An empirical investigation of forecasting methods for ambulance calls - a case study

Mohamed A K Al-Azzani^a, Soheil Davari^b and Tracey Jane England^c

^aDepartment of Economics and Finance, Durham University Business School, Durham, UK; ^bHertfordshire Business School, University of Hertfordshire, Hatfield, UK; ^cSchool of Mathematics, Cardiff University, Aneurin Bevan University Health Board, Newport, UK

ABSTRACT

A primary goal of emergency services is to minimise the response times to emergencies whilst managing operational costs. This paper is motivated by real data from the Welsh Ambulance Service which in recent years has been criticised for not meeting its eight-minute response target. In this study, four forecasting approaches (ARIMA, Holt Winters, Multiple Regression and Singular Spectrum Analysis (SSA)) are considered to investigate whether they can provide more accurate predictions to the call volume demand (total and by category) than the current approach on a selection of planning horizons (weekly, monthly and 3-monthly). Each method is applied to a training and test set and root mean square error (RMSE) and mean absolute percentage error (MAPE) error statistics are determined. Results showed that ARIMA is the best forecasting method for weekly and monthly prediction of demand and the long-term demand is best predicted using the SSA method.

ARTICLE HISTORY

Received 31 May 2019
Accepted 11 June 2020

KEYWORDS

Emergency services;
forecasting; healthcare

1. Introduction

Emergency healthcare is the first point of contact for millions of patients each year with various symptoms such as chest pain, dizziness, and breathing problems. A primary goal of the emergency medical services (EMS) is to minimise the response times to emergencies whilst managing operational costs (Zhou & Matteson et al., 2017). A one-minute reduction in response time leads to an increase of 24% in the survival chance of patients as highlighted in O'Keeffe et al. (2011). An accurate prediction of the emergency demand is essential in ambulance fleet management as it can result in massive savings in terms of patient lives. The high pressure to have an efficient emergency medical system has led to numerous studies in the literature, such as finding the optimal locations of ambulances (Nickel et al., 2016; Leknes et al., 2016), and crew scheduling (Reeves, 2015). However, a significant step in designing an efficient service is the accurate prediction of demand.

This paper is motivated by real data from the Welsh Ambulance Service. The Welsh Ambulance Services Trust (WAST) is the provider of pre-hospital emergency care across Wales with more than 2,840 staff and 709 operational vehicles. In 2015/16, WAST dealt with more than 450,000 verified incidents and experienced an operating cost of £161 million (WAST, 2016). In recent years, the Welsh Ambulance Service Trust (WAST) has attracted negative press reports for missing the eight-minute ambulance arrival response target time. The government-imposed response time

is used as a performance indicator and WAST has been considered as underperforming in relation to the rest of the UK ambulance services. Although the eight-minute target of the government has been retained for life-threatening calls (65% of calls receiving an eight-minute response), public expectations and rising costs call for an accurate forecasting procedure able to predict the daily number of calls under different planning horizons (weekly, monthly, and three-monthly). Having more accurate call volume predictions will enable better staff planning in both the call handling and ambulance crews. Considering different planning horizons will allow WAST to plan for both short-term (weekly rotas) and long-term (quarterly), matching call volume demand and available capacity. A weekly planning horizon allows WAST to plan their weekly staff pattern addressing the likely peaks in call numbers that may occur during the coming week. In WAST as in other NHS organisations, staff rotas are often drawn up for a four-week period so having a monthly planning horizon would help WAST plan their capacity according to the predicted demand. The three-month planning horizon is of most use in the period running up to the end of the year when the NHS experiences winter pressures. This allows WAST to plan for the predicted number of ambulance calls which helps both WAST and the hospital emergency departments deal with their unscheduled care demand.

Robust forecasting will allow sound decisions to be made for capacity and staffing levels. In order to

provide a more reliable and efficient forecasting method, we performed a comparative study of four efficient forecasting procedures, namely Autoregressive Integrated Moving Averages (ARIMA), Holt-Winters (HW), Multiple Linear Regression (MLR), and the non-parametric procedure of Singular Spectrum Analysis (SSA) to recommend the superior one in terms of accuracy across a selection of planning horizons.

