Time-Series Approaches for Forecasting the Number of Hospital Daily Discharged Inpatients

Ting Zhu, Li Luo, Xinli Zhang, Yingkang Shi, and Wenwu Shen

***Abstract*—For hospitals where decisions regarding acceptable rates of elective admissions are made in advance based on expected available bed capacity and emergency requests, accurate predictions of inpatient bed capacity are especially useful for capacity reservation purposes. As given, the remaining unoccupied beds at the end of each day, bed capacity of the next day can be obtained by examining the forecasts of the number of discharged patients during the next day. The features of fluctuations in daily discharges like trend, seasonal cycles, special-day effects, and autocorrelation complicate decision optimizing, while time-series models can capture these features well. This research compares three models: a model combining seasonal regression and ARIMA , a multiplicative seasonal ARIMA (MSARIMA) model, and a combinatorial model based on MSARIMA and weighted Markov Chain models in generating forecasts of daily discharges. The models are applied to three years of discharge data of an entire hospital. Several performance measures like the direction of the symmetry value, normalized mean squared error, and mean absolute percentage error are utilized to capture the under- and overprediction in model selection. The findings indicate that daily discharges can be forecast by using the proposed models. A number of important practical implications are discussed, such as the use of accurate forecasts in discharge planning, admission scheduling, and capacity reservation.**

***Index Terms*—Bed availability, hospital discharges, msarima, srarima, weighted markov chain.**

I. Introduction

THE increases in demand for healthcare products and services from a growing ageing population, increases in morbidity rate from environment deterioration and unhealthy lifestyle behavior, and resource supplement insensitivity, are just some of the factors that continue to put an upward pressure on optimal management of healthcare resources [1]. Rising competition in scarce resources among demands with distinct priorities often make resource planning and scheduling decisions more significant as the share of these resources relative to the operating cost increases and patient care quality declines. As such, hospital managers have to manage resources effectively, while maintaining a high quality of care [1]. Accurate forecasts of future healthcare demand and resource availability is critical for efficient planning and scheduling decisions at all levels, which makes prediction problems about emergency patient flow (see [1]–[3]), occupancy and length of stay (LOS) (see [4]–[6]), patient output and inpatient bed availability (see [7], [8]) have attracted growing attention in healthcare management contexts.

Motivated by overcrowding and service capacity management problems arising in healthcare systems in China, by choosing a typical hospital—West China Hospital (WCH) as cooperated hospital, we consider a time-series forecasting problem of daily discharges to reduce the uncertainty of capacity, so that the decision making of bed reservation and patient admission scheduling can be optimized. WCH is an urban and public tertiary teaching hospital in Chengdu, China. It operates a large inpatient department which has a capacity of about 4800 licensed beds shared by 38 specialty care units as of December 31, 2014. Like most other tertiary hospitals in China, it has been experiencing an increase in the degree of overcrowding (4.91 million outpatient-emergency patients, 0.22 million admissions, and 0.13 million surgeries in 2014), waiting lists for hospitalization have been growing (over 6000 patients waiting to be admitted as of each day).

At the WCH, available bed capacity each day is a competition between backlogged elective requests on waiting list and randomly arriving emergent requests. Admitted elective patients represent about two-third of the total hospital occupancy. At the end of each working day, bed manager devotes a random quota of expected capacity of the next day to backlogged electives in advance; the remainder, can be used to serve emergencies. The expected capacity is the sum of left beds today and the number of discharges the next day. During the next day, available beds are assigned and filled, once scheduled electives and new patients arrive randomly. Since backlogged electives can renege, possibly causing losses of revenue or additional society costs; but scheduling too many of them may result in costs from overflowing or postponing some emergent requests. Generally, beds are extremely limited to satisfy all demand in a timely pattern. In addition to the shortage of bed resource, an additional difficulty stems from the fact that the total bed capacity cannot

Manuscript received May 18, 2015; revised September 19, 2015 and November 3, 2015; accepted December 21, 2015. Date of publication December 23, 2015; date of current version March 3, 2017. This work was supported by the Natural Science Foundation of China (Research on big data-driven resource management of whole society in environment and smart health care under Grant 71532007, Resource dispatching and optimizing research in medical service under Grant 71131006, Design of Sichuan Medicare payments based on patients' medical value maximization and big data under Grant 2015ZR0097, and Hospital resource planning and control optimizing based on patients' demand under Grant 71172197).

T. Zhu, L. Luo, and X. Zhang are with the Department of Industrial Engineering and Engineering Management, Business School of Sichuan University, Chengdu 610065, China (e-mail: 745358376@qq.com; luolicc@163.com; zhangxinli1231@163.com).

Y. Shi and W. Shen are with the West China Hospital of Sichuan University, Chengdu 610065, China (e-mail: syk@mcwcums.com; 573596113@qq.com).

Digital Object Identifier 10.1109/JBHI.2015.2511820

be known with certainty when the quota is allocated each day. So the problem facing bed schedulers is how much of the expected bed capacity to assign for backlogged elective admission requests, and how much for the next day, and so on.

Our investigations suggest that the expected bed capacity is obtained daily by implementing a planned discharge policy in WCH. Group leaders of all specialty care units submit the next day's discharge plan of their patients and the number of remaining beds today to the admission service center (ASC) of WCH before 18:00 each day, based on which bed schedulers of ASC schedule the elective admissions. The number of remaining beds can be known because the duration from 18:00:00 today to 06:59:59 the next day is out of discharge and admission time window; although very few special cases, such as night-death discharges, night-referral discharges, and very few night admissions lead to no more than three beds deviation per day. However, it is of great possibility that the planned patients cannot be discharged for some uncertain reasons, such as changes of their state of illness, postponed MRI/CT examination reports for discharge identification, unsuccessful reimbursement, etc. Admission decisions based on planned discharge information with a current approximate 75% successful planned discharge rate may result in the situation that realized bed capacity cannot satisfy all of the scheduled electives. Hence, the uncertainty of bed capacity mainly results from the uncertainty of daily discharges. In such a case, the overestimate of bed capacity leads to overscheduled elective patients, then leading to high cost-related overflow rate and emergency admissions delaying. This resultantly makes it difficult to evaluate the long-term consequences of decisions taken, and to compare the effects of several possible plans based on accurate bed capacity information. "Is there a way for us to anticipate today how much bed capacity we will have tomorrow?" That question, posed to us by a patient admission administrator was the genesis for the study presented here. This highlights the need for a data-based modeling approach to provide daily discharges forecasts for the entire hospital. However, there are fundamental limitations to what such approaches can achieve: the random fluctuations inherent in random processes limits how well they can be predicted. In this paper, we seek to determine how these limits impact the predictions by proposing three time-series analysis models.

This paper examines the issue of forecasting the number of daily discharges. The objective is to lay out the structure of our methodology, present data and modeling details and results of modeling the daily patient discharges. Seeking to achieve good forecasting performance, we compare three models for forecasting daily discharges. We construct a seasonal regression ARIMA (SRARIMA) model in which seasonal regression is used to capture the day-of-the-week effect and ARIMA model is then fitted to the regression residuals. Moreover, a multiplicative seasonal ARIMA (MSARIMA) model which is a special seasonal ARIMA model to fit seasonal time series is utilized to treat seasonality as stochastically varying. We then compare the forecasts of SRARIMA, MSARIMA with those produced by a combinatorial model based on MSARIMA and weighted Markov Chain. In order to do this, we collect the real hospital data and whether the data can actually be forecast is validated by a white noise test. Furthermore, multiple indicators are introduced to measure the forecasting performance of each of the applied time-series methods. Our results indicate that the combinatorial model outperforms the linear alternatives and may be a viable alternative to the traditional time-series approaches (SRARIMA, MSARIMA). Much study remains to be done to make hospitalwide real-time census prediction a reality in today's complex healthcare delivery systems. However, the findings of this research could be of significant value for practitioners (hospital managers) in taking appropriate decisions.

The remainder of the paper is organized as follows. In the next section, we describe the related study. We present the theory and concept of SRARIMA, MSARIMA, Markov Chain, a combinatorial model based on MSARIMA and weighted Markov chain, and multiple measures used for the performance evaluation in Section III. Section IV presents the methodology and results of empirical experiments. In this section, the experiment results yielded by the proposed models for daily discharge time-series data are presented. We discuss the limitations and future work in Section V. In Section VI, we provide our concluding remarks.

II. Literature Review

Predicting hospital resource availability has been and continues to be a popular topic of research. There have been several studies that have explored various time-series forecasting models for hospital discharges and LOS (see [4]–[6]). This study differs from previous research, in which we attempt to simplify the bed capacity forecasting problem by capturing the fluctuating characteristics of the entire hospital's daily discharges and address the problem of whole hospital discharge census prediction. In 2000, Makridakis and Hibon advocated: "prediction remains the foundation of all science" [9]. As mentioned by Petropoulos *et al.*, predictions are significant for all decision making tasks, from inventory management and scheduling to planning and strategic management [10]. Improved short-term information has the potential to improve the matching of hospital resources to fluctuating demand levels for a number of hospital business processes, such as patient admission control policies [11], reducing periods of under- and overstaffing [12], and management of bed capacity [13].

