Forecasting of emergency department attendances in a tourist region with an operational time horizon

B. Lopez, F. Torrent-Fontbona
University of Girona

J. Roman , J.M. Inoriza
Serveis de Salut Integrats Baix Empordà. Hospital of Palamós

April 21st, 2021

Abstract

Hospital patient waiting times and length of stay are indicators of the quality of emergency department (ED) services, factors that are affected by the number of patient arrivals. It is necessary to accurately estimate ED patient arrivals in order to manage resources effectively. Prediction models, however, are conditioned by the hospital population and its placement (i.e. meteorological conditions). In the particular case of a tourist region, the population has an important amount of variability which challenge EDs. This paper aims to address ED attendances predictions for an hospital placed in a tourist region by means of a new approach that combines multiple linear regression with artificial neural networks and regression tree ensembles, looking for dealing for ED variability and prediction for a week time horizon that enables operational reaction to the ED responsible. The methodology uses exogenous variables such as calendar, weather and socio-economic data to improve the accuracy of these forecasts. Prediction models are built on data for 11-years and the predictions are tested over 1-year. The results showed that the proposed methodology is capable to perform weekly predictions with an error about 5%, demonstrating that it could be used by EDs.

1 Introduction

Emergency department (ED) overcrowding is a well-known worldwide occurrence. Estimates from the Spanish Ministry of Health in 2013 show that for every $1,000$ people there are 562 emergencies in the entire Spanish national health system and of the 5.1 million hospital admissions 57% come from ED [22]. The consequences are long waiting times for patients, reduced service quality and the possibility of increases in mortality rates [32, 33, 37].

In order to improve the service, it is essential to accurately forecast the number of patients that will visit the ED in order to have the required resources. Thus, a better match between supply (ED resources) and demand (patient arrivals) can enhance ED service quality, which in turn, will improve the patient experience and staff morale [42].

One of the causes of this saturation is the discontinuous flow of patients to ED, which is determined by factors related to patients (pathologies, preferences) and other external causes such as working hours, holidays, seasons, weather, pollution or unexpected situations such as natural disasters, accidents, etc. [4, 39]. Therefore, it is reasonable to think that the flow of patients in the ED can be estimated by considering some of these factors. However, prediction models highly depend on population features, which in a tourist region is highly variant. In particular, our research is concerned with a tourist region with a high important amount of ED variability (about 20% of the average).

On the other hand, forecasting should be provided in an operational time horizon so that the ED responsible could set up adequate strategies and resources to deal with the ED demand. Most of the current approaches provide a one-day prediction that has no room for planning [41, 11], while some other ones provide longer forecasting approaches [8] that could eventually imply a too broad planning with an ineffective use of resources. Weekly forecasting seems to be an adequate prediction time horizon to react to the ED demand and review scheduling and resources.

There are several methods available to build models for ED forecasting with successful results: Multiple Linear Regression (MLR) [24], Artificial Neural Networks (ANN) [27], and regression tree ensembles [6, 10]. Recently, hybrid approaches have been raised as an alternative approach [44], and our work is aligned with them.

In this paper, we propose a new method, called TENACE (Tourist EmergeNcy AttendanCes Exit), to deal with ED predictions in a tourist area, that exhibits a robust behaviour when predictions are performed one week ahead. The methodology considers different exogenous variables as causes of ED attendances such as calendar, weather, and socio-economic data, which are of particular interest to our ED problem, at the Hospital of Palamós, which is located in a tourist area. We used 12 years of data from the ED hospital records to built, test and analyse the prediction models.

2 Related work

The literature presents various works on forecasting ED attendances. Usually, these works propose the use of linear methods based on past data about the attendance time-series. For example, two interesting examples are those found in the work of [16, 18]. The authors explore the use of exponential smoothing, Autoregressive Integrated Moving Average (ARIMA), SARIMA (seasonal ARIMA) and GARCH (generalised autoregressive conditional heteroskedasticity) models and pre-processing techniques.

However, patient arrivals time-series have been used alongside working hours, holidays, seasons, weather, pollution or unexpected situations such as natural disasters, accidents, etc. in order to improve the accuracy of ED attendance forecasts [4, 39]. While past ED attendances could represent the population variability, this correlation could determine an unexpected ED attendances variation. For example, [5] proposes to classify days into nine different labels using calendar information and then build a prediction model for each type of day. The authors in [20, 17, 35] opt for including various calendar, weather and pollution information that feed the proposed ARIMA and SARIMA models.

Despite the possible impact on people's health due to weather and pollution, these variables have to be forecast in order to use them to predict ED attendances. The authors in [20, 41] conclude that including weather data does not significantly increase the accuracy of the predictions, but some other authors do [8], thus the place of the ED would also affect the exogenous variables to consider. This paper proposes not only using calendar and weather data to predict ED attendances, but also socio-economic data.

Regarding forecasting time horizon, most of the approaches provide one day prediction [41], and some other larger approaches (14 days in [8], monthly forecasts in [2]). The former provides few time for reacting to the demand and organize transition services, if required, and the later are too broad to manage organize staff shift. Alternatively, This work provides good results at week forecasting, meaning, in a time scale suitable for ED resource scheduling.

While the previous approaches follow linear predictive approaches, the authors in [43] explore the use of ANNs to predict daily ED attendances and compare the performance of ANNs with MLR and nonlinear least square regression. In doing so, the authors use calendar and weather information besides past values of the target time-series. They also pay special attention to analysing the data in order to identify the most relevant variables for ANNs, and in the end achieve the best results with such technique. The presented paper also considers ANNs, but also tree ensembles. Tree ensembles were used in [28] to deal with non-linearity as in this work. However, this paper proposes a new hybrid method that catches linear and non-linear relationships between explanatory variables and ED attendances. Other previous hybrid approaches [44] combine ANN with ARIMA models; while in our approach we combine ANN with regression tree ensembles.

3 Materials and methods

This paper proposes a new method, TENACE, to build ED prediction models based on ED attendances and exogenous variables with calendar, weather and socio-economic information.

3.1 Problem formalisation

Given a set of examples about ED attendance, the learning problem consists of finding models for forecasting attendance. An example of ED forecasting for time t consists of a set of input variables $\vec{X}(t) = [x_1(t), \ldots, x_N(t)]$ labelled with the corresponding output or target variable $y(t)$ (ED attendances): $\langle \vec{X}(t), y(t) \rangle$. For the purpose of this work, time is measured in weeks; however, other time scales (day, month) could be used).

The examples are organised in time series $\langle \vec{X}(t_0), y(t_o) \rangle, \ldots, \langle \vec{X}(t_L), y(t_L) \rangle$ starting at time t_0 and ending at t_L. The variables $x_i(t)$ $i = [1, \ldots, N]$ represent exogenous variables which in this study consist of calendar, socio-economic and weather variables (see Table 1). Calendar data are nominal variables that indicate the month, season, weekday, day of year, if it is a public holiday or not and if it is Easter or a day close to Easter (sometimes they are holidays too). Weather data are numeric variables that provide

Table 1: List of exogenous variables

Calendar data	**Month** $\in [1,12]$: month of the year **Season** $\in [1,4]$: winter, spring, summer, autumn **Weekday** $\in [1,7]$ (only for daily attendances forecast): Monday to Sunday **Holiday** $\in [0,1]$ (only for daily attendances forecast): public holiday or not **Week number** $\in [1,53]$ (only for daily and weekly attendances forecast) **Day of year** $\in [1,366]$ (only for daily attendances forecast) **Easter** $\in [0,9]$: a label for each day from the Sunday before Easter to the Monday after Easter. Label 0 is assigned to other days of the year.
Weather data	**Average, maximum and minimum temperature** of the day, week or month (depending on the aggregation level of the target variable) **Average, maximum and minimum relative humidity** of the day, week or month (depending on the aggregation level of the target variable) **24h solar radiation**: for week and month aggregation levels, it is the average of the daily (24h) radiation of the days of the corresponding week or month.
Socio-economic data	**Unemployment rate** of Catalonia. Annual value. **Population** of the region. Annual value. **Seasonal population** of the regionof the hospital on July 1st and December 31st. **Gross domestic product per capita** of Catalonia every year.

information about the temperature, humidity and solar radiation amount. Socio-economic data are numeric variables about the (annual) population and seasonal (every 3 months) population of the region and the unemployment rate and gross domestic product per capita of the country. The selected exogenous variables have been chosen after analysing the correlation between them and the target variables (ED attendance) and also empirically analysing the benefits of considering them using the different methods studied: MLR, ANN, and tree ensembles.

