Predicting medicine demand in hospitals through stochastic approaches

Daniel Fernando Velez Cardenas

Universidad Nacional de Colombia
Facultad de Ingeniería, Departamento de ingeniería de sistemas
Bogotá, Colombia
2022

Predicting medicine demand in hospitals through stochastic approaches

Daniel Fernando Velez Cardenas

Trabajo de investigación presentado como requisito parcial para optar al título de:

Magister en Ingeniería Industrial

Director (a):

Msc. Jair Eduardo Rocha González

Codirector (a):

Ph.D. Zakaria Yahouni

Línea de Investigación:

Gestión de operaciones

Universidad Nacional de Colombia

Facultad de Ingeniería, Departamento de ingeniería de sistemas

Bogotá, Colombia

2022

A mis padres

Declaración de obra original

Yo declaro lo siguiente:

He leído el Acuerdo 035 de 2003 del Consejo Académico de la Universidad Nacional. «Reglamento sobre propiedad intelectual» y la Normatividad Nacional relacionada al respeto de los derechos de autor. Esta disertación representa mi trabajo original, excepto donde he reconocido las ideas, las palabras, o materiales de otros autores.

Cuando se han presentado ideas o palabras de otros autores en esta disertación, he realizado su respectivo reconocimiento aplicando correctamente los esquemas de citas y referencias bibliográficas en el estilo requerido.

He obtenido el permiso del autor o editor para incluir cualquier material con derechos de autor (por ejemplo, tablas, figuras, instrumentos de encuesta o grandes porciones de texto).

Por último, he sometido esta disertación a la herramienta de integridad académica, definida por la universidad.

Daniel Fernando Vélez Cárdenas

Fecha 28/07/2022

Fecha

Esta investigación será publicada para alcanzar el título de Master en ingeniería en la Universidad Institut polytechnique de Grenoble dentro del convenio de doble titulación que mantiene con la Universidad Nacional de Colombia, Facultad de Ingeniería, Sede Bogotá y ahora se publica en versión idéntica para satisfacer las condiciones de grado en Colombia y para su difusión en el repositorio institucional.

Agradecimientos

Agradezco a mis padres por el continuo apoyo y motivación durante el desarrollo de este trabajo.

Finalmente agradezco a mi director de proyecto, el Ing. Jair Eduardo Rocha por su apoyo y acompañamiento constante en la consecución y ejecución del presente trabajo.

Resumen

Predicción de la demanda de medicamentos en hospitales a través de enfoques estocásticos

Hoy en día, el sector sanitario está cambiando rápidamente. Los hospitales se enfrentan a presupuestos cada vez más limitados y costos elevados. Las actividades logísticas de los hospitales en Francia (gestión de existencias, entrega, etc.) representan uno de los componentes de mayor costo. Los costos logísticos pueden reducirse mediante un sistema optimizado de gestión de inventarios. La optimización del inventario depende en gran medida de la precisión de la predicción de la demanda de medicamentos. El primer objetivo consiste en realizar un estado del arte de los métodos existentes para predecir la demanda de medicamentos en los centros sanitarios. Muchos factores influyen en esta demanda, como su estacionalidad, el tamaño y la ubicación del hospital. En consecuencia, un método estocástico puede ser relevante para captar las fluctuaciones de la demanda. Un segundo objetivo es utilizar los datos históricos de un hospital de Francia para predecir el consumo de medicamentos mediante una cadena de Markov. Se propone un análisis de los resultados experimentales para evaluar la eficacia del método. El resultado podría contribuir a la gestión y el dimensionamiento de los inventarios hospitalarios.

Palabras clave: Gestión de inventarios, predicción de la demanda, farmacia hospitalaria, modelos estocásticos, cadena de Markov.

Abstract

Predicting medicine demand in hospitals through stochastic approaches

Nowadays, the healthcare sector is rapidly changing. The hospitals are facing limited budgets and high costs. The logistics activities of the hospitals in France (stock management, delivery, etc.) represent one of the highest cost components. The logistic costs can be reduced through an optimized inventory management system. The inventory optimization is strongly dependent on the accuracy of the demand prediction of medicines. The first objective consists of making a state of the art of existing methods for predicting medicines demand in healthcare facilities. Many factors influence this demand, such as seasonality, hospital size and location, etc. As a consequence, a stochastic method can be relevant to capture the demand fluctuations. A second objective is to use the historical data of one hospital in France to predict the consumption of medicines using a Markov chain. An analysis of the experimental results is proposed to assess the effectiveness of the method. The result could contribute to the management and dimensioning of hospital inventories.

Keywords: Inventory management, demand forecasting, hospital pharmacy, stochastic models, Markov chain.

Este Trabajo Final de maestría fue calificado en diciembre de 2022 por los siguientes evaluadores:

Carlos Osorio Ramírez, PhD
Profesor Facultad de Ingeniería
Universidad Nacional de Colombia

Contenido

Lista de figuras

Pág.

Lista de tablas

Pág.

Introduction

The healthcare sector is rapidly changing, especially since the 90s. Organizations in the sector are promoting projects in areas such as care-related logistics, information systems and quality of care to cope with characteristics such as increased competition, patient influence and the need for a more efficient and effective service (de Vries, 2011).

The logistics related activities can account for about 46% of a hospital's operating budget in the United States for instance (Landry & Philippe, 2004). The main logistics activities include planning, designing, implementing, and managing material (Pan & Pokharel, 2007). These activities seek to support functions such as inventory management, procurement, and distribution (Pokharel, 2005). One of the highest hospital costs is related to these logistical activities, as an example in the OECD[1] countries it represents 30% and is the second largest cost according to (Volland et al., 2017). More precisely, in the health sector, inventory costs are estimated between 10% and 18% (de Vries, 2011).

The costs related to the logistics in hospitals can be reduced, in some cases even in a half, by implementing efficient logistics management (Volland et al., 2017). That is the reason why hospitals are forced to generate better returns in their internal service. In this case, inventory management becomes more important (de Vries, 2011). In order to decrease the costs associated to inventories, French hospitals are interested in the reduction of the inventory supplies (Aptel & Pourjalali, 2001). Since the optimization of inventory depends highly on the quality of the demand prediction, in this study, we are interested in the prediction models used in inventory management.

[1] The OECD's 38 members are: Austria, Australia, Belgium, Canada, Chile, Colombia, Costa Rica, Czech Republic, Denmark, Estonia, Finland, France, Germany, Greece, Hungary, Iceland, Ireland, Israel, Italy, Japan, Korea, Latvia, Lithuania, Luxembourg, Mexico, the Netherlands, New Zealand, Norway, Poland, Portugal, Slovak

More specifically, **we focus on an activity in hospital logistics, which is inventory management and estimation of medicines demand**. This study begins with an overview of existing methods for forecasting medicine demand in hospital pharmacies. Demand can be stationary if its behavior is constant over time, or it can be non-stationary when it shows a trend. Many factors influence this demand, such as seasonality, the size and location of hospitals, etc. Therefore, a stochastic method may be relevant to capture fluctuations in demand from period to period. The objective is to use historical data from hospitals in France to develop a stochastic method to predict the demand and contrast the results. The result could contribute to the management of hospital inventories.

The research questions that will guide the present study are:

1. What kind of data is needed to predict the demand of medicines in hospitals?
2. What are the stochastic methods used for stock management and particularly in the hospital context? Our problem is similar to the “demand forecasting” What are the machine learning and stochastic methods used for this problem and are they usable for our context.
3. What are the evaluation metrics used for the comparison and validation of demand prediction methods/models in inventory management?

For this purpose, the document is structured as followed. Section 2 presents a literature review about inventory management in hospitals, factors that influence the demand and the exploration of existing literature concerning stochastic models, machine learning and medicine demand forecasting techniques. To this end, keywords were identified from a first search and classified by topics, then a more specific search was performed through a combination of these words. Following that, section 3 corresponds to the methodology followed for the development of this project. Section 4 presents a case study to exemplify the application of the proposed methodology. The results obtained from the application of the proposed method to selected medicines are presented in section 5. Section 6 presents the conclusion and the perspectives.

1. State of the art

There is considerable amount of research on inventory management and demand prediction in the literature. In this section, some research streams that are particularly relevant to our work are reviewed: Inventory management in hospitals and influencing factors and methods used to predict the medicine demand. In the first part, the information will be related to the context of inventory management and the different factors that can affect it. Afterwards, in the second part, a review of different approaches used to deal with data uncertainty is presented.

1.1 Inventory management in hospitals and influencing factors

The inventory management system is an important factor in the performance of the operations in the hospital's system. Inventory management in this context means controlling and managing a variety and quantity of items stored in a hospital pharmacy (Gebicki et al., 2014). Healthcare systems are complex systems linked with several influencing factors such as the demand which is by itself related to other factors like: the size of the hospital, the type of population, etc. These factors influence the performance of inventory systems and must be considered when modeling and analyzing these systems (Saha & Ray, 2019b).