Time-series forecasting is important for healthcare managers to capture features of short-term fluctuations better. Variations of autoregressive (AR) and moving average (MA) methods are utilized to conduct stochastic process analysis [14]; thus, our comparator method is the general class of ARIMA models. Since linear patterns can be perfectly captured by simple ARIMA models or seasonal ARIMA models, most studies use them as the benchmark to test the effectiveness of combined models with mixed results. However, the use of traditional ARIMA models that assume linear relationships between past observations and subsequent forecasts may be problematic and lead to poor forecasts if the data actually exhibits nonlinear patterns [14], so a combinatorial model based on ARIMA and artificial neural networks (ANNs) model is established to improve forecasting accuracy [15]. Identification of the best forecasting techniques for each data set is still "holy grail" in the forecasting field, so empirical comparisons are considered very important [16].

Krogh *et al.* proved that when the single models composing combinatorial forecasting model are accurate enough and diverse enough, the combinatorial forecasting can get better prediction accuracy than a single model [17]. Channouf *et al.* reviewed the study focusing on improving the prediction accuracy by combining different single models and constantly changing combinations of the chosen single model [18]. Using combinatorial model or combining multiple models based on combinations of ARIMA, Markov Chain, support vector machine, ANNs, and fuzzy logic models has shown wide concern. These forecasting models alone all have some limitations, such as inaccurate prediction of linear patterns or untraceable of nonlinear patterns. Moreover, the use of most of these models needs more complex functions between input and output data, or needs an increased number of input variables for both training and utilization, which seriously degrades the prediction performance like overtraining and extrapolation [19]. When it is not possible to find the relevant explanatory variables which are related to the variable being forecast or when the explanatory variables are neither observed nor observable, these models are not useful [20]. High forecasting accuracy of time-series data depends not only on modeling but also on understanding of the characteristics of data, which means that advanced, sophisticated, and simpler extrapolation methods must be associated with specific features of data [10]. Forecasting models alone are not sufficient for decision making [21]. To better fit time-series data with both linear and nonlinear patterns, such as seasonality, trend, cycle, randomness number of observations, interdemand interval, and coefficient of variation, the forecasting model should be adaptively formed based on the features presented from empirical datasets where no theoretical guidance is available.

Recently, Markov chain models have also been extensively studied and applied in various combinatorial forecasting (see [19], [22]–[24]). Many studies modify the forecasting errors of a single prediction model by using Markov chain approaches, which advocate that Markov Chain approaches are useful for time-series data with large random fluctuations (see [25], [26]). Considering the large random fluctuations within healthcare time-series data, many studies make advantage of the state-transition probability matrix of Markov chain to model the inherent random patterns. As we know, it is actually difficult to figure out whether a specific model is effective or a right technique for the unique data before empirical experiments. We can conclude that the model selection mainly depends on the type of data and comparison studies of multiple models or model combinations. In addition, considering the features of large random fluctuations within discharging the patient-flow data, the advantage of Markov chain model on handling this kind of time series, and neither ARIMA model nor Markov chain model can be adequate in modeling and forecasting healthcare time series, we combine ARIMA with Markov chain models, complex autocorrelation structures and features can be interpreted and modeled more accurately.

III. Methods

In this section, we describe three models for forecasting the daily levels of discharges. The first two seasonal ARIMA models serve as a base case against which the more sophisticated combinatorial model is compared.

A. Model Combining SRARIMA

Seasonal regression assumes that the seasonality is deterministic [1]. A seasonal regression model is constructed by first detrending and deseasonalizing the time series, using a least-squares regression of the dependent variable on time, and on explanatory variables for seasonal effects (such as the day-of-the-week effect, the month-of-the-year effect, or the quarter-of-the-year effect etc.), which can be identified from the original data. Hence, the seasonal forecasts of the dependent variable can be obtained. We assume that the effects of trend and seasonality are additive. The seasonal effect forecasts can be denoted as $\hat{X}(k)$ which can be formulated as

$$\hat{X}(k) = \alpha k + \sum_{i=1}^{I} q_{i,k}\beta_i \tag{1}$$

where k is the day index, $q_{i,k}$ is an explanatory variable for seasonal effect, namely the so-called dummy variables with values 0 or 1. Assume it is a day-of-the-week effect, the value of I is 7, the regression model can be written as $\hat{X}(k) = \alpha k + \beta_1 q_{1,k} + \beta_2 q_{2,k} + \cdots + \beta_7 q_{7,k}$. We can denote the seven seasonal dummy variables as

$$q(i,k) = \begin{cases} \mathrm{q}(1,\mathrm{k}) = 1 \text{ for Monday} & \mathrm{q}(1,\mathrm{k}) = 0 \text{ for other days} \\ \mathrm{q}(2,\mathrm{k}) = 1 \text{ for Tuesday} & \mathrm{q}(2,\mathrm{k}) = 0 \text{ for other days} \\ \mathrm{q}(3,\mathrm{k}) = 1 \text{ for Wednesday} & \mathrm{q}(3,\mathrm{k}) = 0 \text{ for other days} \\ \mathrm{q}(4,\mathrm{k}) = 1 \text{ for Thursday} & \mathrm{q}(4,\mathrm{k}) = 0 \text{ for other days} \\ \mathrm{q}(6,\mathrm{k}) = 1 \text{ for Friday} & \mathrm{q}(5,\mathrm{k}) = 0 \text{ for other days} \\ \mathrm{q}(5,\mathrm{k}) = 1 \text{ for Saturday} & \mathrm{q}(6,\mathrm{k}) = 0 \text{ for other days} \\ \mathrm{q}(7,\mathrm{k}) = 1 \text{ for Sunday} & \mathrm{q}(7,\mathrm{k}) = 0 \text{ for other days} \end{cases} \tag{2}$$

α and β_i are the regression coefficients of trend effect and seasonal effect, respectively. The regression residuals are the difference between the observations and the seasonal forecasts formulated above with expression as

$$\varepsilon(k) = X(k) - \hat{X}(k). \tag{3}$$

An ARIMA model is then fitted to the residuals. ARIMA describes the autocorrelation structures of a set of time-dependent random variables, in order to characterize the continuous movements of a forecasting variable, and anticipate the future values of the variable based on past and present values of time-series data. Hence, in an ARIMA model, future value is considered as a linear function of past observations and the corresponding random residuals. The essence of the model is a combination of differencing operation (smoothing) and ARMA model. The general form of the $ARIMA(p,d,q)$ model for fitting $\varepsilon(k)$ is as

$$\Phi(B)\varepsilon(k) = \Theta(B)a(k) \tag{4}$$

where B is the backshift operator defined as $B\tilde{z}_k = \tilde{z}_{k-1}$, in which $\tilde{z} = z - \mu$. $\tilde{z}$ is the value of the time series after subtracting the mean μ of the series. $\Phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \cdots$

$-\phi_p B^p$ and $\Theta(B) = 1 - \theta_1 B - \theta_2 B^2 - \cdots - \theta_q B^q$ are the polynomials in B for AR and MA components, respectively, and $a(k)$ is the error of the residual sequence at time k. When we substitute (2) into (3), the obtained equation for forecasting is as

$$\begin{aligned} X(k) = \ & \hat{X}(k) + \phi_1(X(k-1) - \hat{X}(k-1)) \\ & + \cdots + \phi_p(X(k-p) - \hat{X}(k-p)) \\ & + (1 - \theta_1 B - \cdots - \theta_q B^q)a(k). \end{aligned} \tag{5}$$

By estimating the ARIMA coefficients in (5), forecasts of the subsequent periods can be obtained by substituting the estimates into (5).

B. Multiplicative Seasonal ARIMA Model

Another approach to forecasting seasonal time series is the MSARIMA model. When time-series data possesses both short-term correlations and seasonality effects, and sophisticated correlations between the two kinds of effects exist, it is actually difficult to completely fit the features of given data by only using simple ARIMA model or seasonal ARIMA model, usually MSARIMA model is applied. That is, the underlying processes that generate the time series has the following form $ARIMA(p, d, q) \times (P, D, Q)_S$, which is comprised of the random seasonal model $ARIMA(P, D, Q)_S$ and $ARIMA(p, d, q)$

$$\Phi(B)U(B^S)\nabla^d \nabla_S^D X(k) = V(B^S)\Theta(B)\varepsilon(k) \tag{6}$$

where $X(k)$ and $\varepsilon(k)$ are the observed value and random error at time period k. $\Phi(B)$ and $\Theta(B)$ are mentioned above, $U(B^S)$ and $V(B^S)$ are formulated as

$$U(B^S) = \ 1 - u_1 B^S - u_2 B_S^2 - \cdots - u_P B_S^P \tag{7}$$

$$V(B^S) = \ 1 - v_1 B^S - v_2 B_S^2 - \cdots - v_Q B_S^P. \tag{8}$$

In order to interpret the correlations among different points in the same cycle, $\Phi(B)$ and $\Theta(B)$ are introduced; simultaneously, $U(B^S)$ and $V(B^S)$ are formulated for capturing the correlations among the same point of different cycles. Respectively, $\phi_i (i = 1, 2, \ldots, p), \theta_j (j = 1, 2, \ldots, q), u_m (m = 1, 2, \ldots, P)$, and $v_n (n = 1, 2, \ldots, Q)$ are model parameters. B is the backshift operator, and $\nabla = 1 - B$. Moreover, (p, d, q) represents the AR order, differencing order, and MA order of $ARIMA(p, d, q)$. Similarly, (P, D, Q) represents the AR order, differencing order, and MA order of $ARIMA(P, D, Q)_S$, and S is the number of observations in a seasonal cycle. Therefore, $\Phi(B)\nabla^d X(k)$ represents the correlations among different points in a same cycle and $U(B^S)\nabla_S^D$ represents the correlations among the same point of different cycles. Equation (6) is a general model entailing several important special cases of the ARIMA family of models by changing the values of (p, d, q) and (P, D, Q). The modeling principle of the MSARIMA model can be described as: 1) After trend differencing with order d and seasonal differencing with step length (cycle) S on the raw nonstationary time series, if short-term correlations still exist, correlations exist among different points of the same cycle. The correlations can be extracted by a low-order model $ARMA(p, q)$; 2) after differencing operations, if cyclical effects exist in the time-series and short-term correlations exist among the cyclical effects, which imply that there are correlations among the same point in different cycles. The correlations can be extracted by $ARMA(P, Q)$ with cycle step as unit.