As ED attendances are numeric values, the model to be learnt can be represented by a function $f(\cdot)$ that estimates the output variable $y(t)$ for a given time t (ED attendances at the t time, $t > t_L$), so $\hat{y}(t) = f(\cdot)$. The arguments of this function are the current input variables $x_1(t), \ldots x_N(t)$, and the measurements of past ED attendances $y(t - d_j)$, with a delay d_j, $j \in [1, \ldots, M]$ and $d_M < t_L$, which are assumed to be known when predicting $y(t)$. Equation (1) describes the function archetype for estimating ED attendance.

$$\hat{y}\left(t\right) = f\left(x_1\left(t\right), x_2\left(t\right), \ldots, x_N\left(t\right), y\left(t - d_1\right), y\left(t - d_2\right), \ldots, y\left(t - d_M\right)\right) \tag{1}$$

The complexity of $f(\cdot)$ depends on the method addressed to model it. For example, MLR builds a linear function, while ANN, as designed in this work,

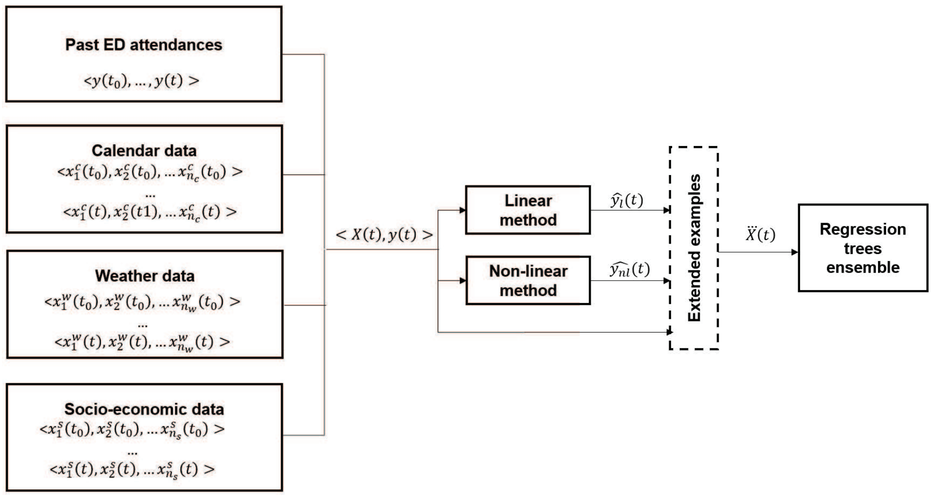

Figure 1: Workflow of the proposed methodology for generating the predictive model. Linear method box represents MLR and non-linear method represents an ANN or a regression tree ensemble.

enables learning nonlinear expressions. On the other hand, the methods are supervised since they use past examples $\langle \vec{X}(t_0), y(t_o)\rangle, \ldots, \langle \vec{X}(t_L), y(t_L)\rangle$ to learn $f(\cdot)$, i.e. trying to reduce the error between $y(t)$ and $\hat{y}(t)$ of the known examples. In doing so, the methods handle the over-fitting problem in different ways, which consists of how to avoid to fit $f(\cdot)$ excessively to the details of the examples and exhibiting a low performance on new instances.

3.2 TENACE method

The TENACE methodology consists of two different *shape* components that capture the linearity and non-linearity relations, plus a regression tree ensemble learner as Figure 1 depicts. The input of the method, $\vec{X}(t)$, consists of past ED attendances and the exogenous variables until time t (see Table 1). Next, the linear and non-linear estimations are computed, $\hat{y}_l(t), \hat{y}_{nl}(t)$ correspondingly. Thus, extended examples $\dddot{\vec{X}}(t)$ are generated, from $\langle \vec{X}(t), y(t)\rangle$ to $\langle \vec{X}(t), y(t), \hat{y}_l(t), \hat{y}_{nl}(t)\rangle$. Finally, a tree ensemble method is used to build a prediction model with the extended examples.

The two *shape* components are separately generated by using a linear and a non-linear machine learning methods, which are trained with the original

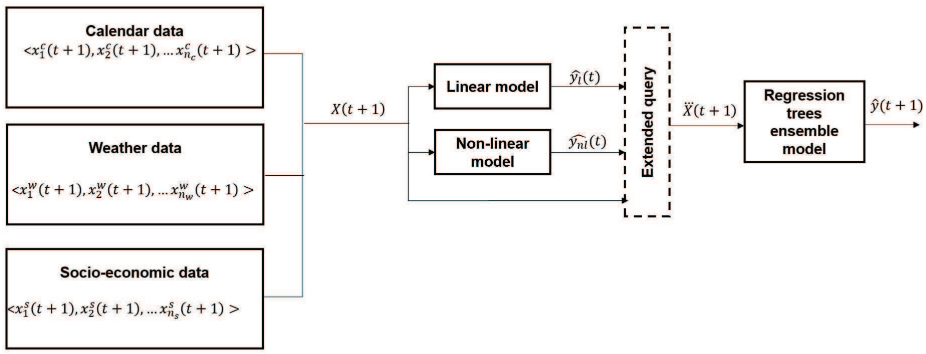

Figure 2: Workflow of the proposed methodology for obtaining a prediction.

examples $\langle \vec{X}(t), y(t)\rangle$. In particular, we used two combinations: the methodology combining MLR and ANN (labelled as $TENACE.MLR+ANN$) and MLR and tree ensembles ($TENACE.MLR+Boosting$). In this latter case, regression tree ensembles are used twice: once for the non-linear component, and second, to build the ensemble model from the extended examples.

The intuition behind the hybridisation is that the ensemble method identifies as a key contribution $\hat{y}_l(t)$ if the data available fits a linear function, and $\hat{y}_{nl}(t)$ otherwise. Moreover, it supports the idea that heterogeneous predictors combine better than homogeneous ones [12]. Moreover, hybrid approaches aimed at combine the major benefits of the techniques: MLR is a popular method but it does not handle complex variable interactions, and it is difficult to interpret regarding the too many predictor variables that compose the model; ANN can manage complex interactions while they do not provide explanations of the models learnt; regression trees provide clinical decision making rules easy to interpret, and they are easy to use but they could also hide some complexity when dealing with ensembles. Therefore, combining these techniques may improve the performance of a sole technique.

Once the models are obtained, the methodology is used to predict next ED attendances, as shown in Figure 2. ED predictions are performed for the next week $\hat{y}_l(t+1)$, but a forecast horizon of several weeks ahead can also be obtained (i.e. $\hat{y}_l(t+1), ...\hat{y}_l(t+h)$, where h is the forecast horizon). The combination of linear and non-linear models in TENACE is expected to be a key to make robust predictions along time in a hospital region with high variability.

3.2.1 MLR

MLR [24] is an often used statistical model in medicine and healthcare to assess the relationship between a number of variables [14]. MLR has achieved good performances in many fields, such as the prediction of the arrivals and departures times of aircraft taxis [31], forecasting photovoltaic energy production [45] or even ED attendances [43].

MLR assumes that function f of Equation (1) is a linear combination of the explanatory variables as it describes Equation (2) where α_i are the regression parameters of input variables and β_i the auto-regression coefficients

$$\hat{y}(t) = \alpha_0 + \sum_{i=1}^{N} \alpha_i x_i(t) + \sum_{j=1}^{M} \beta_j y\left(t - d_j\right) \tag{2}$$

The constructed MLR model may be used to predict future values of the target variable. Furthermore, MLR models give the contribution (α_i and β_j) of each explanatory variable to describe the target variable. Thus, MLR models may be also used to analyse and comprehend which factors have a significant impact.

