Probabilistic forecasting of hourly Emergency Department arrivals

Bahman Rostami-Tabar[a,*], Jethro Browell[b], Ivan Svetunkov[c]

[a] *Cardiff Business School Cardiff University UK.*
[b] *School of Mathematics and Statistics University of Glasgow UK.*
[c] *Centre for Marketing Analytics and Forecasting Lancaster University UK.*

Abstract

An accurate forecast of Emergency Department (ED) arrivals by an hour of the day is critical to meet patients' demand. It enables planners to match ED staff to the number of arrivals, redeploy staff, and reconfigure units. This can have many advantages for healthcare staff and the quality of care delivered to patients. In this study, we develop an innovative model based on Generalised Additive Models and an advanced dynamic model based on exponential smoothing to generate an hourly probabilistic forecast of ED arrivals for a prediction window of 48 hours. We compare the forecast accuracy of these models against appropriate benchmarks, including TBATS, Poisson Regression, Prophet, and simple empirical distribution. We use Root Mean Squared Error (RMSE) to examine the point forecast accuracy and assess the forecast distribution accuracy using Quantile Bias, PinBall Score and Pinball Skill Score. Our results indicate that the proposed models outperform their benchmarks for point and probabilistic forecasts. Our developed models can also be generalised to forecast hourly arrivals in other services such as hospitals, ambulances, or clinical desk services.

Keywords: Emergency Department, Poisson Regression, Probabilistic Forecasting, Generalised Additive Models, Intermittent Exponential Smoothing

1. Introduction

Forecasting Emergency Department (ED) arrivals is critical for informing staffing and scheduling decisions to meet the needs of patients. Accurate ED demand forecasts contribute to a better decision-making process regarding resources allocation and staffing. This is one of the best ways to optimise resources utilisation and minimise related costs. An accurate forecast of patient arrivals is crucial in ED services to depict various courses of action that can result in massive savings in terms of patient lives. Inability to match the staff with the demand might result in patients overcrowding the system, which is a severe problem that causes challenges for the patient flow (Derlet, 2002). Also, it is related to increasing length of stay (Muhammet and Guneri, 2015), low patient satisfaction, unexpected return visits to services, increased health care costs, inaccuracy in electronic medical records (Rostami-Tabar and Ziel, 2022).

Accurate forecasting of arrivals by the hour of the day enables planners to match staff to meet anticipated patients, reconfigure units and redeploy staff. This has many advantages for both patients, staff and the quality of provided services. Hourly forecasts are required to inform the short-term operational planning for the current and the upcoming shifts of the day. This involves the short-term decision-making related to the execution of the delivery process in ED. The combination of an hourly arrival forecast, current staff, being occupied, resource availability and waiting times at ED provide information on the state of the unscheduled care system across the service. Having this complete picture enables the delivery managers to focus on the areas that require intervention to allow the most effective delivery of the service to the patients. However, compared with lower frequency time series forecasting such as monthly, quarterly and yearly, hourly forecasts are challenging because the noise caused by random variation may overshadow any pattern in the time series. Hourly time series generally exhibit multiple seasonal cycles of different lengths: hourly, daily, weekly,

*Corresponding author
Email addresses: rostami-tabarb@cardiff.ac.uk (Bahman Rostami-Tabar), jethro.browell@glasgow.ac.uk (Jethro Browell), i.svetunkov@lancaster.ac.uk (Ivan Svetunkov)

and yearly. They may also express nonstationarity, and their profile may change over time. Therefore, an appropriate forecasting model should consider these features to accurately predict hourly demand admissions.

There are few studies that look at forecasting hourly arrivals in ED and other hospital services using historical time series data and/or predictors such as patient characteristics, weather, holidays and public events. These studies use multiple approaches, including Exponential Smoothing (Svetunkov, 2021a), Autoregressive Integrated Moving Average (ARIMA) (Hyndman and Athanasopoulos, 2021), Autoregressive Conditional Heteroskedasticity (ARCH) (Bollerslev et al., 1994), Vector Autoregressive model (Lütkepohl, 2013), TBATS (De Livera et al., 2011) and Artificial Neural Networks (Hyndman and Athanasopoulos, 2021). We have identified a number of limitations and gaps in the area of ED forecasting that motivated us to undertake this study. These gaps and their importance are discussed below. Most of these studies are limited to only predicting future arrivals as a point forecast (a single number), which does not quantify any uncertainty associated with the number of future arrivals. There are few studies that report uncertainty by presenting prediction intervals, but There is no study generating and evaluating the entire forecast distribution of arrivals. Reporting the uncertainty via the forecast distribution is potentially valuable in this setting and has practical implications for those managing Emergency Departments because the consequences of inadequate staffing are asymmetric, i.e. having more staff than needed is costly, but having less staff than required may lead to worse outcomes for patients. This asymmetry arises because it is preferable to incur a small opportunity cost associated with utilised staff rather than compromise service levels if staff levels are insufficient (Wright et al., 2006). Probabilistic forecasts inform decision-makers about exposure to these risks and potentially enable those risks to be managed more efficiently (Ramos et al., 2013; Rostami-Tabar and Ziel, 2022). Furthermore, if the impact of under- and overstaffing can be quantified, probabilistic forecasts allow *optimal* decisions that balance the cost associated with under- and overstaffing to be calculated. Therefore, in this paper, in addition to generating point forecasts, we also produce and evaluate density forecasts of hourly ED arrivals, comparing several methods for this task. Another drawback of existing studies is that the datasets used are relatively small (e.g. time period of 1-2 years), making it challenging to capture the inter-annual seasonality correctly and to report the forecast accuracy using robust approaches such as time series cross-validation. Such results might not be generalisable. Additionally, most of the forecasting methods used in these publications do not consider the full extent of the multiple seasonality of hourly ED arrivals. Moreover, hourly ED time series may contain low volume values and zeros in some hours of the day, which brings additional challenges to traditional time series forecasting approaches. Finally, all previous publications referenced in this paper are not fully reproducible as underlying data, functions and code are not available.

In this paper, we aim at filling these gaps and generate forecasts for a prediction window of 48 hours. Our contributions to the literature are summarised as follows:

1. We produce probabilistic forecasts, in addition to the point estimation, quantifying uncertainties in future hospital admission, and comparing different forecasting methods using a suite of well-established evaluation metrics;
2. We develop an advanced dynamic model to forecast ED arrivals based on iETS (Svetunkov and Boylan, 2019) and ETSX models with a modification for multiple frequencies, which produced highly-accurate point forecasts;
3. We develop a novel model to produce a probabilistic forecast of ED arrivals based on Generalised Additive Models for Location Scale and Shape, which accounts for i) the bounded and non-Gaussian distribution of arrivals, ii) multiple seasonalities, weather and holiday effects, and iii) variation in forecast uncertainty;
4. We benchmark the accuracy of our model against appropriate models used when multiple seasonality is present, i.e. Prophet, TBATS, Poisson Regression, Exponential Smoothing State Space model (ETS) and the simple empirical distribution of the arrivals;
5. We provide data and code enabling reproduction and refinement of the proposed approach and benchmarks. The proposed approach could also be generalised to forecast hourly requirements for longer horizons and in other services (Al-Azzani et al., 2021), such as inpatient and outpatient care services, the number of attended incidents in ambulance services, or call volumes in clinical desk services.

The rest of the paper is organised as follows: In section 2, we provide a brief overview of hourly ED arrival forecasting; In Section 3, we present the hourly time series of an ED arrival and use various plots to highlight important patterns. In section 4, we describe the modelling approach and benchmark methods. We

then discuss the performance evaluation metrics in section 5; in section 6, we present and discuss our results. Finally, we summarise our findings and present ideas for future research in section 7.

2. Research background: hourly ED forecasting

There is a substantial number of studies that employ models to forecast admissions and arrivals to inform planning and decision making in the healthcare. Areas such as call volume arrivals, ambulance demand and Emergency Department forecasting have received a significant attention. We refer interested readers to some extensive reviews of the relevant literature by Shi et al. (2022), Gul and Celik (2020) and Ibrahim et al. (2016). The time granularity considered in these studies spans from hourly to yearly across different parts of healthcare. However, given the focus of this study, we only discuss hourly ED forecasting.

Hertzum (2017) used linear regression, ARIMA, and Naïve to investigate whether accurate hourly accident and emergency department patient arrivals and occupancy forecasts can be generated using calendar variables. Hertzum (2017) study found that patient arrival variation is larger across the hours of the day than across the days of the week and the months of the year. In terms of the hour of the day, patient arrivals peaked around noon. For days of the week, Monday is the busiest day, while weekends are the quietest ones. July-August are the months with the highest number of patient arrivals, while January and February have the lowest numbers. They indicate that regression and ARIMA models performed similarly in modelling patient arrivals, while ARIMA outperformed regression models in modelling accident and emergency department occupancy.