Based on the earlier conclusions by Wold and Yule of a practical approach to building ARIMA models developed by Box and Jenkins [20], we know that the methodology of ARIMA includes three steps of model identification, parameter estimation, and diagnostic checking [15]. In the model identification step, data transformation and preparation are often needed to make the time-series stationary. A stationary time series has the property that its statistical characteristics, such as mean, variance, and autocorrelation structure are constant over time. Results or plots of autocorrelations and partial autocorrelations of the sample data are the basic tools for stationary identification and order identification of ARIMA model. When trend and heteroscedasticity are observed in the time series, differencing, and log transformation are applied to eliminate the trend and heteroscedasticity to stabilize the variance. Once an initial model is determined, the key difficulty is the estimation of model parameters specifically. The parameters can be identified initially by the characteristics of autocorrelation and partial autocorrelation plots, and then estimated by an optimization procedure according to the Bayesian information criterion (BIC). To recognize whether the chosen model can describe the movement laws of time-series data well, goodness-of-fit test is needed mainly through white noise test. If the residual series is not a white noise sequence, it is necessary to repeat above steps to create a new model.

C. Markov Chain Model

As we know, MSARIMA model has the advantage of fitting time-series trend in linear space, while the disadvantage is that the fitting and forecasting performance of time series at large random fluctuating points is poor. Markov chain approach is a method with respect to stochastic processes with discrete time and state, which is capable of reducing errors obtained from extrapolated prediction. Hence, it is extensively applied to modify predicted values by preliminary models. The primary principle of Markov chain forecasting models is that: 1) Clear about what the states represent; 2) analyzing state changes in Markov processes; 3) utilizing the state-transition probability matrix to fully extract the relationships of state changes; and 4) forecasting the values of state chain based on the extracted relationships. While successful forecasting applications of Markov chain models have been reported in a number of business fields, including finance, accounting, marketing, and production management; to our knowledge, there has been no application of Markov chain models to hospital discharges prediction. The conceptual Markov chain is shown in Fig. 1, and its general form is as follows:

$$p_{k+1} = p_0[V^{(1)}]^{k+1} \tag{9}$$

where p_{k+1} is the state probability distribution of system at time point $(k+1)$, p_0 is the initial unconditional state probability

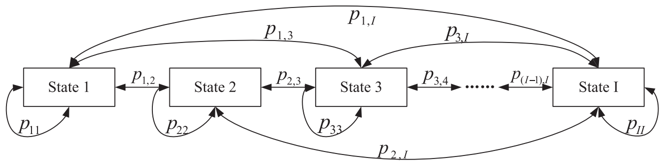

Fig. 1. Conceptual model of Markov chain.

distribution. $V^{(1)}$ is the one step state-transition probability matrix, its general form is as follows:

$$V^{(1)} = \begin{pmatrix} P_{11}^{(1)} & P_{12}^{(1)} & \cdots & P_{1j}^{(1)} \\ P_{21}^{(1)} & P_{22}^{(1)} & \cdots & P_{2j}^{(1)} \\ \vdots & \vdots & \vdots & \vdots \\ P_{i1}^{(1)} & P_{i2}^{(1)} & \cdots & P_{ij}^{(1)} \end{pmatrix} \tag{10}$$

where $P_{\mathrm{ij}}^{(1)} = \frac{n_{\mathrm{ij}}^{(1)}}{N_i}$ indicates the probability that the system is transferred from state E_i at time k to state E_j at time $(k+1)$ by one step. $n_{\mathrm{ij}}^{(1)}$ denotes the frequency that the system is transferred from state E_i to state E_j by one step. N_i denotes the total frequency of state E_i, and it must be true that $\sum_j P_{\mathrm{ij}}^{(1)} = 1$. Since the statistical rules of all kinds of state transition are reflected by the state-transition probability matrix, according to the m-step state-transition probability matrix, we can forecast the future moving trend of the system.

D. Combinatorial Model Based on MSARIMA and Weighted Markov Chain

1) Suppose a Time Series: The training dataset with N observations which has been chosen from the original dataset by using some methods (a fixed ratio of training data and testing data or a k-fold cross-validation method) is expressed as

$$X = \{X(1), X(2), \cdots, X(N)\}. \tag{11}$$

2) Fitting Function: A MSARIMA model can be fitted on X by using the fundamental theory of MSARIMA model introduced in the Section III-B, and the obtained fitting sequence and fitting function are expressed as

$$\hat{X} = \{\hat{X}(1), \hat{X}(2), \cdots, \hat{X}(N)\} \tag{12}$$

$$\hat{X} = F(B, \Phi(B), \Theta(B), U(B^S), V(B^S), \varepsilon(k)). \tag{13}$$

Moreover, the forecasts produced by the fitted MSARIMA model which have the same number with the testing dataset can also be obtained and be expressed as $(X(\hat{N+1}), X(\hat{N+2}), \ldots, X(\hat{N+T}))$.

3) Fitting Performance Measures: According to X and $\hat{X}$, the relative error sequence can be written as

$$Y(k) = \frac{X(k) - \hat{X}(k)}{X(k)} \times 100\% \tag{14}$$

which reflects the volatility of the training sequence around the curve of the fitting sequence, and also the dynamic time-varying degree of the sequence. Furthermore, the changing trend of $Y(k)$ has the features of nonstationary and pure random process, and no aftereffect exists.

4) State Confirming and Categorization: A Markov Chain represents a system of elements moving from one state to another over time, the key is to confirm what the states represent and then consider state categorization. In this research, states represent the relative under- and overprediction degree of fitted values obtained by MSARIMA model against observations. The common utilized methods for state categorization are mean-standard deviation classification method, clustering method, and optimal segmentation method. This research utilizes the mean-standard deviation classification method which is simple and useful in the case of a large sample data to divide the state set of the relative error sequence $Y(k)$ into I distinct states. The general form of state i is $E_i = [\Pi_{i1}, \Pi_{i2}]$. According to the central limit theorem, the state set can be divided into five segmentations, including $(-\infty, \overline{Y} - S]$, $(\overline{Y} - S, \overline{Y} - 0.5S]$, $(\overline{Y} - 0.5S, \overline{Y} + 0.5S]$, $(\overline{Y} + 0.5S, \overline{Y} + S]$, $(\overline{Y} + S, \infty)$. Each state represents the level of the fitting accuracy. Respectively, $\overline{Y}$ and S denote the sample mean and standard deviation of the relative error sequence $Y(k)$, their formulations can be written as

$$\overline{Y} = \frac{1}{N} \sum_{k=1}^{N} Y(k) \tag{15}$$

$$S = \sqrt{\frac{1}{N-1} \sum_{k=1}^{N} (Y(k) - \overline{Y})}. \tag{16}$$

Hence, substituting the obtained $Y(k)$ sequence with N values into (15) and (16), $\overline{Y}$ and S can be calculated, based on which the five states above can be obtained. Once the states are confirmed, each value in the $Y(k)$ sequence can be represented by one of the five states when the value enters in a state interval. So the $Y(k)$ sequence becomes a state sequence, based on which one-step transition frequency matrix and one-step transition probability matrix can be obtained and defined as $V^{(1)}$ referred to Section III-C.

5) Markovian Property Statistical Hypothesis Testing: Whether a random process is a Markov process is needed to be examined by statistical hypothesis tests. Usually discrete Markov chain sequences are tested by using χ^2 statistics. Assume $\hat{P}_j$ represents the ratio between the sum of $(n_{\mathrm{ij}})_{q\times q}$ of the jth column of the transition frequency matrix and the sum of $(n_{\mathrm{ij}})_{q\times q}$ of all columns of the transition frequency matrix. Recall that p_{ij} represents the transition probability from state E_i to E_j (obtained in the 4 step)

$$\hat{P}_j = \frac{\sum_{i=1}^{q} n_{\mathrm{ij}}}{\sum_{i=1}^{q} \sum_{j=1}^{q} n_{\mathrm{ij}}} \tag{17}$$

$$p_{\mathrm{ij}} = \frac{n_{\mathrm{ij}}}{\sum_{j=1}^{q} n_{\mathrm{ij}}}. \tag{18}$$

Since the statistics follow a χ^2 distribution with degree of freedom $(q-1)^2$, choosing a significant level α, we can obtain the value of $\chi^2_\alpha((q-1)^2)$. Denote

$$\chi^2 = 2\sum_{i=1}^{q}\sum_{j=1}^{q} n_i|ln(p_{\text{ij}}/\hat{P}_j)|. \tag{19}$$

According to the obtained $\hat{P}_j$ and p_{ij}, we can calculate the value of χ^2. If $\chi^2 > \chi^2_\alpha((q-1)^2)$, we can confirm that the random sequence is a Markov chain sequence. If not, it cannot be processed as a Markov chain sequence. Suppose in this research the $Y(k)$ sequence is a Markovian process, we can use the weighted Markov Chain method to forecast $(\tilde{Y}(N+1), \tilde{Y}(N+2), \ldots, \tilde{Y}(N+T))$. However, before the forecasting procedures, it is necessary to confirm what the weights are and how to calculate them.