In order to apply MLR to the problem at hand, it is necessary to tackle the MLR assumption that explanatory variables are numeric variables. Nevertheless, calendar variables (e.g. weekday) are nominal. Then, for each nominal variable $x_i(t)$, P_i dummy variables are created, $x_i^1(t), \ldots x_i^{P_i}(t)$ being P_i the number of values of the corresponding nominal variable. These dummy variables (e.g. weekday_monday) take the value of 1 if the nominal value takes the corresponding value (e.g. Monday) and 0 if it does not. To handle the over-fitting problem, cross-validation of the 11-year data has been used to train the model.

The requirements of having numeric variables could not be always an straightforward process in the medical field. Moreover, the resulting model could be difficult to interpret for the clinical staff due to the high amount of variables involved. On the other hand, MLR models cannot catch complex variable interactions (i.e. nonlinear relationships between variables) which may sometimes be significant. Therefore, ANN methods have been proposed as alternative techniques to manage non-linearity, and regression trees due to their facility to derive decision making rules.

3.2.2 ANN

ANNs [27] are inspired by the way the nervous system process information. An ANN is composed of a large number of highly interconnected neurones which are trained to learn by example. While learning in biological systems involves adjustments to the synaptic connections that exist between the neurones, the ANN methods learn by adjusting the connections between neurones. Nowadays, ANNs are widely used for medical applications in various disciplines, mainly managing signals and images: electronic signal analysis, medical image analysis and radiology [26]. ANNs are capable of modelling nonlinear relations between variables, and they are especially useful for forecasting problems as in our case.

ANNs use a topology of nodes, each representing a neuron, to perform the task it has been trained for, as for example, predicting ED attendance. There are different kinds of topologies, but the most used one is the feed forward organisation in which there are several layers of interconnected nodes and, at each layer, every node receives the output variables from the previous layer. Given a stimulus (input variables), the first layer is activated, and the stimulus is propagated through the net until the last layer provides the predicted value. ANNs are considered black-boxes because the presence of hidden layers makes it difficult to interpret the obtained results. However, they achieve good results in complex domains [15, 36, 38].

For ED attendances, Figure 3 shows the topology designed with 10 hidden layers (set empirically), and $N + M$ neurons per layer. The log-sigmoid function has been used as a transfer function for the neurons of all the layers except the last one, for which it is proposed to use a linear transfer function. Cross-validation is also used to avoid the problem of over-fitting.

There are two important weaknesses of ANN. First, the results strongly depend on the topology defined, which is empirically found, and therefore they are not easy to use by non skilled personnel. And second, ANN lack of explanatory capacity, and explanatory models are needed in Medicine in order to design the adequate interventions to manage the prediction outcomes [13].

3.2.3 Regression tree ensembles

A regression tree ensemble [10] consists of a set of regression trees that are used jointly to predict the values of a target variable. A regression tree

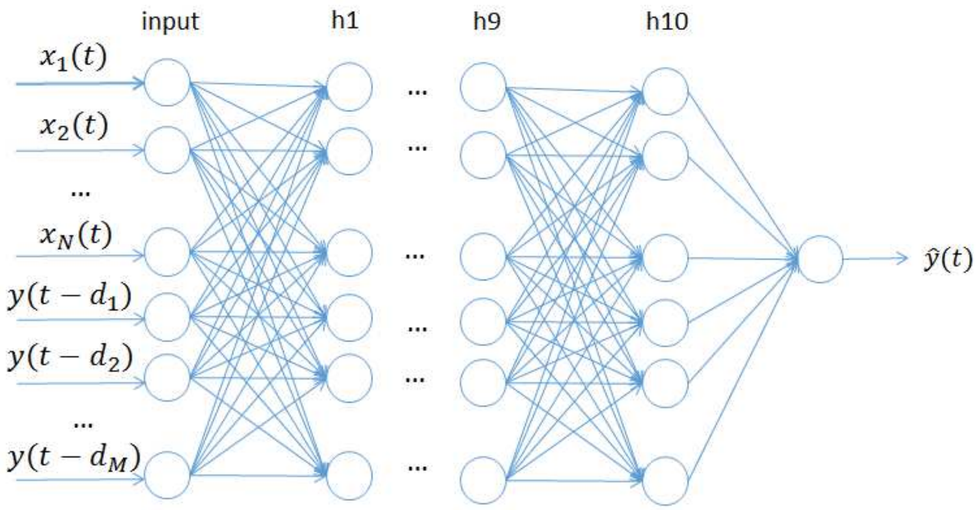

Figure 3: Flow of the artificial neural network designed.

binary recursive partitions the data into two groups according to a certain criteria. The method to build a regression tree is non-parametric, what makes it easy to implement, respect to MLR models which are at times difficult to implement, especially for non-statisticians. Regression trees are also easy to use and interpret. That is, regression trees enable the creation of clinical decision making rules from variables to the predicted outcome according to the criteria selected to split the dataset [19].

Another advantage of regression trees, and unlike MLR and ANNs, is that they are not vulnerable to the scale of the inputs. Therefore, regression trees do not require to normalise or scale input data in order to maximise the performance since they divide the output space in regions which group a set of samples from training data. Moreover, they can manage variable interactions, as well as uneven distribution of variables, while MLR methods do not [19]. Regression trees and regression tree ensembles have been used in applications like electricity consumption forecasts [15, 38], photovoltaic energy production forecast [45] or vessel arrivals estimations [25]. In the medical field, they have been tested for forecasting ED attendance with promising results at long term prediction [8].

This paper proposes to use boosting for ensemble regression trees to cope with the over-fitting problem by de-correlating the different regression trees. According to the standard methodology, boosted tree ensembles for ED attendance are obtained according to the following steps: (i) first a regression tree is obtained using the available data, (ii) then another regression tree is generated only with the data that cannot be well-classified with the previous tree, and (iii) step (ii) is repeated for a given number q of trees. This procedure is used to convert a set of weak learners (i.e. regression trees) into a good learner. The final prediction of the ensemble (g in Figure 4) is obtained by a voting mechanism of the individual outcomes of the regression trees (i.e. $\hat{y}_{\sigma_1}(n), \ldots, \hat{y}_{\sigma_q}(n)$). In our particular implementation, we have set a great number of trees ($q = 1000$) that are using 1-split trees.

Regression trees [6] are decision trees whose nodes correspond to input variables which are ordered according to a given heuristic measure σ. The heuristic measure used in this work is the mean squared error. The set of samples are split recursively according to given values $v_{\sigma_i j}$ of the node variables (in our particular case, the exogenous variables). The output of the regression tree is a (weighted) average of the samples of the region (for example 12.0 is the average of $y(t)$ of the examples with $x_{\sigma_1 1}(n) > v_{\sigma_1 1}$ and $x_{\sigma_1 2}(n) > v_{\sigma_1 2}$ in the top regression tree of Figure 4).

The use of regression trees is not free of drawbacks, however. One of the most important concerns about their application is the fact that they cannot predict beyond the range in the training data (for example, predicting a value of 10 if the trained examples have values up to 9). On the other hand, while the generation of decision making rules from regression trees is simple, the interpretation of ensembles of trees is not so straightforward.

3.3 Dataset

The experimentation has been conducted using 12-year (from 2002 to 2013) emergency attendances data from Hospital of Palamós (Catalunya) with the corresponding calendar, weather[1] and socio-economic [2] data (see Table 1).

The hospital is located in a tourist region with fluctuations in the population depending on the season. Figure 5 shows the number of daily, weekly and monthly ED attendances over 12 years. Table 2 provides a brief statistical description of attendance figures for the 12 years.

[1]Weather data provided by *Servei Meteorològic de Catalunya* (Catalan Meteorological Service).

[2]Socio-economical data provided by *Institut d'Estadística de Catalunya* (Institute of Statistics of Catalunya).

