations between the occurrence or severity of various diseases and the climate, seasons and temperament of men. In modern times, several authors have studied the seasonal and evolving nature of diseases over time (Nelson [15] – Seasonal variation of rheumatic diseases, Schlesinger [16]), a character that differs from one disease to another, and which correlations can be shown with time and climatic variables.

Table 2. Summary of forecasting models for admitted patients.

Autors	Input variables	Output variables	Models	Results	Limitations	Regions
Schweigler et al. [11]	Number of beds occupied at the ED	Emergency occupied beds.	AR MA SARIMA	SARIMA: CIA = -281 ± 27 Log - likelihood = 144 ± 13 AR: CIA = -305 ± 27 Log likelihood = 158 ± 14	Model limited to 24 hours	United States
Boyle et al. [5]	ED visits and admitted patients (Hist.)	Number of admitted patients	ARIMA Exponential smoothing	MAPE = 11% (admission) MAPE = 7% (visit)	No information available on the given health care category	United States
Abraham et al. [6]	ED visits (Hist.)	Number of admitted patients	AR SARIMA	MAPE(AR) = 7% (1 day) MAPE(AR) = 9% (7 days) MAPE(ARIMA) = 5% (1 day) MAPE(ARIMA) = 8% (7 days)		Australia
Peck et al. [13]	ED visits, Arrival method, Acuity level, Fast track, Patient health problem.	Number of admitted patients	Naïve Bayes applied to input variables.	VA bayes Sensibility = 53.48 Specificity = 91.41	Short term predictions	United States

- b- The health care process also differs from one disease to another. Indeed, the treatment for a patient suffering from a heart condition does not require the same resources than a patient with a simple flu. In addition, the proportion of patients admitted depends on the nature of the health problems that brings them to the ED. For example, according to the data from the ED of the hospitals Fleurimont and Hôtel Dieu in Sherbrooke, the probabilities that a patient suffering from a circulatory disorder problem is admitted is statistically more important than a patient with an ear or nose disorder (proportion test: $p = 0.377$ vs. 0.039 (P value = 0.000)).
- c- Finally, knowing the number of patients admitted for each health problem group allows the hospital to better plan its hospitalization resources. This information has a direct and positive impact on waiting time at the emergency; many authors agreed that the transfer process for admitted patients is often the bottleneck of the whole process of the ED [6, 17].

Consequently, the difficulties reside in modeling the number of patient visits to the ED (N_{kt}) and the number of admitted patients (A_{kt}), suffering from a health problem of group k during the period t .

According to “*Le guide de gestion de l’unité d’urgence*” [14], the method used to calculate the resource needs differs for each type of resources. For example, the required number of emergency stretchers implies a

linear function of the number of patients admitted and non-admitted, and this number can be calculated by using the formula (1). For the number of nurses needed, the ratios have been defined either to guarantee a certain rate of patients (e.g. time sorting mean between 5 and 10 min per patient), or to ensure a good service by setting a maximum number of patients treated by a nurse (e.g. 4 to 5 stretchers not monitored by a nurse).

Then the required quantity of the resource r for period t is Q_{rt} , and is expressed as:

$$Q_{rt} = \sum_k [q_{rk}^a \cdot A_{kt} + q_{rt}^{na} \cdot (N_{kt} - A_{kt})], \quad (1)$$

where:

Q_{rt} : Required quantity of resource r for period t .

q_{kt}^a : Required resources r for period t in order to meet the demand for the care of a patient admitted with a health problem of group k .

q_{kt}^{na} : Required resources r for period t in order to meet the demand for the care of a non-admitted patient with a health problem of group k .

N_{kt} : Total number of patients visiting the emergency for period t and suffering from a health problem of group k .

A_{kt} : Total number of patients admitted for period t and suffering from a health problem of group k .

Table 3. Major diagnostic categories.

MDC	Description
0100	Nervous System
0200	Eye
0300	Ear, Nose, Mouth And Throat
0400	Respiratory System
0500	Circulatory System
0600	Digestive System
0700	Hepatobiliary System And Pancreas
0800	Musculoskeletal System And Connective Tissue
0900	Skin, Subcutaneous Tissue And Breast
1000	Endocrine, Nutritional And Metabolic System
1100	Kidney And Urinary Tract
1200	Male Reproductive System
1300	Female Reproductive System
1400	Pregnancy, Childbirth And Puerperium
1500	Newborn And Other Neonates (Perinatal Period)
1600	Blood and Blood Forming Organs and Immunological Disorders
1700	Myeloproliferative DDs (Poorly Differentiated Neoplasms)
1800	Infectious and Parasitic DDs
1900	Mental Diseases and Disorders
2000	Alcohol/Drug Use or Induced Mental Disorders
2100	Injuries, Poison And Toxic Effect of Drugs
2200	Burns
2300	Factors Influencing Health Status
2400	Human Immunodeficiency Virus Infection
2500	Multiple Significant Trauma

3.1 Grouping of health problems

The “Major Diagnostic Category” (MDC) is a grouping method of All Patient Refined Diagnosis Related Groups (APR-DRG). This technique is used in various countries to classify episodes of care in homogeneous groups that presents similar pathologies and treatments equivalent in cost and length of stay; thus, that consume the same hospital resources. The MDC is mainly determined by the primary diagnosis and is encoded 00-25 as described in Table 3.

Working with this grouping approach presents several advantages. First it consists in groups of health problems based on the amount of consumed resources to provide care. Moreover, this method is effective and used by the majority of hospitals. Also, this grouping technique provides an interface with the hospital management system. And finally, estimating the number of patients per MDC allows to plan not only ED resources but also resources needed to hospitalize the admitted patients.

3.2 Explanatory variables

The explanatory variables used in this study are of two types: temporal and climate.

- The temporal variables are the weeks of the year.

- The climate variables are weekly average, maximum and minimum temperature, weekly temperature differences, rain and snow.

3.3 Modeling techniques

This study attempts to compare three different forecasting techniques: linear regression, SARIMA and multivariate SARIMA applied to each MDC.

a) Linear regression:

Modeling the visits into a linear model aims to highlight any correlation between the number of patients per MDC and the predictor variables ($p < 0.05$). To address the interaction between temporal and climatic variables used in this model, the method of partial least square is used.

In a linear model, the number N_{kt} ; visits of patients belonging to the MDC k for period t can be written as:

$$N_{kt} = \sum_i a_{ik} \cdot X_{it} + \varepsilon_{kt}, \quad (2)$$

where X_{it} is the value of the predictor i at the period t and a_{ik} the coefficient of the predictor i and ε_{kt} a correction.

b) SARIMA univariate

The number of visits of patients belonging to a MDC k can be represented by a time series (N_{kt}):

$$N_{kt} = T_{kt} + S_{kt} + \varepsilon_{kt}, \quad (3)$$

where T_{kt} is the trend component, S_{kt} is the seasonality component, and ε_{kt} is the noise.

ARMA is a stochastic process defined by Box-Jenkins (1970) that combines the autoregressive and the moving average processes. This stationary process has been improved to take into account the components of trends and seasonality. The improved process is referred to by ARIMA (Auto-Regression Integrated Moving Average), and denoted by ARIMA (p, d, q) where p is the order of the autoregressive, d is the of the order of differentiation and q the order of the moving average.

If N_{kt} is an ARIMA process, then it is written as follows:

$$\nabla^d \varphi(B) N_{kt} = \theta(B) \varepsilon_{kt}, \quad (4)$$

where ∇d is the operator of differentiation, B an operator such as $N_{k,t} = B \cdot N_{k,t-1}$, and φ and θ are functions of B .

A SARIMA is an ARIMA model with a seasonal component. In this case we add other parameters to the model (4): S, P, D and Q , defining operators and seasonal variables.

$$\varphi(B)_s \vartheta(B^s) \nabla_s^D \nabla^d N_{kt} = \theta(B)_s \theta(B^s) \varepsilon_{kt} \quad (5)$$

Table 4. Age groups.

Groups	1	2	3	4	5	6	7	8	9	10	11	12	13
Ages	00–02	03–05	06–12	13–17	18–24	25–34	35–44	45–54	55–64	65–74	75–84	85–94	+95

c) Multivariate SARIMA

For the multivariate ARIMA, independent variables are incorporated into the formula (5) to form the multivariate model ARIMAX.

$$\emptyset(B)_s \emptyset(B^s) \nabla_s^D \nabla^d N_{kt} = \theta(B)_s \theta(B^s) \varepsilon_{kt} + \sum_i \left(\sum_{j=0}^{L^i} \beta_{t-j}^i \cdot Y_{t-j}^i \right), \quad (6)$$

where Y_{t-j}^i is the independent variable i at period $t-j$ and β_{t-j}^i the corresponding coefficient.