6) Weights of Markov Chain: Since the weights are calculated by normalizing the autocorrelation coefficient of $Y(k)$ at each lag, the autocorrelation coefficients reflect the relationships and correlation degrees of each lag have to be calculated first. The formulations of autocorrelation coefficient and weights of Markov Chain are as follows:

$$r^{(m)} = \frac{\sum_{k=1}^{N-m}[Y(k)-\overline{Y}][Y(k+m)-\overline{Y}]}{\sum_{k=1}^{N}[Y(k)-\overline{Y}]^2} \tag{20}$$

$$w^{(m)} = \frac{|r^{(m)}|}{\sum_{m=1}^{M}|r^{(m)}|} \tag{21}$$

where $r^{(m)}$ denotes the autocorrelation coefficients, the maximum value of m is M, the value of M can be identified by judging whether the autocorrelation coefficient is more than twice Std. Error. The weight vector can be written as:

$$W^{(m)} = (w^{(1)}, w^{(2)}, \cdots, w^{(m)}, \cdots, w^{(M)}). \tag{22}$$

The mth weight value reflects the relative importance degree of the m-step state-transition probability matrix on forecasting the state-transition process.

7) State-Transition Probability Matrix: Since $V^{(1)}$ and M are obtained above, similarly, the m-step transition probability matrix can be calculated as follows:

$$V^{(m)} = \begin{pmatrix} P_{11}^{(m)} & P_{12}^{(m)} & \cdots & P_{1j}^{(m)} \\ P_{21}^{(m)} & P_{22}^{(m)} & \cdots & P_{2j}^{(m)} \\ \vdots & \vdots & \vdots & \vdots \\ P_{i1}^{(m)} & P_{i2}^{(m)} & \cdots & P_{\text{ij}}^{(m)} \end{pmatrix} \tag{23}$$

where $P_{\text{ij}}^{(m)} = \frac{n_{\text{ij}}^{(m)}}{N_i}$ indicates the probability that the system is transferred from state E_i at time k to state E_j at time $(k+m)$ by m steps. $n_{\text{ij}}^{(m)}$ denotes the frequency that the system is transferred from state E_i to state E_j by m steps. N_i denotes the total frequency of state E_i, and it must be true that $\sum_j P_{\text{ij}}^{(m)} = 1$.

8) Forecasts Calculation: The procedure for calculating the forecasts can be seen in Fig. 2 in Appendix.

<table>
<tr><th>Step</th><th>Content</th><th>Method</th></tr>
<tr><td>1</td><td>According to the value of M obtained in the "Weights of Markov Chain" step, determine the probability vectors of the predicted time point k</td><td>According to the states of Y(k) sequence at time points {(k-M), (k-M-1), (k-M-2), … , (k-1)}, we can obtain:
<table>
<tr><th>Time point</th><th>Y</th><th>State</th><th>Lag</th><th>Weight</th><th>Probability vector</th></tr>
<tr><td>k-1</td><td>Y(k-1)</td><td>E(k-1)</td><td>1</td><td>$W^{(1)}$</td><td>$V^{(1)}{}_{E(k-1)}$</td></tr>
<tr><td>k-2</td><td>Y(k-2)</td><td>E(k-2)</td><td>2</td><td>$W^{(2)}$</td><td>$V^{(2)}{}_{E(k-2)}$</td></tr>
<tr><td>--;p …</td><td>…</td><td>…</td><td>…</td><td>…</td><td>…</td></tr>
<tr><td>k-M</td><td>Y(k-M)</td><td>E(k-M)</td><td>M</td><td>$W^{(M)}$</td><td>$V^{(M)}{}_{E(k-M)}$</td></tr>
</table>
After confirming M time points, we can observe their values of Y, and then each value of Y corresponds to a state. According to the lag of the point, the corresponding weight at this lag can be confirmed. Similarly, the probability vector $V^{(1)}{}_{E(k-1)}$ is the row vector at state $E(k-1)$ that is picked out from the one-step transition probability matrix.</td></tr>
<tr><td>2</td><td>According to the M probability vectors obtained in step-1, determine the new probability matrix for predicting the state of time point k</td><td>$V' = \begin{pmatrix} V_{E(k-1)}^{(1)} \\ V_{E(k-2)}^{(2)} \\ \vdots \\ V_{E(k-M)}^{(M)} \end{pmatrix} = \begin{pmatrix} V_{E(k-1)}^{(1)1} & V_{E(k-1)}^{(1)2} & \cdots & V_{E(k-1)}^{(1)q} \\ V_{E(k-2)}^{(2)1} & V_{E(k-2)}^{(2)2} & \cdots & V_{E(k-2)}^{(2)q} \\ \vdots & \vdots & \vdots & \vdots \\ V_{E(k-M)}^{(M)1} & V_{E(k-M)}^{(M)2} & \cdots & V_{E(k-M)}^{(M)q} \end{pmatrix}$
$V_{E(k-1)}^{(1)1}$ is the first element of the row vector $V^{(1)}{}_{E(k-1)}$.</td></tr>
<tr><td>3</td><td>According to the M weights obtained in step-1 and M probability vectors obtained in step-2, calculate the maximum weighted probability vector, and determine the possibly corresponding state of the predicted time point k</td><td>$W^{(m)}V' = \begin{pmatrix} w^{(1)}V_{E(k-1)}^{(1)1} & w^{(1)}V_{E(k-1)}^{(1)2} & \cdots & w^{(1)}V_{E(k-1)}^{(1)q} \\ w^{(2)}V_{E(k-2)}^{(2)1} & w^{(2)}V_{E(k-2)}^{(2)2} & \cdots & w^{(2)}V_{E(k-2)}^{(2)M} \\ \vdots & \vdots & \vdots & \vdots \\ w^{(M)}V_{E(k-M)}^{(M)1} & w^{(M)}V_{E(k-M)}^{(M)2} & \cdots & w^{(M)}V_{E(k-M)}^{(M)q} \end{pmatrix}$
$V = \max\left\{\sum_{m=1}^{M} w^{(m)}V_{E^{(k-m)}}^{(m)1}, \sum_{m=1}^{M} w^{(m)}V_{E^{(k-m)}}^{(m)2}, \cdots, \sum_{m=1}^{M} w^{(m)}V_{E^{(k-m)}}^{(m)q},\right\}$
$= (V_1, V_2, \cdots, V_{j-1}, V_j, V_{j+1}, \cdots, V_q)$
The weighted probability vector V has q probabilities, if the maximum probability is V_j, the predicted state is state E_j.</td></tr>
<tr><td>4</td><td>Calculate the corresponding corrected relative error values, corrected forecasts, of the state at the predicted time point k</td><td>Assume the predicted state is E_j, according to the state interval $E_j \in [\Pi_{j1}, \Pi_{j2}]$, we can obtain the forecasts of $Y(k)$ as:
$\tilde{Y}(k) = \Pi_{j1} \times \frac{V_{j-1}}{V_{j-1}+V_{j+1}} + \Pi_{j2} \times \frac{V_{j+1}}{V_{j-1}+V_{j+1}}$
So the corrected forecasts of $X(k)$ from $X(N+1)$ to $X(N+T)$ as: $\hat{\hat{X}}(k) = X(k) - X(k) \times \tilde{Y}(k)$</td></tr>
</table>

Fig. 2. Calculation procedures of forecasts.

E. Forecasting Accuracy Procedure

Commonly used forecast-accuracy metrics are the *Normalized Mean Squared Error (NMSE)*, the *Mean Absolute Percentage Error (MAPE)*, and the *Direction of the Symmetry Value (DS)* at various forecast lags. NMSE and MAPE are considered as the forecasting accuracy metrics of selected models, and DS as the forecasting ability metric of selected models. They are defined in our research as

$$\text{NMSE} = \frac{1}{T\delta^2}\sum_{k=1}^{T}\left(X(k)-\hat{X}(k)\right)^2 \tag{24}$$

where $\delta^2 = \frac{1}{T-1}\sum_{k=1}^{T}\left(X(k)-\overline{X}(k)\right)^2$.