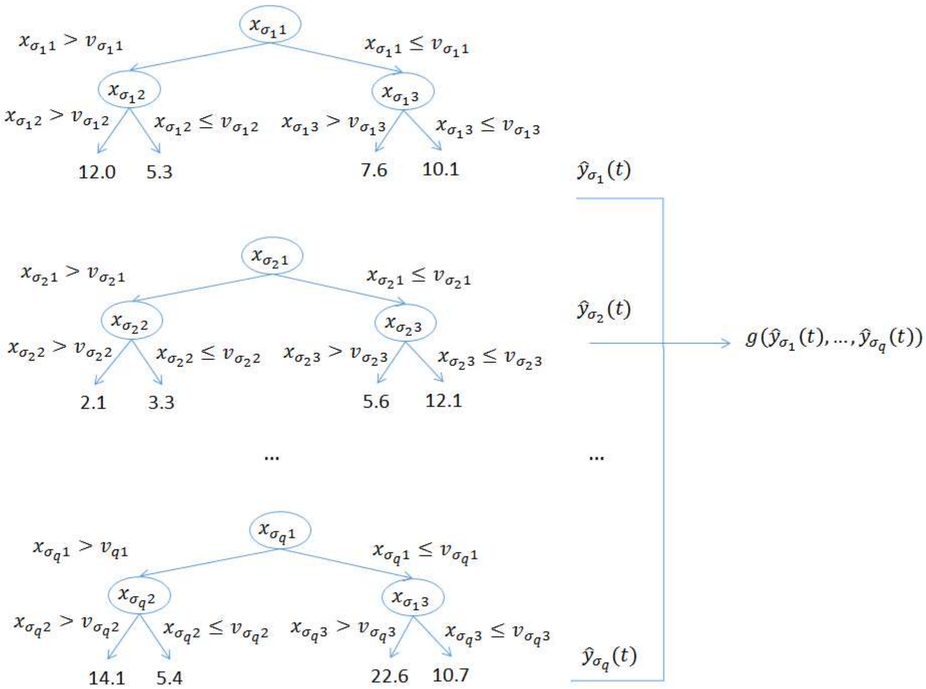

Figure 4: Ensemble of regression trees.

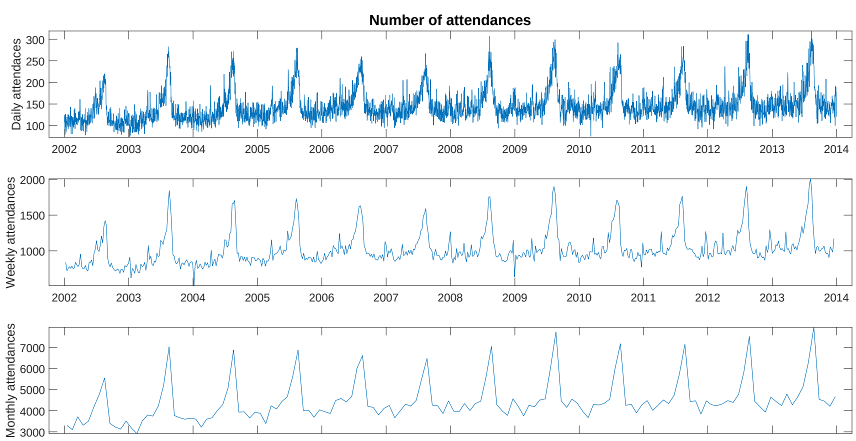

Figure 5: Emergency attendances time-series.

Table 2: Distribution of ED attendances at different time scales

	Daily att.	**Weekly att.**	**Monthly att.**
Average	146.86	1028.4	4470.2
Stand. dev.	36.41	231.43	978.36
Max.	319	2007	7954
Min.	73	513	2920

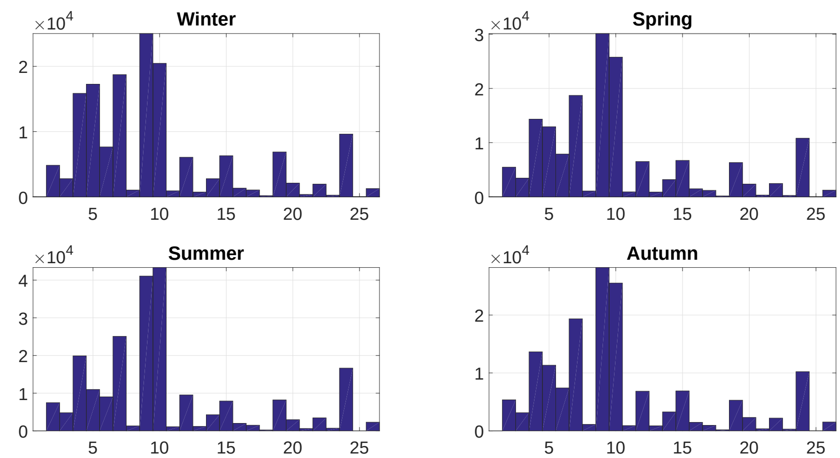

Figure 6: Histograms of diagnoses for each season. X-axis represent the internal codes of the diagnoses.

Despite the significant ED attendances increase in summer, there are not significant differences between the diagnoses made in summer than in the other seasons. As Figure 6 shows, there seems to be an increase in all the diagnoses in summer. Therefore, accurate predictions of ED attendances will have an impact on hospital admissions, which is also a major issue regarding resource management (e.g. number of prepared beds, scheduling of surgeries, etc.).

3.4 Experimentation set up

Eleven years of data, from 2002 to 2012, have been used to train and validate the models, and 1-year data (2013) has been used to test the models for predicting ED attendances one week ahead, the interval time suitable for ED resources planning. The results are analysed in terms of the Mean Average Percentage Error (MAPE)

$$MAPE = 100\frac{1}{T}\sum_{t=1}^{T}\left|\frac{\hat{y}(t) - y(t)}{y(t)}\right|$$

and the Normalised Mean Squared Error (NMSE)

$$NMSE = \frac{1}{T}\sum_{t=1}^{T}\frac{\left(\hat{y}(t) - y(t)\right)^2}{\mathrm{var}\left(y_{training}\right)}$$

where $\hat{y}(t)$ and $y(t)$ are the predicted and measured attendances, var $(y_{training})$ is the variance of the measured attendances used in the training data, and T the number of predicted samples. Therefore, the lower MAPE or NMSE the better.

The proposed methodology is compared to traditional approaches such as MLR, ANN, and more recent ones such as regression tree ensembles and the state-of-the-art method [16] labelled as *Kadri et al. (2014)*, in order to show how TENACE is able to deal with the tourist region variability.

All the methods presented in this paper have been implemented in Matlab [21].

4 Results and analysis

According to the reaction capacity of the ED and the contributions of the TENACE methodology, the results have been analysed according to the variability of data due to tourism, the exogenous variables, the prediction horizon, and the time scales.

and at different time scales (daily, weekly and monthly).

4.1 Results regarding high variability due tourism

Table 3 shows the MAPE and the NMSE obtained at predicting weekly attendances with the data of the tourist region. According to it, the best results are achieved by MLR and TENACE (combining MLR and tree ensembles), followed by the regression tree ensemble method (alone). However,

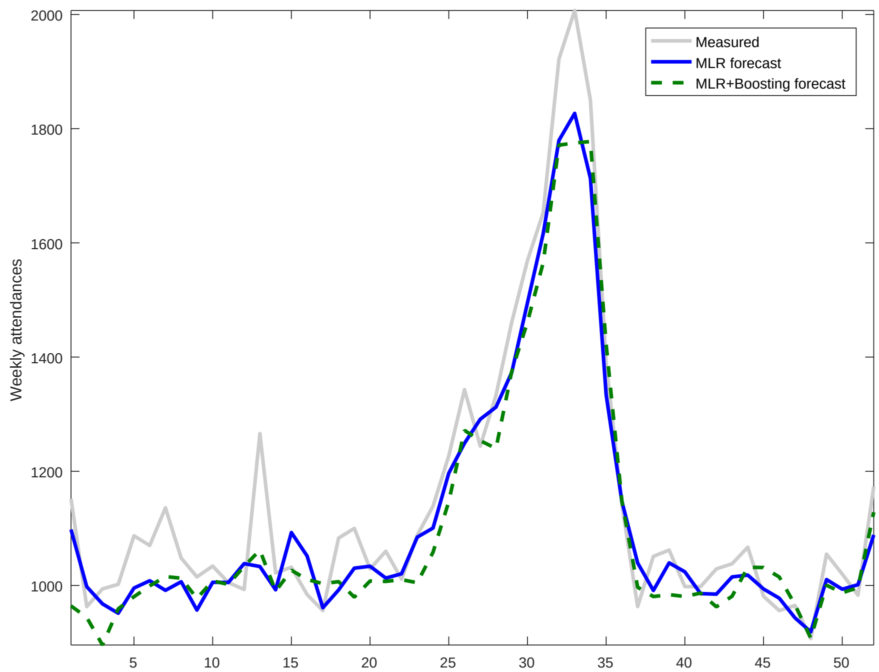

Figure 7: Measured and forecast weekly attendances.

