Healthcare Analytics 3 (2023) 100146

Contents lists available at ScienceDirect

Healthcare Analytics

journal homepage: www.elsevier.com/locate/health

Forecasting daily admissions to an emergency department considering single and multiple seasonal patterns

Adriana Vieira [a,*], Inês Sousa [a,b], Sónia Dória-Nóbrega [c]

[a] Department of Mathematics, University of Minho, Campus de Gualtar, 4710-057, Braga, Portugal
[b] Centre of Molecular and Environmental Biology, University of Minho, Campus de Gualtar, 4710-057, Braga, Portugal
[c] Hospital of Braga, Sete Fontes - São Victor, 4710-243, Braga, Portugal

ARTICLE INFO

Keywords:
Daily data
Emergency medicine
Forecasting
Multiple seasonality
Time series

ABSTRACT

When dealing with several years of daily data, such as the number of daily admissions to a hospital's emergency department (ED), how complex does it get to forecast into the future? With that in mind, this study has two main goals: to explore the differences between several methodologies, considering both single and multiple-seasonal patterns; and to select the most suitable model for the administration of a Portuguese hospital to use while managing their ED. To that end, we first considered the data as a time series with a single weekly seasonal pattern. We then modelled the data using time series regression, linear regression with autoregressive integrated moving average (ARIMA) errors, seasonal ARIMA and exponential smoothing techniques. Second, the data was set to be a time series with weekly and annual seasonal patterns. Then, using Fourier terms, we applied time series regression, linear regression with ARIMA errors and trigonometric exponential smoothing state space models with Box–Cox transformation, ARMA errors, Trend and Seasonal components (TBATS) for the analysis. After selecting the best-fitting models using the Akaike Information Criteria (AIC) values, we forecasted into the future and compared the results using both training and test datasets' root mean square error (RMSE), mean absolute error (MAE) and mean absolute percentage error (MAPE) values. The time series regression model based on seasonal variables and a weekly seasonal pattern gives the best results. However, we decided to use linear regression with ARIMA errors, seasonal variables, and both weekly and annual seasonal patterns. This produces similar results but allows for the annual seasonality to be considered, which is useful when more data is added.

1. Introduction

The Portuguese National Health Service relies heavily on public hospitals, complex systems with multiple areas and departments allocating different activities and resources. The emergency department (ED) is one of the main areas of a hospital. The main focus of this department, as the name implies, is the treatment of situations that need immediate attention.

However, according to the report from the Comissão de Reavaliação da Rede Nacional de Emergência/Urgência [1], in 2010, approximately 46% of admissions to national emergency services were not episodes of an urgent nature. Hence, about 6 million annual admissions to mainland Portugal's public emergency services result from abusive and non-urgent situations. This adds weight to a service that is itself responsible for somewhat unpredictable situations, meaning an overcrowded department, which leads to excessive waiting times and a lower response rate.

Given the financial constraints hospitals face, resource allocation must be balanced between cost and efficiency. It is then clear that hospital administrators must efficiently assign human and material resources to the ED. For this reason, in a central hospital that serves a large population, with the number of admissions to the ED being highly variable, forecasting is essential. In other words, a reliable system for predicting admissions to an ED is crucial.

Previous research has shown that attendance at the ED is conditional on several factors: meteorological [2–10]; seasonal [3,5,6,8,10–12]; epidemiological [10,13,14]; and even environmental [4,10,15,16].

Attia and Edward [2] found no significant effect of environmental conditions on the number of admissions to a paediatric unit's ED. Noble et al. [7], on the other hand, found that the number of patients who went to the ED of three Boston hospitals without using an ambulance (potentially non-urgent cases) was higher when the weather was favourable. To model the number of daily admissions in a veterans hospital, Holleman et al. [6] used both meteorological and calendar variables, concluding that maximum temperature, snow presence, season, day of the week, bank holiday, and week of the month were significant explanatory variables. Attempting to explain the number of

* Corresponding author.
E-mail address: adrimsvieira@gmail.com (A. Vieira).

https://doi.org/10.1016/j.health.2023.100146
Received 27 September 2022; Received in revised form 2 February 2023; Accepted 10 February 2023

admissions into an ED in Denver, Batal et al. [3] discovered maximum temperature and snow presence as statistically significant meteorological variables, and day of the week, the month of the year, season, and the day after a bank holiday as statistically significant calendar variables. Regarding environmental variables, Diaz et al. [4] found, while conducting a study in Madrid, that relative humidity and ozone concentration explained the number of admissions at the ED. Lin et al. [15] found a positive association between the concentration of several air pollutants (ozone, carbon monoxide, nitrogen dioxide among others) and the admissions to the ED of Changsha Central Hospital. Results from a study conducted in the Lisbon Metropolitan Area by Franco et al. [16] showed the existence of a significant correlation between air pollutants and admissions to the ED, with carbon monoxide and ozone leading to a higher number of admissions. As for epidemiological variables, Fuhrmann et al. [14] found a direct positive association between the number of admissions and the number of flu and pneumonia cases in a hospital's ED in North Carolina. A positive correlation between the number of flu cases and the number of admissions to the ED was also concluded in a study conducted in five Milanese hospitals by Murtas et al. [10].

Also, according to Batal et al. [3], Holleman et al. [6], Diehl et al. [5], Wargon et al. [17], Marcilio et al. [18], Vieira and Sousa [12] and Erkamp et al. [19], Mondays have the highest number of emergency room admissions, which decreases throughout the week, with fewer admissions on weekends. In turn, Hitzek et al. [20] concluded exactly the opposite — weekend ED admissions proved to be significantly higher than weekday admissions.

A further challenge in this kind of analysis is the complexity of seasonal patterns in time series data, which is growing due to high-frequency data recording [21]. For example, in a time series with only one year of monthly observations, there is usually only one periodical variation - a monthly seasonal pattern. However, if we take several years' worth of hourly data, the situation is different — daily, weekly, monthly, and annual seasonal patterns are common in this type of data [22].

The issue with such complex seasonal patterns is that only one seasonal pattern can be modelled by the conventional time series methods (e.g. autoregressive integrated moving average (ARIMA), seasonal ARIMA (SARIMA), exponential smoothing). Therefore, other techniques must be employed to handle patterns with several seasonal components or even patterns with non-integer seasonalities (e.g. trigonometric exponential smoothing state space models with Box–Cox transformation, ARMA errors, Trend and Seasonal components (TBATS), linear regression with ARIMA errors and Fourier terms) - [21,23].

Hence, a set of daily data spanning several years is a good example of a time series expected to have multiple seasonal patterns — at least weekly and annual [21]. The data under study fits this description since it was recorded daily across a four-year period.

To model their daily data, the authors of the majority of the studies already mentioned [3,5,6,8,10,13,19,24,25] used linear regression (with or without the inclusion of meteorological, epidemiological, and environmental covariates) or time series models (SARIMA and exponential smoothing models). In addition, Jones et al. [24] and Rocha and Rodrigues [25] also made use of artificial neural network models. In turn, linear regression with ARIMA errors was the modelling approach considered by Vieira and Sousa [12] and Murtas et al. [10]. As a result of how simple it is to model patterns and trends using these methods, which are usually easy to comprehend, they are often used across the literature. However, in each of these studies, the data is modelled using a single seasonal pattern, which may not be ideal.

As already mentioned, this work analyses a dataset on the number of daily admissions to an ED. The data have been recorded from January 1st 2013, until December 31st 2016.

A preliminary analysis of the data showed that the number of daily admissions to the ED had been increasingly changing over the years, having a weekly seasonal pattern, as expected [21,24]. In this case, the more straightforward and standard approach would be to set the frequency of the seasonal pattern to 7. However, as already discussed, daily data frequently include more than one seasonal pattern [21]. In fact, in this particular situation, both weekly and annual seasonal patterns are present in the data. The frequency of the seasonal patterns should then be set as 7 and 365.25 (we use 365.25 as the period instead of 365 to accommodate for leap years). However, as previously stated, not all methodologies allow the inclusion of more than one seasonal pattern or a non-integer frequency. For example, SARIMA and some exponential smoothing approaches only accommodate one seasonal pattern and shorter integer frequencies, which means that the seasonal pattern of 365.25 cannot be used when employing these approaches.

Therefore, we intend to compare the forecasting results when daily data are analysed using various modelling approaches, considering both single-seasonal and multiple-seasonal patterns while paying attention to the nature of the correlation structure. In addition, there is interest in selecting the most appropriate and somehow simple time series model for the hospital administration to use while resource managing.

In Section 2, we give more detailed information about the data. Section 3 describes the statistical methodologies we use to model the data while, in Section 4, we apply the said methodology to the data to understand the differences in the forecasting results. Finally, Section 5 summarises the main conclusions of this paper.

A two-sided significance of 5% was used. Statistical analysis was performed with `R Version 4.0.3` [26].

2. Motivation

2.1. Hospital of Braga

The Hospital of Braga opened in the city of Braga, northwest Portugal, in May 2011. This hospital is the primary reference hospital for approximately 320,000 inhabitants and the secondary reference hospital for about 780,000 patients.