T represents the number of days in the testing set, and $X(k)$ and $\hat{X}(k)$ are the observations and forecasts of daily discharges for day t, respectively. $\overline{X}(k)$ is the mean of all forecasts. Moreover

$$\text{MAPE} = \frac{1}{T}\sum_{k=1}^{T}\left|\frac{X(k)-\hat{X}(k)}{X(k)}\right| \tag{25}$$

$$\text{DS} = \frac{1}{T}\sum_{k=1}^{T} d_i \tag{26}$$

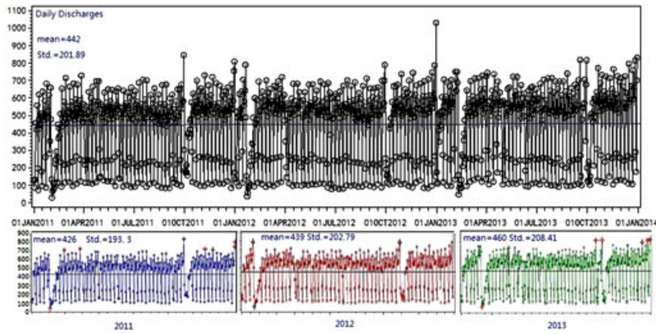

Fig. 3. Time-series data of daily discharges from year 2011 to 2013.

where d_i is designed as follows:

$$d_i = \begin{cases} 1, & (X(k) - X(k-1))(\hat{X}(k) - X(k-1)) \geq 0 \\ 0, & (X(k) - X(k-1))(\hat{X}(k) - X(k-1)) < 0. \end{cases} \tag{27}$$

F. Statistical Hypothesis Testing Procedure

To formally demonstrate the accuracy of the models in this study, the following statistical hypothesis was developed:

H0 *(Null hypothesis)*: There will be no forecasting accuracy differences among forecasts produced by SRARIMA model, MSARIMA model, and those by combinatorial model based on MSARIMA and weighted Markov chain.

To evaluate the hypothesis, we assessed the differences among the forecast accuracy statistics computed for each model. In particular, we used a paired sample *t*-test to evaluate the relative predictive accuracy among three models. Based on prior research, we expect that the nonlinear combinatorial model will yield smaller forecasting errors than the linear models, resulting in the rejection of the null hypothesis.

IV. Empirical Results and Discussion

A. Preliminary Data Analysis

This study used data from the Hospital Information System at the WCH, for the period from January 1, 2011 to December 31, 2013, containing the time of occurrence of each patient discharge with a total number of 484 222 records. We work with the number of discharges on each day, to facilitate the application of time-series models and Markov chain models. Throughout the paper, k denotes the time-index in days and the number of discharges on day k is denoted $X(k)$, for $k = 1, 2, \ldots, S$, where $S = 1096$. The models are fitted to the first 877 observations from January 1, 2011 to May 26, 2013 (N = 877, 80% of all data as training set). The remaining 219 observations from May 27, 2013 to December 31, 2013 are utilized for examining the forecasting ability of fitted models (T = 219, 20% as testing set). The average number of discharges is about 442/day, and the standard deviation is 201.89, which might mainly result from large fluctuations in seasonal effects and special day effects, such as weekends and holidays. Fig. 3 provides a first view of the data; it shows the daily discharge volume. The figure suggests a significant nonstationary, seasonal, and periodical fluctuation without obvious trend. It shows that smaller volumes in the first two weeks of February and the first week of October are significant, possibly resulting from the special holiday effects. And some large values are shown to be significant, which has been remarked in the figure of each year. Moreover, significant weekly periodicity features exist in the time-series data, while great differences exist among different points in a same cycle. The likely explanation is a combination of factors including the differences from outpatient volume, surgery volume, preoperative examination volume and admission volume, or possibly due to different hospital discharge practices for different days of the week, such as fewer discharges over the weekend because of fewer admissions.

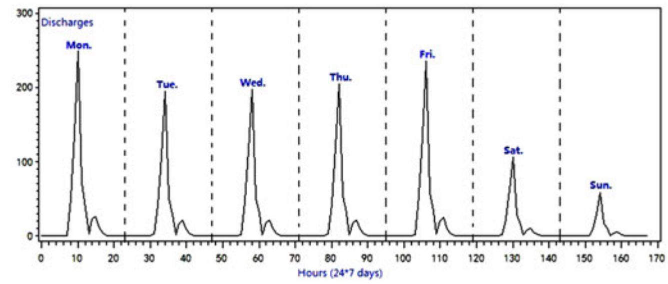

Fig. 4. Average hourly call-volume over the weekly cycle.

Fig. 4 shows average volume by hour over the weekly cycle. The plot reveals a clear hour-of-the-day seasonality; over a 24-h cycle, higher discharge volumes are usually observed from 10:00 to 11:00, and from 14:00 to 15:00; substantially lower volumes in other periods. We can also find that from 18:00:00 today to 06:59:59 the next day the number of discharges is approximate zero over a week. One also observes significant day-of-the-week effects. Closer inspection reveals, not surprisingly, with respect to daily volume, larger discharge activities during Monday and Friday relative to other days of the week. These observations would have to be taken into account when designing admission plans and nurse shift schedules.

B. Models for Daily Discharges

One would expect to see hour-of-day and day-of-week effects in discharges in many hospitals. Our preliminary analysis of the time-series data of WCH confirms the presence of hour-of-day and day-of-week effects. Since we focus on daily discharges data for daily forecasts, here we consider the day-of-week effects only. According to the features of the daily discharge time-series data (see Fig. 3), it is obvious that the daily discharge volume is not only related to the discharge volumes of past one or two days' of the same week, but also related to the discharge volume of the same day of past week; whereas the correlation among the prediction day and the days between the prediction day and the same day of past week is relatively weak. This suggests the following linear model as a first approximation. In this study, all ARIMA modeling is implemented via SAS system (version 9.2), while Markov Chain models are built using EXCEL system (version 2010). Only the one-step-ahead forecasting is considered. Supplementary information is

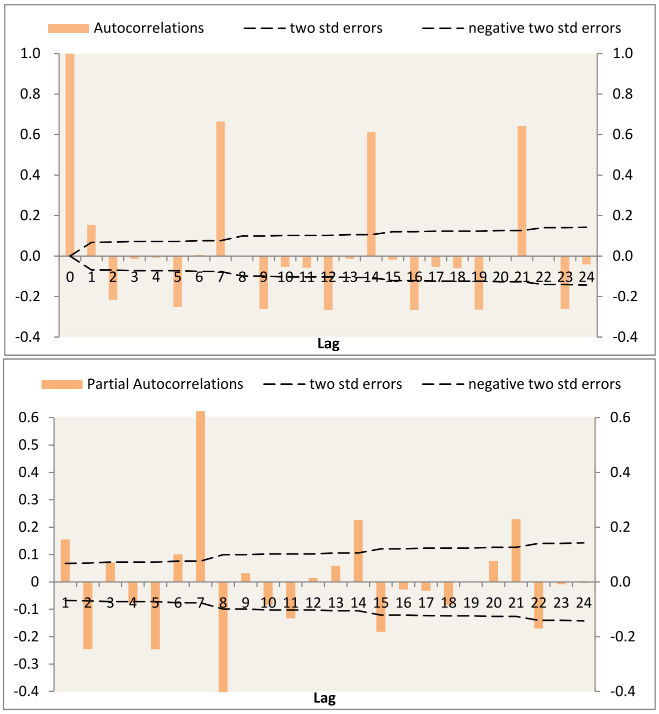

Fig. 5. Autocorrelation coefficients of the original time series.

TABLE I
RESULTS OF SEASONAL REGRESSION MODEL FOR FORECASTING DAILY DISCHARGES

Explanatory variable	Estimate	Std. Error	$t-value$	$p-value$
Time (days)	0.05606	0.0157	3.57	<0.0001
Monday	580.0314	12.5631	46.17	<0.0001
Tuesday	444.43137	12.5717	35.35	<0.0001
Wednesday	467.06332	12.5803	37.13	<0.0001
Thursday	483.53526	12.5889	38.41	<0.0001
Friday	551.8152	12.5975	43.80	<0.0001
Saturday	238.8955	12.5401	19.05	<0.0001
Sunday	105.4823	12.5496	8.41	<0.0001

available from the authors upon requests. To explore whether daily discharges can be forecast, we carry out a white noise test on the time-series data. The results show that at all lag levels, the $p-value$ is less than the significant level 0.001, which means the time series is not random and can be forecast well. Graphic testing method is common utilized for recognizing the stationary of the objective sequence $X(k)$. According to Fig. 5, the plots of autocorrelation and partial autocorrelation coefficients of the daily discharge time series show that there is a cyclical pattern with a mean cycle of about seven days. Only single-seasonality models are included to model the weekly effects but not the daily effects.