Wilcoxon tests[3] do not find significant differences between the performance of these three methods. Moreover, ANN-based methods are the ones that performs the worse. Regarding the TENACE proposal with ANN, the results are quite unsatisfactory, and that seems to be inherent to the use of ANN.

Figure 7 shows the measured and predicted weekly ED attendances for the best methods (MLR and TENACE.MLR-Boosting).

4.2 Sensitivity analysis on exogenous variables

Table 3 shows the results obtained when only calendar data is used instead of all the variables considered in TENACE. In that regard, the use of ex-

[3]Since the absolute values of the errors obtained do not follow a normal distribution, the Wilcoxon test has been chosen to check the differences between time-series errors because this test does not require the assumption of normality.

Table 3: Accuracy on weekly forecast. Best values are in bold face.

	All variables		Calendar	
Method	**MAPE(%)**	**NMSE**	**MAPE(%)**	**NMSE**
MLR	**4.394**	**0.0962**	**5.075**	**0.1267**
ANN	6.939	0.2579	5.809	0.1990
Regression trees ensemble	5.101	0.1230	**5.282**	**0.1239**
Kadri et al (2014)	6.309	0.1660	6.309	0.1660
TENACE.MLR+ANN	8.617	0.5717	7.806	0.1695
TENACE.MLR+Boosting	**5.038**	**0.1130**	**5.587**	**0.1281**

ogenous variables (calendar, weather and socio-economic data) improves the prediction accuracy since all the methods obtain better results. The greatest improvement is achieved by MLR, and then by TENACE.MLR-Boosting.

4.3 Sensitivity analysis on the prediction horizon

Figure 8 and 9 show how MAPE and NMSE evolve when the forecast horizon (in weeks) increases. Figure 8 shows that MLR and TENACE.MLR-Boosting achieve the best results. Wilcoxon tests corroborate that, while the results obtained by tree ensemble, MLR, *Kadri et al. (2014)* and TENACE.MLR-Boosting do not have significant differences for short prediction horizons, these differences become significant, especially between MLR and TENACE.MLR-Boosting, and the others when the forecast horizon increases.

Figure 9 presents similar trends to those showed by Figure 8, but for calendar variables only. In particular, it shows that all methods, except those using ANNs, present similar results and TENACE.MLR-Boosting presents the most robust error against the increase of the forecast horizon.

4.4 Sensitivity analysis on time scales

According to the reaction capacity of the ED, weekly attendances forecasts are the suitable time scale. Nevertheless, we analyse the capacity of the method regarding monthly and daily attendances.

4.4.1 Monthly predictions

Results on monthly prediction are shown in Table 4. MAPE and NMSE values are $3\% - 5\%$ and $0.05 - 0.1$, respectively. The differences among the methods are not significant. The best results are obtained using tree

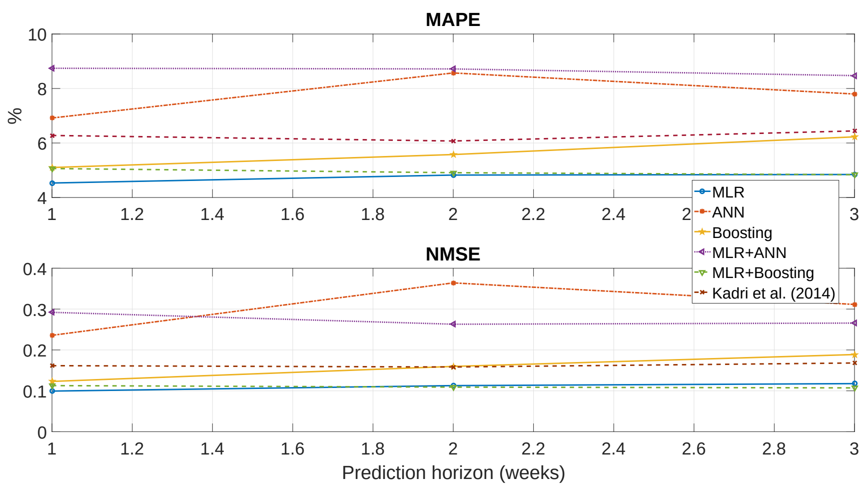

Figure 8: Weekly attendances forecast MAPE and NMSE of the different tested methods regarding forecast length. Calendar, weather and socio-economic data have been used.

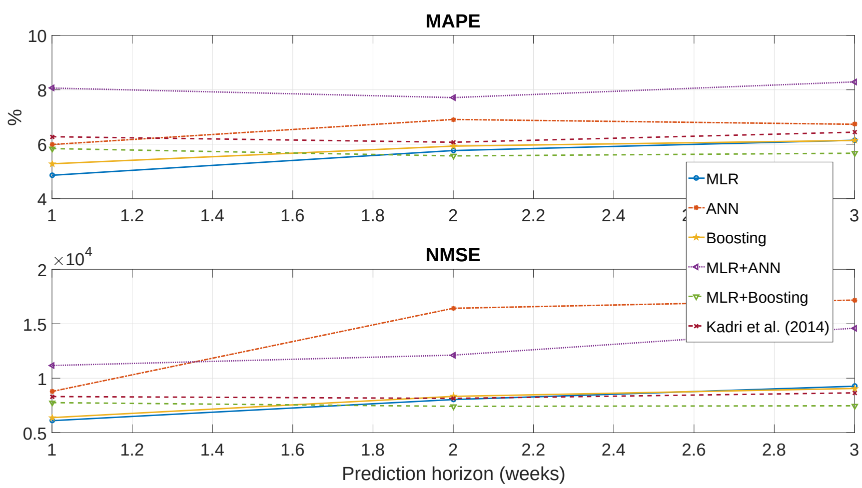

Figure 9: Weekly attendances forecast MAPE and NMSE of the different tested methods regarding forecast length. Only calendar data have been used.

Table 4: Accuracy on monthly forecast. Best values are in bold face.

	All variables		Calendar	
Method	**MAPE(%)**	**NMSE**	**MAPE(%)**	**NMSE**
MLR	**3.980**	0.0782	6.080	0.1399
ANN	5.321	0.1164	4.287	**0.0571**
Regression trees ensemble	4.031	0.0827	**3.663**	0.0800
Kadri et al (2014)	4.363	**0.0734**	4.363	0.0734
TENACE.MLR+ANN	4.948	0.0957	5.135	0.1365
TENACE.MLR+Boosting	4.429	0.1115	4.672	0.0970

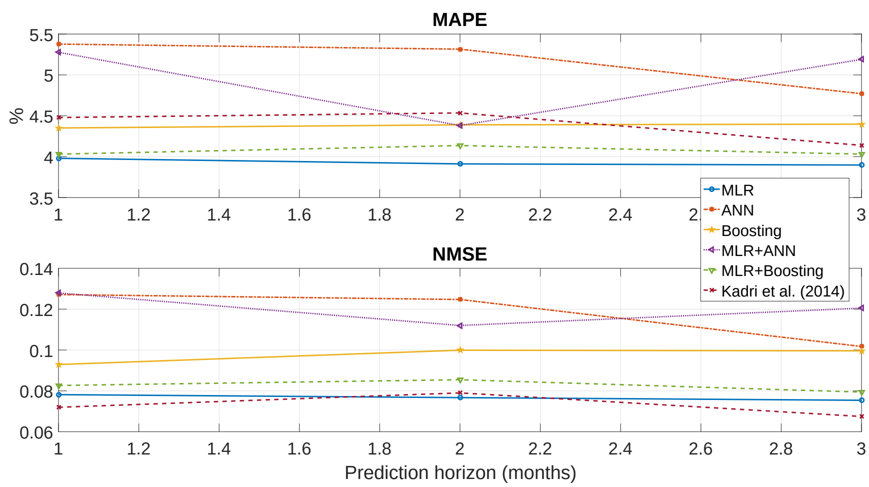

Figure 10: Monthly attendances forecast MAPE and NMSE of the different tested methods regarding forecast length. Calendar, weather and socio-economical data have been used.

ensembles and ANNs without using weather and socio-economic data. Also, *Kadri et al. (2014)* obtains a good NMSE value, with only using past ED attendances values. Therefore, exogenous variables do not have a significant impact when forecasting monthly attendances. On the other hand, TENACE is sensible to this aggregation level.