In an urgent situation, the Hospital of Braga guarantees to care for any patient, Portuguese or not, with no daily limit on the number of patients to be treated. That being said, this public hospital covers an area of about 1.1 million inhabitants, more than 10% of the Portuguese population [27].

The admission of a patient into the hospital can be elective or urgent. When a patient's admission is scheduled in advance, we are in the presence of an elective admission. However, when a patient is admitted through the ED in an unexpected way, the admission is said to be urgent.

At the ED, the admission is made accordingly to the Manchester Triage System [28]. After answering some clinical questions, the patient receives a coloured bracelet according to the situation's urgency. The colour assigned to the patient determines the maximum waiting time (in minutes) he must undergo until the first medical observation.

According to the administration of this facility, each doctor and their team treats, on average, 70 patients each day (some non-medical staff members may be on the team of several doctors). Therefore, an adequate forecast of the daily number of patients results in a more precise selection of staff members.

2.2. Dataset

Between January 1st 2013, and December 31st 2016, 531,164 people were admitted to the ED of the Hospital of Braga, with 79.22% of the situations referring to residents of the hospital's primary area of action and 18.28% referring to residents of the hospital's secondary area of action.

A total of 288,016 (54.22%) patients were female, and 243,148 (45.78%) were male. Their age ranged between 19 and 106 years, with an average of 53.76 years (standard deviation (SD) ± 20.17). Please note that the Hospital of Braga has a separate emergency department

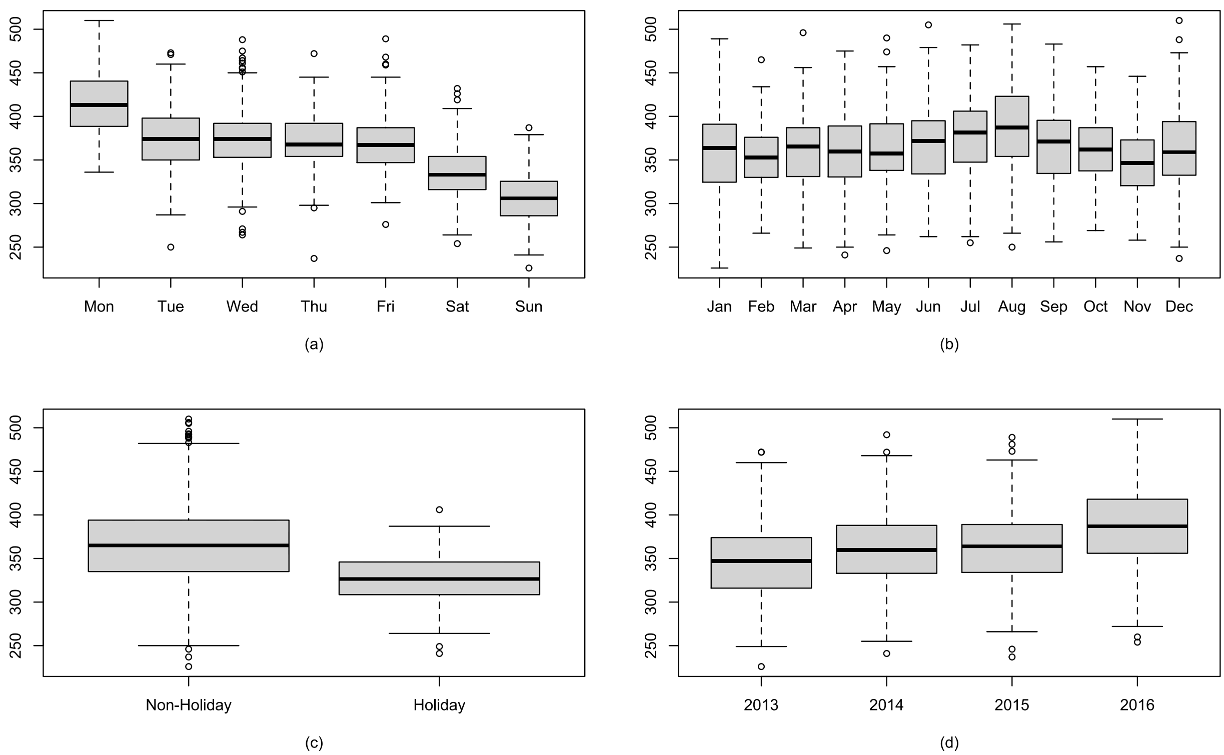

Fig. 1. Boxplots displaying the distribution of daily admissions to the ED of the Hospital of Braga by day of the week (a), month (b), holiday (c), and year (d).

dedicated solely to children, which is why our data does not include under 19-year-old patients.

Only 9.35% of admissions were Immediate or Very Urgent, and 59.74% of admissions were Urgent. On the other hand, 30.55% of admissions to the ED were classified as Standard or Non-urgent.

From a preliminary analysis, we conclude that the number of patients who go to this ED is higher on Mondays, with an average of 415.60 patients (SD ±35.80) and decreases over the week, reaching its minimum on Sundays with an average of 305.70 patients (SD ±29.07). On average, the number of patients admitted to the ED is less on holiday days (mean of 324.75 patients, SD ±35.32) than on non-holiday days (mean of 364.77 patients, SD ±45.72). In terms of the number of patients admitted to the ED by month of the year, July (mean of 380.05, SD ±45.46) and August (mean of 384.93, SD ±52.16) have the higher daily admissions. Also, the daily average of patients entering the ED in 2013 is 345.34 (SD ± 43.11); in 2014 is 359.63 (SD ± 42.38); in 2015 is 362.37 (SD ± 41.34); in 2016 is 386.84 (SD ± 46.93). The same conclusions can be drawn from the analysis of the boxplots in Fig. 1, which show the distribution of daily admissions to the ED by day of the week (Fig. 1(a)), month (Fig. 1(b)), holiday (Fig. 1(c)), and year (Fig. 1 d)).

Fig. 2 shows the daily number of admissions to the ED in grey and a non-parametric estimate of its mean, obtained with the `smooth.spline` function of R, in black. The data appears to show an increasing trend, as well as an annual seasonal pattern.

In order to better understand the time series, we decomposed it using the Multiple Seasonal-Trend decomposition using Loess (MSTL) procedure proposed by Bandara et al. [29]. The MSTL technique is available to implement through the `mstl` function from the R package `forecast` [30]. The decomposition of the time series is displayed in Fig. 3. Aside from the trend component (second panel) and the remainder component (fifth panel), there are two seasonal patterns displayed, one for the week (third panel) and another for the year (fourth panel). Notice the vertical scales. In this situation, the trend component presents a narrower scale than the other components, indicating that the data exhibit a slight linear trend. The existence of the weekly and annual seasonalities can also be verified, with the weekly seasonality being relatively stronger than the annual seasonality.

In light of this, we consider a first scenario in which the data is presented as a time series with a weekly pattern. Additionally, given the existence of an annual seasonal pattern, we consider a second scenario in which the data is a time series with both weekly and annual seasonal patterns.

Note that, when time series data exhibit variation that increases or decreases with the series level, applying a transformation to such data can be advantageous [21]. A widely used family of transformations is the Box–Cox family of transformations [31]. Hence, in both scenarios, we log-transformed the data to smooth the across time variance (Box–Cox transformation with $\lambda = 0$). Remember that, while the Box–Cox family of transformations is introduced as part of the TBATS models in Section 3.5, which searches for and considers a Box–Cox transformation by design, all methodologies used can deal with the transformed time series data. We computed λ using the `BoxCox.lambda` function from the `forecast` package in R [30]. Because the returned value was close to zero, we applied the logarithmic transformation to the data in all scenarios. The Box–Cox transformation outputted by TBATS also uses $\lambda = 0$ (see Section 4.2.3).

From the Kwiatkowski–Phillips–Schmidt–Shin (KPSS) test [32], there is statistical evidence that the series is, in fact, non-stationary ($p < 0.01$), varying with time.

3. Methodology

Using all data available, we divided the analysis into two parts, accordingly to the two main objectives we defined in Section 1.

For the first objective, that is, to compare results considering both single-seasonal and multiple-seasonal patterns, we established a two-step approach.