Choudhury and Urena (2020) used ARIMA, Holt-Winters, TBATS, and neural network methods to forecast hourly accident and emergency department arrivals. ARIMA model was selected as the best fit model. Authors claimed that ARIMA provided high and acceptable hourly ED forecasting accuracy, even outperforming TBATS. Cheng et al. (2021a) developed an ARIMA model for ED occupancy with a seasonal component and exogenous variables, which outperformed a rolling-average benchmark. They also produce prediction intervals, a form of the probabilistic forecast, which were found to be well-calibrated, a necessary property for such forecasts.

Morzuch and Allen (2006) used the Unobserved Components Model (UCM), in which each component of the time series is separately modelled as stochastic. Double-seasonal exponential smoothing and standard Holt-Winters were used to forecast ED arrival for a horizon of 168 hours. The hourly data collected from an ED in Pennsylvania showed no trend and two seasonal cycles: a within-day and a within-week seasonal cycles. The double seasonal model recorded lower RMSEs for all the 168-hour horizons, which was expected due to the strong hourly seasonality of the time series.

McCarthy et al. (2008) employed a Poisson log-linear regression model, including independent variables such as temporal factors (e.g., hour-of-day, day-of-week, type-of-day, season, and calendar year), patient characteristics (i.e., age, gender, insurance status, triage level, mode of arrival, and ambulance diversion status) and climatic factors (i.e., temperature and precipitation) to model patient demand for ED services. The authors produced probabilistic predictions in the form of 50% and 90% prediction intervals for the number of hourly arrivals. Hourly data of ED arrivals in the 1-year study period was modelled and analysed, and it was suggested that the model could be used for forecasting. However, model evaluation was performed in-sample on only one year of data, so it is unclear how this approach would perform in a forecasting setting or compare to simpler approaches. However, the length of the time series in this study was very short (only one year), which did not allow for a rigorous out-of-sample evaluation.

Schweigler et al. (2009) investigated whether time series methods could accurately generate short-term forecasts of ED bed occupancy. A year-long dataset of hourly ED bed occupancy was collected from three facilities. The authors implemented an hourly historical average model, SARIMA model, and sinusoidal model with autocorrelated error for each facility. The historical average model was based on the mean occupancy for each site, for each hour of the day, while the sinusoidal model was based on four parameters: an AR term, a sine coefficient, a cosine coefficient and an intercept. They evaluated the forecast accuracy of four and twelve hours forecast horizons using RMSE. They found that both SARIMA and the sinusoidal models outperformed the historical average (for example, at site 2, the two models improved by 33% the 12-hour forecasts generated by the historical average).

Kim et al. (2014) compared different univariate and multivariate time series forecasting techniques to predict patient volume for a Hospital Medicine programme. The study evaluated linear regression, exponential smoothing, ARIMA, SARIMA, Generalized Autoregressive Conditional Heteroskedasticity (GARCH) and

Vector Autoregressive (VAR) models to forecast for 4 and 24 hours ahead. They used Mean Absolute Percentage Error (MAPE) to report the forecast accuracy. The authors found that the ARIMA outperformed all the other models.

Table 1 summarises the relevant papers.

Table 1: Summary of studies in forecasting hourly arrivals in Emergency Department

Reference	Year	Variable	Horizon	Length of dataset	Method	Metric	Probabilistic	Seasonality
Current study	2023	ED arrivals	48h	5 years	Naïve; Poisson & Quantile Regression; Exponential Smoothing; Prophet; TBATS; Gradient Boosting ; ADAM; GAM	RMSE, Pinball score, skill score, Quantile bias	YES	YES
Cheng et al.	2021	ED visits	1h to 4h	1 year	SARIMAX; Holt-Winters; VAR; ARIMA	MSE, MAE, MAPE, Prediction interval coverage	NO	Single
Choudhury and Urena	2020	ED arrivals	1h to 24h	4 years	ARIMA; Holt-winters; TBATS; ANN	RMSE, ME	NO	Multiple
Asheim et al.	2019	ED arrivals	3h	5 years	Poisson regression	MAPE	NO	Single
Hertzum	2017	ED arrivals	1,2,4,8,24h	3 years	linear regression; SARIMA; Naïve	MAE, MAPE, MASE	NO	Single
Kim et al.	2014	Hospital admission	4h, 24h	3 years	Linear regression; Exponential smoothing; ARIMA; GARCH; VAR	MAPE	NO	Single
Cote et al.	2013	ED arrivals	24h	2 years	Fourier regression	R^2, Standard Error	NO	Single
Chase et al.	2012	ED CUR	30m 1h, 2h, 4h, 8h, 12h	1 year	Binary regression	NA	NO	Single
Schweigler et al.	2009	Bed occupancy	4h, 12h	4 years	Hourly historical average; SARIMA; Sinusoidal model with autocorrelated error	RMSE	NO	Single
Jones et al.	2009	ED census	24h	2 years	VAR; Holt winters	MAE	NO	Single
McCarthy	2008	ED arrivals	n/a	1 year	Poisson log-linear regression model	Prediction interval coverage	Partially	Multiple
Channouf et al.	2007	Ambulance admission	1h, 3h, 6h, 12h, 13h, 14h, 17h, 23h, 24h	2 years	Regression	RMSE	No	Single
Morzuch and Allen	2006	ED arrivals	168h	3 years	Double Exponential Smoothing; Additive Holt Winter	RMSE	No	Multiple

Asheim et al. (2019) developed a Poisson time-series regression model with continuous day-of-week and week-of-year effects to implement a real-time system that could forecast ED arrivals on 1, 2, 3 hours ahead. Measuring the accuracy using the MAPE, Asheim et al. (2019) noticed that significant improvement happened when the time of notification was incorporated into the model, especially in the one-hour horizon.

Cheng et al. (2021b) used one year of ED visits time series to evaluate the Rolling Average, SARIMAX, ARIMA, VAR and Holt-Winter to forecast ED occupancy up to 4-hours ahead. The forecast accuracy is evaluated using Mean Squared Error (MSE), Mean Absolute Error (MAE) and MAPE for point forecast and coverage for prediction intervals of 80% and 95%. They show that SARIMAX provides a more accurate forecast of hourly ED occupancy.

According to the studies mentioned above, it can be said that they have shown complications in forecasting hourly patient accident and emergency department visits, and the application of forecasting hourly patient visits is not well established. Some of the studies claimed that the accuracy of forecasting models on hourly accident and emergency department data is low (Boyle et al., 2012; Hertzum, 2017), while others mentioned that the accuracy of ED hourly forecast is at an acceptable level (Choudhury and Urena, 2020; McCarthy et al., 2008; Schweigler et al., 2009).

There are a few limitations in the literature which encourage us to undertake this research and examine different forecasting approaches:

(i) Current approaches to forecast hourly ED arrivals do not fully consider the feature of data such as multiple seasonal cycles and changing profile over time;
(ii) Almost all research studies produce point forecasts and, at best, report prediction intervals. There is a lack of studies presenting the entire forecast distribution of hourly ED arrivals that better represent the uncertainty of future arrivals, providing a holistic picture of future demand for a planner;
(iii) most studies are not reproducible, as it is almost impossible to reapply the approaches without the help of the authors of those papers;
(iv) studies are limited in terms of the length of historical data used for training purposes and forecast performance evaluation and
(v) some studies in this area lack a rigorous experimental design, i.e. they do not use benchmark methods or report forecast accuracy.

3. Preliminary analysis

Data used in this study comprises counts of patients' arrival times at one of the largest ED units in the UK between April 2014 and February 2019, extracted from the ED administrative database of the hospital. We aggregated the patients' arrival times to obtain hourly arrivals, which are used in this study. Figure 1 illustrates the distribution of arrivals for each hour of the day and the day of the week. Although the data is noisy, it reveals some systematic patterns.