Figure 10 and 11 show the evolution of MAPE and NMSE according to the prediction horizon. They show that MLR, $MLR + Boosting$ and *Kadri et al. (2014)* obtain similar values, but better than the other methods.

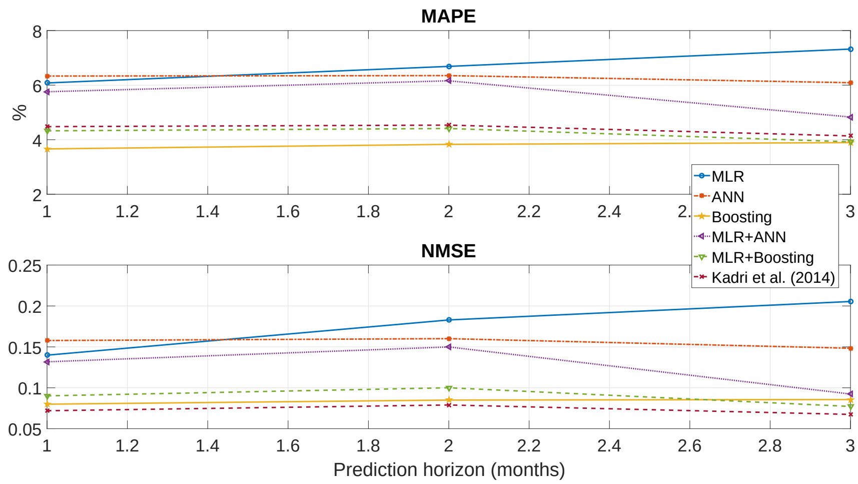

Figure 11: Monthly attendances forecast MAPE and NMSE of the different tested methods regarding forecast length. Only calendar data have been used.

4.4.2 Daily predictions

Table 5 shows the MAPE and the NMSE obtained predicting daily attendances. In this case, the state-of-the-art method [5] labelled as *Boyle et al. (2012)* has been also considered, since it provides predictions at this time scale. The best results are obtained for TENACE.MLR+ANN, followed by TENACE.MLR+Boosting and MLR. When only calendar information is used, the accuracy slightly decreases in comparison when weather and socio-economic information is used. However, there are not significant differences, according to Wilcoxon tests, between the results achieved with and without weather and socio-economic information. Moreover, results achieved by the state-of-the-art methods, *Boyle et al. (2012)* and *Kadri et al. (2012)*, are significantly worse than the methods proposed here. Regarding *Kadri et al. (2012)*, this is due to the fact that it does not use exogenous variables.

Figure 12 shows the measured and forecast daily attendances using the methods with the best MAPE: TENACE and MLR. It illustrates that the trained models are capable of catching the attendance behaviour except for some big local peaks or dips.

Figures 13 and 14 represent the MAPE and NMSE regarding the forecast

Table 5: Accuracy on daily forecast. Best values are in bold face.

	All variables		Calendar	
Method	**MAPE(%)**	**NMSE**	**MAPE(%)**	**NMSE**
MLR	7.82	0.2232	**8.214**	0.2411
ANN	8.03	0.2333	8.898	0.2697
Regression trees ensemble	8.17	0.2503	8.442	0.2610
Boyle et al (2012)	12.37	0.3760	12.37	0.3760
Kadri et al (2014)	12.46	0.5715	12.46	0.5715
TENACE.MLR+ANN	**7.51**	**0.2041**	8.301	**0.2386**
TENACE.MLR+Boosting	7.80	0.2268	8.261	0.2467

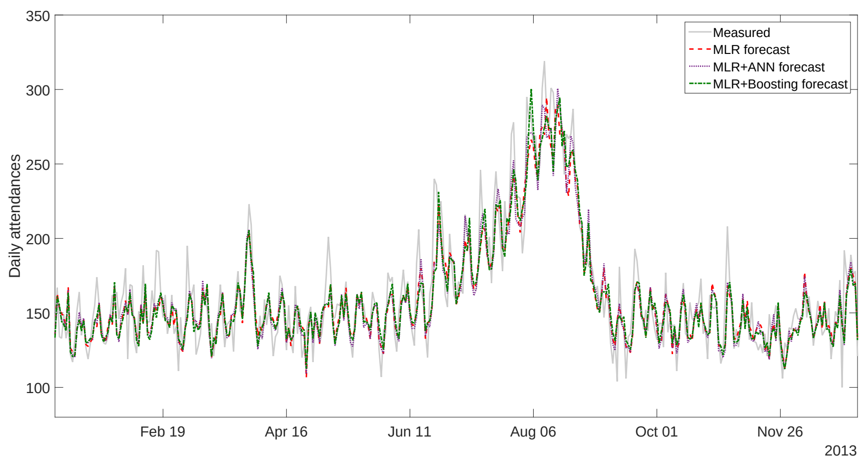

Figure 12: Measured and forecast daily attendances.

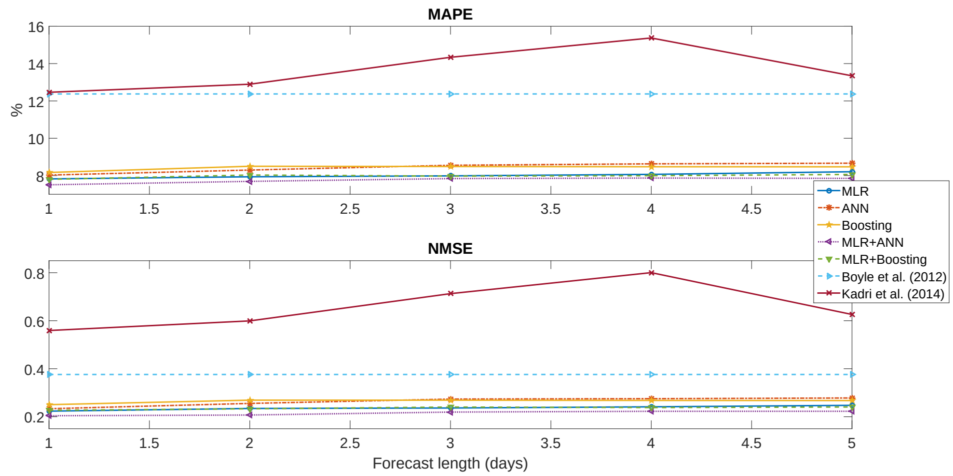

Figure 13: Daily attendances forecast MAPE and NMSE of the different tested methods regarding forecast length. Calendar, weather and socio-economical data have been used.

horizon[4] using weather and socio-economic information or without it, respectively. According to them, there is a slight, but not significant, reduction of the accuracy as the forecast horizon increases. Moreover, $MLR + ANN$ and $MLR + Boosting$ significantly outperform not only *Boyle et al. (2012)* and *Kadri et al. (2012)*, but also ANN and tree ensembles alone when the forecast horizon increases.

4.5 Discussion

The proposed method, TENACE, have been analysed predicting ED attendances time-series that have a significant peak in summer (but not always with the same magnitude and at the same day/week/month) and have annual non-constant increases (see Figure 5). Therefore, the predicted time-series are non-stationary (have variant average and standard deviation), have an important amount of variability (the average standard deviation magnitude is about 20% of the average) and the seasonality they have do not follow exact periods of time. These features are considered as detrimental when forecast-

[4]A prediction horizon of five days is considered as enough to change caretaker shifts according to *Hospital de Palamós* ED manager.