An important factor in the inventory management is the demand (Saha & Ray, 2019b). In hospitals the demand for health supplies is strongly linked to the doctor's recommendations based on the patient's condition (Abdulsalam et al., 2018). The consumption of healthcare supplies is non-stationary, the demand may depend on aspects such as the number of patients, the patient's conditions, the stage of treatment, among others. Some studies assume demand as independent and constant. However, to perform an analysis in a more real scenario, it is necessary to consider a stochastic demand determined by a probability function (Attanayake et al., 2014).

The classification of inventory problems in healthcare sector depends on the nature of the associated factors: the problems can be either deterministic or uncertain (Saha & Ray, 2019a). In the case of deterministic problem, the influencing factors are known over the period, this kind of inventory problems are not frequent in hospitals but may be applied to general medical items like syringes, gloves, intravenous fluids, and vaccines (Saha & Ray, 2019a). The other case is uncertainty; Healthcare system faces conditions of uncertainty such as the patient's clinical conditions and response to treatment, availability of providers, demand for medicines and lead time (Addis et al., 2015). The demand character distinguishes stochastic inventory models from deterministic models. Demand is fixed in deterministic models while in stochastic models it is of probabilistic nature, which means it is a random variable with a probability distribution (Polanecký & Lukoszová, 2016). Several factors that fluctuate at random way, such as the patient's condition, the patient's uncertain reaction to therapy, the length of stay at hospital and the transition from one type of hospital care unit to another at various phases of treatment, can all have a substantial impact on demand (Saha & Ray, 2019a). These factors are random and change with time, so the demand of required medicines is highly uncertain (Saha & Ray, 2018).

1.2 Methods for data uncertainty

The methods used to deal with the data uncertainty depends on the degree of knowledge about the data. In this study, we consider the case where a probability distribution is known. According to (Saha & Ray, 2019b) most of the healthcare inventory issues with complete knowledge of the probability distribution of random variables are modeled using stochastic techniques and are commonly formulated using Mixed-Integer Programming (MIP) to optimize inventory including policies such as reorder points.(Roni et al., 2016) formulates a model using MIP for a single-item inventory system in a hospital facing both regular and emergency demands to obtain an optimal inventory policy for medicines.

MIP is also used to construct an inventory system for drugs in intensive care units and surgical supplies in the operating room. Furthermore, there are a variety of stochastic inventory models formulated as stochastic programming in the literature (Saha & Ray, 2019b). In Stochastic Programming (SP), the parameter values are characterized by probability distributions, the assumption that required data are known and constant is relaxed. In hospital pharmacy, the inventory model for medicines to minimize the expected total inventory costs with the service level and space constraints are formulated as SP to minimize the expected total number of orders (Saha & Ray, 2019b). The model has many variables, non-linear and stochastic constraints. The service level is a stochastic constraint, that is, the probability of no shortage of medicines.

1.3 Forecasting methods

1.3.1 Traditional methods

The demand for healthcare items is a significant element affecting healthcare inventory systems (Saha & Ray, 2019a). Traditional methods for predicting demand in a hospital exist, and several of them can also anticipate the stationary demand for healthcare elements. (Lopez Ramirez et al., 2014) applies different traditional methods such as simple moving mean, exponential smoothing and ARIMA models on no changing demand. They apply causal models like simple and multiple linear regression when exist a cause-effect relation between a dependent variable and independent variables. A forecasting exercise is performed by (Varghese et al., 2012) in two hospitals using in most cases ARIMA models for several medicines with no significant trends or non-stationary patterns in the demand. However, healthcare item consumption is usualy non-stationary and uncertain (Vila-Parrish et al., 2008).

1.3.2 Machine learning methods

In recent years, machine learning approaches have gained popularity, especially when it comes to integrate parameter estimation like the demand and inventory optimization (Goltsos et al., 2021). Research in pharmacology identify effective and accurate methods for constructing a predictive model: linear regression, random forest method, construction a time series prediction using a neural network, and support vector regression (Goltsos et al., 2021). The demand for most medicines depends on the periods of increased risk of spreading diseases, in this case, seasonal trends need to be considered when developing the model. A study using real drug consumption data in Rwanda is conducted by (Villegas et al., 2018). The authors applied machine learning methods to predict the demand for 10 selected medicines. According to their findings, machine learning methods can be used to predict the demand for medicines and random forests has the best performance. Machine learning can also be used to compare and to select the best prediction model among a set of predictive models for non-stationary demand (Villegas et al., 2018).

1.3.3 Stochastic methods

A commonly used stochastic approach to problems of uncertain inventories with random variables whose probability distribution is known is the Markov decision process which use functional stochastic dynamic programming equations in most of the cases (Saha & Ray, 2019b). Markov chain is a particular type of stochastic process with discrete or continuous time and states (Anderson & Goodman, 1957) this is commonly utilized in different fields including inventory management to forecast the demand of products.

There are some studies on inventory management in hospitals that seek to predict medicine demand. According to (Saha & Ray, 2019a) existing medicine inventory models in the healthcare area assume demand is unaffected by external factors such as changing patient conditions and the medicine's uncertain reaction. Saha and Ray (Saha & Ray, 2019a) use a Markovian demand approach to incorporate such external factors, modeling a patient condition-based medication demand process which allows to capture the uncertain patient conditions (in terms of treatment stages and type of care units). (Hermosilla et al., 2020) propose models to predict the demand for cardiovascular drugs using Markov chain. They consider models with 3 and 4 states. Each state represents patient conditions. The authors found that the fluctuations in consumption level can be modeled using Markov chains. These methods can be used to prioritize patients with greater consumption levels in critical inventory situations. A Markov chain model is defined by (Kocer, n.d.) to model and estimate the intermittent demand in products.

The reviewed studies on methods that predict drug demand are summarized in Table 1-1. In this table, the objectives, the characteristics of the data, the performance measures used, and the limitations and advantages of each method/paper are given.

Table 1-1: Methods used in inventory management to predict medicines demand.

Method classification	Reference	Method(s)	Data	Performance measures	Limitations	Advantages
Traditional methods	Ramírez et al. (2014)	Simple moving average	Monthly	MAE (Mean Absolute Error) Typical deviation	-Better results in non-variable demand -Cannot capture demand variation well	-Easy to apply -Easy to understand
		Exponential smoothing				
		Simple and Multiple linear regression				
	Hyndman and Athanasopoulos (2018)	Average	Monthly	MAD (Mean Absolute Deviation) MSE (Mean Squared Error) MASE (Mean Absolute Scaled Error) MAPE (Mean Absolute Percentage Error)		
		Naïve				
		Naïve seasonal				
	Varghese et al. (2012)	Simple moving average	Weekly	MAE		
		Cumulative average				
		Exponential smoothing				
		Naïve				
Machine learning methods	A G Kravets et al. (2018)	Linear regression	Monthly	MSE MAPE MSE	-Require many observations -Require good data -Not easy to apply -Not easy to understand	-Powerful in Big data
		Random forest				
	Mbonyinshuti et al. (2021)	Linear Regression	Monthly	RMSE R-square		
		Artificial neural network				
		Random forest				
Stochastic processes	Hermosilla et al. (2020)	Markov chain - several numbers of states	Monthly	MAP	-Better with many observations -Need good quality in data	-Capture variation in demand well
	Saha and Ray (2019)	Markov chain based in stage of treatment		Parameters of inventory		
	Kocer (2013)	Modified Markov chain model		MASE		

Overall, the health sector is in constant change as the logistics related activities. Especially the inventory management, represents an important branch to optimize the efficiency of hospitals logistics. Consequently, hospitals, particularly in France, are the focus of our research subject. The inventory management problems have commonly a stochastic nature since the demand depends on several factors such as the patient's condition, the patient's uncertain reaction to treatment, the period of stay, and the transition between units according to the treatment phases. This is the motivation of this study.

In the present study, we will focus on Markov chain model to propose a model to capture well the medicine demand fluctuations.

2. Methodology

The Markov chain technique is used to capture consumption fluctuations from one period to another. The objective of this section is to propose a Markov chain model that can predicts the medicine demand in futures periods. In the following section, we start by presenting some basics about Markov chain, then the proposed methodology is described.