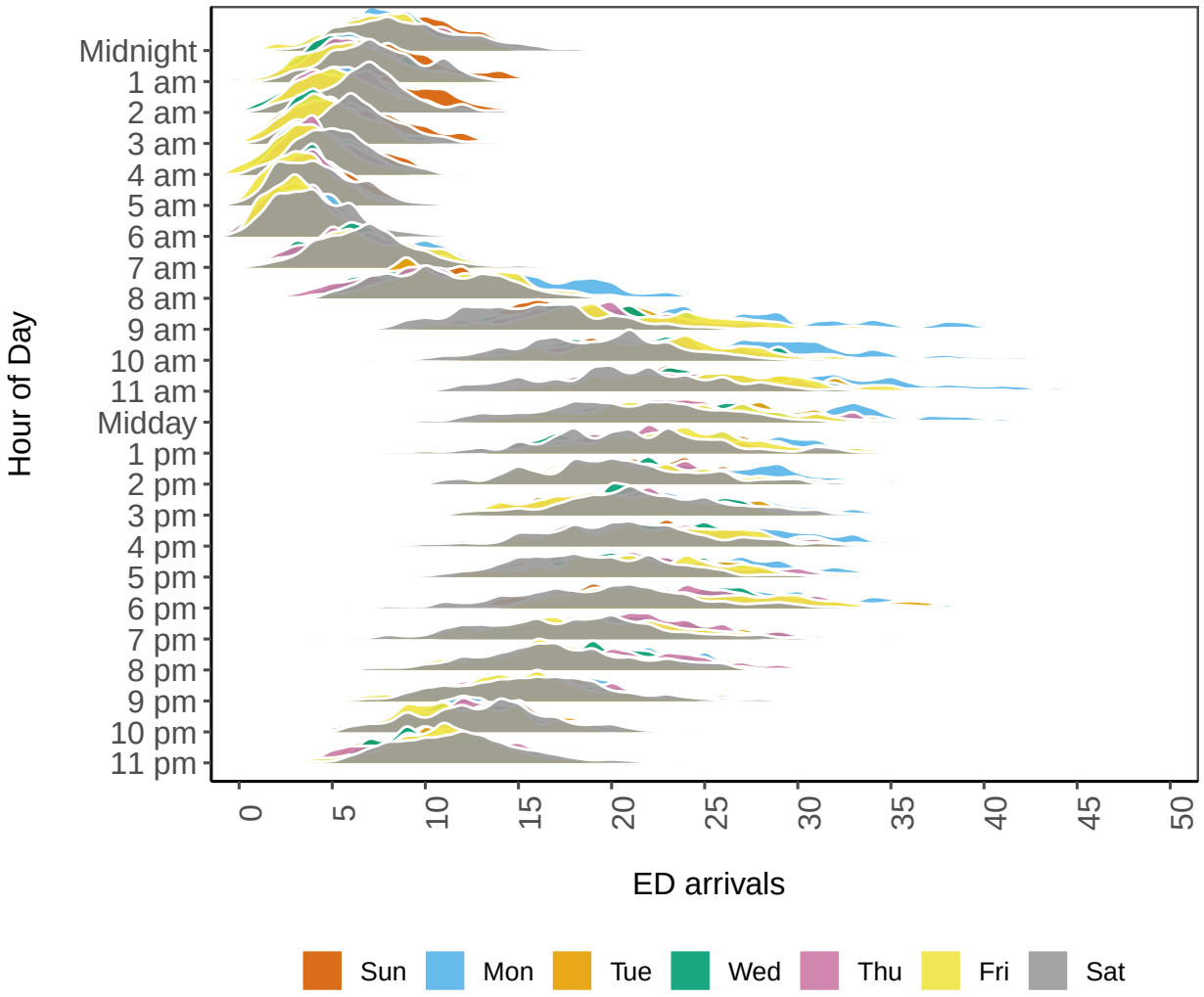

Figure 1: Distrubution of admission by hour-of-day and day-of-week. Most days have a distinct pattern, such as relatively high arrivals on Monday mornings and in the early hours of Saturday and Sunday morning.

It is clear that the number of arrivals has a sub-daily structure, similar to the one summarised by Hertzum (2017). The ED arrivals decrease between midnight and early morning and then increase until the evening, decreasing after that again. It is also clear that ED service gets systematically more visits on Mondays between 8 a.m. and 5 p.m. Moreover, the number of arrivals around midnight is slightly higher for Saturdays and Sundays.

Figure 1 also highlights significant skewness for almost every hour of the day that varies with time-of-day, which should be accounted for in forecasting methods. Some skewness might be related to holidays and special events. It is also clear that arrivals are less volatile between midnight and early morning.

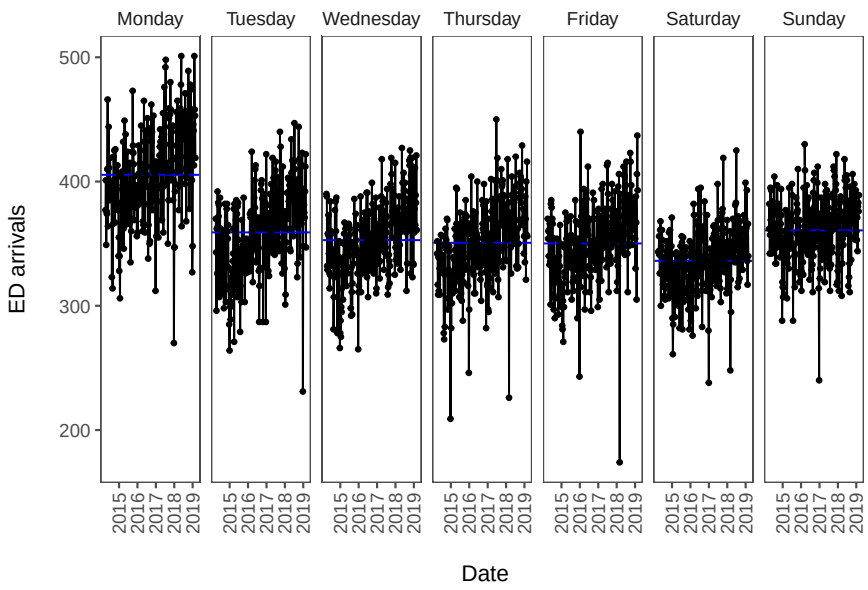

Figure 2: Subseries plot: day of weekly arrivals

Figure 2 illustrates the daily subseries plot, with the x-axis representing the date and y-axis the ED arrivals. Each individual plot illustrates how arrivals change over time for a each day of week from Monday to Saturday. The blue line shows the average arrival for the given day. It is clear that ED arrivals on Mondays are higher than on other days. This is followed by Saturday. This indicates that there are significantly more arrivals on Mondays and Saturdays compared to the rest of the week. This might be due to the closure of General Practitioners outpatient clinics over the weekends.

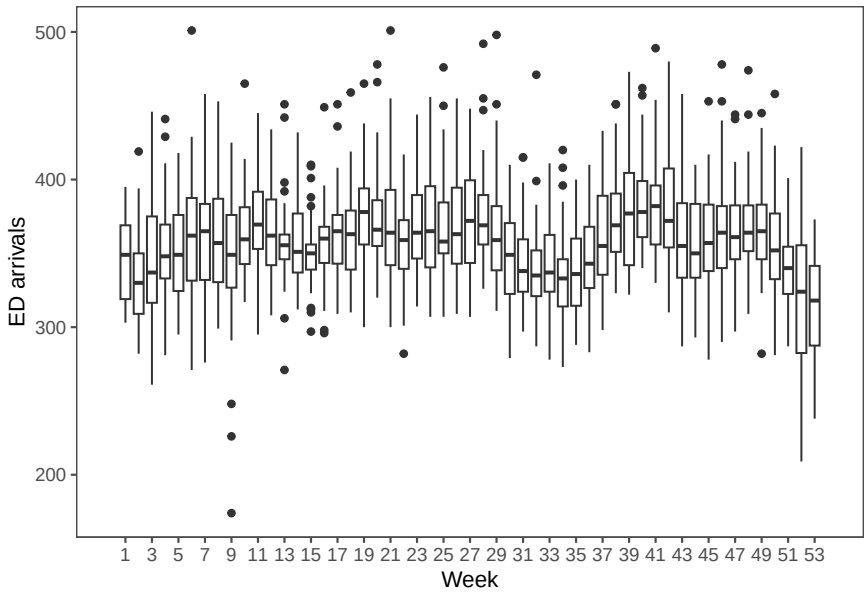

Figure 3: Arrivals by week-of-year. There is an annual trend of reduced arrivals durnng the summer and winter holiday period.

Figure 3 highlights the week of year seasonality in the ED arrivals. We observe that arrivals are significantly lower from week 29 to 35, corresponding to the Summer period. Moreover, the number of arrivals is lower at the beginning and the end of the year. The arrivals increase from week 36 and peak in weeks 39-42. This corresponds to the September - October period.

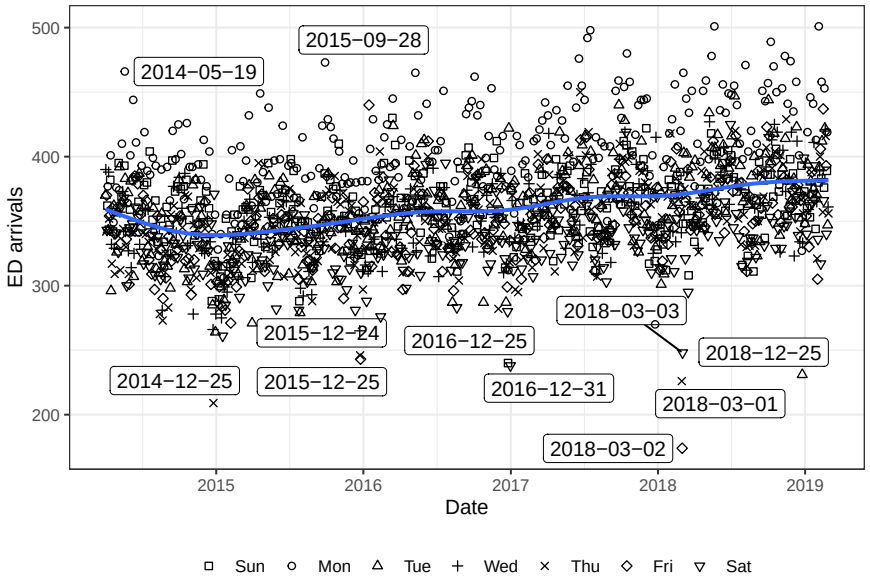

Figure 4: Daily arrivals over the entire dataset with moving average trend (blue line) and dates of outliers labled. Low outliers invariables correspond to public holidays.