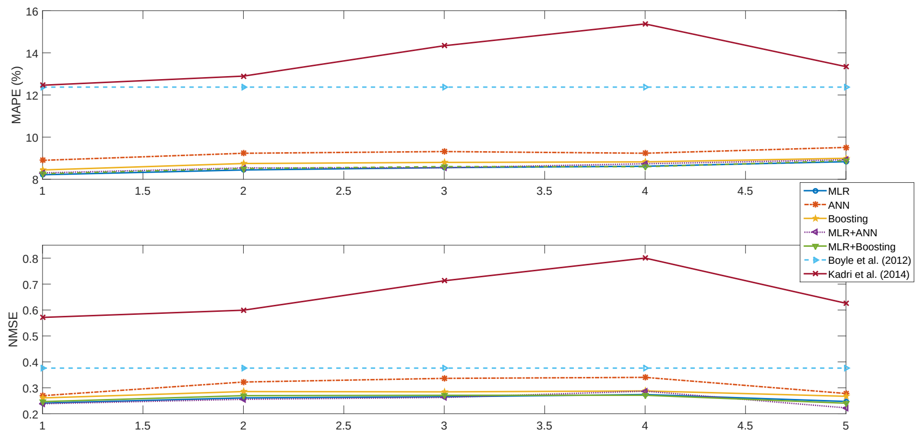

Figure 14: Daily attendances forecast MAPE of the different tested methods regarding forecast length. Only calendar data have been used.

ing a time-series, but the results achieved in this paper are very good. MLR usually obtains good results, meaning that catching linear relations between variables is a good choice for forecasting this kind of time-series. However, $MLR + Boosting$ achieves good results for all cases and it is specially robust against the increase of the prediction horizon.

Moreover, weather and socio-economic information has not been proved to significantly improve the accuracy. This conclusion is aligned with [20]. On the other hand, socio-economic and weather data must be predicted to further predict ED attendances, which could be a cause of additional ED forecast imprecision [41].

Regarding the impact of the results on the ED management, the method proposed achieves a MAPE lower than 5% at the weekly level, the time lapse in which the ED reviews the scheduling of its resources. The application of the results is expected to improve the ability to estimate the number and type of resources that a hospital needs at a particular time, and as a result improve patient care resources by being prepared for patients beforehand. It can also help to optimise the economic cost of hospitals, as by anticipating patient needs, the hospital can make the necessary adjustments according to the demands of the population and avoiding incurring unexpected expenses. Currently, Hospital of Palamós weekly checks the ED situation and if a sustained high demand is detected, mechanisms such as reviewing scheduled surgeries with patient admission, stopping patient admissions for studies, opening of new beds, and ED staff reinforcements are activated. Therefore, weekly forecasts could improve ED management by activating these mechanisms before the high demand situation happens. The robustness of TENACE regarding the forecast horizon, at the weekly and daily level, opens also an opportunity to improve surgery activity scheduling and patient admissions scheduling that could be affected by ED attendances.

4.6 Limitations

ED forecast also depends on returning patients that are not taken into account in this study. However, and according to the recent study [34], returning patients analysis should be performed in cooperation with the hospitals of a neighbourhood, since one person could first attend to the ED of a hospital and become a returning patient in a second one. Therefore, the returning visit information should be considered in a future work in collaboration with neighbourhood hospitals.

Other types of ED attendances have been gathered in GEMSA and CCMU classifications [1]. Therefore, GEMSA and CCMU categories can offer new ways of generating specific forecasting models (e.g. one per category) with more robustness to epidemic periods [1].

Finally, the use of ED forecasting models should be put in context, focussing on the pragmatic applications of the forecasts. This means that ED personnel should be able to feed the forecasting models [3]. In that regard, TENACE can be used as a starting point to predict ED overflow and to schedule permanent and transitory staff accordingly [23].

5 Conclusions and future work

The problem of emergency department overcrowding can be alleviated by the use of predictive tools to improve the planning of their resources if forecasting are provided at the appropriate time horizon. Such tools need to take into account the discontinuous flows in EDs that are conditioned by the hospital population. For example, hospitals located in tourist areas are subject to important population fluctuations throughout the year. This paper proposes a new approach, TENACE, to predict ED attendances in tourist regions with such features.

TENACE combines linear and non-linear methods, so that linear models follow ED time series while non-linear approaches deal with peaks and other sources of variability. The methods is fed with external factors such as calendar, weather and socio-economic data. The methods have been tested on data from the Hospital of Palamós, which is distinctive due to its location in a tourist region. Data for 11-years (from 2002 to 2012) have been used for model training and validation, and tests have been carried out on data for one year (2013). The results show that TENACE exhibits a performance similar to MLR and regression trees for predicting ED attendances one week ahead, but it exhibits a robust behaviour when the forecast horizon is enlarged, enabling a better management of ED planning strategies.

For future work, there are other factors in addition to calendar, weather and socio-economic data, like the evolution of influenza or important and crowded events, that affect the number of patient arrivals that need to be used in the prediction models [9, 30], or environmental factors [7]. In addition, the quality of ED services is tied not only to the number of hospital attendances, but also the number of admissions, which can be in turn be favoured with triage ED predictions of disposition outcomes [29, 40]. Therefore, it is also important to forecast hospital admissions which are not usually correlated with ED attendances.

acknowledgements

The work was carried out with support from the Generalitat de Catalunya 2017 SGR 1551.

References

[1] Afilal, M., Yalaoui, F., Dugardin, F., Amodeo, L., Laplanche, D., Blua, P.: Forecasting the Emergency Department Patients Flow. Journal of Medical Systems (2016). DOI 10.1007/s10916-016-0527-0

[2] Becerra, M., Jerez, A., Aballay, B., Garcés, H.O., Fuentes, A.: Forecasting emergency admissions due to respiratory diseases in high variability scenarios using time series: A case study in Chile. Science of the Total Environment **706**(134978) (2020). DOI 10.1016/j.scitotenv.2019.134978

[3] Bergs, J., Heerinckx, P., Verelst, S.: Knowing what to expect, forecasting monthly emergency department visits: A time-series analysis. International Emergency Nursing **22**(2), 112–115 (2014). DOI 10.1016/j.ienj.2013.08.001

[4] Boyle, A., Beniuk, K., Higginson, I., Atkinson, P.: Emergency Department Crowding: Time for Interventions and Policy Evaluations. Emergency Medicine International **2012**, 1–8 (2012). DOI 10.1155/2012/838610

[5] Boyle, J., Jessup, M., Crilly, J., Green, D., Lind, J., Wallis, M., Miller, P., Fitzgerald, G.: Predicting emergency department admissions. Emergency Medicine Journal **29**(5), 358–365 (2012). DOI 10.1136/emj.2010.103531

[6] Breiman, L., Friedman, J., Stone, C.J., Olshen, R.A.: Classification and regression trees. CRC Press (1984)

[7] Chen, J., Chen, H., Wu, Z., Hu, D., Pan, J.Z.: Forecasting smog-related health hazard based on social media and physical sensor. Information Systems **64**(Mar), 281–291 (2017). DOI 10.1016/j.is.2016.03.011

[8] Duwalage, K.I., Burkett, E., White, G., Wong, A., Thompson, M.H.: Forecasting daily counts of patient presentations in Australian emergency departments using statistical models with timevarying predictors. Emergency Medicine Australasia **32**(4), 618–625 (2020). DOI 10.1111/1742-6723.13481

[9] Ekström, A., Kurland, L., Frrokhnia, N., Castrén, M., Nordberg, M.: Forecasting Emergency Department Visits Using Internet Data. Annals of Emergency Medicine **65**(4), 436–442.e1 (2015). DOI 10.1016/J.ANNEMERGMED.2014.10.008

[10] Elith, J., Leathwick, J.R., Hastie, T.: A working guide to boosted regression trees. Journal of Animal Ecology **77**(4), 802–813 (2008). DOI 10.1111/j.1365-2656.2008.01390.x

[11] McAllan, Fern J., EgertonWarburton, D., O'Reilly, G., Weiland, T.J., Jelinek, G.A.: Planning for the future: Modelling daily emergency department presentations in an Australian capital city. Emergency

Medicine Australasia **31**(5), 750–755 (2019). DOI 10.1111/1742-6723.1324

[12] Gashler, M., Giraud-Carrier, C., Martinez, T.: Decision Tree Ensemble: Small Heterogeneous Is Better Than Large Homogeneous. In: 2008 Seventh International Conference on Machine Learning and Applications, pp. 900–905. IEEE (2008). DOI 10.1109/ICMLA.2008.154

[13] Green, M., Ekelund, U., Edenbrandt, L., Bjork, J., Forberg, J.L., Ohlsson, M., .: Explaining artificial neural network ensembles: A case study with electrocardiograms from chest pain patients. In: Proceedings of the ICML/UAI/COLT 2008 Workshop on Machine Learning for Health-Care Applications, Helsinki, Finland, (2008).