2.1 Markov chain basics

A stochastic process is a succession of random variables $\{X_t = X_1, X_2, \dots, X_t\}$ with t ∈ T (time states). A discrete-time Markov chain is stochastic process $\{S_n = s_1, s_2, \dots, s_n\}$ in which the future state S_{t+1} only depends on the current state S_t (Du et al., 2020), that is:

$$P(S_{t+1} \mid S_{1,} S_{2,} S_{3,} \dots, S_{t,}) = P(S_{t+1} = i \mid S_n = j) = p_{ij} \quad (1)$$

where St represents the state of the consumption data series at time t and p_{ij} the probability of transition between the state *i* at time *t* and the state j at time *t+1*. There are n states and the probability transition matrix P_{ij} contains the transition probabilities. It is defined as $\boldsymbol{P} = [p_{ij}]_{nxn}$, and satisfies the following properties:

$$0 \leq p_{ij} \leq 1 \; and \quad (2)$$

$$\sum_{j=1}^{t} p_{ij} = 1 \quad (3)$$

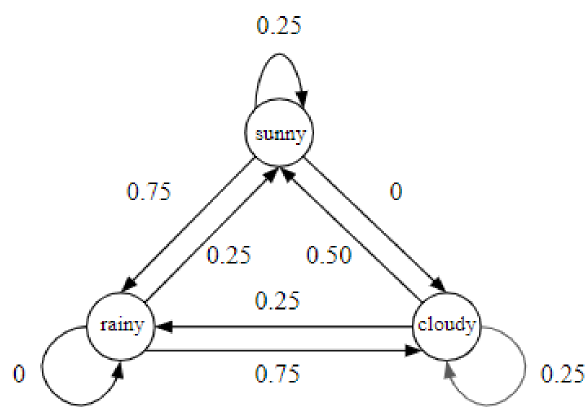
The probability matrix can be used to calculate the next state:

$$S_{t+1} = P * S_t \tag{4}$$

Where S_{t+1} is the state probability distribution vector in the state S_j at time *t+1*

To illustrate the concept, we use the states of the weather as observed phenomenon (stochastic process). Suppose that Sn takes values in 3 states: {sunny, rainy, cloudy}. It is possible to have many observations in a period of time and recurrent phenomena can be identified. For example, the possibility of a sunny day being followed by a rainy day. From this, the transition probabilities between the different weather states can be calculated and can be represented using the following figure. For instance, in Figure 2-1 the probability of having a sunny day followed by a rainy day is 75%.

Figure 2-1: Example of Markov chain model for weather

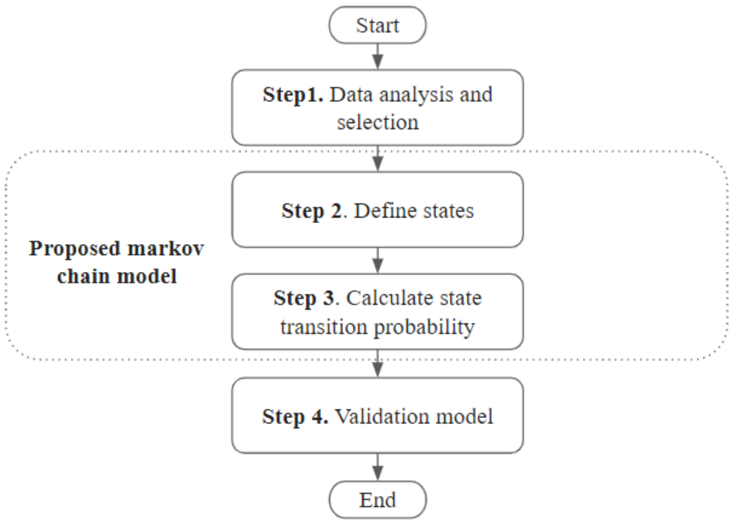


Source: Author

2.2 Methodology proposed

The proposed steps of our Markov-Chain-based methodology are summarized in the following:

Figure 2-2: Proposed methodology for forecasting consumption - Markov chain Model



Source: Author

2.2.1 Step 1: Data analysis and selection

Data related to consumption of medicines in 6 hospitals in France for the years 2016, 2017 and 2018 are available. This data consists of the consumptions of medicines aggregated by months in the hospital of Montpellier. This data has been collected and preprocessed in a doctoral work in progress (Koala et al., 2022). The original database had 65450 observations (consumption of a medicine for one month) and 111 columns (variables) of which only 5 were retained and renamed to reduce the number of variables that are not useful for our study.

The variables include the quantity of medicine (aggregated by months), the name of medicine consumed, year, and month of consumption.

The Montpellier data set has 65450 observations and 2396 medicines. The following criteria were used to select the used medicines:

1. Medicines with consumption values for all months in the years 2016, 2017 and 2018 are selected (without missing values).
   From the application of the criteria 1, a database with 979 medicines is generated.
2. Medicines with a high level of consumption, specifically greater than 500 units are selected. We applied this criterion to avoid predicting values of medicines with low consumptions.

3. The last criterion consists of selecting medicines with a large variation between their maximum and minimum consumption value. For this criterion, the ratio between the minimum value and the maximum value was calculated and medicines with a ratio less than 0.5 were taken. This criterion was applied to avoid studying medicines where the consumption is steady.

After applying these criteria, the resulting dataset has 156 medicines. Table 2-1 shows the top ten most consumed medicines in terms of quantity consumed.

Table 2-1: Top 10 most consumed medicines in the hospital of Montpellier for 2016, 2017 and 2018

Medicines	Quantity consumed			
	2016	**2017**	**2018**	**Total**
HYDROXYZINE 25 MG COMP PELL SEC	158,760	167,080	123,800	449,640
DIFFU K 600 MG GELULE	104,860	130,930	148,600	384,390
BEVITINE 100MG AMP 2 ML INJ	93,780	96,960	92,115	282,855
SOLUPRED 20MG COMPRIME ORODISPERSIBLE	67,255	68,130	70,350	205,735
MIRTAZAPINE 15MG COMP ORODISP	81,208	69,702	50,508	201,418
TERBUTALINE SOL INH 5MG UNIDOSE 2 ML	66,450	59,790	58,710	184,950
FUROSEMIDE 20 MG AMP 2 ML INJ	64,412	60,972	51,945	177,329
PRAZEPAM 10 MG COMPRIME SECABLE	65,250	60,210	48,460	173,920
CELLUVISC COLLYRE 4 MG UNIDOSE	47,705	60,374	60,010	168,089
IPRATROPIUM INH AD 0.5MG DOSE 2ML	60,600	52,650	51,910	165,160

2.2.2 Step 2: Define states of Markov Chain model

The states can be defined from the data sequence by grouping each data point into intervals of consumption. The number of states is a parameter in the proposed model. The size of the intervals in which the values will be classified can vary according to the number of states defined. In the present study, the proposed model is built using two configurations, the first one with 3 states and the second one with 4 states. At the same time, the size of each of the intervals also depends on the method used to group data. The proposed four methods are described below:

Method 1

The states are determined by dividing the data series in quantiles. The data is divided into equal sized groups. It means that the number of points in each state/interval is the same. The three states are represented as follow:

$$[0, Q1], (Q1\ , Q2], (Q2, \infty) \tag{5}$$

where Q1, Q2 and Q3 are respectively the first, second and third quartile.

Method 2

The grouping of the data is performed taking the average value of consumption and applying a percentage tolerance α above and below this value. So, the distance between the extremes of the grouping intervals is calculated by applying the percentage alpha to the average value of the consumption data as shown below

$$size\ =\ average\ *\ \alpha \tag{6}$$

This method generates the following three states/intervals.

$$[0, average - size], [average - size, average + size], [average\ +\ size, \infty) \tag{7}$$

Method 3

In this method, the extremes for the intervals are calculated taken the maximum and minimum values of consumption in data series and dividing by the numbers of groups desired. Limit is the distance between the extremes of the intervals and is defined in formula (8). The result is the distance between each limit to form the consumption intervals as defined in formula (9). In contrary to method 1, the number of values in each interval/state is not necessarily the same.

$$limit = \left\lceil \frac{(max_{quantity} - min_{quantity})}{n} \right\rceil \quad (8)$$

where n is the number of states. The intervals will be:

$$[0, i * limit],]i * limit, (i + 1) * limit] \; for \; i = 1, \dots, n - 1 \quad (9)$$

Method 4

According to (Du et al., 2020) a common machine learning technique can be used which is K-means. K-means clustering takes the data sequence and classifies each data point into a specific cluster. The algorithm is described below:

Step 1. Define the initial number of cluster centers

Step 2. Assign datapoints to clusters

Step 3. Update cluster centroids

Step 4. Repeat step 2 and 3 until the inertia is the least

Now that the four methods are defined. The states are generated and the probability transition between each pair of states is defined in step 3.

2.2.3 Step 3. Calculate the transition probability matrix

The probability transition P_{ij} between states S_i and S_j can be determined using the frequencies of transitions following expression (Du et al., 2020):

$$P_{ij} = \frac{n_{ij}}{\sum_{j=1}^{K} nij} = P_{ij}\{X = S_j | X = S_i\} \quad (10)$$

where n_{ij} is the number of transitions between the state *i* and the state *j* in one step, and k is the total number of the states.