The time unit to consider will be a period that will satisfy the following conditions: quality of the data for good decision-making, accuracy of the model and a horizon that provides time to react.

The records related to epidemic diseases and occasional events in time, such as influenza, will not be considered in this research because of their unique and unpredictable characters.

3.4 Number of patients admitted

A model for the number of admitted patients is proposed. Similarly to the model used to predict the ED visits, this model relies on a MDC orientated approach. Peck et al. [13] worked on a model that defines the probabilities of admission for a given patient. Among the predictor variables, they considered the health problem showed by the patient at the triage stage. The model they developed used to predict the number of admitted patients after the triage step. It aroused our attention because it highlights the great influence of the health problem nature on the probabilities for a patient to be admitted.

In this study, we designate the stochastic variable τ_k to express the probabilities for a patient belonging to a MDC k to be admitted and A_{kt} represents the number of patients admitted. Then A_{kt} is written as:

$$A_{kt} = \tau_k N_{kt} \quad (7)$$

where N_{kt} is the number of ED visits of patients belonging to a MDC k at period t .

In order to draw the real-time status of an ED (total number of patients waiting and the number of patients with potential probabilities to be admitted), this study attempts to identify correlations between the admission rate and the age group for each MDC k . The age groups are the same active groups that have been used in the Emergency management system, and are described in the Table 4.

3.5 Model evaluation

Several types of quality measurement models have been proposed in the literature. The most popular ones are the Least squares, mean absolute error (MAE), the mean absolute error percentage (MAPE) and the mean square error (MSE).

In order to make a comparison between the models established in this work and the models developed in the literature, the MSE and the MAPE are considered and are expressed as follows:

$$\text{MSE}_k = \frac{\sum_t (N_{kt} - \check{N}_{kt})^2}{\text{DL}}, \quad (8)$$

$$\text{MAPE}_k = \frac{1}{T} \sum_{t=1}^T \left| \frac{N_{kt} - \check{N}_{kt}}{N_{kt}} \right|. \quad (9)$$

where N_{kt} is the number of patient visits observed at period t , \check{N}_{kt} is the estimated number of patient visits at period t , T is the number of periods and the DL is the freedom degree.

3.6 Choice of the time unit

The time unit to be chosen for ED visit prediction models is not unanimous in the literature; some authors mention climate factors and other sociological factors. The time units often used by authors are the month [8,18], the day [5–7,9,10,12] and the hour [11,19–21].

Modeling the number of ED visits based on the week is a field of research that has not been explored enough. Choosing the week as the time unit for a predictive model presents two main advantages: it provides enough data to create analyzable time series and it consists in an ideal planning horizon for managers. Scheduling for a shorter period than a week does not allow enough flexibility to be responsive and to secure the resources needed. This study explores the variation in the number of visits within a week, and for each MDC.

3.7 Data sources

The data used in this study come from the information system of the Centre Hospitalier de l'Universitaire de Sherbrooke (CHUS). The CHUS includes four EDs and host all together nearly 87 000 patients each year. The data collected represent nearly 380 000 records for the years 2008–2011.

For confidentiality purposes, the anonymity of the patients is respected in the database provided by the hospital, which makes it impossible to identify them.

Table 5. Prediction models per MDC.

Linear regression			SARIMA			SARIMAX			
MDC	MSE	MAPE	Models	MSE	MAPE	Models	Variables ($p < 0.05$)	MSE	MAPE
0100	68.2	8.0%	(100)(000)	104.3	10.4%	(100)(000)	N_{Snow}, W_i	132.8	9.2%
0200	24.9	16.2%	(111)(000)	36.1	21.2%	(111)(000)	$T_{\text{max}}, MT_{\text{max}}, N_{\text{Rain}}$	50.6	18.5%
0300	160.9	9.90%	(500)(100)	244.5	12.6%	(500)(100)	$N_{\text{Rain}}, N_{\text{Snow}}, W_i$	207.1	8.9%
0400	272.8	8.7%	(200)(110)	330.3	7.4%	(200)(000)	$T_{\text{max}}, N_{\text{Snow}}, W_i$	225	6.9%
0500	135.8	6.8%	(111)(100)	189.3	8.4%	(111)(100)	N_{Rain}	156.6	6%
0600	247.0	7.2%	(300)(001)	285.6	7.5%	(300)(001)	N_{Snow}, W_i	344.4	6.7%
0700	10.7	26.3%	(100)(000)	14.9	31.5%	(100)(000)	N_{Snow}, W_i	20.9	29.1%
0800	278.8	6.5%	(100)(001)	408.5	8.1%	(100)(001)	DJC, N_{Snow}, W_i	397.0	6.3%
0900	74.2	12.3%	(101)(111)	71.89	12.1%	(100)(101)	N_{Snow}, W_i	72.16	9.3%
1000	8.6	23.9%	(101)(000)	12.7	30.7%	(101)(000)	DJC, N_{Snow}, W_i	13.11	23.6%
1100	61.4	8.8%	(101)(001)	99	11%	(101)(001)	$T_{\text{min}}, N_{\text{Snow}}, W_i$	88.2	8.4%
1200	7.8	46.2%	(100)(000)	11	64.7%	(100)(000)	DJC	15.8	49.4%
1300	23.7	14.2%	(201)(000)	31.8	17.6%	(201)(000)	N_{Snow}, W_i	34.8	14.1%
1400	Refers to Childbirths								
1500	Refers to Childbirths								
1600	16.5	22.0%	(101)(000)	22.93	27.5%	(101)(000)	N_{Snow}, W_i	24.4	21.8%
1700	6.7	40.1%	(500)(000)	7.61	48.5%	(500)(000)	N_{Snow}, W_i	8.4	38.6%
1800	56.5	26.5%	(111)(001)	36.18	19%	(100)(101)	N_{Snow}, W_i	37.5	16.4%
1900	62.2	10.6%	(101)(000)	93.4	13.7%	(101)(000)	N_{Snow}, W_i	90.1	10.4%
2000	4.2	49.2%	(111)(001)	6.95	68.2%	(111)(001)		7.6	
2100	23.7	14.2%	(101)(100)	31.5	17.2%	(101)(100)	N_{Snow}, W_i	15.1	9.2%
2200	4.8	50.8%	(111)(000)	6.23	65%	(111)(000)		8.6	53.6%
2300	625.0	12.9%	(111)(101)	321.6	9.5%	(111)(101)	W_{51}	324	7.4%
2400	Refers to infections caused by the immunodeficiency virus, rare								
2500	179.2	11.0%	(300)(001)	265.6	14.1%	(300)(001)	$N_{\text{Rain}}, N_{\text{Snow}}, W_i$	252.3	10.6%

Climate data is retrieved from the database of the National Archives of climate information and from Canada Environment: <http://climat.meteo.gc.ca>.

4 Results

4.1 Analysis of the number of ED visits as time series

Drawing graphs of weekly visits for each MDC shows the different behaviors and provides relevant evolutionary and seasonal information. This information will be used to guide the Box-Jenkins technique to model series.

The MDC 0300 series' graph which corresponds to the diseases and disorders of the ear, nose, mouth, throat and craniofacial bone shows a seasonal pattern for each year with peaks during the first weeks of the year, which are explained by the presence of the winter cold in the area of Sherbrooke during these periods.

This behavior is the same for MDC 0400 series corresponding to diseases and disorders of the respiratory system.

The MDC 0400 series of graphs which corresponds to the diseases and disorders of the skin, subcutaneous tissue and breast also shows a seasonal pattern for each year but with peaks in the neighboring weeks of the week 28, periods when the temperature is warmer.

4.2 Analysis of predictive models of ED visits

According to Table 5, which summarizes the results of the analysis, it is difficult to claim that a certain forecasting model of ED visits is better than another for all the MDCs. This can be explained by the influence of different temporal and climatic factors on the given disease. Indeed, the MDCs 0100, 0200 and 2500 are best represented by linear regression models, while the remaining time series are best represented by ARIMAX models.