In the first scenario, we considered the data as a time series with a single weekly seasonal pattern (seasonal period of 7). We applied four methodologies — time series regression, SARIMA, exponential smoothing and linear regression with ARIMA errors. Regarding the

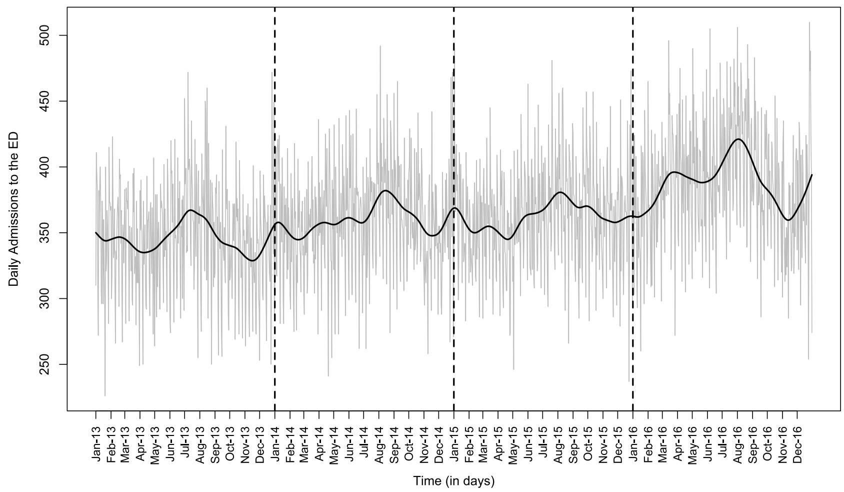

Fig. 2. Daily Admissions to the ED of the Hospital of Braga from January 1st 2013 until December 31st 2016 (grey lines) with a non-parametric estimate of its mean (black solid line). The vertical black dashed lines indicate the first day of 2014, 2015 and 2016.

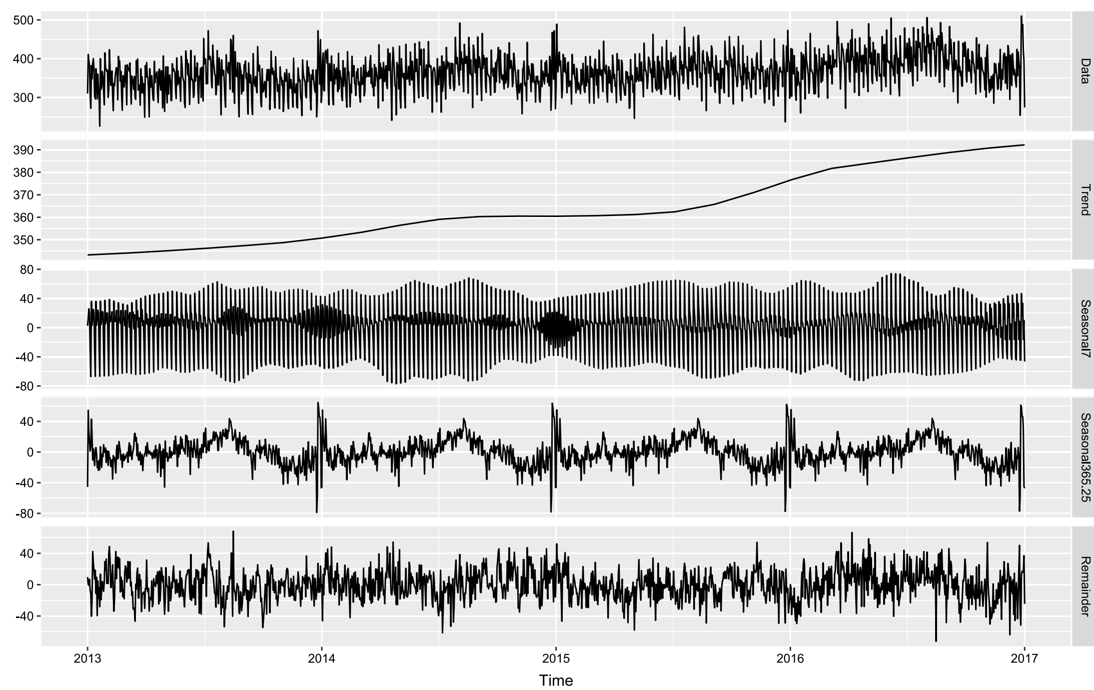

Fig. 3. Number of daily admissions to the ED of the Hospital of Braga from January 1st 2013 to December 21st 2016 (1st panel) and its four additive components — trend component (2nd panel), weekly and annual seasonal components (3rd and 4th panels, respectively), and remainder component (5th panel).

second scenario, the data was set to be a time series with two seasonal patterns, weekly and annual (seasonal periods of 7 and 365.25). We used the standard time series regression, linear regression with ARIMA errors and TBATS models for the analysis.

For each methodology within each scenario, we compared the models' goodness of fit using the Akaike Information Criteria (AIC - Akaike [33]) values, thus choosing the most suitable models. We then forecasted into the future and compared the results against the actual observed data. In order to do this, we used three measures: the mean absolute percentage error (MAPE - Coleman and Swanson [34]), the root mean square error (RMSE - Barnston [35]) and the mean absolute error (MAE - Sammut and Webb [36]).

Given such comparison, it was then possible to choose a model (or models) the hospital administration could use to forecast the daily number of admissions at their ED.

Over the next subsections, we describe the methodologies used to model the data. But first, let us consider y_t, $t = 1, \dots, n$ as the response variable at time t (in this case the response variable refers to the daily number of admissions at the ED in day t).

3.1. Time series regression

Time series regression [21] explains and models a time series assuming a linear association with other time series. This method allows one to model, with the introduction of covariates, both trend and seasonal variation, with easily interpretable results. Such models are written as

$$y_t = \beta_0 + \beta_1 x_{1,t} + \beta_2 x_{2,t} + \cdots + \beta_k x_{k,t} + \epsilon_t, \quad (1)$$

where, $x_1, x_2, \dots, x_k$ represent the k covariates, and $\beta_1, \beta_2, \dots, \beta_k$ each associated regression parameter. The $\beta_j, j = 1, \dots, k$ regression parameter measures the effect, on average, of its corresponding covariate over the response, after considering the effects of the $k - 1$ remaining covariates. The parameter β_0 represents the predicted value for y_t when all covariates are set to 0 or take the baseline category. As for ϵ_t, it describes the error at each time point t.

The regression parameters are estimated by the least squares method. Errors should have mean zero and be uncorrelated. Having the errors be normally distributed with a constant variance σ^2 is also helpful, though not necessary, for estimating prediction intervals. Also, the covariates should not be correlated.

To test the existence of autocorrelation in the residuals, we used the Breusch–Godfrey test, with the null hypothesis being "there is no autocorrelation in the residuals".

To fit the linear regression model to the time series data, we used the `tslm` function from the `forecast` package in R [30]. This function is similar to the `lm` function, generally used in linear regression. However, this is better suited for working with time series data since it offers additional features, such as the ability to introduce a Box–Cox transformation parameter as an argument.

In this case, we introduced the covariate time (t - number of days since January 1st) as a way to model the linear trend that appears to be present in the data.

We considered three seasonal variables: day of the week, month of the year and whether a given day is a holiday or not. These variables were created as dummy (zero–one) variables, where one represents the occurrence of the event. As holidays we considered: New Year's Day, Carnival Tuesday, Good Friday, Easter Sunday, Freedom Day, Labour Day, Corpus Christi, Portugal Day, Assumption Day, Republic Day, All Saints Day, Restoration of Independence, Immaculate Conception Day and Christmas Day.

As meteorological variables, we used the daily maximum temperature in degrees Celsius (data from Instituto Dom Luiz - FCUL [37]). In terms of epidemiological factors, we took into account the number of flu cases in the previous week (data from Administração Regional de Saúde do Norte). We assumed that both data sets were time series with a weekly pattern.

Notice that including meteorological or epidemiological variables in the final model will imply that, when forecasting, this information has to be forecasted *a priori*. In practice, this can result in higher inaccuracies in predictions. For that reason, models with or without these covariates were fitted and compared.

Note: An alternative to using seasonal variables is to use Fourier terms. This is especially useful for long seasonal periods. Usually, fewer predictors are needed with Fourier terms versus categorical variables, especially when the seasonal period is large. The maximum number of Fourier terms allowed is $m/2$, where m denotes the frequency of the seasonal pattern - [21]. The K number of Fourier terms to be included in the regression equation is the one that minimises the model's AIC value. A time series regression model with Fourier terms is written as

$$y_t = \beta_0 + \sum_{j=1}^{k} \beta_j x_{j,t} + \sum_{k'=1}^{K} \left[\alpha_{k'} \sin\left(\frac{2\pi k' t}{m}\right) + \gamma_{k'} \cos\left(\frac{2\pi k' t}{m}\right) \right] + \epsilon_t, \quad (2)$$

where $\alpha_{k'}$ and $\gamma_{k'}$, $k' = 1, \dots, K$ represent the amplitudes of the sine and cosine waves, respectively.

3.2. Seasonal autoregressive integrated moving average (SARIMA)

The non-seasonal model of order (p, d, q) - ARIMA(p, d, q) - results of combining of an autoregression model of order p (AR(p)), a moving average model of order q (MA(q)) and the differencing of degree d.

As indicated, the ARIMA model is a non-seasonal model. A seasonal component, may however be considered. ARIMA models with a seasonal component are known as Seasonal Autoregressive Integrated Moving Average (SARIMA) models [38].