In the present study the state transition matrix is composed of historical data of consumption by months. The probability transition from one state to another describes the probability of changing the state in the next month.

2.2.4 Step 4. Validation of the proposed model

To validate our model, we choose to create it based only on historical consumption data between 2016 and 2017. We call it the prediction model. Ideally, this model should allow to predict the consumptions for every year such as 2018. Therefore, we apply it in 2018 and compare the prediction error/gap. The application here consists of generating the predicted model of 2018 using the same states of the prediction model. Then the real probabilities of 2018 are compared with the probabilities of the prediction model.

In the literature, there are a variety of ways for evaluating predicting error metrics. For this purpose, coefficient of determination (R^2), Mean Square Error (MSE), and Mean of Absolute Error (MAE) were selected for evaluating the predicted model with the real model constructed using historical data. The closer the R^2 is to 1, the greater the prediction value's fit to the observation value, and the better the model's prediction performance. The average squared difference between the estimated values and the actual values are measured by the MSE. A MSE value closer to zero reflects a smaller difference between the original values and the estimated values. The mean of absolute errors between predicted and actual values is measured by the MAE. The R^2, MSE, and MAE can be calculated through equation (11), (12), and (13) respectively.

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \tag{11}$$

$$MSE = \frac{1}{n} - \sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \tag{12}$$

$$MAE = \frac{1}{n} - \sum_{i=1}^{n}|y_i - \hat{y}_i| \tag{13}$$

R^2, MSE and MAE will be used to evaluate the prediction of consumption in medicines.

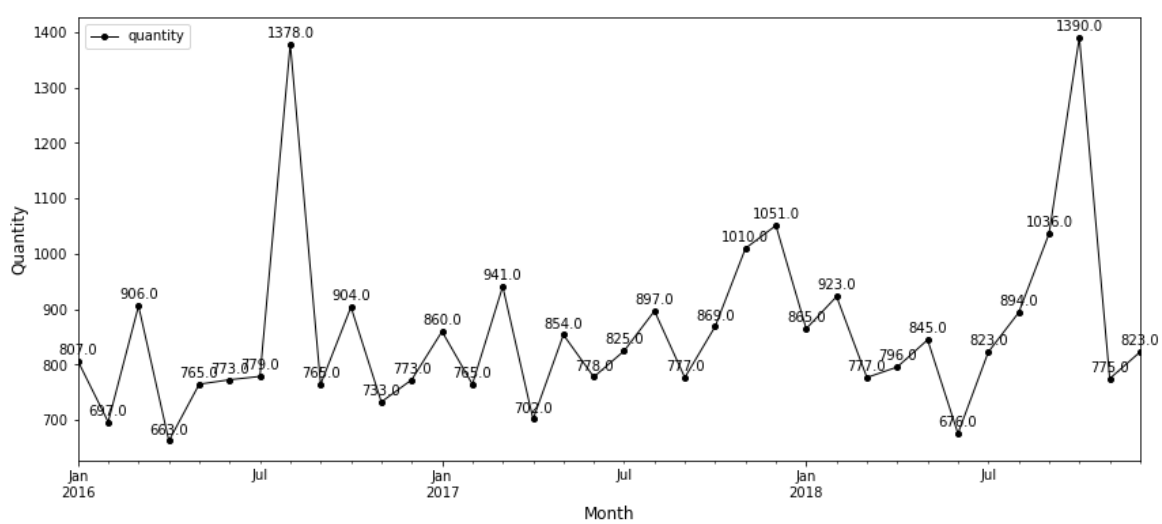
3. Case study

In this section, a case study is presented to illustrate the proposed Markov chain model. For this purpose, the consumption of a particular medicine is selected and presented as an example of illustration. We mentioned that the consumption of 2016 and 2017 is used to build the Markov chain model and the different four methods defined to group the values in intervals of consumption are applied. Finally, the constructed model is applied and tested on 2018 data using the same consumption intervals/states.

Step 1. Data analysis and selection

The consumption of VASELINE OFFICINALE POMMADE TUBE 20G medicine at Montpellier hospital are selected. Figure 3 below shows the variation of the consumption on three years.

Figure 3-1: Consumption of VASELINE OFFICINALE POMMADE TUBE 20G between January 2016 and December 2018



Source: Author

According to Figure 3-1, this medicine has a stationary demand because during the observation time (3 years) it is not possible to observe a trend in its behavior and the values vary within a constant range over time. This means that for this time period the values are between the maximum (1390) and the minimum (663).

Step 2. Define states of Markov Chain model

For the present case study, the number of states was set to 3. However, the number of states can be varied to increase or decrease the size of the intervals in which the consumption values will be classified. In the present study the proposed model is also applied with a 4-state configuration.

For three states, each monthly consumption value between 2016 and 2017 can be assigned to one of the three consumption intervals: L “Low consumption”, M “Medium consumption” and, H: “High consumption”. First, the application of each of the proposed methods is performed using the 2016 and 2017 data to obtain the consumption state intervals and to feed the Markov chain model. Secondly, the model generated from 2016 and 2017 is applied to 2018 using the consumption state intervals previously established.

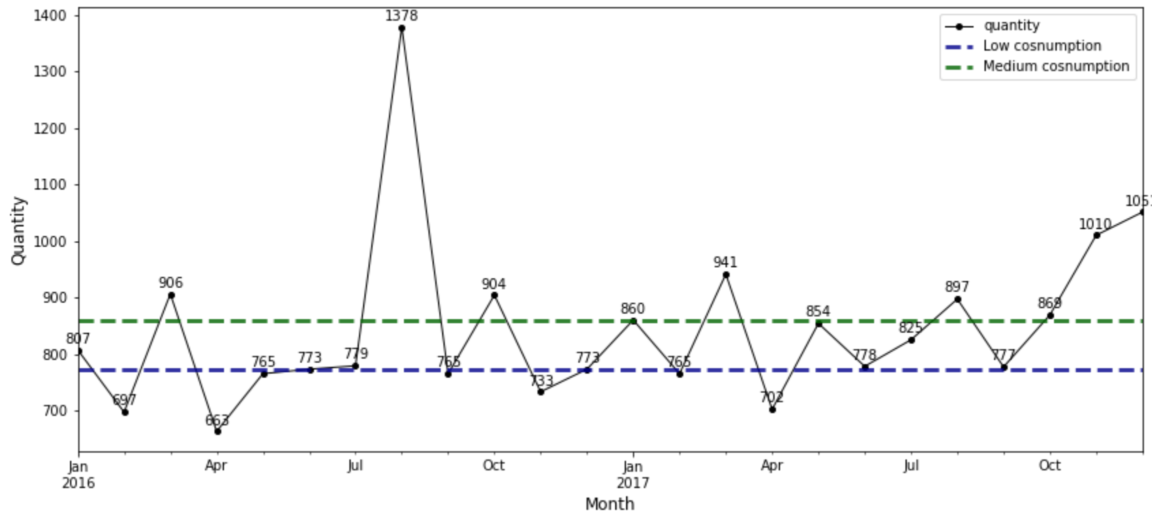
Proposed Markov chain model

The methods described in section 3 will be applied below. The results of the application are presented in Table 3-1.

Method 1

The grouping of each historical value of consumption in 2016 and 2017 for Vaseline using method 1 is represented in Figure 3-2. The values were classified into three consumption ranges. Below the blue line, the values represent a low level of consumption. Between the blue line and the green line, the values represent a medium level of consumption. And, above the green line, the values represent a high level of consumption. Table 3-1 shows the limit values for each of the 3 consumption intervals.