In the heart of any emergency service, the key dilemma is uncertainty. However, once uncertainty is reduced, the scope for efficiency can be enhanced. The main objective of this paper was to understand the stochastic nature of ambulance demand faced by WAST throughout the year and incorporate forecasting models that exceed the accuracy of in-house forecasting methods. The overall contribution of this paper is as a case-study which illustrates the benefit of time series approaches in predicting the call volume in a busy ambulance setting. Further contributions of this paper are:

- Firstly, it builds and validates forecasting models that exceed the accuracy of the current practice.
- Secondly, it helps us to discover the additional benefits of modelling each emergency type alone as opposed to at the aggregate level (one size fits all).

The novelty in this paper lies in the incorporation of distribution-free machine learning methods such as Singular Spectrum Analysis (SSA) to forecasting time series in addition to the conventional parametric methods.

The rest of this paper is organised as follows: [Section 2](#) provides a summary of the relevant publications to our research and defines the contribution of the study described here. The analysis of the WAST data is provided in [Section 3](#) and the forecasting models are presented in [Section 4](#). [Section 5](#) deals with the discussion of the results. The paper concludes with a summary and future research areas in [Section 6](#).

2. Literature review

Matteson et al. (2011) suggest, at the time of their paper, that the current practice for forecasting is often rudimentary with averages used to predict the future number of calls. However, the authors comment that averages which use a small number of data points can produce noisy estimates which may lead to cost and efficiency implications. The authors suggest that more formal time series methods may be able to account for the variation and provide better estimates. In this paper we examine traditional and non-parametric models to see whether they provide better

estimates than the current average-based approach used within WAST.

Regression models have been popular in predictions within the healthcare sector. Kamenetzky et al. (1982) estimated the demand for emergency transportation services with four independent variables, namely the population in the area, employment in the area, and two indicators of socio-economic status. Recently, Lowthian et al. (2011) studied the ambulance calls in Melbourne and the impact of population growth and ageing on this demand. They compared the performance of the log-linear regression model with the linear model and showed that the demand from people above 85 would increase in a period of ten years. Salimi et al. (2016) is another study applying regression models in the healthcare sector where the authors employed a linear Poisson regression model to quantify the association between emergency ambulance dispatches and ambient atmospheric conditions. Channouf et al. (2007) also consider forecasting approaches applied to historical daily and hourly ambulance call data. They consider regression and ARIMA models in their case study. In their daily predictions, the regression model which considers the day of the week and month of the year, performs well over a short planning horizon whilst the ARIMA performs better over a two-week window.

ARIMA models have been one of the widely used methods of forecasting in the literature and are often used when there are evidences of non-stationarity in the data. ARIMA models have been applied in numerous applications such as energy consumption (Sen et al., 2016), air pollution (Kumar & Jain, 2010), and retail sales (Ramos et al., 2015). There are studies of the prediction of ambulance demand in the literature such as Wong and Lai (2014) where the inclusion of temperature in the ARIMA model brought about a reduction of 10% in the forecast error. In another study, Zuidhof (2010) compared the performance of ARIMA with Holt-Winter and regression models for ambulance services in Amsterdam. Recently, Zhou and Matteson et al. (2017) applied a Kernel density estimator to the data in Melbourne, Australia which performed better than an industry practice, an unwarped kernel density estimation and a time-varying Gaussian mixture model. Moreover, some studies have dealt with a hybridisation of ARIMA models and other tools such as neural networks such as Zhang (2003).

The sensitivity of the conventional time series models to violation of their assumptions has led to a recent rise in the interest to apply non-parametric models which offer numerous advantages and are not restricted by the parametric methods assumptions. Neural networks have been a popular non-parametric method used for

forecasting in different fields such as tourism demand modelling (Constantino et al., 2016), call volume prediction (Jalal et al., 2016), renewable energy systems (Sheela & Deepa, 2013), software defect prediction (Arar & Ayan, 2015), and the rice trade (Pakravan et al., 2011). The literature on using Singular Spectrum Analysis (SSA) as a non-parametric method is not as rich as the other methods. Table 1 provides a list of recent publications in this field with their application area. As is shown, there has been an increase in the interest to apply SSA with a promising prediction performance in a variety of fields. The popularity and wide range of applications of SSA can be attributed to the fact that it is not dependent on parametric assumptions such as linearity, stationarity, and normality. The flexibility of SSA makes it a useful method in real-world scenarios as it enables users to model without the need for data transformations which would otherwise result in a loss of information according to Hassani et al. (2009). For further insights regarding SSA, one can refer to Rukhin (2002) and references therein. This holds true for the conventional methods as well. Interested readers can refer to various sources in the literature for further information. For example, Makridakis et al. (1998) provides a very useful introduction to the traditional forecasting methods discussed in this paper.