This technique does not include the use of external variables and is not as easy to interpret and identify as the previous one. However, SARIMA models are capable of modelling a wide range of seasonal data. A SARIMA model is formed by including seasonal terms in the ARIMA models and it is written as

ARIMA $(p, d, q)(P, D, Q)_{[m]}$,

where (p, d, q) account for the non-seasonal part of the model and (P, D, Q) describe the seasonal part of the model. The characters p and P represent the number of non-seasonal and seasonal autoregressive terms; q and Q indicate the number of non-seasonal and seasonal moving average terms; d and D represent the non-seasonal and seasonal differences that must be performed to transform the time series into stationary. The frequency of the seasonal pattern is denoted by m.

We assume $\phi_1, \phi_2, \dots, \phi_k$ as the parameters related to the non-seasonal autoregressive part of the model; $\theta_1, \theta_2, \dots, \theta_k$ as the parameters related to the non-seasonal moving average part of the model; $\Phi_1, \Phi_2, \dots, \Phi_k$ as the parameters related to the seasonal autoregressive part of the model; and $\Theta_1, \Theta_2, \dots, \Theta_k$ as the parameters related to the seasonal moving average part of the model.

To fit the SARIMA models we used the `Arima` function, with the `seasonal` argument, from the `forecast` package in R [30]. SARIMA models allow the introduction of recurrent patterns, like the weekly pattern (i.e. 7 days) we assumed our first-scenario time series to have. However, it does not accommodate for non-integer seasonal patterns, and actually, in R, the `Arima` function only allows seasonal periods up to 350. Estimation of parameters is achieved by maximum likelihood. The errors should have mean zero and be uncorrelated. As with the time series regression, having the errors be normally distributed with a constant variance σ^2 is useful when estimating prediction intervals.

To evaluate the existence of autocorrelation in the residuals, we used the Ljung–Box test, with the null hypothesis being "there is no autocorrelation in the residuals".

3.3. Exponential smoothing

Exponential smoothing is a class of forecasting techniques for smoothing time-series data using the exponential window function [21]. Unlike the moving average models, which weigh past observations equally when they fall within the moving average window, exponential smoothing uses weights that decrease exponentially over time. In other words, the more recent the observation, the higher the associated weight.

According to the characteristics of the time series under study, there are several forms of exponential smoothing methods. We will focus on the exponential smoothing approach proposed by Hyndman et al. [39] - the Error, Trend, Seasonal (ETS) models. The ETS models are a class of time series models with an underlying state space model that comprises a level component, a trend component, a seasonal component and an error component. One of the most significant benefits of this technique is that it is fully automated and simple to implement. However, as with ARIMA models, the inclusion of other relevant data, like external variables, is not allowed.

Each state space model is labelled as ETS$(\cdot,\cdot,\cdot)$ for (Error, Trend, Seasonal). As possibilities for each component we are assuming Error = {A,M}, Trend = {N,A_d,A} and Seasonal = {N,A,M}. Here, the letters A stand for additive, A_d for additive damped, M for multiplicative, and N for none. This allows one to consider several exponential smoothing models.

We considered only models where Seasonal≠N, meaning a seasonal pattern is taken into account. Such models fall under the Holt-Winters' method [40,41]. Also, since we transformed the data, the `ets` function from the `forecast` package in R [30] does not allow the use of multiplicative models. Additionally, since they have been proven successful across time-series data analysis, we considered models with an additive damped trend [42]. Therefore, in this work, we only took into consideration the following models

- ETS(A,A,A)

$$
\begin{aligned}
y_t &= \ell_{t-1} + b_{t-1} + s_{t-m} + \epsilon_t \\
\ell_t &= \ell_{t-1} + b_{t-1} + \alpha\epsilon_t \\
b_t &= b_{t-1} + \beta\epsilon_t \\
s_t &= s_{t-m} + \gamma\epsilon_t
\end{aligned} \tag{3}
$$

- ETS(A, A$_d$, A)

$$
\begin{aligned}
y_t &= \ell_{t-1} + \phi b_{t-1} + s_{t-m} + \epsilon_t \\
\ell_t &= \ell_{t-1} + \phi b_{t-1} + \alpha\epsilon_t \\
b_t &= \phi b_{t-1} + \beta\epsilon_t \\
s_t &= s_{t-m} + \gamma\epsilon_t
\end{aligned} \tag{4}
$$

- ETS(A, N, A)

$$
\begin{aligned}
y_t &= \ell_{t-1} + s_{t-m} + \epsilon_t \\
\ell_t &= \ell_{t-1} + \alpha\epsilon_t \\
s_t &= s_{t-m} + \gamma\epsilon_t
\end{aligned} \tag{5}
$$

Each model comprises an equation that describes the observed data (y_t), and some state equations that describe how the level (ℓ_t), trend (b_t) and seasonal (s_t) components change over time t. Each model is also described by a set of parameters α (smoothing parameter for the level of the series), β (smoothing parameter for the trend component), γ (smoothing parameter for the seasonal component), ϕ (damping parameter) and σ^2 (variance of the error component, that should be a white noise process).

The state space equations for each existing ETS model are fully disclosed in [39]. As hinted above, to fit the ETS model we used the `ets` function from the `forecast` package in R [30].

When fitting a model, both the parameters (α, β, γ and ϕ), and the initial states (ℓ_0, b_0, s_0 and $s_{-1}, \dots, s_{-m+1}$, (m represents the seasonal pattern considered) can be specified. However, if not specified, their estimation is achieved by maximum likelihood. Note that, regarding the initial seasonal states alone, there are $m-1$ parameters to estimate. Hence, for a larger seasonal pattern, estimation becomes almost impossible. Actually, in R, the seasonal frequency is limited to 24.

When considering an additive model, the errors are assumed as uncorrelated, with a normal distribution with mean zero and constant variance σ^2. To evaluate the existence of autocorrelation in the residuals, we used the Ljung–Box test, with the null hypothesis being “there is no autocorrelation in the residuals”.

3.4. Linear regression with ARIMA errors

Linear regression with ARIMA errors combines two of the most used statistical techniques (Linear Regression and ARIMA) into a single model for forecasting time-series data.

The combination of both ARIMA and linear regression methods allows the inclusion of both time-subject aspects and other pertinent information (e.g. the effect of external variables) into one modelling method [21].

Hence, linear regression with ARIMA errors should be considered when introducing both covariates and serial correlation.

A linear regression with ARIMA errors is defined by the linear regression Eq. (1) (see Section 3.1). In this equation, the assumed uncorrelated error term, ϵ_t, is substituted for an error term, δ_t, that can be autocorrelated. This autocorrelated error term is assumed as an ARIMA process. Hence, the linear regression with ARIMA errors model is given by

$$y_t = \beta_0 + \beta_1 x_{1,t} + \dots + \beta_p x_{p,t} + \delta_t \tag{6}$$

$$\delta_t = \phi_1\delta_{t-1} + \dots + \phi_p\delta_{t-p} + \epsilon_t - \theta_1\epsilon_{t-1} - \dots - \theta_q\epsilon_{t-q}, \tag{7}$$

where $\beta_0, \beta_1, \dots, \beta_p$, $\phi_1, \phi_2, \dots, \phi_p$ and $\theta_1, \theta_2, \dots, \theta_q$ denote the regression parameters, the non-seasonal autoregression parameters and the non-seasonal moving average parameters, respectively. $\delta_t \sim ARIMA(p,d,q)$ and $\epsilon_t \sim N(0,\sigma^2)$.

The parameters of linear regression with ARIMA errors are estimated simultaneously, usually by the least squares method, as with time series regression. However, this time the sum of squared ϵ_t is the one to be minimised and not the sum of squared δ_t. Maximum likelihood can also be used to estimate the parameters.

Notice that if any seasonal difference is taken into account, all variables in the regression model will also be differentiated before the model is estimated.

As with time series regression: the covariate time (t) was introduced to model the trend that appears to be present in the data; day of the week, month of the year and holidays were introduced to model the seasonal component (Fourier terms are also capable of modelling the seasonal component of the data and were used); models with or without meteorological and epidemiological variables were fitted. See Section 3.1 for further details about the regression component of the model.

In order to fit a linear regression with ARIMA errors in R, one should use the `xreg` argument of the `Arima` or the `auto.arima` function from the `forecast` package [30].

To evaluate the existence of autocorrelation in the residuals, which should behave like a normally distributed white noise process, we used the Ljung–Box test, with the null hypothesis being “there is no autocorrelation in the residuals”.

3.5. Trigonometric exponential smoothing state space model with Box–Cox transformation, ARMA errors, Trend and Seasonal components (TBATS)

A TBATS model is an innovations state space model that allows for multiple non-integer seasonality cycles, contrary to the BATS (acronym for Box–Cox transformation, ARMA errors, Trend, and Seasonal components) model [43]. In reality, the TBATS model is an improvement on both the BATS model and the methodology described in Section 3.3. In this model, T denotes the trigonometric terms (Fourier terms) for seasonality, B denotes the Box–Cox transformation, A denotes the ARMA errors, T denotes the trend term, and S denotes the seasonal periods.