Figure 3-2: Intervals of consumption for 2016 and 2017 of Vaseline medicine with method 1

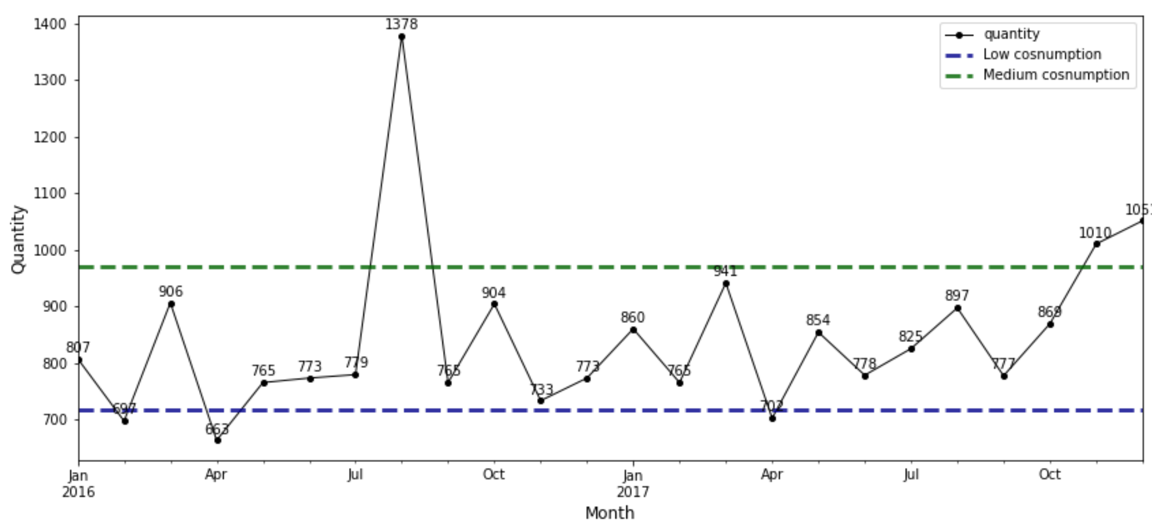


Source: Author

Method 2

The intervals for method 2 are calculated using the equations (6) and (7) with a percentage tolerance α = 15%. The average value of consumption is 845 and the intervals are represented in the following figure.

Figure 3-3: Intervals of consumption for 2016 and 2017 of Vaseline medicine with method 2

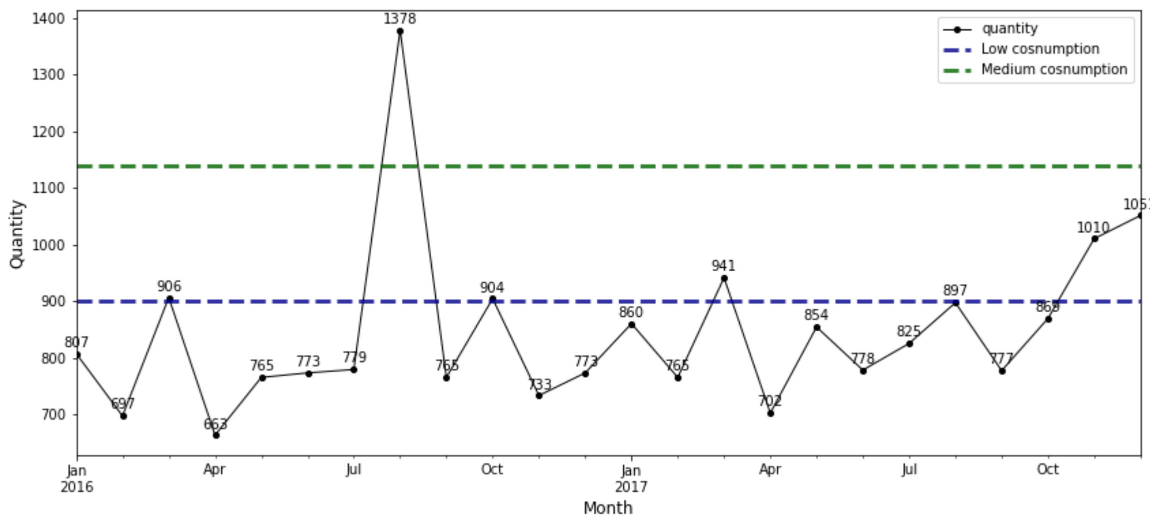


Source: Author

Method 3

To find the values of the limits in method 3, the maximum and minimum values are 1378 and 663 respectively. In our case study the number of states used is n = 3. Using equation (8) and (9), the limits can be calculated as shown below.

Figure 3-4: Intervals of consumption for 2016 and 2017 of Vaseline medicine with method 3.

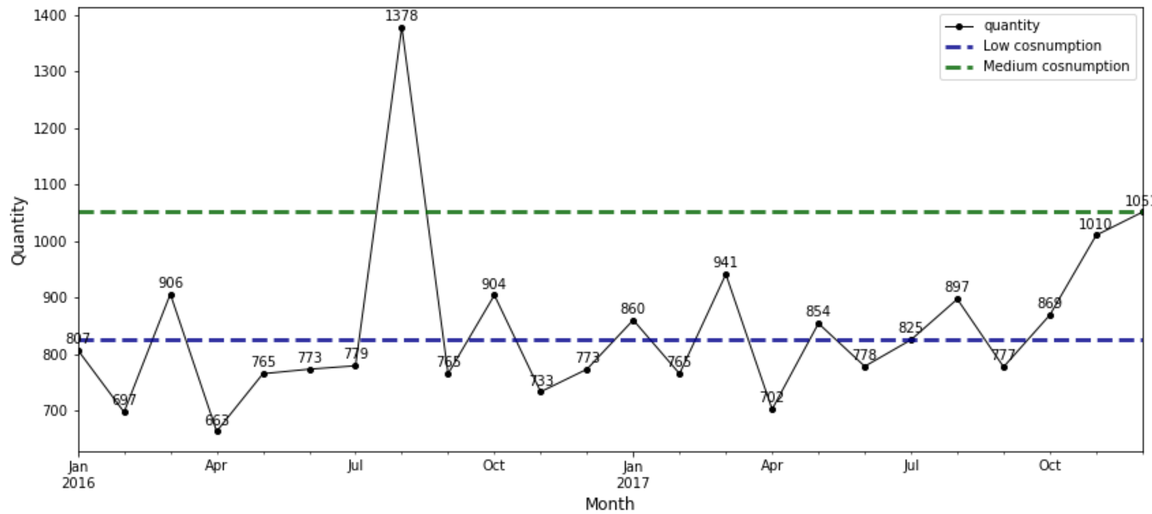


Source: Author

Method 4

Using the k-means algorithm the consumption intervals in figure 3-5 were found. For the case study the optimum number of centroids was 3.

Figure 3-5: Intervals of consumption for 2016 and 2017 of Vaseline medicine with method 4.



Source: Author

Table 3-1: Intervals of consumption for the proposed methods

	States	Method 1	Method 2	Method 3	Method 4
Intervals	Low	[0 - 773]	[0 - 717]	[0 - 901]	[0 - 825]
	Medium	(773 - 860]	(717 - 971]	(901 - 1139]	(825 - 1051]
	High	> 860	> 971	> 1139	> 1051

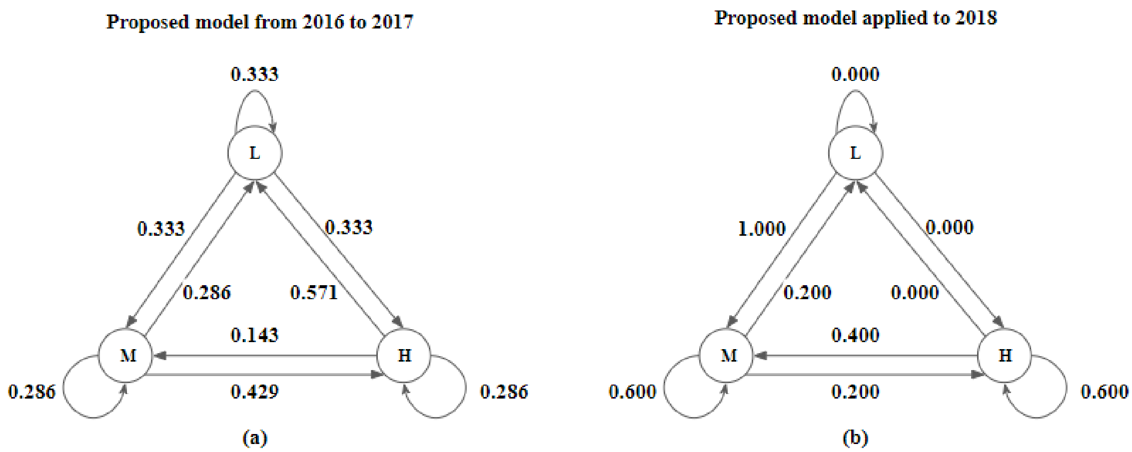
After applying each of the methods to group the data, it can be observed that the size of the intervals varies depending on the method implemented.

Step 3. Calculate the transition probability matrix

To build the model for 2018, the same definition of states/intervals was applied. The proposed Markov chain model generated from the application of each method on the historical data between 2016 and 2017, and the model resulting from the application of the model in 2018 are shown below for each method.