Finally, Figure 4 presents the time plot of daily arrivals. Each point represents one day, and points are shape-coded by day-of-week to show the weekly cycle. The figure shows more arrivals on Mondays than on other weekdays, agreeing with the previous findings. Moreover, we can observe a long-term trend line and the significant effect of some holidays. We can see that arrivals near Christmas and New Year's day are significantly lower than other days of the year. Moreover, Figure 4 allows us to identify the impact of special events on arrivals. For instance, we see that arrivals are significantly low for 01-03 of March 2018. These days correspond to the Storm Emma with heavy snowfall that resulted in travel disruption, mass power outages, and schools closed in the UK (Wales, n.d.).

Based on this analysis and the literature review, we should consider models that can take the following into account:

- Hour-of-day, day-of-week, and week-of-year seasonalities,
- Long-term trend (or a slowly changing level),
- Calendar events, such as holidays,
- Lags of calendar events to accommodate the potential changes in demand the next day after a holiday,
- Other events, such as sporting fixtures,
- Temperature effects.

We propose several forecasting models that account for the structures outlined above.

4. Model building

4.1. Naive

We start with one of the simplest forecasting approaches used in practice - assuming that in the next few hours, everything will be the same as in the similar hours of a similar day in the past. This is called "Naïve". In our case, given that we need a distribution of values, we will use a modified approach, where the empirical distribution of the hourly arrival time series is used to forecast the future arrival distribution (La Salle et al., 2021). We consider the empirical distribution of all available historic data (Benchmark-1) and the empirical distribution of the most recent year of historic data on a rolling basis (Benchmark-2) to capture potential changes in behaviour over time.

4.2. Poisson Regression

Regression is one of the most popular forecasting methods that use explanatory variables to predict a variable of interest (in our case, the ED arrivals). The classical linear regression model is formulated as

$$y_t = \mathbf{x}_t'\boldsymbol{\beta} + \epsilon_t, \tag{1}$$

where $\mathbf{x}_t$ is the vector of explanatory variables, $\boldsymbol{\beta}$ is the vector of parameters, ϵ_t is the error term, which is typically assumed to follow Normal distribution with zero mean and a fixed variance, and t is the time index. However, in the context of healthcare and ED arrivals, the assumption of Normality is unrealistic because the number of admitted patients is an integer and non-negative. So the linear regression model should be substituted by some other model. One of the models that is frequently used in practice is the Poisson regression (see, for example, McCarthy et al., 2008), which can be summarised as

$$\begin{aligned} y_t &\sim \text{Poisson}\left(\lambda_t\right) \\ \log \lambda_t &= \mathbf{x}_t'\boldsymbol{\beta} \end{aligned} \quad . \tag{2}$$

The logarithm in (2) is needed to ensure that the parameter of Poisson distribution is always positive. This model can be estimated via maximisation of the likelihood function based on Poisson mass function . There is no single correct answer when selecting explanatory variables for the model, and the decision needs to be made for each specific case. In our experiment, we will only include dummy variables, capturing a variety of calendar events:

1. Hour of the day,
2. Day of the week,
3. Week of the year,
4. Holidays (such as Christmas, New Year etc.),
5. 24 hours lags of holidays.

The variables *1–3* allow to model the seasonal patterns on the appropriate level of detail throughout the year, while *4* covers the changes in admittance due to calendar events. Finally, *5* is needed to capture the potential phenomenon of change in admittance after the holiday (e.g. people might try not to go to the hospital on Christmas eve and thus go the next day). This model assumes that all these effects are deterministic and do not change over time. Still, the exponentiation in (2) introduces an interaction effect between dummy variables so that the 3 pm on Monday in January will be different from 3 pm on Monday in July, although the parameters for the hour of day and day of the week are fixed and do not change over time. We use the `alm()` function from the `greybox` package (Svetunkov, 2021b) for R (Team, 2021) for the experiments and denote this model as "Poisson Regression".

4.3. ETS - Exponential Smoothing

Hyndman et al. (2008) developed a state space approach for exponential smoothing models. The model can have a set of elements, including different types of Error, Trend and Seasonal components (thus *ETS*). Given the popularity of the ETS model in the forecasting community, we decided to include the basic ETS(A,N,A) model with the seasonal component with a frequency of 24 (hour of the day) as a benchmark. This was done using the `adam()` function from the `smooth` package (Svetunkov, 2021c) for R and denoted as *ETS*. This model does not capture the day of week or week of year effects, does not include explanatory variables, but its seasonal component and level change over time. This model is included as a benchmark, only to see how the other models perform compared to it.

4.4. Prophet

Prophet is a forecasting procedure created by Facebook (Taylor and Letham, 2018) that accounts for multiple seasonality, piecewise trend and holiday effects. Prophet is robust to missing data and shifts in the trend and typically handles outliers well. Prophet works well on daily data seen in Facebook. It is robust and automated, making it easy to learn for beginners. The implementation may be less flexible than other methods. The model itself relies on the Multiple Source of Error state space model, initially proposed by Kalman (1960). The model is incorporated using the corresponding implementation of the Fable package in

R. We use the `prophet()` function from the `fable` package (O'Hara-Wild et al., 2020). Note that the input data is assigned with an hourly and daily seasonality. This method has been adopted in some healthcare service providers in the United Kingdom to produce forecasts, therefore we have included it as one of our benchmarks.

4.5. TBATS

De Livera et al. (2011) proposed a model to deal with time series exhibiting multiple complex seasonalities called "TBATS". It includes a Box-Cox Transformation, ARMA model for residuals and a trigonometric expression of seasonal terms. The latter gives the model more flexibility to deal with fractional seasonality and reduces the parameters of the model when the frequencies of seasonalities are high. We fit a TBATS model using the `tbats()` function from the `forecast` package in R (Hyndman et al., 2020).

4.6. Quantile Regression and Gradient Boosting Machines

Quantile regression allows the production of density forecasts without assuming a fixed distributions shape controlled by a small number of parameters, such as the Poisson distribution with parameter λ. By producing forecasts of multiple quantiles the full predictive distribution can be constructed. Quantile regression is particularly useful where data do not follow a simple distribution, the distribution shape changes over time with some covariate, which may be the case with ED arrivals.

Gradient Boosting Machines (GBM) are a tree-based machine learning model for regression and classification. Here we produce probabilistic forecasts of ED arrivals via multiple quantile regression using GBMs as implemented in the R package `gbm` (Greenwell et al., 2020). GBMs are a best-in-class algorithm for similar regression problems characterised by modest volumes of training data and possible interactions between input features, and are therefore an appealing choice for ED arrival forecasting. Here GBMs are fit with the following features: hour of the day, day of the week, school holiday, temperature and day of the year. Hyperparameters are chosen via grid search on the training data, and were chosen as follows: 500 trees, an interaction depth of 2, and a shrinkage 0.1. An advantage of GBMs is their ability to learn interactions in comparison to the additive models presented later which require the user to specify possible interactions between inputs. Here, GBM is included as a reference for the performance of *out-of-the-box* machine learning models applied with minimal effort (Ridgeway, 2007). We also explored the possibility of performing linear regression with additive models, but these did not perform as well as GBM so are omited for brevity.

4.7. ADAM: multiple seasonal iETSX

Svetunkov (2022) proposed a framework for dynamic models called the Augmented Dynamic Adaptive Model (ADAM). This framework encompasses ARIMA (Box and Jenkins, 1976), ETS (Hyndman et al., 2008) and regression, supporting multiple frequencies, non-normal distributions and intermittent demand (Svetunkov and Boylan, 2019). Based on this framework, we use the ETS(M,N,M) model with frequencies 24 (hour of the day) and 168 (hour of the week), adding dummy variables for the week of the year, holidays and lagged holidays. This way, we update the hour of day and day of week seasonal indices, keeping the week of year one fixed, thus reducing the number of estimated parameters. Given that the data exhibits randomly occurring zeroes, we use the direct probability model of Svetunkov and Boylan (2019) to treat those values. Finally, given the skewness of the empirical distribution observed in the preliminary analysis, we use the Gamma distribution for the error term. This model can be formulated as a set of the following equations:

$$
\begin{aligned}
&y_t = o_t z_t \\
&\log z_t = \log l_{t-1} + \log s_{1,t-24} + \log s_{2,t-168} + \mathbf{x}_t' \boldsymbol{\beta} + \log\left(1 + \epsilon_t\right) \\
&\log l_t = \log l_{t-1} + \log(1 + \alpha \epsilon_t) \\
&\log s_{1,t} = \log s_{1,t-m} + \log(1 + \gamma_1 \epsilon_t) \\
&\log s_{2,t} = \log s_{2,t-m} + \log(1 + \gamma_2 \epsilon_t) \\
&o_t \sim \text{Bernoulli}\left(\mu_{a,t}\right) \\
&a_t = l_{a,t-1}\left(1 + \epsilon_{a,t}\right) \\
&l_{a,t} = l_{a,t-1}(1 + \alpha_a \epsilon_{a,t}) \\
&\mu_{a,t} = \min(l_{a,t-1}, 1)
\end{aligned}
\quad , \tag{3}
$$