Usually, a TBATS model is designated as $\left(\lambda, \{p,q\}, \phi, \left\{\langle m_1, k_1\rangle, \dots, \langle m_T, k_T\rangle\right\}\right)$ and can be described by the following equations

- Box–Cox transformation

$$y_t^{(\lambda)} = \begin{cases} \frac{y_t^{\lambda}-1}{\lambda}, & if \quad \lambda \neq 0 \\ log(y_t), & if \quad \lambda = 0 \end{cases} \tag{8}$$

- Seasonal periods

$$y_t^{(\lambda)} = l_{t-1} + \phi b_{t-1} + \sum_{i=1}^{T} s_{t-m_i}^{(i)} + d_t \tag{9}$$

- Trend

$$
\begin{aligned}
l_t &= l_{t-1} + \phi b_{t-1} + \alpha d_t \\
b_t &= (1-\phi)b + \phi b_{t-1} + \beta d_t
\end{aligned} \tag{10}
$$

- ARMA error

$$d_t = \sum_{i=1}^{p} \phi_i d_{t-i} + \sum_{j=1}^{q} \theta_j \epsilon_{t-j} + \epsilon_t \tag{11}$$

- Trigonometric terms (Fourier terms)

$$
\begin{aligned}
s_t^{(i)} &= \sum_{j=1}^{k_i} s_{j,t}^{(i)} \\
s_{j,t}^{(i)} &= s_{j,t-1}^{(i)} \cos \upsilon_j^{(i)} + s_{j,t-1}^{*(i)} \sin \upsilon_j^{(i)} + \gamma_1^{(i)} d_t \\
s_{j,t}^{*(i)} &= -s_{j,t-1}^{(i)} \sin \upsilon_j^{(i)} + s_{j,t-1}^{*(i)} \cos \upsilon_j^{(i)} + \gamma_2^{(i)} d_t
\end{aligned} \tag{12}
$$

Table 1
Description of each time series regression model and respective AIC value.

Model	Used covariates	AIC
Model 1	Time, day, month, holiday, maximum temperature and number of flu cases	−2583.41
Model 2	Time, day, month, holiday and maximum temperature	−2578.37
Model 3	Time, day, month, holiday and number of flu cases	−2577.83
Model 4	Time, day, month and holiday	−2571.68

where $y_t^{(\lambda)}$ denotes the Box–Cox transformation with parameter λ performed on the time series; l_t and b_t describe how the level and short-run trend change over time; b denotes the long-run trend; ϕ represents the damping parameter; $s_t^{(i)}$ denotes the ith seasonal component at time t; $m_1, \ldots, m_T$ are the list of the seasonal periods and $k_1, \ldots, k_T$ are the corresponding number of Fourier terms used for each seasonality; d_t denotes an ARMA(p, q) process to model the errors; and ϵ_t is a Gaussian white-noise process with zero mean and variance σ^2; α, β and γ_1, γ_2 represent the smoothing parameters and $\upsilon_j^{(i)} = 2\pi j/m_i$, $i = 1, \ldots, T$; $s_{j,t}^{*(i)}$ represent the change of the ith seasonal component over time.

As in the case of ETS models, this model is fully automated and simple to implement. Also, compared to other methods, such as linear regression with ARIMA errors, the TBATS approach allows the seasonality to change slowly over time. However, as with ETS and ARIMA models, the inclusion of other relevant data, like external variables, is not allowed. In contrast with the ETS and ARIMA models, TBATS models have the ability to handle multiple seasonal patterns in addition to non-integer frequencies such as the 365.25 periodicity we utilised to account for the annual seasonal pattern [44].

To fit the TBATS model we used the `tbats` function from the `forecast` package in R [30]. Again, to evaluate the existence of autocorrelation in the residuals, which should also behave like a white noise process, we used the Ljung–Box test, with the null hypothesis being "there is no autocorrelation in the residuals".

4. Results

Let us consider the first three years (2013, 2014 and 2015) of data on the admissions to the Hospital of Braga's ED as our training dataset and data from 2016 as our test dataset.

We started by modelling the training dataset using the various methods discussed in Section 3, considering a single seasonal pattern (weekly) and multiple seasonal patterns (weekly and annual) separately. Then, to assess the models' goodness of fit, we compared the AIC values of the models considered for each modelling technique. Having chosen a maximum of two models (with the lowest AIC values) within each method, we compared the models between methods (single seasonal pattern vs multiple seasonal patterns) using the RMSE, MAE and MAPE values of both training and test datasets.

It should be noted that the AIC can be helpful for selecting between models within the same methodology. However, it should not be used to compare models that utilise different methods since the likelihood is often computed differently [21].

4.1. Single seasonal pattern

4.1.1. Time series regression

The models used in this subsection are described in Table 1. In addition, the AIC values for each model are also listed in Table 1.

As described in Section 3.1, we introduced the variable time (t) to model the trend present in the data. On the other hand, to model seasonality, we introduced as covariates the day of the week, the month of the year, and whether a given day is a holiday or not. Also, the maximum daily temperature and the number of flu cases the week before were introduced in the time series regression model. All four models present all covariates as statistically significant. Also, all models reveal statistical evidence of existing autocorrelation in the residuals ($p < 0.0001$ for the Breusch–Godfrey test).

Model 1 presents the lower value of AIC (−2583.41). Therefore it was the chosen time series regression model assuming a single seasonal pattern to forecast into 2016.

Notice that, if using Model 1 to forecast future admission numbers, the hospital administration needs to have information on both meteorological and epidemiological variables. These would need to be forecasted, which leads to more error. However, regarding the maximum daily temperature, the hospital could use the information on the weather forecast given by the Portuguese Institute of the Sea and the Atmosphere (Instituto Português do Mar e da Atmosfera in Portuguese). As for the number of flu cases, the information is not so easy to obtain or predict. We forecasted the flu-related data using the `auto.arima` for this model. Hence, in case the hospital's administration decides against using the number of flu cases or the maximum temperature, we also considered Model 4 (AIC = −2571.68) to forecast into the future.

Table 2
Description of each SARIMA model and respective AIC value.

Model	SARIMA terms	AIC
Model 5	ARIMA(2,0,2)(0,1,2) $_{[7]}$	−2535.39
Model 6	ARIMA(2,0,2)(1,1,1) $_{[7]}$	−2533.73
Model 7	ARIMA(2,0,2)(1,1,2) $_{[7]}$	−2536.14
Model 8	ARIMA(2,0,1)(0,1,2) $_{[7]}$	−2537.26
Model 9	ARIMA(3,0,2)(0,1,2) $_{[7]}$	−2538.11
Model 10	ARIMA(2,0,3)(0,1,2) $_{[7]}$	−2540.05

Table 3
Description of each ETS model and respective AIC value.

Model	ETS components	AIC
Model 11[a]	ETS(A,N,A)	2045.54
Model 12[b]	ETS(A,A$_d$,A)	2049.90
Model 13	ETS(A,A,A)	2052.64

[a]Using the automatic selecting argument of the `ets` function, with a non-damped trend.
[b]Using the automatic selecting argument of the `ets` function, with a damped trend.

4.1.2. SARIMA models

We started by modelling the trend present in the data with a seasonal difference at lag 7. It was not necessary to differentiate again ($p = 0.1$ for the KPSS-test).

After that, we analysed the ACF and PACF plots to identify the orders of seasonal and non-seasonal autoregressive and moving average terms. Also, the `auto.arima` function of the package `forecast` in R was implemented. However, the model obtained with `auto.arima` failed the Ljung–Box test ($p = 1.25\text{e}{-11}$, there is statistical evidence that the residuals are correlated), and for that, it was disregarded.

The models taken into consideration are listed in Table 2. The AIC values for each model are also included in Table 2.

All models presented $p > 0.05$ for the Ljung–Box test, which means there is statistical evidence that the residuals are not correlated. The model with the lowest AIC value (−2540.05) - Model 10 (ARIMA(2,0,3)(0,1,2) $_{[7]}$) - was taken into consideration. This model also accounted for significant parameters overall (the first moving average parameter did not reveal statistical significance), while most of the remaining models reveal no statistical significance for the second parameter of the moving average process.

4.1.3. Exponential smoothing

As was already mentioned in Section 3.3, the ETS models are entirely automated. Therefore we started by using the automatic selecting argument of the `ets` function (package `forecast`) in R, with both damped and non-damped trend component. We specified a third model (Model 13) for comparison. The used models are described in Table 3, alongside their respective AIC values.

Table 4
Description of each linear regression with ARIMA errors model and respective AIC value.

Model	Used covariates	ARIMA terms	AIC
Model 14	Time, day, month, holiday, maximum temperature and number of flu cases	ARIMA(2,0,0)	−2638.21
Model 15	Time, day, month, holiday and maximum temperature	ARIMA(2,0,1)	−2638.62
Model 16	Time, day, month, holiday and number of flu cases	ARIMA(2,0,0)	−2634.77
Model 17	Time, day, month and holiday	ARIMA(2,0,1)	−2633.88

Table 5
Description of each linear regression model with $K = 62$ Fourier terms and respective AIC value.