Method 1

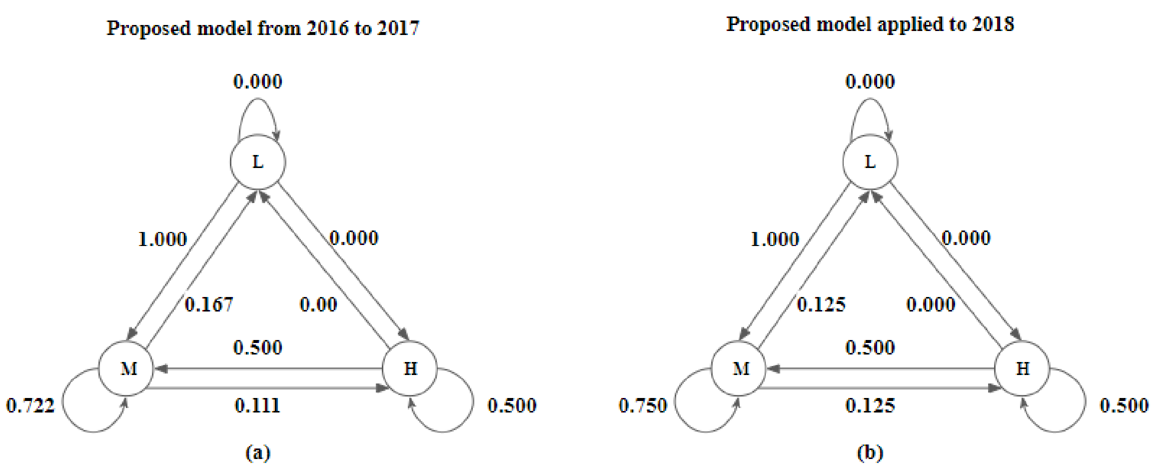
Figure 3-6: Proposed Markov chain model from 2016 to 2017 (a) applied to 2018 (b) for method 1



Source: Author

Method 2

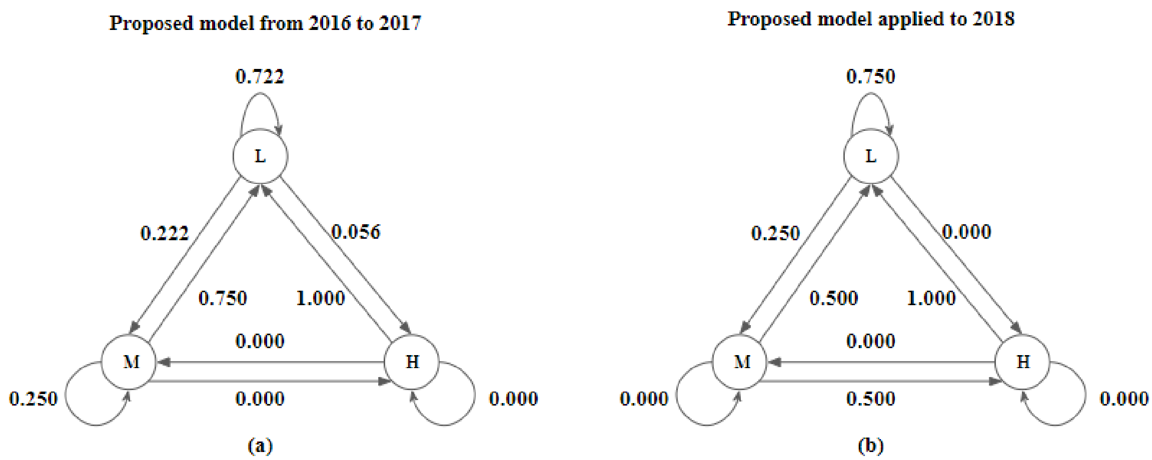
Figure 3-7: Proposed Markov chain model from 2016 to 2017 (a) applied to 2018 (b) for method 2.



Source: Author

Method 3

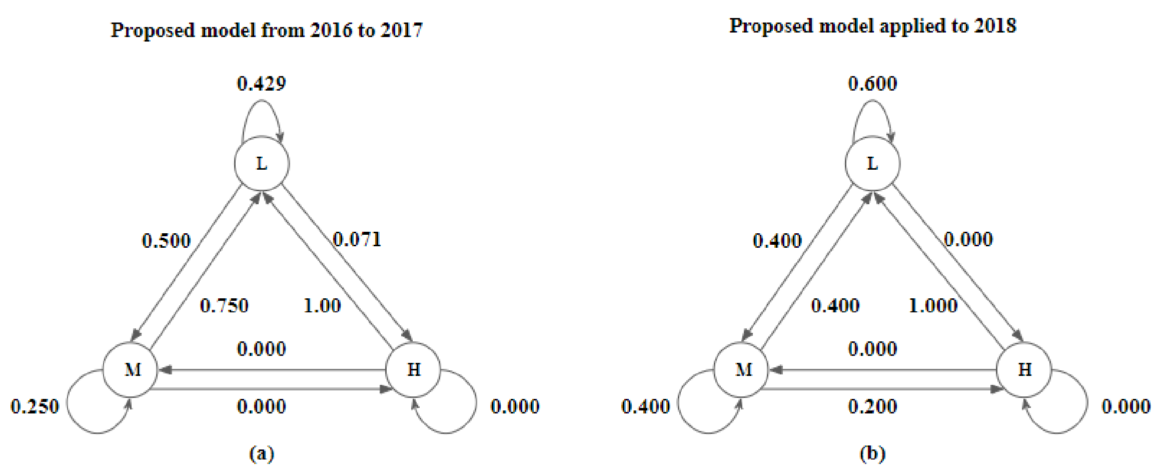
Figure 3-8: Proposed Markov chain model from 2016 to 2017 (a) applied to 2018 (b) for method 3.



Source: Author

Method 4

Figure 3-9: Proposed Markov chain model from 2016 to 2017 (a) applied to 2018 (b) for method 4.



Source: Author

For each of the methods used to build the proposed model for 2016 and 2017, the transition probabilities between the different consumption levels can be observed. The probabilities of transition between the same pair of consumption levels vary depending on the method being used. For instance, the probability of moving from the "Low consumption" to the "High consumption" state for the 4 methods are different as shown in figures 3-6, 3-7, 3-8 and 3-9. This is because the size of the consumption intervals varies according to the method used to group the values.

Something similar occurs when applying the proposed model on 2018 data. In this case, the intervals were fixed in the 2016 and 2017 model, which makes the differences persist.

Step 4. Validating the proposed model

Table 3-2 shows the error metrics comparing the proposed model based on historical data between 2016 and 2017 and the model resulting from the application in 2018 using the 4 proposed methods to group the consumption values.

Table 3-2: Error metrics for the proposed model applying the 4 methods with 3 states to Vaseline

Methods	R^2	MSE	MAE
Method 1	-0.371	0.382	0.350
Method 2	**0.997**	**0.017**	**0.009**
Method 3	0.663	0.205	0.123
Method 4	0.740	0.159	0.115

The results of the case study show that the proposed model using method 2 to define states has the best value which is 0.997 for R2. Method 2 also has the best accuracy from the values of MSE and MAE that are 0.017 and 0.009, respectively. Method 1 has the worst results for R^2, MSE, MAE. Therefore, the proposed model using method 2 generates the best results for predicting Vaseline medicine consumption in 2018.

4. General results

Models with 3 States

The performance of the proposed 3-state models for the selected medicines (156) is shown in Table 4-1. Medicines that were validated using the metrics are highlighted.

The best values for R^2 correspond to those higher than 90%. For MSE and MAE the best values correspond to the closest to zero.

Table 4-1: Error metrics for the proposed model applying the 4 methods with 3 states to medicines selected

Method	$\mathbf{R^2}$ (>0.9)	**MSE** (<0.1)	**MAE** (<0.1)
Method 1	0	0	0
Method 2	VASELINE OFFICINALE POMMADE TUBE 20G	CETIRIZINE 10 MG COMP PELL SEC VASELINE OFFICINALE POMMADE TUBE 20G	CETIRIZINE 10 MG COMP PELL SEC VASELINE OFFICINALE POMMADE TUBE 20G ESIDREX 25 MG COMPRIME SECABLE RAMIPRIL 2.5 MG COMPRIME
Method 3	SOLUPRED 20MG COMPRIME ORODISPERSIBLE	0	NEBIVOLOL 5 MG COMPRIME QUADRISEC ESIDREX 25 MG COMPRIME SECABLE SOLUPRED 20MG COMPRIME ORODISPERSIBLE DIAZEPAM 10 MG COMPRIME SECABLE

Method 4	0	0	ESIDREX 25 MG COMPRIME SECABLE ACIDE FOLIQUE 5 MG COMPRIME DIAZEPAM 10 MG COMPRIME SECABLE3

The results in table 4-1 show that methods 2 and 3 have in general the best performance in predicting the consumption of the selected medicines. The number of medicines within each interval of metrics values is similar. Method 2 performed better. It has 1 medicine whit R^2 value higher than 0.9. Regarding MSE and MAE, method 2 has 2 medicines with values less than 0.1 for MSE and has 4 medicines with values less than 0.1 for MAE. For method 3, the number of medicines into the first interval of values for R^2 is 1 and for MSE and MAE are 0 and 4 respectively. We can conclude that method 2 can be used to predict the consumption for Vaseline, method 3 for Solupred, etc.