For all the MDCs, the SARIMAX provides better information than simple SARIMA models. The level of snow, the maximum and minimum weekly average temperature and weeks of the year are often factors that are significant in SARIMAX models.

For models with a small amount of data, such as low weekly visits (MDCs 0700, 1000, 1200, 1600, 1700 and 2200), none of the three models offers a satisfactory result with a MAPE beyond 20%.

To compare our method with those proposed in the literature, the weighted mean absolute percentage error (WMAPE) was chosen as the indicator of accuracy of the sum of multiple time series. The MDCs are sorted in descending order by number of visit means. The WMAPE which is the MAPE of several different time series at a

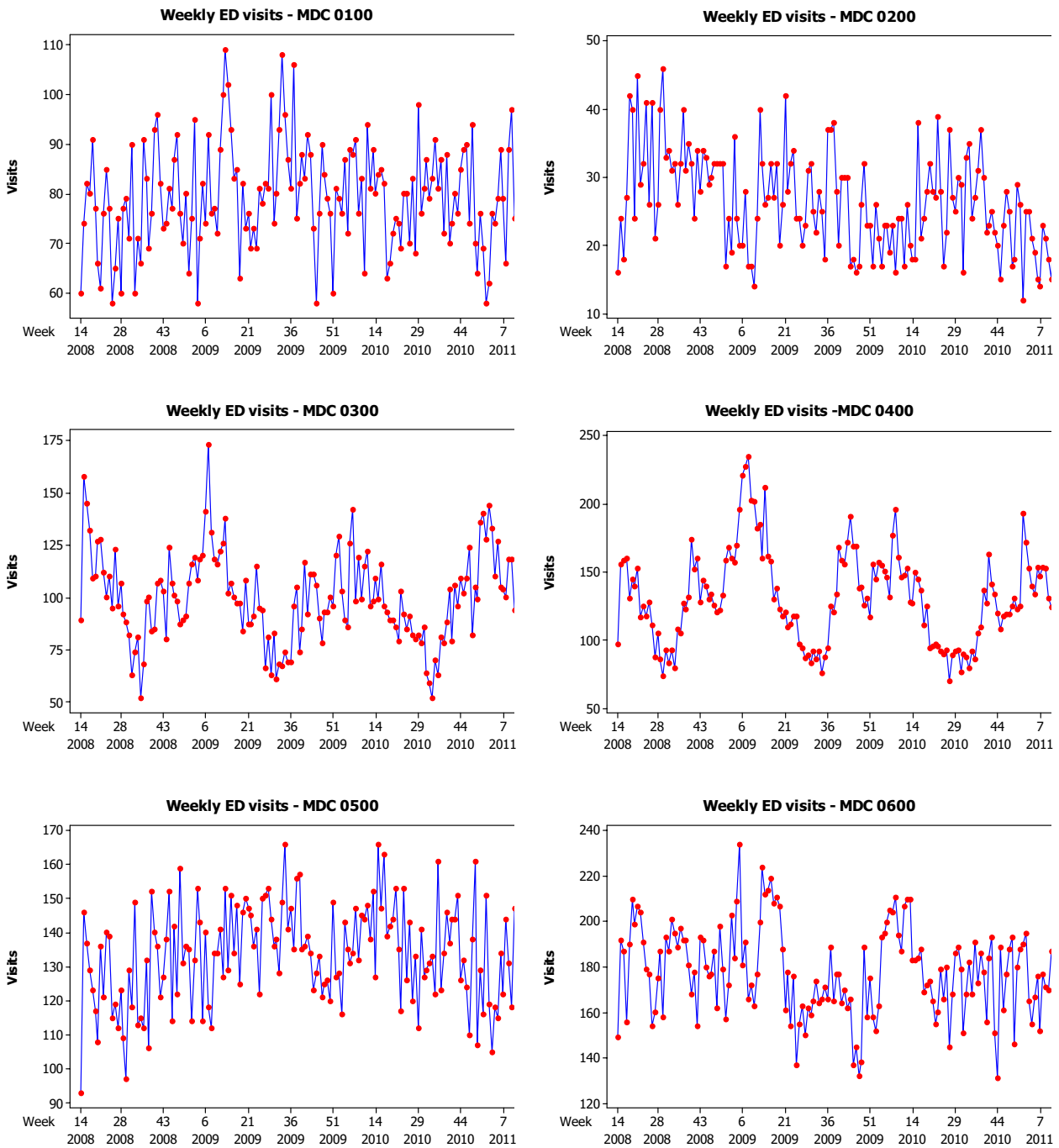


Fig. 1. ED visits per analyzed MDC.

level i is determined by the following formula:

$$(WMAPE)_i = \frac{\sum_{j=1}^i N_j \cdot MAPE_j}{\sum_{j=1}^i N_j}, \quad (10)$$

where

i is the number of considered MDCs (MDCs classified in decreasing average number of weekly visits).

N_j is the average number of visits to a MDC with a ranking j (ranking in descending average number of weekly visits)

$MAPE_j$ is the average percentage absolute errors for a MDC with a ranking j (descending ranking average weekly visits).

Figure 2 presents a combined view of the MDCs cumulative average number of weekly visits (in descending order) and the corresponding adjusted average MAPE. According to these graphs, the proposed methodology in this

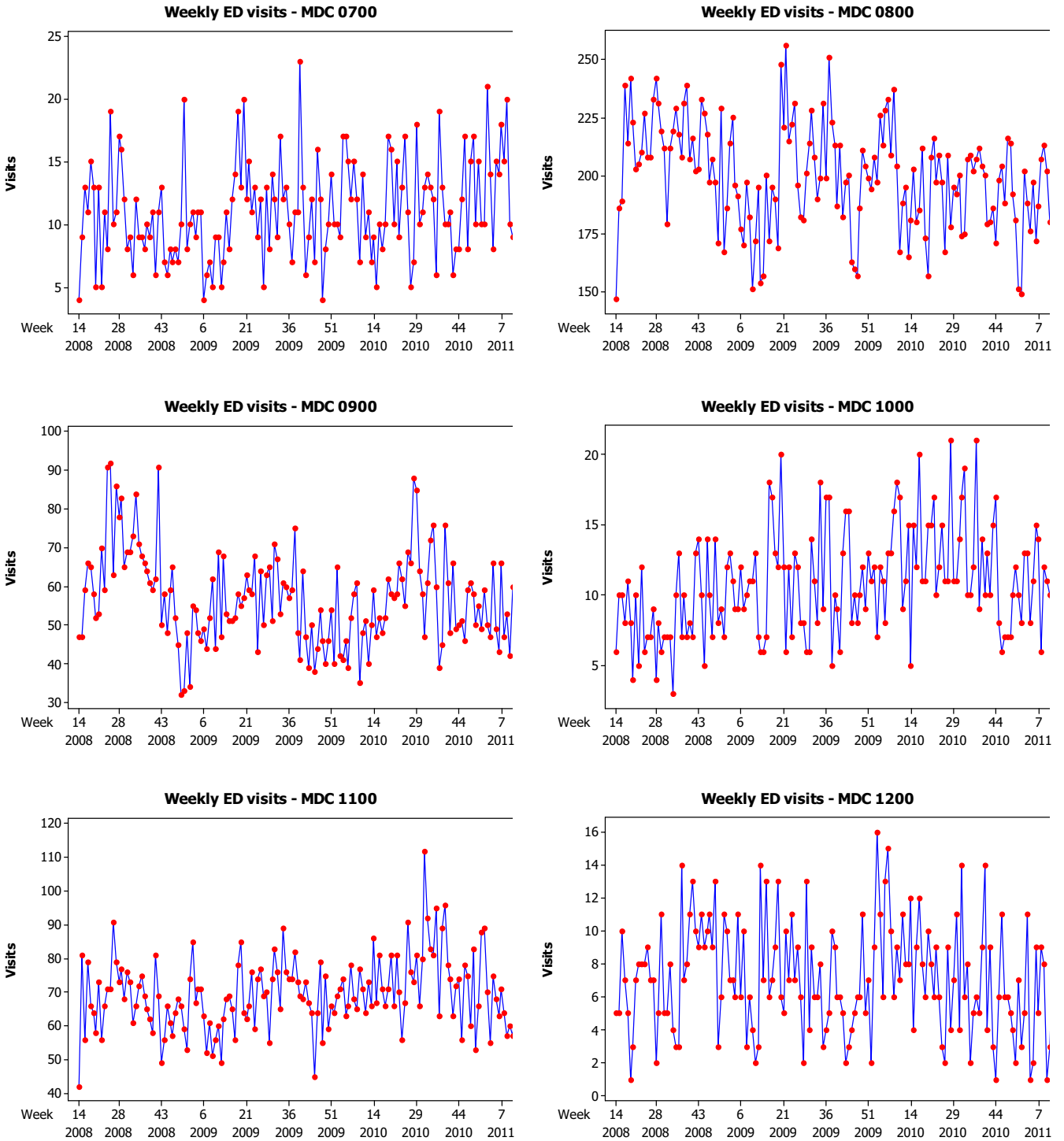


Fig. 1. Continued.

book offers a MAPE of 8.7% to 95% of the total number of patient visits. Although the MDC oriented methodology offers a total MAPE of less than 10%.