Model	Used covariates	AIC
Model 18	Time, day, month, holiday, maximum temperature and number of flu cases	−2639.23
Model 19	Time, day, month, holiday and maximum temperature	−2639.51
Model 20	Time, day, month, holiday and number of flu cases	−2632.03
Model 21	Time, day, month and holiday	−2631.51

All models reveal statistical evidence of existing autocorrelation in the residuals ($p < 0.0001$ for the Ljung–Box test). In order to compare modelling approaches and then forecast, we selected Model 11 since it presented the smaller AIC value (2045.54).

4.1.4. Linear regression with ARIMA errors

As in time series regression, here we introduce the time variable (t) to model the trend present in the data, and the day of the week, the month of the year and whether a given day is a holiday or not as covariates to model seasonality. We also introduced the daily maximum temperature and the number of influenza cases in the previous week. Having the regression part of the model, we used the `auto.arima` function to select the ARIMA components of the error term.

The models considered are described in Table 4. In addition, the AIC values for each model are also listed in Table 4.

All four models reveal statistical evidence of existing autocorrelation in the residuals ($p < 0.0001$ for the Ljung–Box test).

Model 15 presents the lower value of AIC (−2638.62). Therefore it was the chosen model assuming a single seasonal pattern to forecast into 2016.

As in time series regression, if using Model 15 to forecast future admissions numbers, the hospital administration needs information on the meteorological variables. Again, the hospital could use the information on the weather forecast given by the Portuguese Institute of the Sea and the Atmosphere (Instituto Português do Mar e da Atmosfera in Portuguese). We also considered Model 17 (AIC = −2633.88) to forecast the future in case the hospital decides against using the maximum temperature as a covariate.

4.2. Multiple seasonal patterns

4.2.1. Time series regression

Considering the time series with a weekly seasonal pattern, we included the longer annual seasonality by introducing Fourier terms in the regression model. We selected the number of Fourier terms by minimising the AIC.

Hence, the models considered, with $K = 62$ Fourier terms, are described in Table 5. Also, the AIC values for each model are listed in Table 5.

The variable time (t) was again introduced to model the trend present in the data. In addition, the covariates day of the week, month of the year and whether a given day is a holiday or not were included to model seasonality. We also introduced as possible covariates the daily maximum temperature and the number of influenza cases in the previous week.

Model 19 presents the lower value of AIC (−2639.51). Therefore it was the chosen model assuming both weekly and annual seasonal patterns to forecast into 2016. Nevertheless, if the hospital administration does not want to use meteorological or epidemiological covariates, we assume model 21 for comparison (AIC = −2631.51). All models reveal statistical evidence of existing autocorrelation in the residuals ($p < 0.02$ for the Breusch–Godfrey test).

4.2.2. Linear regression with ARIMA errors

Again, considering the time series with a weekly seasonal pattern, we included the longer annual seasonality by introducing Fourier terms in the regression model. We selected the number of Fourier terms by minimising the AIC.

Hence, the models considered, with $K = 5$ Fourier terms, are described in Table 6. Also, the AIC values for each model are listed in Table 6.

The variable time was introduced as a way to model the trend present in the data. The covariates day of the week, month of the year and whether a given day is a holiday or not were included to model seasonality. We also introduced as possible covariates the daily maximum temperature and the number of influenza cases in the previous week.

Model 23 presents the lower value of AIC (−2647.52). Therefore it was the chosen model assuming both seasonal patterns to forecast into 2016. But, again, if the hospital administration does not want to use meteorological or epidemiological covariates, we assume Model 25 for comparison (AIC = −2641.05). All models still reveal statistical evidence of existing autocorrelation in the residuals ($p < 0.0001$ for the Ljung–Box test).

4.2.3. TBATS

As mentioned in Section 3.5, the TBATS models are fully automated. Therefore, after setting the seasonal periods as 7 and 365.25 we used the automatic function `tbats` (package `forecast`) in R to select the model. The fitted model, Model 26, is given by TBATS $(0, \{2, 2\}, -, \{\langle 7, 3\rangle, \langle 365.25, 1\rangle\})$, and has an AIC value of 14883.77. Additionally, the model reveals statistical evidence of existing autocorrelation in the residuals ($p < 0.02$ for the Ljung–Box test).

4.3. Single seasonal pattern vs. multiple seasonal patterns

In this subsection, we are comparing the models selected (using the AIC values) in Sections 4.1 and 4.2. In order to do that, we used the RMSE, the MAE and the MAPE values of both training and test datasets. This allows us to understand the models' fitting capacity and to evaluate their forecasting performance.

Table 6
Description of each linear regression with ARIMA errors model, with $K = 5$ Fourier terms and respective AIC values.

Model	Used covariates	ARIMA terms	AIC
Model 22	Time, day, month, holiday, maximum temperature and number of flu cases	ARIMA(1,0,0)	−2645.73
Model 23	Time, day, month, holiday and maximum temperature	ARIMA(1,0,0)	−2647.52
Model 24	Time, day, month, holiday and number of flu cases	ARIMA(1,0,0)	−2639.57
Model 25	Time, day, month and holiday	ARIMA(1,0,0)	−2641.05

Table 7
Description of each model under consideration, along with the training and test data's corresponding RMSE, MAE, and MAPE values.

	Methodology	Model	Dataset	RMSE	MAE	MAPE
Single seasonal pattern (weekly)	Time series regression with seasonal, meteorological and epidemiological variables	Model 1	Training	25.62	19.76	5.62%
			Test	33.56	26.82	6.83%
	Time series regression, with only seasonal variables	Model 4	Training	25.80	19.78	5.63%
			Test	33.04	26.37	6.70%
	SARIMA	Model 10	Training	26.03	19.96	5.67%
			Test	42.15	33.78	8.46%
	Exponential Smoothing	Model 11	Training	26.76	20.54	5.86%
			Test	40.90	32.55	8.16%
	Regression with ARIMA errors, seasonal and meteorological variables	Model 15	Training	24.92	19.17	5.46%
			Test	33.93	27.07	6.86%
	Regression with ARIMA errors, and seasonal variables	Model 17	Training	24.97	19.14	5.46%
			Test	33.14	26.42	6.72%
Multiple seasonal patterns (weekly and annual)	Time series regression with Fourier terms, seasonal and meteorological variables	Model 19	Training	22.45	17.49	4.96%
			Test	34.31	27.77	7.09%
	Time series regression with Fourier terms, and seasonal variables	Model 21	Training	22.53	17.49	4.96%
			Test	33.59	26.99	6.91%
	Regression with ARIMA errors, Fourier terms, seasonal and meteorological variables	Model 23	Training	24.65	19.05	5.42%
			Test	33.95	27.29	6.92%
	Regression with ARIMA errors, Fourier terms, and seasonal variables	Model 25	Training	24.73	19.07	5.43%
			Test	33.04	26.39	6.72%
	TBATS	Model 26	Training	26.35	20.21	5.76%
			Test	37.02	29.53	7.45%

Remember that we considered data on the number of admissions to the ED from 2013, 2014 and 2015 as our training dataset and data from 2016 as our test dataset. The values of interest for each model, given the single or multiple seasonal patterns, can be seen in Table 7.

Concerning the methods used within the single seasonal pattern framework, notice that the linear regression with ARIMA errors models fit the training data a little bit better than the other approaches. On the other hand, the standard time series regression model with only seasonal variables provides slightly more accurate forecasts on the test data. In fact, both standard time series regression and linear regression with ARIMA errors models using only seasonal variables produce more accurate forecasts. In turn, when using the SARIMA and ETS models to fit the training dataset, they both produce better results than the time series regression with seasonal, meteorological and epidemiological variables. However, when it comes to forecasting the test data, the SARIMA and ETS models are the ones with the worst performance.

Also, note that the SARIMA models are the only ones that show uncorrelated residuals. However, they are the ones who produce the least accurate forecasts. Keep in mind that, the existence of correlation in the residuals implies that there is still some information in the data the models could not explain. However, it does not imply one is not able to produce forecasts. Actually, the forecasts from a model with autocorrelated residuals are unbiased; the prediction intervals are the ones that are larger than they would be if the residuals were uncorrelated [21].

As for the techniques within the framework of multiple seasonal patterns, the time series regression models fit the training data better than the other methods, followed by linear regression with ARIMA errors models. Concerning the forecasts over the test data, the time series regression and linear regression with ARIMA errors models using only seasonal variables produce more accurate forecasts. On the other hand, the TBATS model reveals worst predictive power than the other techniques. This is possibly due to the non-inclusion of external variables like whether a day is a holiday or not [45].

In general, incorporating both weekly and annual seasonal patterns is shown to be effective in terms of goodness of fit over the training data. However, that does not always seem to be the case, at least in terms of their ability to forecast the test findings. In fact, in terms of the accuracy of the forecasts, the best performing model is the one using standard time series regression with only seasonal variables (Model 4), closely followed by the models using linear regression with ARIMA errors and only seasonal explanatory variables, both with a single (Model 17) or multiple seasonal patterns (Model 25).