In validation process, probabilities whose maximum value is 1 are being compared. This significantly influences the performance metrics. For example, the MAE takes the average squared difference between the estimated values and the actual values. By observing the differences between the probability values of the proposed model and the applied model in figure 3-7 for method 2, it is possible to see that the differences will always be in the range between 0 and 1. A low difference between the actual probability value and the predicted value makes the MAE always generate an apparently good result. The same is true for R^2 and MAE. For this reason, later in this analysis we will compare the actual states with the predicted states.

It is possible to analyze the consumption behavior of the medicines that have the best values for R^2, MSE and MAE. To observe the variability of consumption for these medicines the ratio between the minimum and maximum quantity consumed is calculated and shown in table 4-2. As well as the method that can be used to predict consumption in each case.

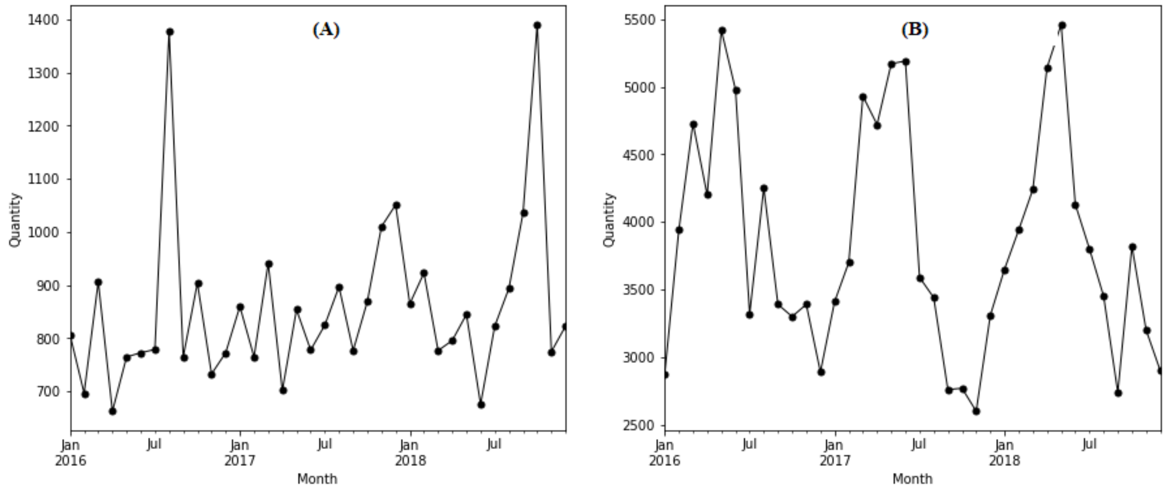
Table 4-2: Medicines with better values for R2, MSE and MAE

R^2 (>0.9)				**MSE** (<0.1)				**MAE** (<0.1)				**Medicine**	**Ratio**
Method													
1	2	3	4	1	2	3	4	1	2	3	4		
	*				*				*			Vaseline officinale pommade tube 20g	0.48
					*				*			Cetirizine 10 mg comp pell sec	0.48
											*	Acide folique 5 mg comprime	0.48
									*			Ramipril 2.5 mg comprime	0.47
		*										Solupred 20mg comprime orodispersible	0.46
										*		Nebivolol 5 mg comprime quadrisec	0.42
										*	*	Diazepam 10 mg comprime secable	0.42
									*	*	*	Esidrex 25 mg comprime secable	0.41

The medicines with the highest consumption values among the top 10 are Vaseline and Ceterizine. They have ratio values of 0.48. According to performance metrics, method 2 can be used to predict the consumption of medicines with high volatility.

In addition, another important characteristic to analyze for these medicines is that it is possible to observe a seasonal behavior in their consumption between 2016 and 2018. Figure 4-1 shows the consumption for Vaseline and Cetirizine.

Figure 4-1: Consumption of Vaseline (A) and cetirizine (B) between 2016 and 2017.



Source: Author

Vaseline and Cetirizine has some periodic fluctuations in comsuption. For Vaseline it is possible to observe that consumption starts to increase in July of 2016 and 2018. Reaches a high consumption value around September and then decreases to a typical consumption value. In the case of Cetirizine, it is observed that for all three years (2016, 2017 and 2018) consumption begins to increase during the first months of the first half of the year. It remains at a high level for about two months and decreases in the second half of all years.

The analysis was performed by taking only two of the medicines with the highest volatility values. However, a more in-depth analysis of consumption behavior over a longer period of time could yield important results on the relationship between volatility and seasonality and the performance results obtained for the proposed model.

As explained the used metrics do not allow to validate precisely the models. Another type of analysis can be used. Consumption levels for real data in 2017 can be compared with consumption levels predicted by the model for 2018 as shown in table 4-3 for Vaseline medicine. For this objective, method 2 will be used and the initial state will be the month of December 2017 with a consumption value of 773, which does signify a medium consumption level. Using the Markov chain model proposed in Figure 3-7 (A) from method 2. It is observed that starting at an average consumption state, there is a 72 %

probability of staying at the same value. Thus, predicted consumption state for the month of January 2018 will probably be "Medium". In the same way, the consumption values for the following months can be predicted as follows

Table 4-3: Comparison of actual and predicted values for 3 states using method 2 for Vaseline medicine

State by month	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dic
Predicted	M	M	M	M	M	M	M	M	M	M	M	M
Real data	M	M	M	M	M	L	M	M	H	H	M	M

Table 4-3 shows that the model proposed in this study predicts the right variations for 9 months. Errors are highlighted in red.

In contrast to the performance metrics used in the evaluation of the proposed model, the above comparison allows us to observe how accurate the Markov model is in predicting consumption intervals on a month-to-month.

Based on this result, it can be considered to include a level of adjustment above and below each of the consumption intervals defined by method 2 that allows a wider range of values to be included. In such case, it is worth to mention that our prediction is not accurate because it predicts a range of values and not the exact one.

Models with 4 States

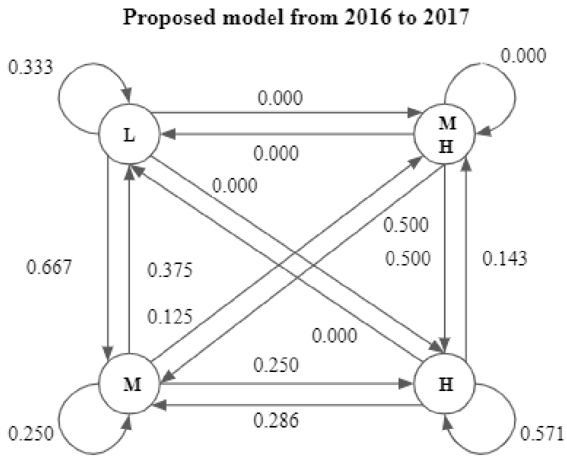
To understand how the models behave on 4 states, table 4-4 show the results (method 2 can be used only for an odd number of states). None of the methods were able to find a medicine with a good R^2 or MSE values. The method 3 has a superior performance in this case but only for MAE. It has 4 medicines with a value less than 0.1. We select Cetirizine for further analysis.

Table 4-4: Error metrics for the proposed model applying the 4 methods with 4 states to medicines sele

Method	R^2 (>0.9)	**MSE** (<0.1)	**MAE** (<0.1)
Method 1	0	0	0
Method 3	0	0	CETIRIZINE 10 MG COMP PELL SEC MAGNESIUM CHLORURE 10% AMP 10ML INJ OXYBUPROCAINE COLL 0.4% UNIDOSE 0.4 ML CISATRACURIUM 10 MG AMP 5 ML INJ
Method 4	0	0	0

Consumption levels can also be compared in the case of 4 states. The new consumption range/state will be designated as medium-high consumption (MH). The medicine used for this part of the analysis is cetirizine. The initial state will be the month of December 2017 with a consumption value of 3310, which is a medium consumption level. The proposed model with 4 states for cetirizine is presented in figure 13 and table 9 shows the states of real predicted data.

Figure 4-2: Proposed Markov chain model from 2016 to 2017 with method 3 of Cetirizine medicine.



Source: Author

Table 4-5: Comparison of actual and predicted values for 4 states using method 3 for cetirizine medicines

State by month	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dic
Predicted	L	M	L	M	L	M	L	M	L	M	L	M
Real data	M	M	MH	H	H	MH	M	M	L	M	L	L

According to the Markov Chain model shown in Figure 4-5, when starting at a medium consumption level there is a 37.5% probability of going to a low consumption level, comparatively this probability is higher than that of staying in the same state or going to the medium-high and high consumption states. Therefore, the predicted state for January 2018 will be Low. The consumption values for the following months are shown in the above table. Table 9 shows that the predicted states correspond to the real states for only 5 months.