where α, β, γ_1, γ_2 and α_a are the smoothing parameters, defining how adaptive the components of the model should be, l_t is the level component for the demand sizes, $s_{1,t}$ and $s_{2,t}$ are the seasonal components, $\boldsymbol{\beta}$ is the vector of parameters for the explanatory variables, o_t is the binary variable, which is equal to one, when demand occurs and to zero otherwise, $l_{a,t-1}$ is the level component for the occurrence part of the model, and $(1+\epsilon_t) \sim \Gamma(s^{-1}, s)$, where $s = \frac{1}{T}\sum_{t=1}^{T} e_t^2$ is the scale of the distribution. Finally, a_t is an unobservable series, underlying the occurrence part of the model and $(1+\epsilon_{a,t})$ is an unobservable error term for a_t. Svetunkov and Boylan (2019) discuss how to estimate such a model. We expect this model to perform on par with the Poisson regression, potentially outperforming it in some instances, due to the dynamic nature of the model (level and seasonal components). Although the data is integer-valued, we expect that Gamma distribution will be a good approximation for it. If integer-valued quantiles are needed, then rounding up can be done for them (see Appendix 8.1 for the explanation). This model is implemented in the `adam()` function from the `smooth` package (Svetunkov, 2021c) for R and is denoted in our experiment as "ADAM-iETSX".

4.8. GAMLSS

Suppose we assume that our predictive distribution follows a given parametric distribution, as in Poisson regression discussed above. In that case, the forecasting task becomes one of predicting the future values of that distribution's parameters. We then can use Generalised Additive Models for Location, Scale and Shape (GAMLSS). These are the distributional regression models where the parameters are modelled as additive function of explanatory variables. This provides a powerful and flexible framework for probabilistic forecasting, provided that suitable distribution and additive model structures can be found. In practice, this means employing expert judgement and experimenting with various distributions and evaluating their suitability using available training data.

Let $F_t(y_t)$ be a predictive cumulative probability distribution of y_t. In a distributional regression context, $F_t(y_t)$ is modelled via a parametric model, $F(y_t|\boldsymbol{\theta}_t)$, where $\boldsymbol{\theta}_t$ is an m-dimensional vector of parameters. In a GAMLSS framework of Rigby and Stasinopoulos (2005) the elements $j = 1, ..., m$ of $\boldsymbol{\theta}_t$ are modelled as

$$g_j(\theta_{j,t}) = \mathbf{A}_{j,t}\boldsymbol{\beta}_j + \sum_i f_{j,i}(\boldsymbol{x}_t^{S_{j,i}}), \quad \text{for} \quad j = 1, \dots, m, \tag{4}$$

where g_j is a monotonic link function, $\mathbf{A}_{j,t}$ is the t-th row of the design matrix $\mathbf{A}_j$, $\boldsymbol{\beta}_j$ is a vector of regression coefficients, $\boldsymbol{x}_t$ is a d-dimensional vector of covariates and $S_{j,i} \subset \{1, \dots, d\}$ **is ....** If $S_{j,i} = \{1, 3\}$, then following our notation $\boldsymbol{x}_t^{S_{j,i}}$ is a two dimensional vector formed by the first and third elements of $\boldsymbol{x}_t$. Each $f_{j,i}$ is a smooth function, constructed as

$$f_{j,i}(\boldsymbol{x}^{S_{j,i}}) = \sum_{k=1}^{K_{j,i}} b_k^{ji}(\boldsymbol{x}^{S_{j,i}})\beta_k^{ji}, \tag{5}$$

where b_k^{ji} are spline basis functions of dimension $|S_{j,i}|$, while β_k^{ji} are regression coefficients. The smoothness of each $f_{j,i}$ is controlled via ridge penalties, the definition of smoothness being dependent on the type of effect and penalty being used. See Wood (2017) for a detailed introduction to *GAM/GAMLSS* models, smoothing splines bases and penalties.

As our data are counts, the natural starting point is the Poisson distribution with an additive model for $\log \lambda_t$ of the form

$$\log(\lambda_t) = \sum_{i=1}^{7} \beta_i \delta(D_i(t) - i) + \sum_{j=1}^{7} D_j(t) f_j(H(t)) + t f_{\mathrm{Y}}(Y(t)) + f_{\mathrm{Temp}}(Y(t), C_t) \quad . \tag{6}$$

The functions $H(t)$, $D(t)$ and $Y(t)$ return the hour of the day (1–24), day of the week (1–7), and day of the year (1-366) at time t, respectively, and C_t is the temperature at time t. This model is called Poisson-1 in discussions below.

However, experiments on the training data reveal that calibration of forecasts based on the Poisson distribution is poor, suggesting that the shape of the distribution is unsuitable for the present application. In particular, we observe that forecast uncertainty appears to vary depending on the time of day and possibly other explanatory variables. Therefore, we consider more flexible, two-parameter distributions to specify

additive models for both location and scale parameters, specifically the truncated Normal distribution, with truncation at 0. The resulting density forecasts are given by

$$F_t(y_t, \mu_t, \sigma_t) = \frac{\Phi\left(\frac{y_t-\mu_t}{\sigma_t}\right) - \Phi\left(\frac{-y_t}{\sigma_t}\right)}{1 - \Phi\left(\frac{-y_t}{\sigma_t}\right)} \tag{7}$$

with additive models

$$\mu_t = \sum_{i=1}^{10} \beta_i D_i^+(t) + \sum_{j=1}^{10} D_j^+(t) f_j(H(t)) + t f_{\mathrm{Y}}(Y(t)) + f_{\mathrm{Temp}}(Y(t), C_t)$$

$$log(\sigma_t) = \sum_{i=1}^{10} D_i^+(t) f(H(t))$$

for the mean and variance parameters. This model is referred to as NOtr-1 below.

Furthermore, we consider an extension to the additive models for λ_t and μ_t above to incorporate school and public holidays into D. These models are labelled Poisson-2 and NOtr-2. We also performed experiments with the truncated t distribution (Ttr-2) and negative binomial distribution (NBI-2), but these did not result in forecasts as well calibrated as the truncated normal.

5. Forecast performance evaluation

In order to assess the performance of models, we evaluate predictive quantiles at probability levels 0.05 to 0.95 in steps of 0.05 and conditional expectations for 0 to 48 hours ahead produced by each model. We forecast up to 48 hours because this is the operational horizon in the ED, for which it is possible to make short-term changes in the shifts for nurses and doctors. The forecasts are produced every 12 hours for the holdout of 365 days in a rolling origin fashion (Tashman, 2000), resulting in 727 origins. Based on these values, several error measures are calculated to evaluate the performance of models in terms of specific quantiles and expectation. The latter is measured via Root Mean Squared Error (RMSE):

$$\mathrm{RMSE} = \sqrt{\frac{1}{h}\sum_{j=1}^{h} e_{t+j}^2}, \tag{8}$$

where h is the forecast horizon and e_{t+j} is the point forecast error j steps ahead.

The objective of density forecasts is to be as sharp as possible while remaining reliable/calibrated (Gneiting et al., 2007). A forecast is said to be sharp if the predictive distribution has a relatively small spread, indicating low uncertainty, which is valuable to decision-makers provided the forecast is calibrated. Calibration, also called reliability, is the property that forecast probabilities match the observed frequency of realisations. If a forecast is calibrated, then, for example, 20% of observations should fall below the $\alpha = 0.2$ predictive quantile (with some tolerance based on the finite sample size). This property is necessary for forecast probabilities to be used in quantitative decision-making. Calibration is typically evaluated visually using reliability diagrams, which plot the nominal coverage, α, against observed frequency mean ($\mathbf{1}(y_t \leq q_{\alpha,t})$). We use several scores to assess the quantile performance of models.

First, to measure quantile performance, we need to calculate the pinball score, which is a strictly proper score used to evaluate quantile forecasts and is the discrete form of the Continuous Rank Probability Score (Hyndman and Athanasopoulos, 2021). It rewards sharpness and penalises miscalibration, so it measures overall performance. However, calibration should still be verified separately. Furthermore, The Pinball Score for an individual quantile matches the loss function minimised in a quantile regression model. The Pinball Score is given by

$$\mathrm{Pinball} = \frac{1}{T|\mathcal{A}|}\sum_{\alpha\in\mathcal{A}}\sum_{t=1}^{T}\left(q_{\alpha,t} - y_t\right)\left(\mathbf{1}(y_t \leq q_{\alpha,t}) - \alpha\right), \tag{9}$$

where $\mathcal{A} = \{0.05, 0.1, ..., 0.95\}$ is the set of quantiles being estimated.

To compare model performance, and the significance of any apparent difference in performance, we will use skill scores, which can be calculated for any metric via:

$$\text{Skill} = \frac{M_{\text{ref}} - M}{M_{\text{ref}}} \tag{10}$$

where M is the metric's value for the method being considered, M_{ref} is the metric's value for a reference method. The skill score shows by how many percent the reference approach is worse than the one under consideration. We will use bootstrap re-sampling of skill scores to determine if the differences in forecast performance (i.e. positive or negative skill) are significantly different from zero (Efron, 1981). Here we use the best performing simple benchmark, Naive (explained in Subsection 4.1), as the reference model and employ a block-bootstrap with blocks of length 24h to account for temporal correlations of the underlying data (Bergmeir et al., 2016; Hongyi Li and Maddala, 1996).