The prediction models were not able to provide a good accuracy for the MDC 2300 with a MAPE = 9%. The next MDC is 0500 has a lower MAPE = 6% which helped to decrease the WMAPE.

4.3 Analysis of the daily number of visits to the ED during the week

The analysis of variance (ANOVA) applies to the number of ED visits per day of the week and shows that the majority of the MDCs do not have equal number of visits during the different days of the week. Plotting Box Plots graphs show different interactions between

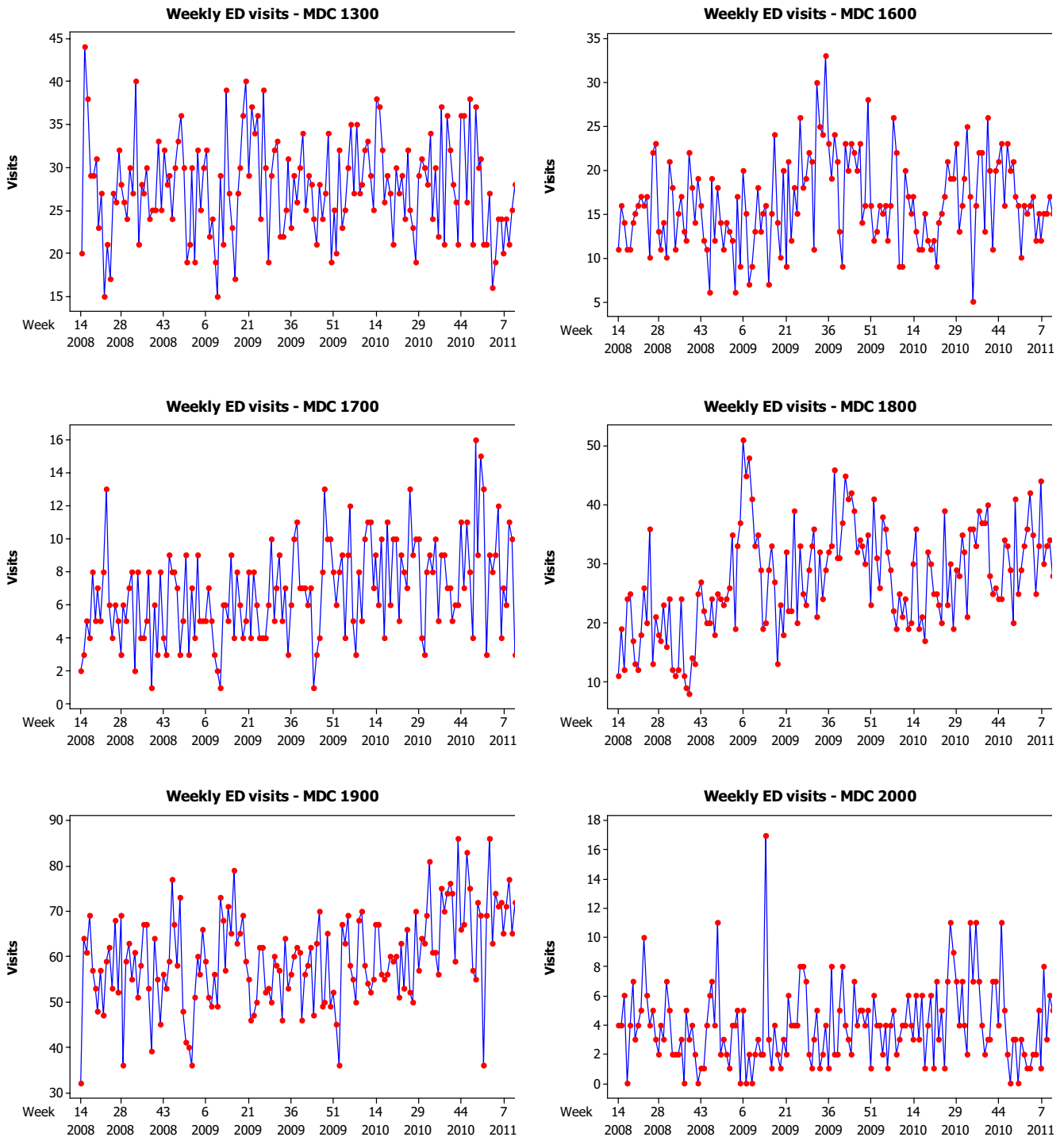


Fig. 1. Continued.

the number of ED visits and the days of the week. Indeed, for MDCs such as 0500 (Diseases and Disorders of the circulatory system), the number of visits during the weekend days are lower than the numbers of visits during the weekdays. Whereas for a MDCs such as 0300 (Diseases and disorders of the ear, nose, mouth, throat and craniofacial bones), the traffic is more frequent during the weekends.

4.4 Admission rate analysis

4.4.1 Prediction of weekly admission rates

Figure 3 shows that there are significant differences regarding admission rates between τ_k and the MDCs k . The MDCs with high admission rates are disorders related to the hepatobiliary system and to heart problems with respective average admission rates of 60.8% and 64.3%,