A SRARIMA model by combining seasonal regression and ARIMA has been fitted to the 877 training dataset to capture the day-of-week effects, the results of a seasonal regression model are shown in Table I. Moreover, the fitting performance measures are presented in Table II. As shown in Tables I and II, the variance explained by day-of-week effects for daily discharges is quite large: an R^2 and $AdjustedR^2$ of 0.9399 and

TABLE II
PERFORMANCE MEASURES OF SEASONAL REGRESSION MODEL

R-squared	0.9399	Adjusted R-squared	0.9394
Root MSE	117.6775	Dependent Mean	434.2452
Coefficient Variance	27.0993	Durbin-Watson	1.1111
F-statistic	1699.54	p-value of F-statistic	$< .0001$

TABLE III
PARAMETER ESTIMATION OF THE RESIDUALS OF A SEASONAL REGRESSION MODEL'S REGRESSION RESIDUALS

Parameter	Estimate Statistics	Std. Error	$t-value$	$p-value$
MA1,1	−0.19129	0.07282	−2.63	<0.0001
AR1,1	0.29079	0.07096	4.1	<0.0001

0.9394, respectively, which are very high, which indicates that it is of great value for capturing different discharge practices for different days of the week. The regression residual sequence of the fitted seasonal regression model is then modeled to an ARIMA model. The autocorrelation and partial autocorrelation plots showed weak autocorrelations for high lags, by using the BIC criteria, an ARIMA(1,0,1) model was fitted. The white noise test results show that p-value at all lags is more than 0.05, and the estimated parameters and t test results are shown in Table III, indicating a good fitting of the model. Based on the results of seasonal regression in Table I, the SRARIMA equation used for forecasting as

$$\begin{aligned} X(k) = {} & 0.05606k + 580.03143q_{1,k} + 444.43137q_{2,k} \\ & + 467.06332q_{3,k} + 483.53526q_{4,k} + 551.8152q_{5,k} \\ & + 238.89553q_{6,k} + 105.48234q_{7,k} \\ & + 0.29079(X(k-1) - \hat{X}(k-1)) \\ & + a(k) + 0.19129a(k-1). \end{aligned} \tag{28}$$

The error in fit and the error in forecast of the fitted SRARIMA model have been presented in Fig. 6. By analyzing the performance of the fitted SRARIMA model, we figure out that a high degree of fit can be achieved, and it is well capable of forecasting the periods outside the one it was trained on, while large fluctuations still occur in special days which cannot be forecast well. Other measures of model quality will be compared with other models below.

From the modeling results of the SRARIMA model above, we can confirm that significant and mathematically very large day-of-the-week effects exist. To extract the day-of-the-week effects, a seasonal differencing operation is conducted in a seven-step first. The corresponding autocorrelation coefficients of the differencing sequence are shown in Fig. 7, by comparing Figs. 5 and 7, we can figure out that there is a significant seasonal pattern in the original training sequence. However, the differencing cannot capture the seasonality completely, as it still presents short-term correlation and seasonal patterns at several lags. In this case, MSARIMA models are often applied to address these

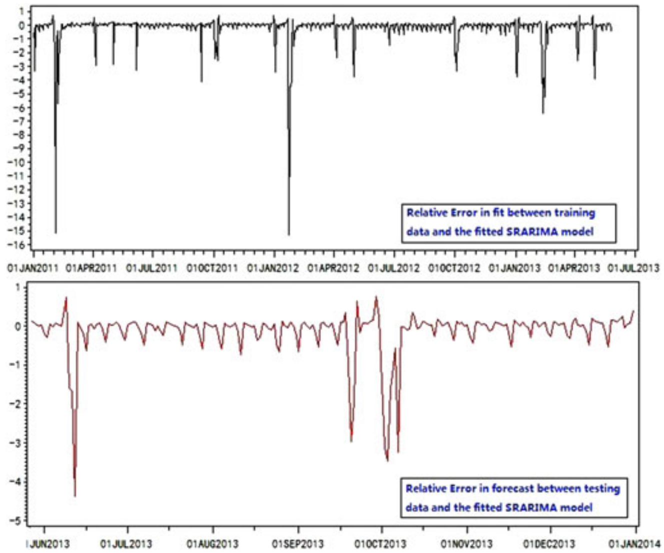

Fig. 6. Fitting and forecasting performances of the fitted SRARIMA model.

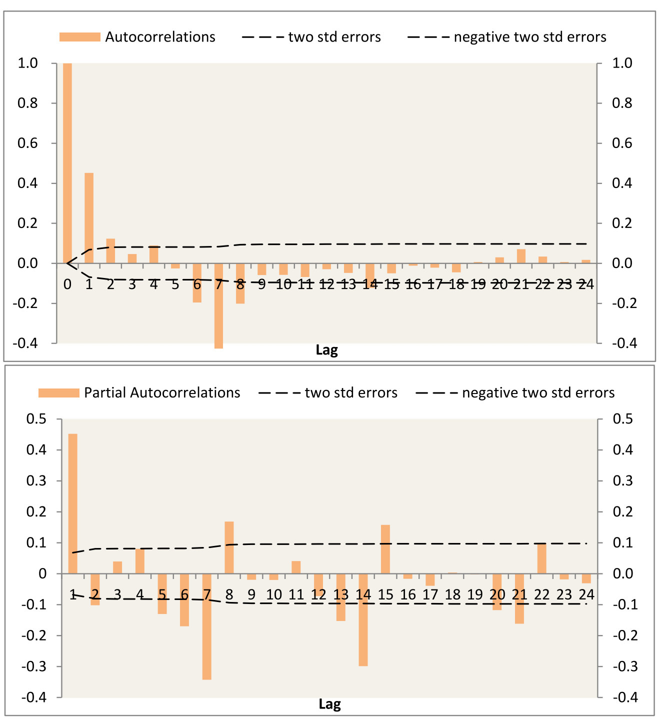

Fig. 7. Autocorrelation coefficients of the differencing time series.

seasonal time series. We examined several MSARIMA models on the training data. The best MSARIMA model is built, in terms of having lowest autocorrelation in the residuals and lowest variance of the residuals, as $ARIMA(1,0,4) \times (0,1,1)_7$, the formulation is written as

$$\hat{X}(k) - \hat{X}(k-7) = \frac{1 - 0.0.42817B}{1 + 0.09726B^4}(1 - 0.85007B^7)\varepsilon(k). \tag{29}$$

The white noise test of the residual sequence show that p-value at all lags is more than 0.05, and the significant test of estimated parameters are conducted to test the fitted model.

TABLE IV
PARAMETER ESTIMATION OF THE FITTED MSARIMA MODEL

Parameter	Estimate Statistics	Std. Error	$t-value$	$p-value$
MA1,1	−0.09726	0.03386	−2.87	<0.0001
MA2,1	0.85007	0.01791	47.47	<0.0001
AR1,1	0.42817	0.03073	13.93	<0.0001

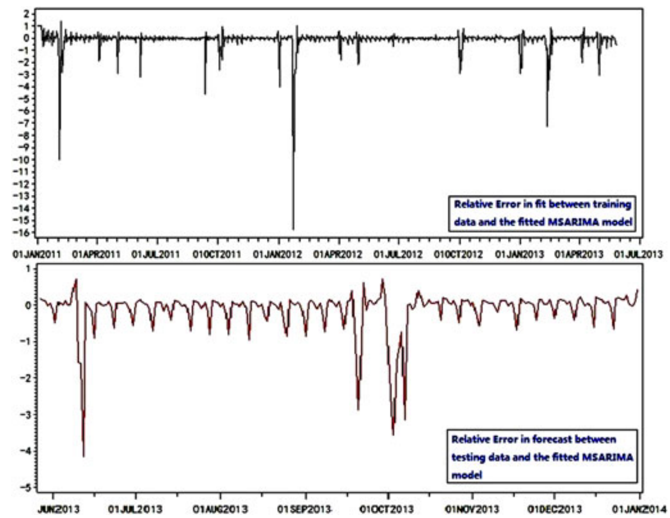

Fig. 8. Fitting and forecasting performances of the fitted MSARIMA model.

The results are shown in Table IV. According to the results of χ^2 test, the p values at each lag are more than the significant level 0.05, which means that the null hypothesis representing residuals are independent and not correlated cannot be rejected. The residual sequence is white noise sequence. Moreover, the p values of all the parameters are less than the significant level 0.001, which indicates that the fitting effects of the model is good. The overall fitting and forecasting performances are displayed in Fig. 8, from which we can intuitively figure out that a higher degree of fit may be achieved by the MSARIMA model than the SRARIMA model; while higher forecast errors than the SRARIMA model. Moreover, error in fit and error in forecast increase at the large random fluctuating points just like those of the SRARIMA model.

To revise the fitting errors and improve the forecasting accuracy of models, this paper proposes a combinatorial model based on MSARIMA and weighted Markov chain. To use the Markov chain approach, it is necessary to test the Markovian property of the random process. Considering the uncertainty of state division boundaries, we divide the state set into five segmentations by calculating the sample mean and standard deviation of relative error values $Y(k)$, respectively, are 0.1415 and 0.8643, based on which the state categorization result is displayed in Table V. Since the dimension of the state-transition matrix q is 5, the value of $\chi^2_\alpha(16)$ is 26.296 at a significant level 0.05. By using (17)–(19), the value of χ^2 is calculated as 435.078, which is larger than 26.296. Hence, the random process is a Markov process. According to (20) and (21), the values of $r^{(m)}$ and

TABLE V
STATE CATEGORIZATION RESULTS OF RELATIVE ERROR SEQUENCE

State Number	Interval	State Interpretation	Frequency	Ratio
$E1$	$-0.92905 \leq$ Y(k) $\leq$ -0.72277	Overestimated	7	0.0081
$E2$	$-0.72277 <$ Y(k) $\leq$ -0.29062	Slightly overestimated	33	0.0380
$E3$	$-0.29062 <$ Y(k) $\leq$ 0.57367	Moderate deviation	768	0.8848
$E4$	0.57367 $<$ Y(k) $\leq$ 1.00582	Slightly underestimated	16	0.0184
$E5$	1.00582 $<$ Y(k) $\leq$ 15.21728	Underestimated	44	0.0507

TABLE VI
AUTOCORRELATION COEFFICIENTS AND WEIGHTS OF EACH LAG (M=4)

m	1	2	3	4
$r^{(m)}$	0.37917	0.19671	0.07771	0.11027
$w^{(m)}$	0.49639	0.25752	0.10173	0.144436