In contrast, the TBATS model (Model 26) with two seasonal patterns and the SARIMA (Model 10) and exponential smoothing (Model 11) models using only one seasonal pattern performed worse when modelling the training data and forecasting the test data.

5. Discussion and conclusions

In terms of the most straightforward exploratory analysis, we found that similar to prior studies [3,5,6,10,17,18], Mondays have the highest number of ED admissions, which decreases throughout the week, with fewer admissions on weekends. This, together with the fact that fewer people are admitted to the ED on holidays, leads to the same conclusion drawn by [3]: people may prefer to postpone their visit to the ED rather than lose their free time. It is also likely that people will become ill as a result of their holiday and weekend indulgences. Also, the number of admissions is higher in summer months than in winter months, contrary to what Batal et al. [3] and Holleman et al. [6] discovered, but in line with what Diehl et al. [5] concluded.

As for the inclusion of meteorological (maximum temperature) and epidemiological (number of flu cases) variables in the several models

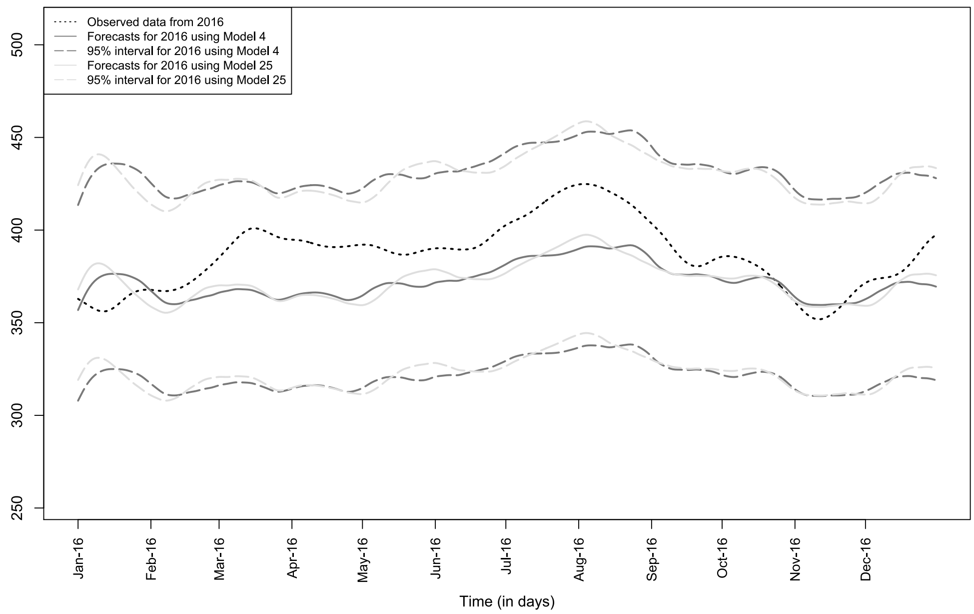

Fig. 4. Forecast of daily admissions to the Emergency Department for the year of 2016 using Model 4 (solid dark grey line) and using Model 25 (solid light grey line) against the real values (black dotted line). The dashed grey lines represent the lower and upper values of the 95% intervals (dashed dark grey line for Model 4 and dashed light grey line for Model 25).

that allow for covariates, the conclusion seems to be in line with that of Batal et al. [3], Holleman et al. [6], Jones et al. [24] and Vieira and Sousa [12] - although the inclusion of such variables in the models translates into a marginally better fit to the training data, it provides less accurate forecasts of the test data. For that reason, and since the actual difference in the fitting of the training data itself is almost non-existing, we would refrain from using models that include meteorological or epidemiological variables.

Although methods taking into account both weekly and annual seasonal patterns reveal better results when fitting the training data, methods taking only a weekly seasonal pattern result in marginally better forecasts. In fact, the most accurate forecasts are given by the simpler model – Model 4 – which utilises standard time series regression based on seasonal variables and a weekly seasonal pattern. In fact, this is consistent with what is found in the literature [3,24]. Model 4 forecasts the daily number of admissions to the ED of the Hospital of Braga with an accuracy of 6.70%. The linear regression with ARIMA errors model with a weekly seasonal pattern (Model 17) forecasts with an accuracy of 6.72%, followed by the linear regression with ARIMA errors model with both weekly and annual seasonal patterns (Model 25), which also has a MAPE of 6.72%.

The non-regression-based techniques revealed the worst performances. Even the most sophisticated method (TBATS), which is entirely automated, considers both seasonalities and allows such seasonalities to fluctuate over time, produced less accurate outcomes (MAPE of 7.45%).

Therefore, in terms of model selection, Model 25 (linear regression with ARIMA errors, seasonal variables, weekly and annual seasonal patterns) is the model we advise the hospital administration to use. This model provides good results in terms of accuracy and goodness of fit, and it allows for the annual seasonality to be taken into account, which is useful in the long run when more years of daily data are to be added.

Alternatively, Model 4 (time series regression with seasonal variables and a weekly seasonal pattern) can also be of consideration. It performs the best in terms of forecast and is easier to use and understand.

Fig. 4 shows the forecasts for daily admissions to the ED of the Hospital of Braga for the year 2016 using Model 4 and using Model 25 against the actual observed values, as well as their respective 95% confidence intervals. Graphically, Model 25 seems to follow the seasonal pattern of the actual data slightly better than Model 4.

By considering and comparing several forecasting techniques that account for single and multiple seasonal patterns, our work fulfils a lacuna in the literature regarding high-frequency recorded daily data containing complex seasonal patterns.

For future work, we would like to test this model on other hospitals in the country since the model is already being implemented in Braga, with good feedback from the administration. Also, it would be interesting to study the number of admissions not daily but hourly and to study the number of admissions by medical area within the ED.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The data that has been used is confidential.

Acknowledgements

The authors would like to thank the administration of the Hospital of Braga for providing the data to support this study. The data are owned by the Hospital of Braga. Restrictions apply to the availability of these data, which were used under a restricted licence for this study. The data and computer code are not publicly available due to privacy or ethical restrictions.

The authors would also like to thank the Administração Regional de Saúde do Norte for providing the data on the number of flu cases. Restrictions apply to the availability of these data.

Funding

This work was supported by the PhD grant 128191/2016 funded by FCT I.P., by the Ministério da Ciência, Tecnologia e Ensino Superior (MCTES), Portugal to the first author.

References

[1] Comissão de Reavaliação da Rede Nacional de Emergência/Urgência, Reavaliação da Rede Nacional de Emergência/Urgência - Relatório, Technical Report, CRRNEU, 2012.

[2] M. Attia, R. Edward, Effect of weather on the number and the nature of visits to a pediatric ED, Am. J. Emerg. Med. 16 (4) (1998) 374–375, http://dx.doi.org/10.1016/s0735-6757(98)90131-0.

[3] H. Batal, J. Tench, S. McMillan, J. Adams, P. Mehler, Predicting patient visits to an urgent care clinic using calendar variables, Acad. Emerg. Med. 8 (1) (2001) 48–53, http://dx.doi.org/10.1111/j.1553-2712.2001.tb00550.x.

[4] J. Diaz, J. Alberdi, M. Pajares, R. Lopez, M. Lage, A. Otero, A model for forecasting emergency hospital admissions: effect of environmental variables, J. Environ. Health 64 (3) (2001) 9–15.

[5] A. Diehl, M. Morris, S. Mannis, Use of calendar and weather data to predict walk-in attendance, South. Med. J. 74 (6) (1981) 709–712, http://dx.doi.org/10.1097/00007611-198106000-00020.

[6] D. Holleman, R. Bowling, C. Gathy, Predicting daily visits to a walk-in clinic and emergency department using calendar and weather data, J. Gen. Intern. Med. 11 (4) (1996) 237–239, http://dx.doi.org/10.1007/BF02642481.

[7] J. Noble, M. LaMontagne, C. Bellotti, H. Wechsler, Variations in visits to hospital emergency care facilities: ritualistic and meteorological factors affecting supply and demand, Med. Care 9 (5) (1971) 415–427, http://dx.doi.org/10.1097/00005650-197109000-00005.

[8] Y. Zhang, J. Zhang, M. Tao, J. Shu, D. Zhu, Forecasting patient arrivals at emergency department using calendar and meteorological information, Appl. Intell. 52 (2021) 11232—11243, http://dx.doi.org/10.1007/s10489-021-03085-9.

[9] P.H. Chau, K.K.-L. Lau, X.X. Qian, H. Luo, J. Woo, Visits to the accident and emergency department in hot season of a city with subtropical climate: association with heat stress and related meteorological variables, Int. J. Biometeorol. 66 (2022) 1955—1971, http://dx.doi.org/10.1007/s00484-022-02332-z.

[10] R. Murtas, S. Tunesi, A. Andreano, A.G. Russo, Time-series cohort study to forecast emergency department visits in the city of milan and predict high demand: a 2-day warning system, BMJ Open 12 (2022) e056017, http://dx.doi.org/10.1136/bmjopen-2021-056017.