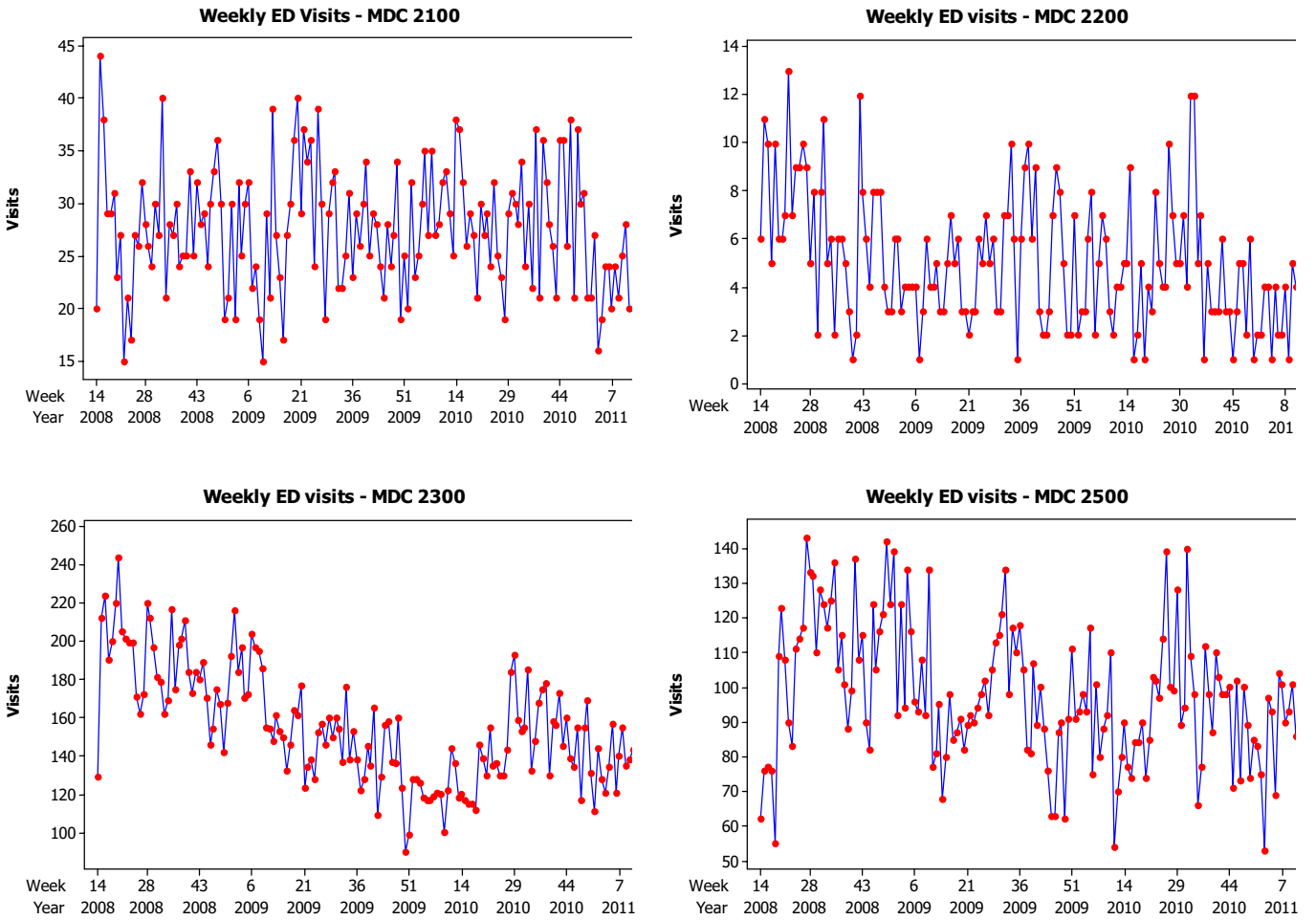


Fig. 1. Continued.

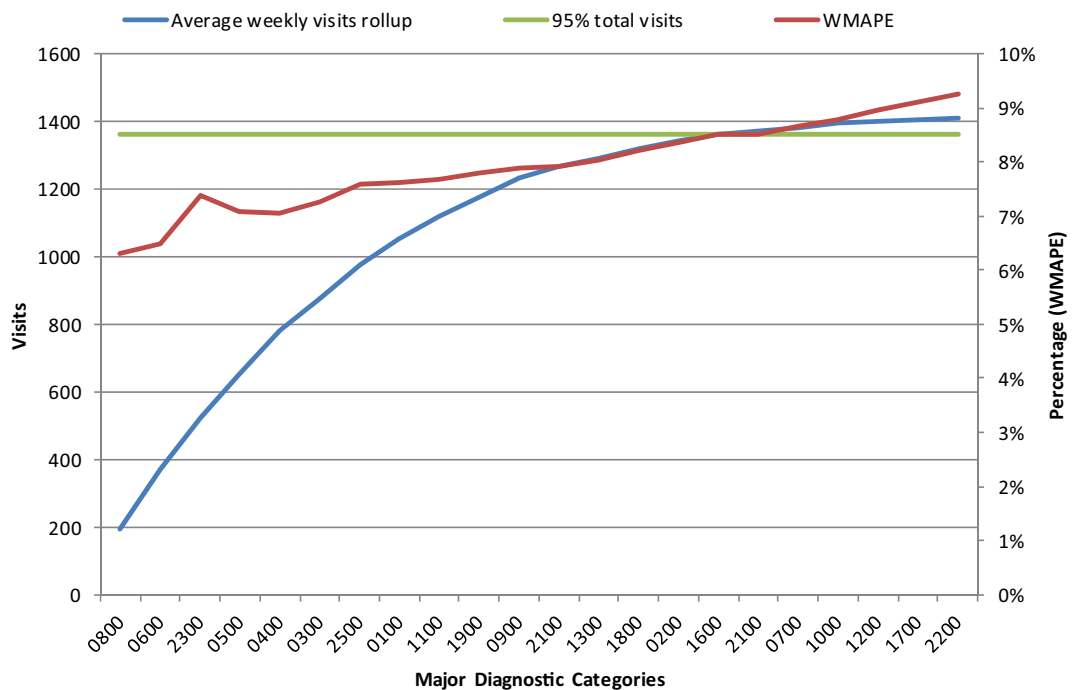


Fig. 2. Weekly visits and WMAPE.

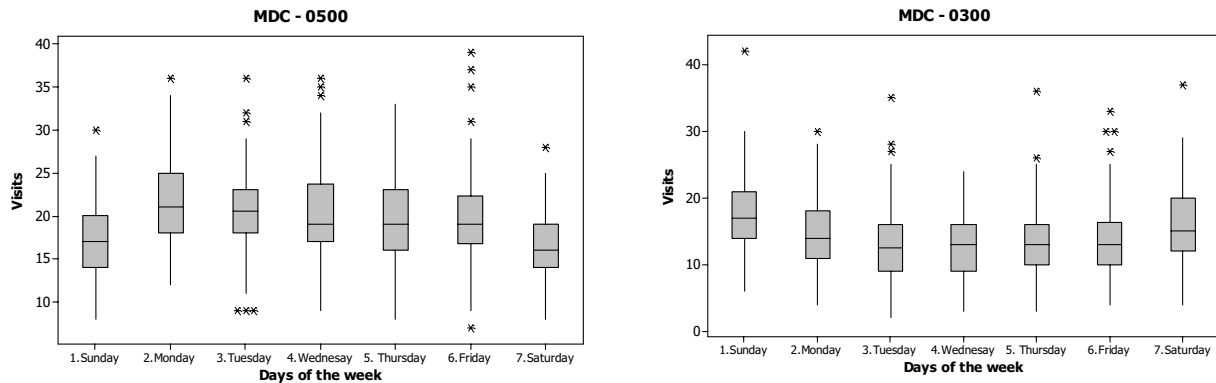


Fig. 3. Visits for each day of the week.

Table 6. Admission rates’s distributions τ_k per MDC.

MDC	Average %	Standard deviation	Fitted distribution	P-value	Comments
0100	26.6	5.7	Normal	0.076	
0200	4.0	2.2			
0300	2.2	1.3	Box-Cox	0.657	Lambda = 1
0400	33.7	5.4	Normal	0.647	
0500	37.8	4.9	Normal	0.098	
0600	23.3	3.4	Normal	0.453	
0700	60.8	16.4	Box-Cox	0.190	Lambda = 2
0800	11.9	2.7	Normal	0.831	
0900	9.9	4.3	Box-Cox	0.944	Lambda = 0.5
1000	40.6	17.0	Box-Cox	0.396	Lambda = 0.5
1100	20.6	5.7	Box-Cox	0.115	Lambda = 0.5
1200	17.2	9.1	Log-Normal	0.066	
1300	10.7	5.9	Gamma	0.200	
1400					
1500					
1600	32.5	12.0	Box-Cox	0.408	Lambda = 0.5
1700	48.9	18.2	Weibull	0.084	
1800	25.8	8.9	Normal	0.306	
1900	41.6	7.1	Normal	0.328	
2000	38.8	17.2		<0.005	
2100	14.7	108	Box-Cox	0.807	Lambda = -0.1
2200	26.1	13.9	Johnson Trans.	0.281	
2300	14.7	3.3	Normal	0.525	
2400					
2500	9.1	3.6	Normal	0.081	
Hearth prob.	64.3	7.3	Normal	0.425	

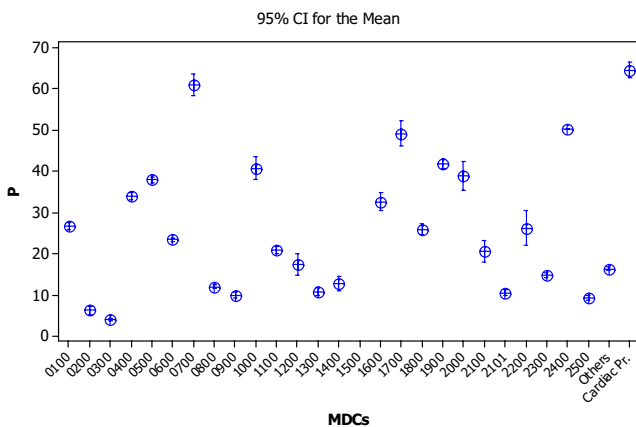


Fig. 4. Admission rate intervals.

while health problems associated with disorders of the ear, nose, mouth, throat and craniofacial bones have the lowest admission rate with an average of 2.2%.

Analyzing admission rates, Table 7 shows that these rates can be modeled by different distributions laws depending on the MDC in question.