Finally, we have calculated the computational time for one iteration on the first rolling origin to compare the speed of each function. All functions were re-estimated on each iteration. ADAM and Poisson regression estimated the parameters, taking the ones obtained in the initial model application to the data in the first origin as the pre-initials. This allowed to speed up the computation for these two models. The initial estimation of ADAM took approximately one hour and 25 minutes. Each step in the experiment took the time shown in Section 6.

6. Results

The data is portioned into training (from 2014-04-01 to 2018-02-28) and test (from 2018-03-01 to 2019-02-28) sets, with all model development and hyper-parameter tuning performed using training data only. The rolling origin advances in 12-hour steps and the forecast horizon is set for 48 steps ahead.

Figure 5 presents Pinball score aggregated across forecasting horizons for each quantile. It shows that the difference in performance among the models mainly comes from the middle of the distribution and somewhat from the upper tail. There is very little difference in performance for the lower tail (except for TBATS, which has a consistently higher Pinball value than the other models). This is interesting and reassuring that the better models are better at probabilities that matter more to decision-makers.

Probabilistic forecasts are evaluated following the principle of *sharpness subject to calibration*, meaning that the sharper forecast is prefered provided that it is calibrated. Mis-calibrated forecasts are unsuitable for use in decision-making, so they should be excluded. Calibration is evaluated visually in Figure 6, which highlights a systematic negative bias across all probability levels in many models, with only the truncated normal and t family GAMLSS models (NOtr-1, NOtr-2, Ttr-2) and ADAM-iETSX models showing good calibration across most probability levels. Notably, both benchmarks exhibit nega- tive quantile bias as they struggle to capture the long term trend of increasing arrivals. This could result in poor staffing decisions. This is because the empirical distribution of whole data fails to characterise how arrivals in ED may change over time.

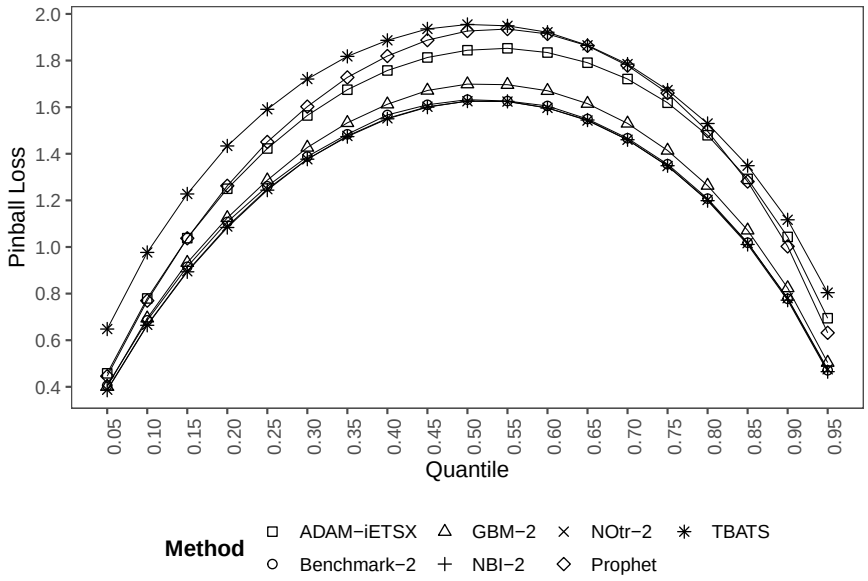

Figure 5: Pinball loss values by quantile. Benchmark-2, NBI-2 and NOtr-2 have similar performance with most improvement of the latter two over the benchmark comming from the lower quantiles. All other models have a greater Pinball score (worse performance) than Benchmark-2 across all quantiles.

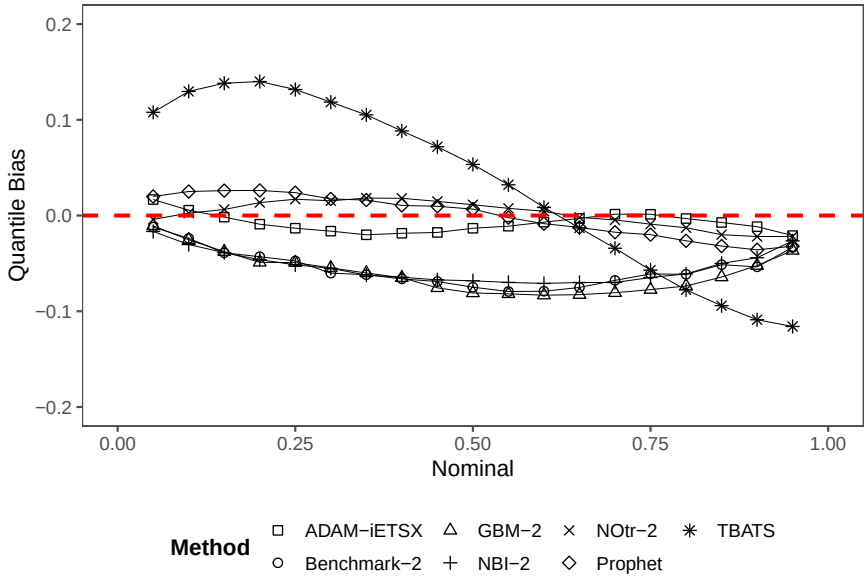

Figure 6: Quantile bias illustrates the difference between the nominal exceedance of prediceted quantiles and the observed frequency of exceedance. If a predictive model is calibrated, quantile bias should be approximately zero accross all probability levels. Here we see that the predictive distributions produced by TBATS are underdispersed (too narrow/over confident); those produced by NBI-2, GBM-2 and Benchmark-2 bias and underpredict accross all probability levels; and that forecasts produced by ADAM-iETSX, NOtr-2 and Prophet are well calibrated.

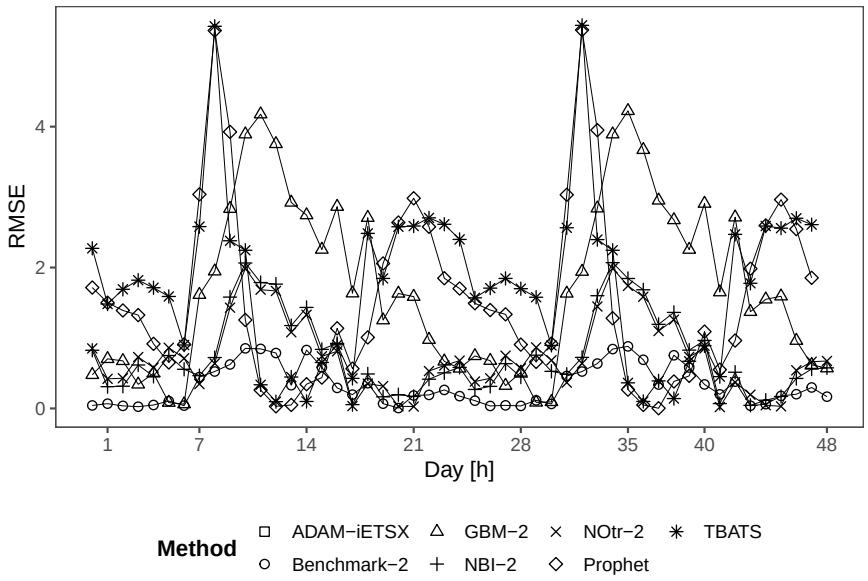

Figure 7: RMSE values over the forecast horizon for forecasts issues at midnight. Attendance during the morning is more challenging to predice resulting in greater RMSE during these hours for all models.

Figure 7 reports the RMSE for each forecast horizon. It illustrates the times of day that are harder to predict – morning pick-up and afternoon peak. We can see that some models perform much better than others on specific forecast horizons. For example, Benchmark does an excellent job in predicting 1 - 5 steps ahead ED arrivals and, in general, doing well in forecasting arrivals in the night. In fact, the Benchmark model performs consistently well in terms of the conditional mean, not making as huge mistakes as, for example, Prophet, TBATS and GBM do for some lead times.