TABLE VII
RESULTS OF PAIRED SAMPLE TEST

Pair	95% Confidence Interval	$t-value$	DF	$p-value$
SRARIMA–MSARIMA	$(-0.00468, 0.01578)$	1.069	218	$0.268 > 0.05$
SRARIMA–Combinatorial	$(0.03993, 0.09300)$	4.300	218	$0.000 < 0.05$
MSARIMA–Combinatorial	$(0.08721, 0.24507)$	4.149	218	$0.000 < 0.05$

$V^{(1)}$

$$V^{(1)} = \begin{pmatrix} 0 & 0.2857143 & 0.7142857 & 0 & 0 \\ 0.0303030 & 0.0909091 & 0.5757576 & 0.0909091 & 0.2121212 \\ 0.0039113 & 0.0234681 & 0.9530639 & 0.0065189 & 0.0130378 \\ 0.0625 & 0.125 & 0.3125 & 0.1875 & 0.3125 \\ 0.0454545 & 0.0909091 & 0.25 & 0.1136364 & 0.5 \end{pmatrix}$$

$V^{(2)}$

$$V^{(1)} = \begin{pmatrix} 0 & 0 & 0.5714286 & 0.1428571 & 0.2857143 \\ 0.0303030 & 0.0606061 & 0.6060606 & 0.0606061 & 0.2424242 \\ 0.002611 & 0.0287206 & 0.9386423 & 0.0130548 & 0.0169713 \\ 0.0625 & 0.125 & 0.5 & 0 & 0.3125 \\ 0.0681818 & 0.1363636 & 0.3863636 & 0.0454545 & 0.3636364 \end{pmatrix}$$

$V^{(3)}$

$$V^{(1)} = \begin{pmatrix} 0 & 0 & 0.8571429 & 0 & 0.1428571 \\ 0 & 0 & 0.6875 & 0.09375 & 0.21875 \\ 0.002611 & 0.035248 & 0.921671 & 0.0130548 & 0.0274151 \\ 0.0625 & 0 & 0.6875 & 0.0625 & 0.1875 \\ 0.0681818 & 0.1363636 & 0.5 & 0.0227273 & 0.2727273 \end{pmatrix}$$

$V^{(4)}$

$$V^{(1)} = \begin{pmatrix} 0 & 0 & 0.8571429 & 0 & 0.1428571 \\ 0.030303 & 0.0909091 & 0.7575758 & 0 & 0.1212121 \\ 0.0013072 & 0.0313725 & 0.9111111 & 0.0169935 & 0.0392157 \\ 0.0625 & 0.0625 & 0.75 & 0 & 0.125 \\ 0.0697674 & 0.0930233 & 0.6046512 & 0.0465116 & 0.1860465 \end{pmatrix}$$

Fig. 9. State-transition probability matrix of each lag.

$w^{(m)}$ are calculated and shown in Table VI, the maximum lag that the value of $r^{(m)}$ is larger than twice standard error is 4, so M is 4. According to the results of Table VII, the state of relative error at each time point can be determined, and, thus, state transition probability matrix from $one-step$ to $M-step$ can be calculated, and they are displayed in Fig. 9.

Based on the state transition probability matrix at each step, the weights of Markov chain, and the calculating procedures in Fig. 2, the forecasts of $Y(k)$ denoted as $(\tilde{Y}(N+1), \tilde{Y}(N+2), \ldots, \tilde{Y}(N+T))$ can be calculated, the corrected forecasts of the daily discharge sequence denoted as $(\hat{\hat{X}}(N+1), \hat{\hat{X}}(N+2), \ldots, \hat{\hat{X}}(N+T))$ at each corresponding testing day can be obtained. Fig. 10 is displayed to compare the overall forecasting performances of the three models. It is obvious that the highest degree of forecasting accuracy is achieved by the combinatorial model with the relative error values fluctuating around zero,

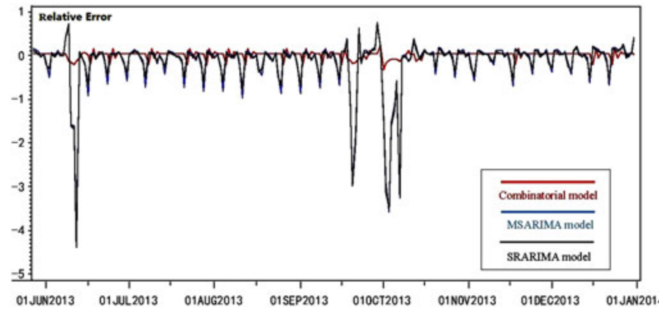

Fig. 10. Forecasting performance comparison of the three fitted models.

TABLE VIII
PERFORMANCE MEASURES

Performance measures	The combinatorial model	SRARIMA model	MSARIMA model
NMSE	0.0691	0.3194	0.3354
MAPE	0.0807	0.2657	0.2890
DS	88.53%	83.03%	83.49%
Variance of absolute relative error	0.0029	0.3257	0.3111
Absolute relative error $\leq 8\%$	73.52%	42.01%	34.70%
Absolute relative error $\leq 15\%$	84.93%	67.12%	64.84%
Absolute relative error $\leq 20\%$	99.09%	74.89%	76.26%

while the MSARIMA model performs the lowest degree of forecasting accuracy. Moreover, other approaches to comparing the model quality of the three models are applied, such as statistical hypotheses and performance measures.

The paired sample t-test is used to test the statistical hypotheses, the objective is to test whether significant differences exist among the relative error values in forecast of the three utilized models. The results are shown in Table V. From the results of the p-values, we can conclude that there are significant differences between the combinatorial model and the MSARIMA model, between the combinatorial model and the SRARIMA model, while no significance differences between the MSARIMA model and the SRARIMA model. The use of performance measures (NMSE, MAPE, and DS) confirm the conclusions mentioned above, and the comparison results are shown in Table VIII. From the results, we can convince that the combinatorial model significantly outperforms the linear SRARIMA and MSARIMA models in short-term daily discharge prediction, which can also be indicated by Fig. 10.

V. DISCUSSIONS

The proposed models can forecast daily discharges in the short term well, and there are considerable improvements in forecasting performance by combining the weighted Markov chain model with traditional linear alternative time-series models. We discuss the research from three perspectives: model performances, model implications, and limitations.

1) We can conclude that discharges are predictable, with a very large amount of variance explained by the trend and the day-of-the-week, which suggests that adjusting for

the day-of-the-week effect before fitting the time series can improve the quality of fitted models. To validate this hypothesis, experiments including a seasonal regression of the daily discharges on covariates (time and day of the week), and several MSARIMA with a seven-day-step seasonal differencing were conducted, we have found that the best performing linear models of the daily discharge time-series data, SRARIMA, and MSARIMA, could explain most of the variance for one day ahead, and some of the variance for up to one week ahead. From Figs. 6 and 8, although it seems that the SRARIMA model slightly outperforms the MSARIMA, the effect is not likely to be significant. Statistical hypothesis is needed to be conducted to figure out the fitted models' forecasting ability differences. Moreover, they provide no useful information about the inherent randomness and fluctuations in the time series, which is reflected by large errors at specific similarly testing days. Since the variability in discharges is likely due to both the natural variability in discharges and LOS, and due to the artificial variability in hospital practices, such as discharges increase with scheduled surgeries and high bed occupancy level. The linear SRARIMA model and MSARIMA model cannot capture these nonlinear patterns and features well, and it only captures short-range dependencies, while long-memory effects cannot be detected. A model that can detect the nonlinear patterns in the data is needed. Our results indicate that the combinatorial model based on the best performing MSARIMA and weighted Markov chain outperforms the single SRARIMA model and the single MSARIMA model for short-term prediction of daily discharges. In other words, these results highlight the usefulness of Markov chain models in providing better forecasts of time-series data with large random fluctuations over the traditional linear alternative (see Fig. 10, Tables VII and VIII). As we know, there is a variety of time-series forecasting approaches as mentioned in the literature review section, linear or nonlinear, the selection of models should mainly depend on data characteristics, so the art of model selection is to make a greatest degree of match between data characteristics and model.

2) While obtaining both good fitting and forecasting performance models, to better understand the practical implications of these fitted models on bed capacity reservation decision making, it is necessary to learn how forecasts of daily discharges-oriented daily bed availability can impact the real work of bed managers. As stated in Section I, bed manager devotes a random quota of expected bed capacity to the backlogged electives. In the case of a daily varying allocation quota, the forecasts of bed capacity would not change the decision on it, whereas affect the actual number of scheduled elective admissions, which may result in overscheduled patients or overreserved beds both at high penalty costs. Hence, the accuracy degree of daily discharges-oriented daily bed availability plays a significant role on bed capacity reservation decision optimizing. In this research, although better approaches to obtain the forecasts of daily discharges can provide more alternatives for decision makers to make a better decision, the optimizing effect cannot be verified and validated against that by using expected discharge data from implementing a planned discharge policy in WCH (a policy also used in NUH in Singapore [27]). Additional future study focuses on comparing the effects resulting from these predicted values between actual values should be considered.