4.5 Short term admission rate prediction

This exercise aims to establish a forecasting tool for the collect of short term data regarding the number of patients waiting in the ED waiting room with potential possibilities to be admitted. In the studies of Peck et al. [13], a correlation study was performed for each MDC between

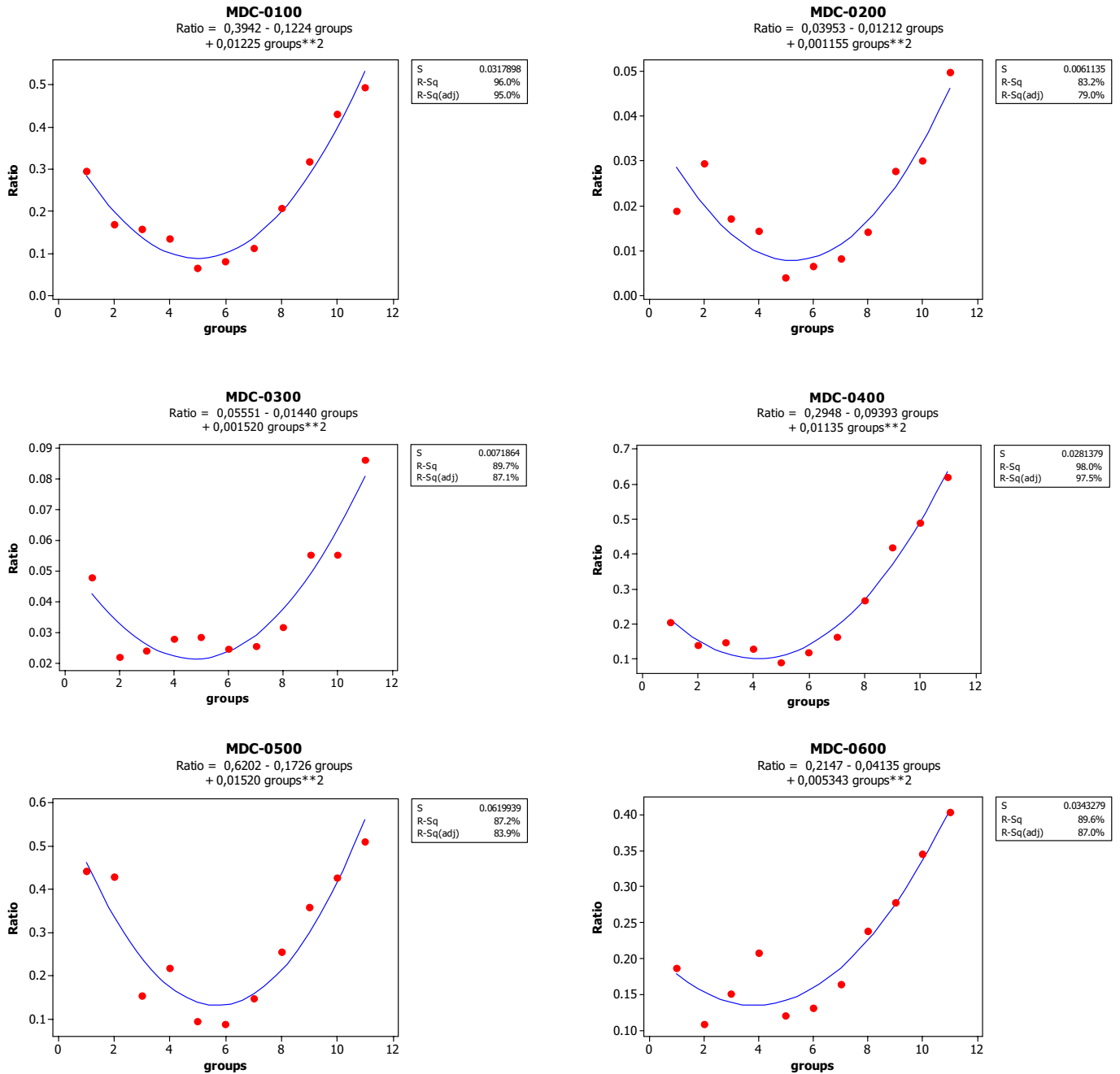


Fig. 5. Correlations: age groups vs. admission ratios per MDC.

age groups and the weekly percentage of admitted patients. Figure 5 shows good quadratic correlations ($R^2 > 87\%$) between age groups and admission ratios for the MDCs with more than 85% of total ED visits. Only the MDCs 1000 and 2200 have low correlation ($R^2 < 50\%$).

Managers and staff complain of a lack of visibility of the Emergency Department status. In this study, we were able to show some strong correlations between the admission rate and the combined variables (MDC, Age groups). This finding could help managers to trace a real time status of the ED after triage, and then help

the admission departments and specialist doctors to react efficiently according to the situations.

4.5.1 Limitations and practical considerations

The aim of the proposed models in this paper is to estimate the number of patients visiting the ED and the number of those who will be admitted. The time unit is the week and not the day as proposed in other works. Choosing the week as time unit avoid having time-series

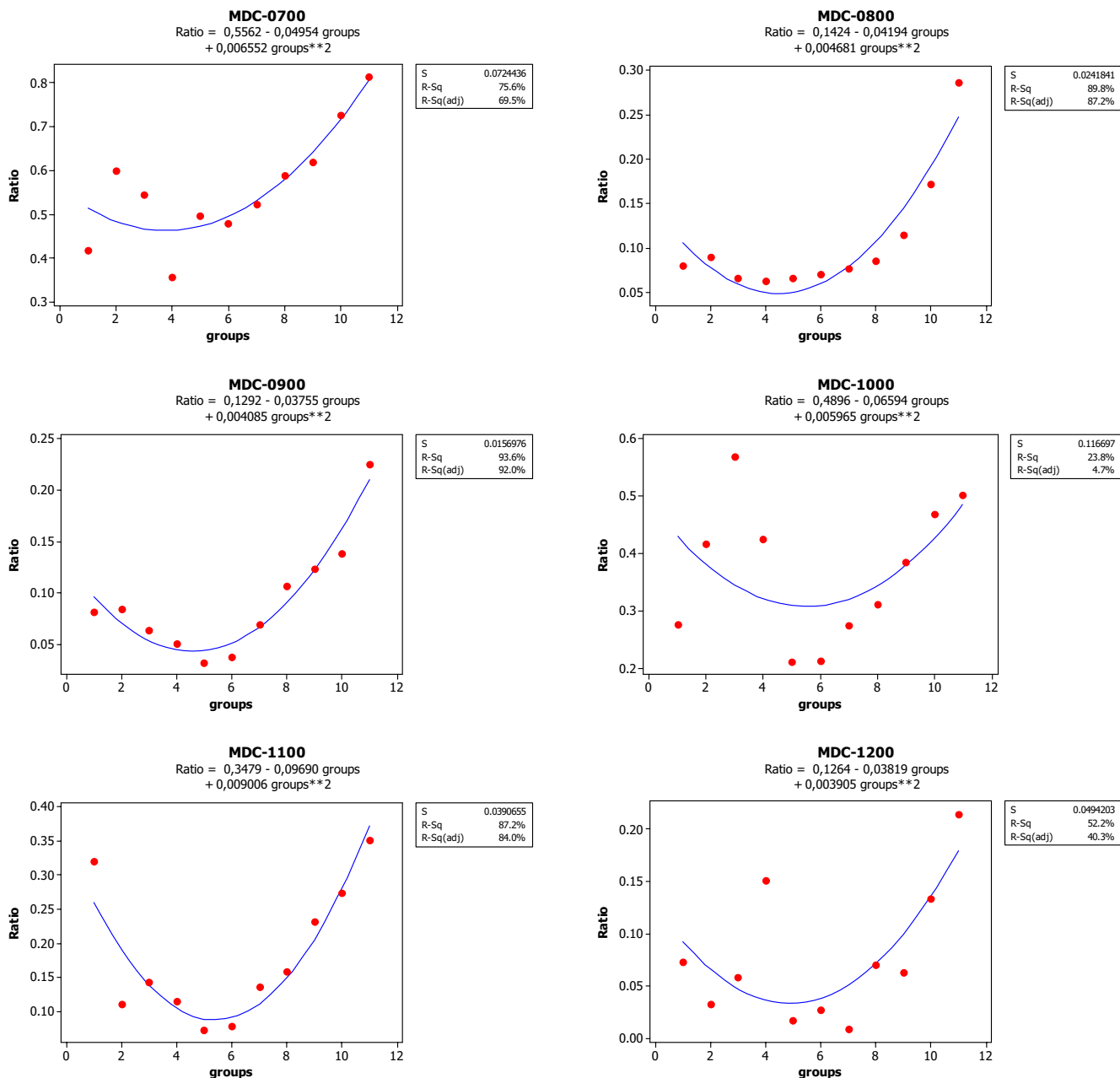


Fig. 5. Continued.

low values that are difficult to treat and also avoid generating non-accurate models. However the week could also be too long and not adequate to make short-term schedules (daily). This limitation may be improved by using weight factors of each day of the week.