Table 2: Summary of studies in hourly emergency care forecasting

Method	Quantile Bias	Pinball	RMSE	Time (minutes)
NOtr-1	0.0098967	1.222583	0.2675957	451.6620471
ADAM-iETSX	0.0104673	1.417260	0.0896228	92.9348605
NOtr-2	0.0118522	1.208561	0.2675957	86.5895462
Ttr-2	0.0140221	1.210108	0.3324146	956.5532849
Prophet	0.0193799	1.447037	0.2955460	20.6755021
ETS	0.0194389	1.434862	0.0121247	10.7175205
Poisson-1	0.0372137	1.204920	0.0095263	1.4763353
Poisson-2	0.0373884	1.188109	0.0082932	5.0768588
NBI-2	0.0540725	1.206241	0.3830272	1.2791870
Benchmark-2	0.0557392	1.217429	0.2592800	0.0947247
GBM-2	0.0600153	1.261690	1.7770897	602.4317496
TBATS	0.0855702	1.536080	0.4859770	273.0558176
Regression-Poisson	0.0929416	1.293524	0.8490258	67.1401641
Benchmark-1	0.1047874	1.254491	1.0042634	0.3874450

In above figures (Figures 5- 7), we only present the performance of the top seven methods. However, we have evaluated the performance of 14 methods from the test period. Evaluation metrics and computational time (in minutes) for all methods are presented in Table 2. They are ordered by Quantile Bias. The five models identified above have a Quantile Bias of 0.014 or less, which is substantially lower than the next group of forecasts with Quantile Biases of 0.037 and above, ETS being the only exception with a value of 0.019.

One more thing to notice is that the ADAM-iETSX model with rounded up quantiles did not perform better than the simpler one with continuous ones (Table 2). This implies that the rounding is not necessary

in general, but if integer values are needed (for example, to decide how many nurses to have), then using the continuous model and then rounding up the quantiles could be considered a reasonable strategy.

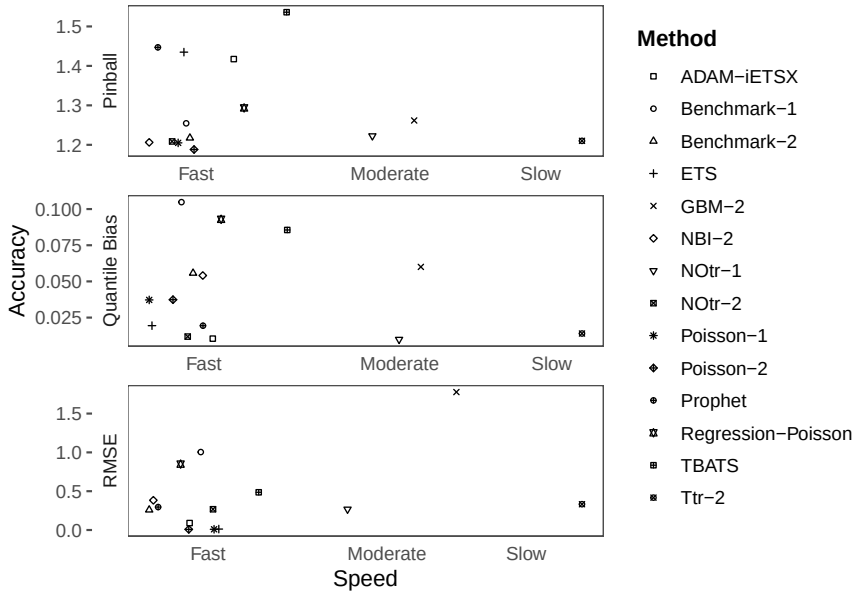

Figure 8: Running time vs. forecast performance

Finally, Figure 8 reports the forecast performance of each approach versus the computational time required to generate the forecast for a given forecasting horizon of 48 hours, presented in Table 2. The X-axis shows the speed of each method presented as slow, moderate or fast. We observe that there is no clear association between speed and accuracy improvement of the models used in this study.

We found that most of the methods considered in this paper fall into the fast category with a different range of performance, depending on the type of error measure. Among the fastest methods, ETS is very competitive when assessed using Quantile Bias and RMSE. Moreover, Poisson-1 and Poisson-2 provide accurate forecasts when evaluated by RMSE and Pinball.

Figure 8 shows that NOtr-2 is the fastest method that provides consistent accuracy assessed via all three accuracy measures. The figure also indicates that while the t family GAMLSS model (Ttr-2) is the slowest method, it has a very good performance across the three presented accuracy measures. Therefore, if the speed is not a major constraint when generating the forecasts, one may employ this approach to generate forecasts.

\textcolor{blue}{The main benefit of the applied models is the use of probabilistic forecasts to inform decision making. Probabilistic forecasts contain all potential future outcomes and help planners to achieve more efficient decisions by not only predicting the most likely outcome but also quantifying the probability of all possible outcomes including extremely high or extremely low arrivals in the Emergency Departments. This information enables decision makers to manage risk associated with low-probability-high-impact events. Based on these forecasts, hospitals can decide how many nurses to have for each shift to make the work of the ED more efficient, e.g. to meet the service level targets set by the National Health Service (NHS) in the United Kingdom.

Practically speaking, the running time can be an important aspect for managers depending on the frequency of generating forecasts. If it is high, one may employ models that do not require a lot of computational time, such as Poisson regression (Poisson 1), sacrificing the accuracy of forecasts. If the frequency is lower and running time is not a big concern, a model like GAMLSS (Ttr-2) should be used.

Probabilistic forecasts of hourly ED arrivals can benefit ED process because it provides the timing and the magnitude of the unlikely scenarios with huge impact on the service delivery, which are fundamental for capacity planning. Probabilistic forecasts can be used to better manage risks of under and over resource allocation, which consequently can reduce both costs and risks for patients, staff and the service as a whole.

7. Conclusion

Short-term forecasting of arrivals at emergency departments is an important element of hospital staff and resource management. Furthermore, due to the asymmetric impact of an excess resource shortage, especially in emergency departments, quantifying forecast uncertainty is also of value as it enables planners to manage associated risks. In this paper, we have developed methods for producing a probabilistic forecast of hourly arrivals up to 48 hours ahead, comparing different state-of-the-art approaches.

Two approaches produced highly accurate, calibrated probabilistic forecasts: a time series model and a model based on distributional regression. The first is ADAM-iESTX, an extension of exponential smoothing incorporating multiple seasonalities, explanatory variables and assuming a Gamma predictive distribution. The second, labeled NOtr-2, regressed the two parameters of a truncated (at zero) normal distribution on the date and time features and temperature. Both approaches produced calibrated probabilistic forecasts, but the point prediction produced by ADAM-iESTX had a lower RMSE than NOtr-2, while NOtr-2 produced forecasts with a lower Pinball score. This suggests that the latter may be preferred if the whole distribution is used in decision-making.

Having compared the performance of a wide range of methods, we make the following observations: the choice of distribution assumed for probabilistic forecasts and choice of model features are as if not more important than the type of model employed; methods based on quantile regression, which do not assume a parametric distribution for forecasts, do not perform as well as those based on parametric distributions; and the best performing models handled the non-negative and skewed nature of the data automatically without the need for post-processing. These observations reflect the characteristics of the data, which is representative of ED arrivals, but determining the extent to which they generalise is beyond the scope of this paper. Furthermore, methods based on continuous-valued distributions are not adversely affected by the fact that the data are integer-valued. Rounding up predictive quantiles to the next integer does not make predictions worse.

Finally, we have found that out-of-the-box models, which require minimal tuning or manual development, do not perform as well as well-considered statistical approaches. The popular TBATS, Prophet and Gradient Boosting Machine algorithms perform poorly compared to ADAM-iETSX and NOtr-2, and even the benchmarks. Of the models requiring a modest amount of user input and expertise, exponential smoothing (ETS) was found to perform well. ETS produces reasonably well-calibrated forecasts, in contrast to the poorly calibrated benchmarks and has highly accurate point forecasts. However, its probabilistic forecasts were considerably worse than NOtr-2 in terms of Pinball score.

The dataset used in this study does not include the period of the COVID-19 pandemic. During COVID-19, the dynamics of ED arrivals has changed substantially. This means that any forecasting model used for ED arrivals forecasting during that period would need to be modified to reflect those changes for that specific period. One of the simplest modifications would be to include a set of dummy variables, capturing different stages of the pandemic. However, this is outside of the scope of this paper and can be considered as a direction for future research.

Probabilistic forecasting opens the door to more sophisticated resource management in healthcare settings by providing decision-makers with uncertainty information and enabling quantitative risk management. Linking forecasts of arrivals with upstream (ambulance call-outs) and downstream (length of stay, medical outcomes) analytics offers an opportunity to improve forecasting skills and may also be necessary to maximise benefits through more holistic decision-making.

Further research is required to investigate the practical benefits of probabilistic forecasts in healthcare and how they can inform planning and decision making. This may require employing discrete event simulation or application to the newsvendor problems. While this study has focused on hourly short-term forecasting, producing longer-term daily forecast (e.g. 180-270 days ahead) is often required by planners to support winter planning in ED and Ambulance services which requires more investigation. Moreover, more research is needed in the forecasting of other important variables such as length of stay, bed occupancy and waiting time, in addition to patient arrivals and admissions. This may require considering the dynamics among various services, including General Practitioners, Emergency Departments, Ambulance and Fire & Rescue services.

8. Appendices

8.1. Quantiles of rounded up random variables

Before proceeding with the proof, we need to define the quantiles of the continuous and rounded up random variables:

$$P\left(y_t < k\right) = 1 - \alpha, \tag{11}$$

and

$$P\left(\lceil y_t \rceil \leq n\right) \geq 1 - \alpha, \tag{12}$$

where n is the quantile of the distribution of rounded up values (the smallest integer number that satisfies the inequality (12)) and k is the quantile of the continuous distribution of the variable.