3) In addition, only one day ahead discharges forecasting is considered without other prediction intervals discussed. Since the objective of this research is to obtain the daily accurate bed capacity for reservation purpose by forecasting daily discharges. Available bed capacity the next day is not only related to discharges the next day but also related to the left beds today. As unknown the left beds at the end of each day several days ahead, but only known the daily discharges several days ahead, and capacity reservation decision is made with one day a unit, it is not helpful for capacity reservation decision making several days ahead. So, only one day ahead modeling rather than several days ahead was considered, which is the main limitation of the generalization of the research. Possible lines of future research can add in the data of left beds each day to the setting of this research, to model several days ahead bed availability and support corresponding days ahead bed capacity reservation decision making. Second, it is the lack of more discussions on alternative models and reasons why they are not applicable. Third, the number of data *(training data)* used for model creation and the number of previous days *(testing data),* the fitted model used for forecasting have not been discussed sufficiently. We only applied a fixed ratio of the two types of data but not a k-fold cross-validation method. Considering large samples produce more stable estimates when estimating model coefficients, we use 4/5 of all data, and three models obtain relative robust results. In future work, k-fold cross-validation methods can be utilized for the comparison of model performance. Fourth, estimated model parameters reflect the behavior of the particular hospital-WCH, and at a particular duration in time. Such models cannot be applied directly with the same parameters to all the hospitals, instead require parameter reestimation. Fifth, daily bed availability forecast can be obtained by forecasting daily discharges resulting from the specific planned-admission behavior of WCH.

VI. Conclusion

We compared the forecast accuracy of a SRARIMA model, a MSARIMA model, and that of a combinatorial model by combining a weighted Markov Chain model with the fitted MSARIMA model for short-term daily discharges forecasting. Our results indicate that the combinatorial model outperforms the best fitting SRARIMA model and the MSARIMA model for each value examined one-day ahead. The results can be used to support the decisions of admission management and inpatient capacity planning. While this research suggests using MSARIMA–Markov chain-type models for better forecasts of time-series data with large random fluctuations, some concerns exist with regards to the following aspects:

1) Based on the single precision correction fitting results of the linear MSARIMA model, we model and modify the fitted residual sequence ($Y(K)$) of the MSARIMA model directly, by using a weighted Markov Chain model first in order to realize the double precision correction, and then backward compute fitting values of the combinatorial model. The advantage of this model is that the second forecasting values of the model can be corrected directly by modifying $Y(K)$, no need for introducing to the original model to compute the forecasts. This method not only improves the prediction accuracy but also reduce the computational workload.
2) According to the features of different scales (the same day of different weeks, and different days of the same week) of time-series data, the proposed model detects the inherent features (cycle, seasonality, short-range dependencies) of the objective time series. Compared with the simple integration of linear time series, the combinatorial model can capture the nonlinear patterns efficiently by utilizing a weighted Markov Chain model, and thus it has a better explanation ability to the sequence with large random fluctuations.

Overall, the results indicate that the combinatorial model outperforms the single SRARIMA model and the MSARIMA model for daily bed availability prediction. The results can be used to support the decisions of patient admission scheduling and bed capacity planning. While the research suggests using a weighted Markov Chain model for better forecasts, a concern exists with regards to the fact that: since the combinatorial model increases the number of parameters, it must be a corresponding increase in sample size in order to obtain good prediction precision. In addition, to the best of our knowledge, the model proposed in this research is utilized in healthcare contexts for the first time, future optimization on calculation processes is needed. The limitations stated in Section V should be aware of when applying this model. Of course, this paper provides evidence on the applicability of a combinatorial model based on MSARIMA models and Markov Chain models as a better decision support tool over the linear alternative to forecast short-term hospital bed availability.

References

[1] G. Abraham, G. B. Byrnes, and C. A. Bain, "Short-term forecasting of emergency inpatient flow," *IEEE Trans. Inform. Technol. Biomed.*, vol. 13, no. 3, pp. 380–388, May 2009.

[2] S. A. Andrew, M. P. Joy, and J. Pearson, "Forecasting demand of emergency care," *Health Care Manage. Sci.*, vol. 5, no. 4, pp. 297–305, 2012.

[3] M. Xu, T. C. Wong, and K. S. Chin, "Modeling daily patient arrivals at emergency department and quantifying the relative importance of contributing variables using artificial neural network," *Decision Support Syst.*, vol. 54, no. 3, pp. 1488–1498, 2013.

[4] D. Tandberg and C. Qualls, "Time series forecasts of emergency department patient volume, length of stay, and acuity," *Ann. Emergency Med.*, vol. 23, no. 2, pp. 299–306, 1994.

[5] M. Mackay, "Choice of models for the analysis and forecasting of hospital beds," *Health Care Manage. Sci.*, vol. 8, no. 3, pp. 221–230, 2005.

[6] S. J. Littig and M. W. Isken, "Short term hospital occupancy prediction," *Health Care Manage. Sci.*, vol. 10, no. 1, pp. 47–66, 2007.

[7] M. V. Pauly and P. Wilson, "Hospital output forecasts and the cost of empty hospital beds," *Health Services Res.*, vol. 21, no. 3, pp. 403–428, 1986.

[8] P. B. Cerrito and J. C. Cerrito, "PMC1 forecasting with time series models to examine the availability of inpatient beds," *Value Health*, vol. 10, no. 3, pp. A184, 2007.

[9] S. Makridakis and M. Hibon, "The M3-Competition: results, conclusions and implications," *Int. J. Forecasting*, vol. 16, no. 4, pp. 451–476, 2000.

[10] F. Petropoulos, S. Makridakis, V. Assimakopoulos, and K. Nikolopoulos, "Horses for Courses in demand forecasting," *Eur. J. Oper. Res.*, vol. 237, no. 1, pp. 152–163, 2014.

[11] D. Clerkin, P. J. Fos, and F. E. Petry, "A decision support system for hospital bed assignment," *Hospital Health Services Admin.*, vol. 40, no. 3, pp. 386–400, 1994.

[12] D. L. Ledersnaider and B. S. Channon, "SDM95-reducing aggregate care team costs through optimal patient placement," *J. Nursing Admin.*, vol. 28, no. 10, pp. 48–54, 1998.

[13] M. Mackay, "Practical experience with bed occupancy management and planning systems: an Australian view," *Health Care Manage. Sci.*, vol. 4, no. 1, pp. 47–56, 2001.

[14] Q. Cao, B. T. Ewing, and M. A. Thompson, "Forecasting medical cost inflation rates: A model comparison approach," *Decision Support Syst.*, vol. 53, no. 1, pp. 154–160, 2012.

[15] G. P. Zhang, "Time series forecasting using a hybrid ARIMA and neural network model," *Neurocomputing*, vol. 50, pp. 159–175, 2003.

[16] R. Fildes and S. Makridakis, "The impact of empirical accuracy studies on time series analysis and forecasting," *Int. Stat. Rev.*, pp. 289–308, 1995.

[17] A. Krogh and J. Vedelsby, "Neural network ensembles, cross validation, and active learning," *Adv. Neural Inform. Process. Syst.*, pp. 231–238, 1995.

[18] N. Channouf, P. L'Ecuyer, A. Ingolfsson, and A. N. Avramidis, "The application of forecasting techniques to modeling emergency medical system calls in Calgary, Alberta," *Health Care Manage. Sci.*, vol. 10, no. 1, pp. 25–45, 2007.

[19] S. A. Pourmousavi Kani and M. M. Ardehali, "Very short-term wind speed prediction: a new artificial neural network–Markov chain model," *Energy Convers. Manage.*, vol. 52, no. 1, pp. 738–745, 2011.

[20] W. T. Lin, "Modeling and forecasting hospital patient movements: Univariate and multiple time series approaches," *Int. J. Forecasting*, vol. 5, no. 2, pp. 195–208, 1989.

[21] Purwanto, C. Eswaran, and R. Logeswaran, "A dual hybrid forecasting model for support of decision making in healthcare management," *Adv. Eng. Softw.*, vol. 53, pp. 23–32, 2012.

[22] A. D. Sahin and Z. Sen, "First-order Markov chain approach to wind speed modelling," *J. Wind Eng. Ind. Aerodyn.*, vol. 89, no. 3, pp. 263–269, 2001.

[23] H. Kantz, D. Holstein, M. Ragwitz, and N. K. Vitanov, "Markov chain model for turbulent wind speed data," *Physica A, Stat. Mech. Appl.*, vol. 342, no. 1, pp. 315–321, 2004.

[24] A. Shamshad, M. A. Bawadi, W. M. A. Wan Hussin, T. A. Majid, and S. A. M. Sanusi, "Markov chain model for turbulent wind speed data," *Physica A, Stat. Mech. Appl.*, vol. 342, no. 1, pp. 315–321, 2004.

[25] G. W. Harrison, A. Shafer, and M. Mackay, "Modelling variability in hospital bed occupancy," *Health Care Manage. Sci.*, vol. 8, no. 4, pp. 325–334, 2005.

[26] H. F. Xie, T. J. Chaussalet, and P. H. Millard, "A model-based approach to the analysis of patterns of length of stay in institutional long-term care," *IEEE Trans. Inform. Technol. Biomed.*, vol. 10, no. 3, pp. 512–518, Jul. 2006.

[27] P. Y. Shi, M. C. Chou, J. G. Dai, Ding Ding, and Joe Sim, "Models and insights for hospital inpatient operations: Time-dependent ED boarding time," *Manage. Sci.*, 2015.

Authors' photographs and biographies not available at the time of publication.