In practice, predicting climatic variables over one week is not an easy task. The use of these variables as predictors may affect the quality of the linear regression models and the SARIMAX proposed in this work.

It should also be taken into consideration that this study did not include patients who visit the ED and leave before being seen by a doctor. This departure is often associated with long waiting periods. The number of patients

who left the ED can be very important and significantly affect the quality of the proposed predictive models.

Finally, another limitation of this study is that the patients were treated in a regional teaching hospital in the province of Quebec, where almost all citizens have national health insurance with unrestricted access to emergency care. This should be considered when generalizing the findings of the study to other countries.

5 Conclusion

This study highlights the advantages offered by grouping the patients into Major Diagnostic Categories for patterns of ED visits and admission models.

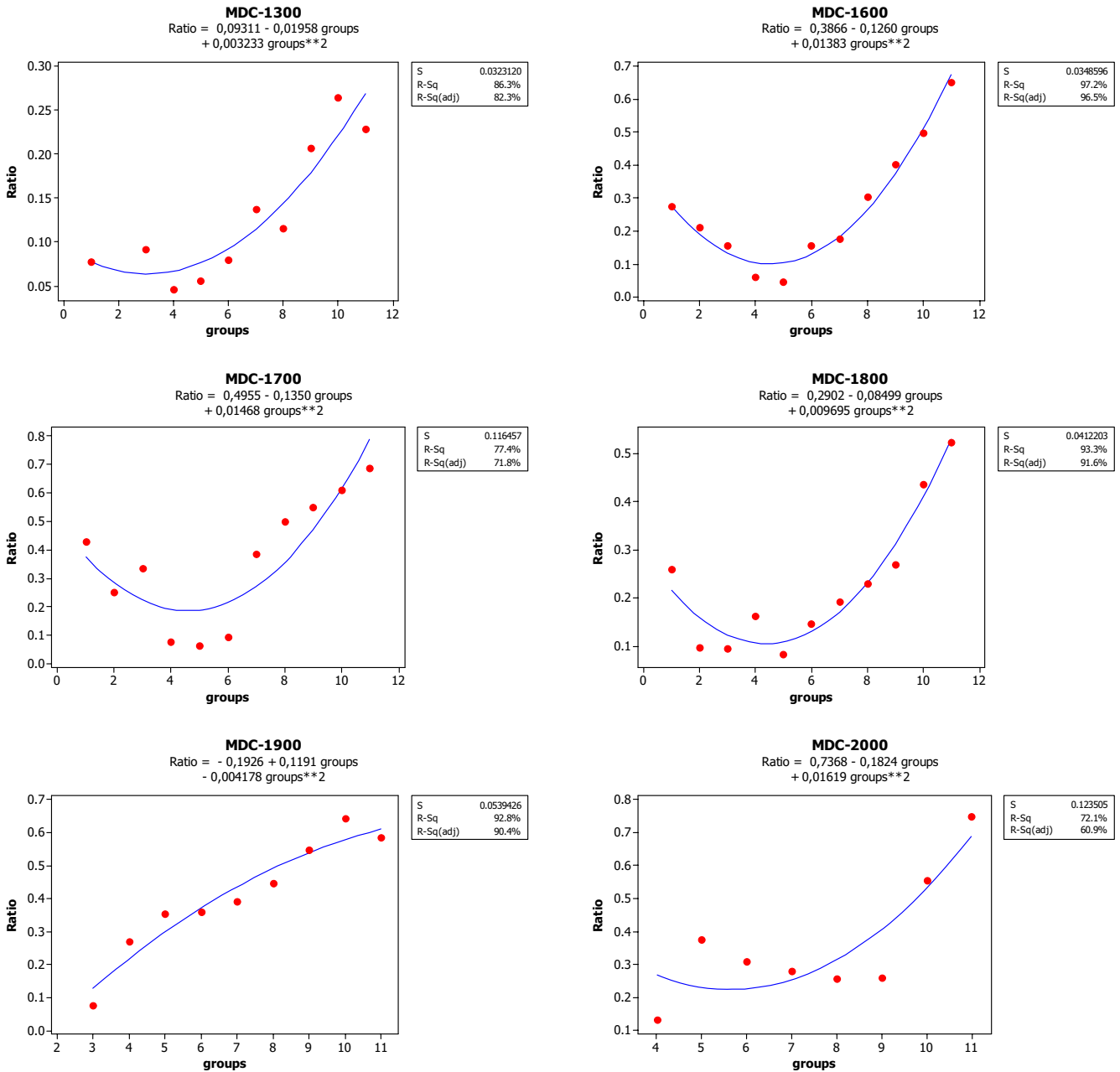


Fig. 5. Continued.

In the first part of the study, three techniques are deployed to model the number of patients visiting ED for each MDC, namely SARIMA univariate linear regression, and multivariate SARIMA (SARIMAX). For the majority of the MDCs the multivariate SARIMA (SARIMAX) provides the best results with the lowest MAPE. The use of these different modeling techniques highlights the behavior of the time series of the number of patient visits to the ED.

Modeling the ED visits per MDC helps managers to better plan resources needed to meet demand. This MDC grouping approach does not affect the quality of the model in estimating the total number of visits.

The second part of this work aims to establish a model to estimate the rate of admitted patients. The number of admitted patients is a highly appreciated information by managers and when this number is allocated to each MDC, it allows a more accurate calculation of the necessary quantity of resources. Proper planning of the resources needed for hospitalizations has a positive effect on the admission of ED process, often identified as a bottleneck in the whole process of emergency. The number of patients admitted is not modeled as a time series as some authors has done [6], but as a product of the time series N_{kt} and the admission ratio τ_k , independent stochastic variables. It is shown that some of these variables are identifiable to normal distribution laws.