In order to prove that $n = \lceil k \rceil$, we need to use the following basic property:

$$\lceil y_t \rceil \leq n \iff y_t \leq n, \tag{13}$$

which means that the rounded up value will always be less than or equal to n if and only if the original value is less than or equal to n. Taking into account (13), the probability (12) can be rewritten as:

$$P\left(y_t \leq n\right) \geq 1 - \alpha. \tag{14}$$

Note also that the following is true:

$$P\left(\lceil y_t \rceil \leq n - 1\right) = P\left(y_t \leq n - 1\right) < 1 - \alpha. \tag{15}$$

Taking the inequalities (11), (12), (14) and (15) into account, the following can be summarised:

$$P\left(y_t \leq n - 1\right) < P\left(y_t < k\right) \leq P\left(y_t \leq n\right), \tag{16}$$

which is possible only when $k \in (n-1, n]$, which means that $\lceil k \rceil = n$. So the rounded up quantile of continuous random variable y_t will always be equal to the quantile of the descritised value of y_t:

$$\lceil Q_\alpha(y_t) \rceil = Q_\alpha\left(\lceil y_t \rceil\right). \tag{17}$$

It is also worth noting that the same results can be obtained with the floor function instead of ceiling, following the same logic. So the following equation will hold for all y_t as well:

$$\lfloor Q_\alpha(y_t) \rfloor = Q_\alpha\left(\lfloor y_t \rfloor\right). \tag{18}$$

Reproducibility

R code to produce all results in this paper is available at

References

Al-Azzani, M.A., Davari, S., England, T.J., 2021. An empirical investigation of forecasting methods for ambulance calls-a case study. Health Systems 10, 268–285.

Asheim, A., Bjørnsen, L.P.B.-W., Næss-Pleym, L.E., Uleberg, O., Dale, J., Nilsen, S.M., 2019. Real-time forecasting of emergency department arrivals using prehospital data. BMC emergency medicine 19, 42.

Bergmeir, C., Hyndman, R.J., Benítez, J.M., 2016. Bagging exponential smoothing methods using STL decomposition and box-cox transformation. International Journal of Forecasting.

Bollerslev, T., Engle, R.F., Nelson, D.B., 1994. ARCH models. Handbook of econometrics 4, 2959–3038.

Box, G., Jenkins, G., 1976. Time series analysis: forecasting and control. Holden-day, Oakland, California.

Boyle, J., Jessup, M., Crilly, J., Green, D., Lind, J., Wallis, M., Miller, P., Fitzgerald, G., 2012. Predicting emergency department admissions. Emergency Medicine Journal 29, 358–365.

Cheng, Q., Argon, N.T., Evans, C.S., Liu, Y., Platts-Mills, Timothy F., Ziya, S., 2021b. Forecasting emergency department hourly occupancy using time series analysis. The American Journal of Emergency Medicine 48, 177–182.

Cheng, Q., Argon, N.T., Evans, C.S., Liu, Y., Platts-Mills, Timothy F., Ziya, S., 2021a. Forecasting emergency department hourly occupancy using time series analysis. The American Journal of Emergency Medicine 48, 177–182. doi:https://doi.org/10.1016/j.ajem.2021.04.075

Choudhury, A., Urena, E., 2020. Forecasting hourly emergency department arrival using time series analysis. British Journal of Healthcare Management 26, 34–43.

De Livera, A.M., Hyndman, R.J., Snyder, R.D., 2011. Forecasting time series with complex seasonal patterns using exponential smoothing. Journal of the American Statistical Association 106, 1513–1527.

Derlet, R.W., 2002. Overcrowding in emergency departments: Increased demand and decreased capacity. Annals of emergency medicine 39, 430–432.

Efron, B., 1981. Nonparametric estimates of standard error: The jackknife, the bootstrap and other methods. Biometrika 68, 589–599.

Gneiting, T., Balabdaoui, F., Raftery, A.E., 2007. Probabilistic forecasts, calibration and sharpness. Journal of the Royal Statistical Society: Series B (Statistical Methodology) 69, 243–268.

Greenwell, B., Boehmke, B., Cunningham, J., Developers, G., 2020. Gbm: Generalized boosted regression models.

Gul, M., Celik, E., 2020. An exhaustive review and analysis on applications of statistical forecasting in hospital emergency departments. Health Systems 9, 263–284.

Hertzum, M., 2017. Forecasting hourly patient visits in the emergency department to counteract crowding. The Ergonomics Open Journal 10.

Hongyi Li, G., Maddala, 1996. Bootstrapping time series models. Econometric reviews 15, 115–158.

Hyndman, R., Athanasopoulos, G., Bergmeir, C., Caceres, G., Chhay, L., O'Hara-Wild, M., Petropoulos, F., Razbash, S., Wang, E., Yasmeen, F., 2020. forecast: Forecasting functions for time series and linear models.

Hyndman, R.J., Athanasopoulos, G., 2021. Forecasting: Principles and practice. OTexts.

Hyndman, R.J., Koehler, A.B., Ord, J.K., Snyder, R.D., 2008. Forecasting with Exponential Smoothing. Springer Berlin Heidelberg.

Ibrahim, R., Ye, H., L'Ecuyer, P., Shen, H., 2016. Modeling and forecasting call center arrivals: A literature survey and a case study. International Journal of Forecasting 32, 865–874.

Kalman, R.E., 1960. A New Approach to Linear Filtering and Prediction Problems. Journal of Basic Engineering 82, 35. doi:10.1115/1.3662552

Kim, K., Lee, C., O'Leary, K., Rosenauer, S., Mehrotra, S., 2014. Predicting patient volumes in hospital medicine: A comparative study of different time series forecasting methods. Northwestern University, Illinois, USA, Scientific Report.

La Salle, J.L.G., David, M., Lauret, P., 2021. A new climatology reference model to benchmark probabilistic solar forecasts. Solar Energy 223, 398–414.

Lütkepohl, H., 2013. Vector autoregressive models, in: Handbook of Research Methods and Applications in Empirical Macroeconomics. Edward Elgar Publishing.

McCarthy, M.L., Zeger, S.L., Ding, R., Aronsky, D., Hoot, N.R., Kelen, G.D., 2008. The challenge of predicting demand for emergency department services. Academic Emergency Medicine 15, 337–346.

Morzuch, B.J., Allen, P.G., 2006. Forecasting hospital emergency department arrivals. 26th Annual Symposium on Forecasting, Santander, Spain.

Muhammet, G., Guneri, A.F., 2015. Forecasting patient length of stay in an emergency department by artificial neural networks. Journal of aeronautics and space technologies (Havacilik ve uzay teknolojileri dergisi) 8, 1–6.

O'Hara-Wild, M., Hyndman, R., Wang, E., Caceres, G., 2020. fable: Forecasting models for tidy time serie.

Ramos, M.H., Van Andel, S.J., Pappenberger, F., 2013. Do probabilistic forecasts lead to better decisions? Hydrology and Earth System Sciences 17, 2219–2232.

Ridgeway, G., 2007. Generalized boosted models: A guide to the gbm package. Update 1, 2007.

Rigby, R.A., Stasinopoulos, D.M., 2005. Generalized additive models for location, scale and shape. Journal of the Royal Statistical Society: Series C (Applied Statistics) 54, 554–507.

Rostami-Tabar, B., Ziel, F., 2022. Anticipating special events in emergency department forecasting. International Journal of Forecasting 38, 1197–1213. doi:https://doi.org/10.1016/j.ijforecast.2020.01.001

Schweigler, L.M., Desmond, J.S., McCarthy, M.L., Bukowski, K.J., Ionides, E.L., Younger, J.G., 2009. Forecasting models of emergency department crowding. Academic Emergency Medicine 16, 301–308.

Shi, M., Rostami-Tabar, B., Gartner, D., 2022. Forecasting for unschduled care services: A literature review (Working Paper), Under Review in European Journal of Operational Research.
Svetunkov, I., 2021b. Greybox: Toolbox for model building and forecasting.
Svetunkov, I., 2021c. Smooth: Forecasting using state space models.
Svetunkov, I., 2021a. Forecasting and analytics with ADAM.
Svetunkov, I., 2022. Forecasting and analytics with ADAM.
Svetunkov, I., Boylan, J.E., 2019. Multiplicative state-space models for intermittent time series. doi:10.13140/RG.2.2.35897.0624
Tashman, L.J., 2000. Out-of-sample tests of forecasting accuracy: An analysis and review. International Journal of Forecasting 16, 437–450. doi:10.1016/S0169-2070(00)00065-0
Taylor, S.J., Letham, B., 2018. Forecasting at scale. The American Statistician 72, 37–45.
Team, R.C., 2021. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.
Wales, B., n.d. Wales returning to normal as snow thaws and temperatures rise.
Wood, S.N., 2017. Generalized additive models: An introduction with r. chapman; hall/CRC.
Wright, P.D., Bretthauer, K.M., Côté, M.J., 2006. Reexamining the nurse scheduling problem: Staffing ratios and nursing shortages. Decision Sciences 37, 39–70.