[11] L. Zibners, B. Bonsu, J. Hayes, D. Cohen, Local weather effects on emergency department visits: A time series and regression analysis, Pediatr. Emerg. Care 222 (2) (2006) 104–106, http://dx.doi.org/10.1097/01.pec.0000199561.34475.29.

[12] A. Vieira, I. Sousa, Regressão linear com erros correlacionados na previsão de admissões em urgência hospitalar, in: M. Salgueiro, P. Vicente, T. Calapez, C. Marques, M. Silva (Eds.), Programa e Resumos do XXIII Congresso da Sociedade Portuguesa de Estatística, ISBN: 978-972-8890-41-4, 2017, pp. 248–249.

[13] O. Araz, D. Bentley, R. Muelleman, Using Google Flu Trends data in forecasting infuenza-like-illness related ED visits in Omaha, Nebraska, Am. J. Emerg. Med. 32 (9) (2014) 1016–1023.

[14] C. Fuhrmann, M. Sugg, C. Konrad II, A. Walker, Impact of extreme heat events on emergency department visits in north carolina (2007–2011), J. Community Health 41 (1) (2016) 146–156, http://dx.doi.org/10.1007/s10900-015-0080-7.

[15] H. Lin, Y. Long, Y. Su, K. Song, C. Li, N. Ding, Air pollution and hospital admissions for critical illness in emergency department: a tertiary-center research in Changsha, China, 2016–2020, Environ. Sci. Pollut. Res. Int. 29 (15) (2022) 21440—21450, http://dx.doi.org/10.1007/s11356-021-17295-8.

[16] P. Franco, C. Gordo, E.M. Costa, A. Lopes, Air pollution and emergency hospital admissions—Evidences from Lisbon Metropolitan Area, Portugal, Appl. Sci. 10 (22) (2020) 7997, http://dx.doi.org/10.3390/app10227997.

[17] M. Wargon, E. Casalino, B. Guidet, From model to forecasting: a multicenter study in emergency departments, Acad. Emerg. Med. 17 (9) (2010) 970–978, http://dx.doi.org/10.1111/j.1553-2712.2010.00847.x.

[18] I. Marcilio, S. Hajat, N. Gouveia, Forecasting daily emergency department visits using calendar variables and ambient temperature readings, Acad. Emerg. Med. 20 (8) (2013) 769–777, http://dx.doi.org/10.1111/acem.12182.

[19] N. Erkamp, D. van Dalen, E. de Vries, Predicting emergency department visits in a large teaching hospital, Int. J. Emerg. Med. 14 (1) (2021) 34, http://dx.doi.org/10.1186/s12245-021-00357-6.

[20] J. Hitzek, A. Fischer-Rosinský, M. Möckel, S. Kuhlmann, A. Slagman, Influence of weekday and seasonal trends on urgency and in-hospital mortality of emergency department patients, Front. Public Health 10 (2022) 711235, http://dx.doi.org/10.3389/fpubh.2022.711235.

[21] R. Hyndman, G. Athanasopoulos, Forecasting: Principles and Practice, second ed., OTexts, Melbourne, Australia, 2018, Accessed on March 27th 2021.

[22] M. Sousa, A.M. Tomé, J. Moreira, Long-term forecasting of hourly retail customer flow on intermittent time series with multiple seasonality, Data Sci. Manag. 5 (3) (2022) 137–148, http://dx.doi.org/10.1016/j.dsm.2022.07.002.

[23] I. Naim, T. Mahara, A. Idrisi, Effective short-term forecasting for daily time series with complex seasonal patterns, Procedia Comput. Sci. 132 (2018) 1832–1841, International Conference on Computational Intelligence and Data Science.

[24] S. Jones, A. Thomas, R. Evans, S. Welch, P. Haug, G. Snow, Forecasting daily patient volumes in the emergency department, Acad. Emerg. Med. 15 (2) (2008) 159–170, http://dx.doi.org/10.1111/j.1553-2712.2007.00032.x.

[25] C.N. Rocha, F. Rodrigues, Forecasting emergency department admissions, J. Intell. Inf. Syst. 56 (3) (2021) 509–528, http://dx.doi.org/10.1007/s10844-021-00638-9.

[26] R.C. Team, R: A Language and Environment for Statistical Computing, R Foundation for Statistical Computing, Vienna, Austria, 2021, URL: https://cran.r-project.org/.

[27] Direção Geral de Saúde, Rede de Referenciação Hospitalar de Emergência/Urgência - Relatório, Technical Report, ISBN: 972-9425-99-X, 2001.

[28] K. Mackway-Jones, J. Marsden, J. Windle, Emergency Triage-Manchester Triage Group, John Wiley & Sons, Chichester West Sussex, UK, ISBN: 978-1-118-29906-7, 2014.

[29] K. Bandara, R.J. Hyndman, C. Bergmeir, MSTL: A seasonal-trend decomposition algorithm for time series with multiple seasonal patterns, Int. J. Oper. Res. (2021) http://dx.doi.org/10.1504/ijor.2022.10048281.

[30] R.J. Hyndman, G. Athanasopoulos, C. Bergmeir, G. Caceres, L. Chhay, M. O'Hara-Wild, F. Petropoulos, S. Razbash, E. Wang, F. Yasmeen, forecast: Forecasting functions for time series and linear models, 2022, URL: https://pkg.robjhyndman.com/forecast/, R package version 8.18.

[31] G. Box, D. Cox, An analysis of tranformations, J. R. Stat. Soc. Ser. B Stat. Methodol. 26 (2) (1964) 211–252, http://dx.doi.org/10.1111/j.2517-6161.1964.tb00553.x.

[32] D. Kwiatkowski, P.C.B. Phillips, P. Schmidt, Y. Shin, Testing the null hypothesis of stationarity against the alternative of a unit root, J. Econometrics 54 (1–3) (1992) 159—178, http://dx.doi.org/10.1016/0304-4076(92)90104-Y.

[33] H. Akaike, Likelihood of a model and information criteria, J. Econometrics 16 (1) (1981) 3–14, http://dx.doi.org/10.1016/0304-4076(81)90071-3.

[34] C. Coleman, D. Swanson, On MAPE-R as a measure of cross-sectional estimation and forecast accuracy, J. Econ. Soc. Meas. 32 (4) (2007) 219–233, https://escholarship.org/uc/item/73s8831w.

[35] A.G. Barnston, Correspondence among the correlation, RMSE, and heidke forecast verification measures; refinement of the heidke score, Weather Forecast. 7 (4) (1992) 699–709, http://dx.doi.org/10.1175/1520-0434(1992)007<0699:CATCRA>2.0.CO;2.

[36] C. Sammut, G.I. Webb (Eds.), Mean absolute error, in: Encyclopedia of Machine Learning, Springer US, Boston, MA, 2010, p. 652, http://dx.doi.org/10.1007/978-0-387-30164-8_525, Online ISBN 978-0-387-30164-8.

[37] Instituto Dom Luiz - FCUL, Observações meteorológicas de superfície - estação clássica, Last update of data: 2020-03-11, Available at: http://dados-met-idl.campus.ciencias.ulisboa.pt.

[38] P. Brockwell, R. Davis, Introduction to Time Series and Forecasting, second ed., Springer-Verlag, New York USA, ISBN: 0-387-95351-5, 2002.

[39] R. Hyndman, A. Koehler, R. Snyder, S. Grose, A state space framework for automatic forecasting using exponential smoothing methods, Int. J. Forecast. 18 (3) (2002) 439–454, http://dx.doi.org/10.1016/S0169-2070(01)00110-8.

[40] C. Holt, Forecasting Seasonals and Trends by Exponentially Weighted Moving Averages *(O.N.R. Memorandum No. 52)*, Carnegie Institute of Technology, Pittsburgh USA, 1957, http://dx.doi.org/10.1016/j.ijforecast.2003.09.015.

[41] P. Winters, Forecasting sales by exponentially weighted moving averages, Manage. Sci. 6 (3) (1960) 324–342, http://dx.doi.org/10.1287/mnsc.6.3.324.

[42] E. Gardner, E. McKenzie, Forecasting trends in time series, Manage. Sci. 31 (10) (1985) 1237–1246, http://dx.doi.org/10.1287/mnsc.31.10.1237.

[43] A. De Livera, R. Hyndman, R. Snyder, Forecasting time series with complex seasonal patterns using exponential smoothing, J. Amer. Statist. Assoc. 106 (496) (2011) 1513–1527, http://dx.doi.org/10.1198/jasa.2011.tm09771.

[44] Á. Cuevas, R. Ledo, E. Quilis, Seasonal adjustment of the spanish sales daily data, SERIEs 12 (2021) 687—708, http://dx.doi.org/10.1007/s13209-021-00251-7.

[45] R. Hyndman, Forecasting with daily data, 2013, https://robjhyndman.com/hyndsight/dailydata/. Accessed: 2021-12-01.