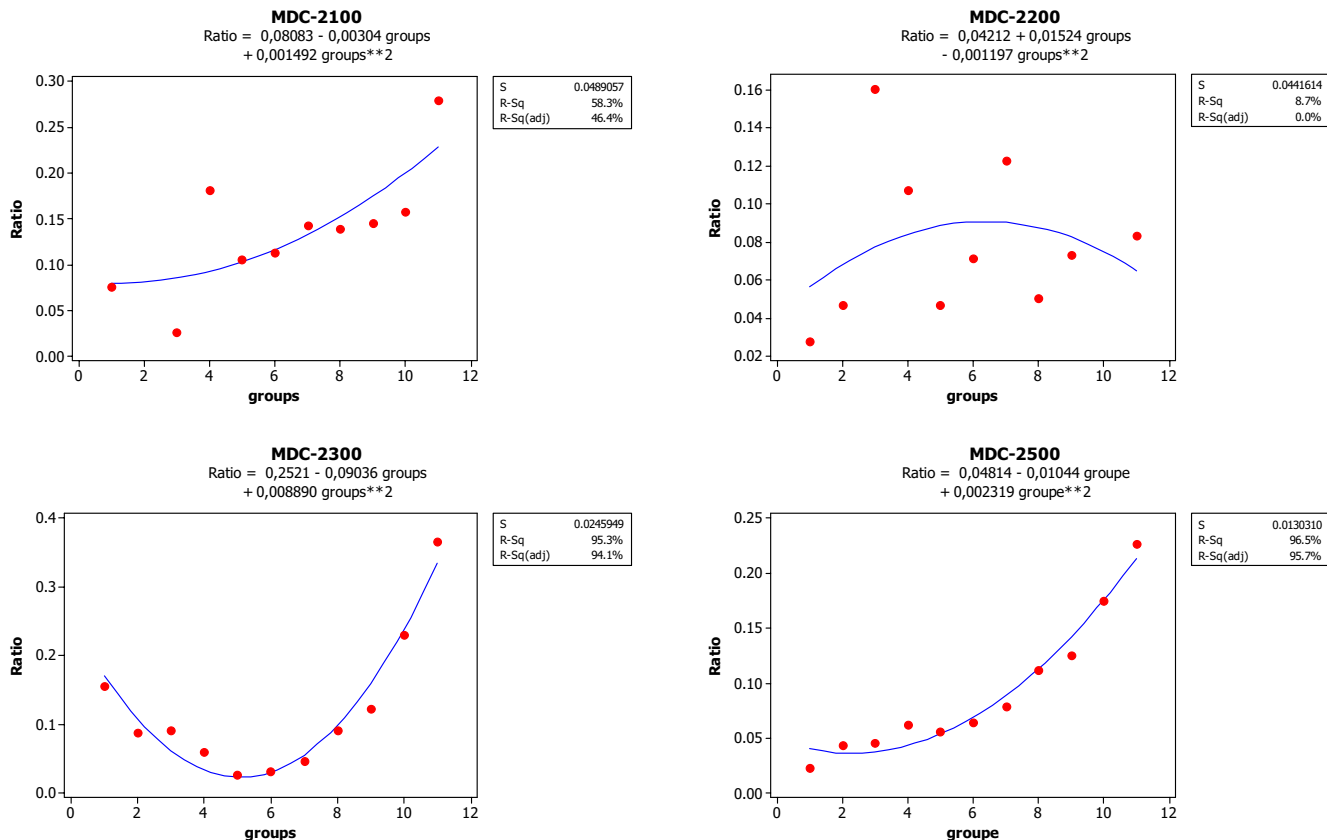


Fig. 5. Continued.

The study also finds significant correlations between the probabilities for a patient to be admitted and the age group he belongs to. This correlation could be used to set the real-time status of ED in order to improve the coordination between all the ED staff members.

References

1. A.G. Perry, K.M. Moore, L.E. Levesque, C.W.L. Pickett, M.J. Korenberg, *A Comparison of Methods for Forecasting Emergency Department Visits for Respiratory Illness Using Telehealth Ontario Calls* (Canadian Public Health Association, 2010)
2. S. Sadeghi, A. Barzi, N. Sadeghi, B. King, A Bayesian model for triage decision support, *Int. J. Med. Inform.* **75**, 403–411 (2006)
3. P. Walsh, P. Cunningham, S.J. Rothenberg, S. O'Doherty, H. Hoey, R. Healy, An artificial neural network ensemble to predict disposition and length of stay in children presenting with bronchiolitis, *Eur. J. Emerg. Med.* **11**, 259–264 (2004)
4. Yan Sun, Bee Hoon Heng, Seow Yian Tay, and Eillyne Seow, Predicting hospital admissions at emergency department triage using routine administrative data (*Soc. Acad. Emerg. Med.*, 2011)
5. J. Boyle, M. Jessup, J. Crilly, D. Green, J. Lind, M. Wallis, P. Miller, G. Fitzgerald, Predicting emergency department admissions, *Emerg. Med. J.* **29**, 358–369 (2012)
6. G. Abraham, G.B. Byrnes, C.A. Bain, "Short-term forecasting of emergency inpatient flow", *IEEE Trans. Inf. Technol. Biomed.* **13**, 380–388 (2009)
7. M. Wargon, E. Casalino, Bertrand Guidet, MD, *From model to forecasting: A multicenter study in emergency departments* (the Society for Academic Emergency Medicine, 2010)
8. C.-F. Chen, W.-H. Ho, H.-Y. Chou, S.-M. Yang, I.-T. Chen, H.-Y. Shi, Long-term prediction of emergency department revenue and visitor volume using autoregressive integrated moving average model, *Comput. Math. Methods Med.* **2011**, 395690 (2011)
9. Hye Jin Kam, Jin Ok Sung, Woong Park, *Prediction of daily patient numbers for a regional emergency medical center using time series analysis* (The Korean Society of Medical Informatics, 2010)
10. Y. Sun, B.H. Heng, Y.T. Seow and E. Seow, Forecasting daily attendances at an emergency department to aid resource planning, *BMC Emerg. Med.* **9**, 1 (2009)
11. L.M. Schweigler, J.S. Desmond, M.L. McCarthy, K.J. Bukowski, E.L. Ionides, J.G. Younger, *Forecasting models of emergency department crowding* (Society for Academic Emergency Medicine, 2009)
12. S.J. Spencer, T. Alun, R.S. Evans, S.J. Welch, P.J. Haug, G.L. Snow, *Forecasting daily patient volumes in the emergency department* (The Society for Academic Emergency Medicine, 2008)
13. J. Peck, S. Gaehde, J. Benneyan, S. Graves, D. Nightingale, *Using prediction to improve patient flow in a health care delivery chain*

14. Guide de gestion de l'unité d'urgence, ministère de la santé et des services sociaux, 2000
15. R.J. Nelson, G.E. Demas, S.L. Klein, L.J. Kriegsfeld, F. Bronson, *Seasonal fluctuations in disease prevalence* (Cambridge University Press, 2002, pp. 58-88)
16. N. Shlesinger, M. Schlesinger, Seasonal variation of rheumatic diseases, *Discov. Med.* **5**, 64–69 (2005)
17. J.R. Broyles, J.K. Cochran, A queuing-base statistical approximation of hospital emergency department boarding, in *Proc. of the 41st International Conference on Computers & Industrial Engineering*
18. Champion R, Kinsman LD, Lee GA, et al. Forecasting emergency department presentations. *Aust. Health Rev.* **31**, 83–90 (2007)
19. N.R. Hoot et al., *Forecasting emergency department crowding: An external multicenter evaluation* (The American College of Emergency Physicians, 2009)
20. M.L. McCarthy, S.L. Zeger, R. Ding, D. Aronsky, N.R. Hoot, G.D. Kelen, *The challenge of predicting demand for emergency department services* (The Society for Academic Emergency Medicine, 2008)
21. B.J. Morzuch, P. Geoffrey Allen, Forecasting hospital emergency department arrivals, in *Proc. of 26th Annual Symposium on Forecasting Santander, Spain, June 11–14, 2006*
22. Association québécoise d'établissements de santé et de service sociaux, *Guide de gestion de l'urgence* (2006)
23. N.R. Hoot, D. Aronsky, *Systematic review of emergency department crowding: Causes, effects, and solutions* (The American College of Emergency Physicians, 2008)
24. D.W. Spaite, F. Bartholomeaux, J. Guisto, E. Lindberg, B. Hull, A. Eyherabide, S. Lanyon, E.A. Criss, T.D. Valenzuela, C. Conroy, Rapid process redesign in a university-based emergency department: Decreasing waiting time intervals and improving patient satisfaction, *Ann. Emerg. Med.* **39**, 168–177 (2002)
25. D.R. Eitel, S.E. Rudkin, M.A. Malvey, J.P. Killeen, J.M. Pines, Improving service quality by understanding emergency department flow: a white paper and position statement prepared for the american academy of emergency medicine, *J. Emerg. Med.* **38**, 70–79 (2010)
26. H. Beaulieu, J.A. Ferland, B. Gendron, P. Michelon, A mathematical programming approach for scheduling physicians in the emergency room, *Health Care Manag. Sci.* **3**, 193–200 (2000)
27. Ministère de la Santé et des Services sociaux du Québec, Rapport annuel de gestion 2010–2011, 2011
28. Ordre des infirmières et infirmiers du Québec, Le triage à l'urgence: lignes directrices pour l'infirmière, 2007
29. S.J. Welch, B.R. Asplin, S. Stone-Griffith, S.J. Davidson, J. Augustine, J. Schuur, *Emergency department operational metrics, measures and definitions: Results of the second performance measures and benchmarking summit* (The American College of Emergency Physicians, 2010)
30. Institut canadien d'information sur la santé, Temps d'attente et soins de santé au Canada : ce que nous savons et ce que nous ignorons, 2006
31. K. Grumbach, D. Keane, A. Bindman, Primary care and public emergency department overcrowding, *Am. J. Public Health* **83**, 372–378 (1993)
32. S.J. Spencer, R.E. Scott, T.L. Allen, A. Thomas, P.J. Haug Shari, J. Welch C, G.L. Snow, A multivariate time series approach to modeling and forecasting demand in the emergency department, *J. Biomed. Inform.* **42**, 123–139 (2009)
33. Ordre des infirmières et infirmiers du Québec, *Lignes directrices pour l'infirmière pour le triage à l'urgence*, Zotac Nvidia GeForce GTX 660 2GB DDR5 HDMI and Dual DVI
34. User Guide, Rats 7, Estima, 2007