The MAE yields a good result for cetirizine. However, when looking at the comparison between the actual and predicted states for each month, the model was correct for only 5 months of the year. From this, it can be concluded that using MAE as a metric to evaluate the performance of the proposed is not enough because probabilities with a range of values between 0 and 1 are being compared. However, when comparing actual and predicted consumption states, it is possible to conclude that this metric is not as appropriate.

According to the above analysis, the main results of the present study are summarized as follows:

I. For 3 states method 2 is the best one for predicting medicine demand for the selected medicines, especially for the medicines Vaseline and Cetirizine, which have similar characteristics related to a high variability and the presence of seasonality in their consumption.
II. In the case of 4 states, method 3 was the only method to give some results with MAE values closer to 0.
III. Comparison between actual and predicted consumption states on the months of the year is useful to complement the validation of the model with the selected performance metrics.

5. Conclusión and Future Work

Uncertainty of factors such as patient's condition, the length of stay in hospitals and transition between different care units impact medicine demand and require better inventory management. Predicting medicine demand can provide an important input to make decisions in a better management of the inventory.

Therefore, Markov chain was used for the proposition of four models that can be used for predicting medicine demand. The states of these models represent intervals of consumption in a specific period (month). Historical data of medicine consumption in Montpellier hospital were used. Based on performance evaluation, some of the proposed Markov chain models can be appropriate as a forecasting tool for some of the selected medicines.

Some limitations of this study are: (1) The present study used monthly data available for two years to propose the model. This means that the model was proposed with 24 observations. More data could better capture fluctuations in drug consumption. (2) The proposed model can only model stationary consumptions and not trendy ones. However, if consumption behavior changes significantly in the future (the minimum or maximum values changes for instance), the model will have to be adjusted to the new conditions. For example, if consumption has a large increase, the parameters in the methods used to group the data into intervals will have to change.

For future research, new methods to group the data can be implemented to improve the performance of the proposed model. Even the variation in the parameters used to calculate the intervals in each method can generate interesting results. In addition, the application of the proposed model in the other hospitals can generate interesting results to validate the model with more data. Models that include more variables about influencing factors such as patient condition and number of care units can be built to improve the accuracy. For example, the model developed by Saha and Ray (Saha & Ray, 2019a) can be explored in depth to capture better the function.

Bibliografía

Abdulsalam, Y., Gopalakrishnan, M., Maltz, A., & Schneller, E. (2018). The impact of physician-hospital integration on hospital supply management. *Journal of Operations Management*, *57*, 11–22. https://doi.org/10.1016/j.jom.2018.01.001

Addis, B., Carello, G., Grosso, A., Lanzarone, E., Mattia, S., & Tànfani, E. (2015). Handling uncertainty in health care management using the cardinality-constrained approach: Advantages and remarks. *Operations Research for Health Care*, *4*, 1–4. https://doi.org/10.1016/j.orhc.2014.10.001

Anderson, T., & Goodman, L. (1957). Statistical Inference About Markov Chains. *The Annals of Mathematical Statistics*, *28*. https://doi.org/10.1214/aoms/1177707039

Aptel, O., & Pourjalali, H. (2001). Improving activities and decreasing costs of logistics in hospitals: A comparison of U.S. and French hospitals. *The International Journal of Accounting*, *36*(1), 65–90. https://doi.org/10.1016/S0020-7063(01)00086-3

Attanayake, N., Kashef, R. F., & Andrea, T. (2014). A simulation model for a continuous review inventory policy for healthcare systems. *2014 IEEE 27th Canadian Conference on Electrical and Computer Engineering (CCECE)*, 1–6. https://doi.org/10.1109/CCECE.2014.6901005

de Vries, J. (2011). The shaping of inventory systems in health services: A stakeholder analysis. *International Journal of Production Economics*, *133*(1), 60–69. https://doi.org/10.1016/j.ijpe.2009.10.029

Du, H., Zhao, Z., & Xue, H. (2020). ARIMA-M: A New Model for Daily Water Consumption Prediction Based on the Autoregressive Integrated Moving Average Model and the

Markov Chain Error Correction. *MDPI Water*, *12*(3), 760. https://doi.org/10.3390/w12030760

Gebicki, M., Mooney, E., Chen, S.-J., & Mazur, L. M. (2014). Evaluation of hospital medication inventory policies. *Health Care Management Science*, *17*(3), 215–229. https://doi.org/10.1007/s10729-013-9251-1

Goltsos, T. E., Syntetos, A. A., Glock, C. H., & Ioannou, G. (2021). Inventory – forecasting: Mind the gap. *European Journal of Operational Research*, S0377221721006500. https://doi.org/10.1016/j.ejor.2021.07.040

Hermosilla, A., Carmagnola, R., Sauer, C., Redondo, E., & Centurion, L. (2020). *Demand forecasts for chronic cardiovascular diseases medication based on Markov chains*. *5*(2), 6.

Koala, D., Yahouni, Z., Alpan, G., & Si Mohand, D. (2022). Correlation Analysis of Factors Impacting Health Product Consumption in French Hospitals. *10th IFAC Conference on Manufacturing Modelling, Management and Control: MIM*.

Kocer, U. U. (n.d.). FORECASTING INTERMITTENT DEMAND BY MARKOV CHAIN MODEL. *International Journal of Innovative Computing, Information and Control*, 13.

Landry, S., & Philippe, R. (2004). How Logistics Can Service Healthcare. *Supply Chain Forum: An International Journal*, *5*(2), 24–30. https://doi.org/10.1080/16258312.2004.11517130

Lopez Ramirez, A. J., Jurado, I., Fernandez Garcia, M. I., Isla Tejera, B., Del Prado Llergo, J. R., & Maestre Torreblanca, J. M. (2014). Optimization of the demand estimation in hospital pharmacy. *Proceedings of the 2014 IEEE Emerging Technology and Factory Automation (ETFA)*, 1–6. https://doi.org/10.1109/ETFA.2014.7005057

Pan, Z. X. (Thomas), & Pokharel, S. (2007). Logistics in hospitals: A case study of some Singapore hospitals. *Leadership in Health Services*, *20*(3), 195–207. https://doi.org/10.1108/17511870710764041

Pokharel, S. (2005). Perception on information and communication technology perspectives in logistics: A study of transportation and warehouses sectors in Singapore. *Journal of Enterprise Information Management*, *18*(2), 136–149. https://doi.org/10.1108/17410390510579882

Polanecký, L., & Lukoszová, X. (2016). Inventory Management Theory: A Critical Review. *Littera Scripta*, *9*(2), 11.

Roni, M. S., Eksioglu, S. D., Jin, M., & Mamun, S. (2016). A hybrid inventory policy with split delivery under regular and surge demand. *International Journal of Production Economics*, *172*, 126–136. https://doi.org/10.1016/j.ijpe.2015.11.015

Saha, E., & Ray, P. K. (2018). Inventory Management and Analysis of Pharmaceuticals in a Healthcare System. In P. K. Ray & J. Maiti (Eds.), *Healthcare Systems Management: Methodologies and Applications: 21st Century Perspectives of Asia* (pp. 71–95). Springer. https://doi.org/10.1007/978-981-10-5631-4_7

Saha, E., & Ray, P. K. (2019a). *Patient condition-based medicine inventory management in healthcare systems*. http://www.tandfonline.com/doi/epub/10.1080/24725579.2019.1638850?needAccess=true

Saha, E., & Ray, P. K. (2019b). Modelling and analysis of inventory management systems in healthcare: A review and reflections. *Computers & Industrial Engineering*, *137*, 106051. https://doi.org/10.1016/j.cie.2019.106051

Varghese, V., Rossetti, M., Pohl, E., Apras, S., & Marek, D. (2012). Applying Actual Usage Inventory Management Best Practice in a Health Care Supply Chain. *International Journal of Supply Chain Management*, *1*(2), 10.

Vila-Parrish, A. R., Ivy, J. S., & King, R. E. (2008). A simulation-based approach for inventory modeling of perishable pharmaceuticals. *2008 Winter Simulation Conference*, 1532–1538. https://doi.org/10.1109/WSC.2008.4736234

Villegas, M. A., Pedregal, D. J., & Trapero, J. R. (2018). A support vector machine for model selection in demand forecasting applications. *Computers & Industrial Engineering*, *121*, 1–7. https://doi.org/10.1016/j.cie.2018.04.042

Volland, J., Fügener, A., Schoenfelder, J., & Brunner, J. O. (2017). Material logistics in hospitals: A literature review. *Omega*, *69*, 82–101. https://doi.org/10.1016/j.omega.2016.08.004