[14] Hidalgo, B., Goodman, M.: Multivariate or Multivariable Regression?. American Journal of Public Health **103**(1), 39–40 (2013). DOI 10.2105/AJPH.2012.300897

[15] Jurado, S., Nebot, A., Mugica, F., Avellana, N.: Hybrid methodologies for electricity load forecasting: Entropy-based feature selection with machine learning and soft computing techniques. Energy **86**, 276–291 (2015). DOI 10.1016/j.energy.2015.04.039

[16] Kadri, F., Harrou, F., Chaabane, S., Tahon, C.: Time Series Modelling and Forecasting of Emergency Department Overcrowding. Journal of Medical Systems **38**(9), 107 (2014). DOI 10.1007/s10916-014-0107-0

[17] Kam, H.J., Sung, J.O., Park, R.W.: Prediction of Daily Patient Numbers for a Regional Emergency Medical Center using Time Series Analysis. Healthcare Informatics Research **16**(3), 158 (2010). DOI 10.4258/hir.2010.16.3.158

[18] Kim, K., Lee, C., O'Leary, K., Rosenauer, S., Mehrotra, S.: Predicting patient volumes in hospital medicine: A comparative study of different time series forecasting methods. Tech. rep., Northestern University (2014)

[19] Lewis, R.J.: An Introduction to Classification and Regression Tree (CART) Analysis. Presented at the 2000 Annual Meeting of the Society for Academic Emergency Medicine in San Francisco, California (2000)

[20] Marcilio, I., Hajat, S., Gouveia, N.: Forecasting Daily Emergency Department Visits Using Calendar Variables and Ambient Temperature Readings. Academic Emergency Medicine **20**(8), 769–777 (2013). DOI 10.1111/acem.12182

[21] MATLAB:version 7.10.0 (R2010a). Natick, Massachusetts, The MathWorks Inc., (2010).

[22] Ministerio de Sanidad Servicios Sociales e Igualdad: Informe anual del sistema nacional de salud 2013 (2014)

[23] Murtas, R., Decarli, A., Russo, A.G.: Forecasting emergency admissions in the city of Milan to predict overflow: a 2-day warning system. European Journal of Public Health **30**(Supplement_5), (2020). DOI 10.1093/eurpub/ckaa165.588

[24] Neter, J., Kutner, M.H., Nachtsheim, C.J., Wasserman, W.: Applied linear statistical models. Irwin Chicago (1996)

[25] Pani, C., Fadda, P., Fancello, G., Frigau, L., Mola, F.: A data mining approach to forecast late arrivals in a transhipment container terminal. Transport **29**(2), 175–184 (2014). DOI doi:10.3846/16484142.2014.930714

[26] Patel, J., Goyal, R.: Applications of Artificial Neural Networks in Medical Science. Current Clinical Pharmacology **2**(3), 217–226 (2007). DOI 10.2174/157488407781668811

[27] Patterson, D.: Artificial neural networks: theory and applications. Prentice Hall PTR Upper Saddle River, NJ, USA (1998)

[28] Poole, S., Grannis, S., Shah, N.H.: Predicting Emergency Department Visits. AMIA Joint Summits on Translational Science proceedings. AMIA Joint Summits on Translational Science **2016**, 438–45 (2016)

[29] Raita, Y., Goto, T., Faridi, M.K., Brown, D.F.M., Camargo, C.A., Hasegawa, K.: Emergency department triage prediction of clinical outcomes using machine learning models. Critical Care **23**(1), (2019). DOI 10.1186/s13054-019-2351-7

[30] Ram, S., Zhang, W., Williams, M., Pengetnze, Y.: Predicting asthma-related emergency department visits using big data. IEEE Journal of Biomedical and Health Informatics **19**(4), 1216–1223 (2015). DOI 10.1109/JBHI.2015.2404829

[31] Ravizza, S., Chen, J., Atkin, J.A.D., Stewart, P., Burke, E.K..: Aircraft taxi time prediction: Comparisons and insights. Applied Soft Computing Journal **14**(PART C), 397–406 (2014). DOI 10.1016/j.asoc.2013.10.004

[32] Richardson, D.: Increase in patient mortality at 10 days associated with emergency department overcrowding. Medical journal of Australia **184**(5), 213–216 (2006)

[33] Sánchez, M., Miró, Ò., Coll-Vinent, B., Bragulat, E., Espinosa, G., Gómez-Angelats, E., Jiménez, S., Queralt, C., Hernández-Rodríguez, J., Alonso, J.R., Millá, J.: Saturación del servicio de urgencias: factores asociados y cuantificación. Medicina Clínica **121**(5), 167–172 (2003). DOI 10.1016/S0025-7753(03)73892-5

[34] Shy, B.D., Loo, G.T., Lowry, T., Kim, E.Y., Hwang, U., Richardson, L.D., Shapiro, J.S.: Bouncing Back Elsewhere: Multilevel Analysis of Return Visits to the Same or a Different Hospital After Initial Emergency Department Presentation. Annals of Emergency Medicine (2017). DOI 10.1016/j.annemergmed.2017.08.023

[35] Sun, Y., Heng, B., Seow, Y., Seow, E.: Forecasting daily attendances at an emergency department to aid resource planning. BMC Emergency Medicine **9**(1), 1 (2009). DOI 10.1186/1471-227X-9-1

[36] Taylor, J.W.: Triple seasonal methods for short-term electricity demand forecasting. European Journal of Operational Research **204**(1), 139–152 (2010). DOI 10.1016/j.ejor.2009.10.003

[37] Trzeciak, S.: Emergency department overcrowding in the United States: an emerging threat to patient safety and public health. Emergency Medicine Journal **20**(5), 402–405 (2003). DOI 10.1136/emj.20.5.402

[38] Tso, G.K., Yau, K.K.: Predicting electricity energy consumption: A comparison of regression analysis, decision tree and neural networks. Energy **32**(9), 1761–1768 (2007). DOI 10.1016/j.energy.2006.11.010

[39] Tudela, P., Mòdol, J.: La saturación en los servicios de urgencias hospitalarios. Emergencias **27**, 113–120 (2015)

[40] Vaghasiya, M.R., Murphy, M., O'Flynn, D., Shetty, A.: The emergency department prediction of disposition (EPOD) study. Australasian Emergency Nursing Journal **17**(4), 161–166 (2014). DOI 10.1016/j.aenj.2014.07.003

[41] Wargon, M. and Guidet, B. and Hoang, T. D. and Hejblum, G.: A systematic review of models for forecasting the number of emergency department visits. Emergency Medicine Journal **26**(6), 395–399 (2009). DOI 10.1136/emj.2008.062380

[42] Wargon, M., Casalino, E., Guidet, B.: From Model to Forecasting: A Multicenter Study in Emergency Departments. Academic Emergency Medicine **17**(9), 970–978 (2010). DOI 10.1111/j.1553-2712.2010.00847.x

[43] Xu, M., Wong, T., Chin, K.: Modeling daily patient arrivals at Emergency Department and quantifying the relative importance of contributing variables using artificial neural network. Decision Support Systems **54**(3), 1488–1498 (2013). DOI 10.1016/j.dss.2012.12.019

[44] Yucesan, M., Gul, M., Celik, E.: A multi-method patient arrival forecasting outline for hospital emergency departments. International Journal of Healthcare Management **13**(S1), 283–295 (2020). DOI 10.1080/20479700.2018.1531608

[45] Zamo, M., Mestre, O., Arbogast, P., Pannekoucke, O.: A benchmark of statistical regression methods for short-term forecasting of photovoltaic electricity production, part I: Deterministic forecast of hourly production. Solar Energy **105**(Jul), 792–803 (2014). DOI 10.1016/j.solener.2013.12.006