Whilst many of studies mentioned compare ARIMA, Holt-Winters and SSA, few compare them with regression models or discuss their accuracy over a range of planning horizons. We aim to examine whether one method outperforms the others or if certain methods are more suited to a given forecasting task (e.g. weekly, monthly, 3-monthly). As SSA is a relatively new approach we wanted to see how it performed in comparison to the more traditional forecasting approaches. In summary, the motivation of this paper was to determine a strategic approach to forecasting ambulance calls over a selection of planning horizons.

Ibrahim et al. (2016) provide an extensive review of the forecasting models that have been used to predict

the volume of call arrivals at a call centre. The authors comment on the importance of accurate call volume predictions ahead of determining staff rotas. They also comment on the complexity of the system being modelled and how using appropriate forecasting techniques can lead to more efficient operational decisions. One final point that they mention in their conclusion is that there often exists a gap between academia and industrial practice with a company being unaware of the forecasting techniques that are available to them. We therefore wanted to examine whether more traditional methods could provide a better solution for WAST.

3. Data analysis

The data file contained 209,411 records. Since this is highly sensitive data, it was free of any personal or postcode information. The data recorded the time the call was received (to the nearest second) for the three financial years 2012–2015 (April 2012 – March 2015). The data included the ambulance call volumes received along with how the call was received (999, care helpline, police). The data also included a description of the incident and the incident category as well as the priority type. We also used a separate data set of binary data for regression analysis, namely, public and school holidays, and major sports events. Average daily temperatures were also included in the data and regression model.

Understanding the ambulance demand pattern is an essential step in the prediction process. This section gives an overall view of the call volume demand on an hourly, daily, and monthly basis. Table 2 shows descriptive statistics for the daily call volume and per emergency category.

Moreover, the possible effect of weather variables on the demand for emergency services were investigated using correlation analysis. This included the average daily temperature, and the type of weather such as fog, rain, snow, and sleet. Besides, the correlation of demand on consecutive days was explored to examine the hypothesis that demand is affected by the previous day's demands.

Table 1. Recent applications of SSA.

Author	Year	Location	Benchmark	Application are
Shen and Huang	2008	US	Standard industry practice	Call centre
Hassani et al.	2009	UK	ARIMA and Holt-Winters	UK industrial production
Mamoudvand et al.	2013	Iran	Hyndman and Ullah (2007) model	Mortality rates
Vile et al.	2012	UK	Holt-Winters and ARIMA	Ambulance demand
Hassani et al.	2013	UK	Dynamic factor model	Inflation dynamics
Beneki and Yarmohammadi	2014	Iran	Neural network	Daily exchange rates
Gillard and Knight	2014	UK	Mean method	Ambulance demand
Xiao et al.	2014	UK	ARIMA and neural networks	Air transport demand
Hassani et al.	2015	UK	ARIMA	Tourist arrivals
Silva and Hassani	2015	US	Exponential smoothing, ARIMA	US trade

Table 2. Descriptive statistics.

Data	Mean	Median	Standard Error	Min	Max
Total Demand	191.4	191.0	18.95	121.0	261.0
Emergency A	78.62	79.00	11.25	41.0	120.0
Emergency C	112.8	112.0	13.39	66.0	173.0

This study focuses on two call categories, namely category (A) and category (C) calls. While a category (A) call is a life-threatening and time-critical one with death as a likely outcome, category (C) calls are neither serious nor life-threatening. It should be noted however that even a category (C) call needs a quick rather than a late response which might make the patient status worse. In our study, we separate the analysis of the two categories due to the difference in their significance and pattern. Figure 1 depicts the emergency calls on an hourly basis and is based on the categories (A) and (C). As is shown, while the peak volume of calls for category (A) calls is between 09:00 AM and 11:00 AM, the peak for category (C) calls is between noon and 02:00 PM. The monthly volume of calls

between 2012 and 2015 is also shown in Figure 2 which shows the increasing number of calls towards the end of each year for 2012 and 2013. Generally, whilst the highest demands are observed in December and early January, demands in August and November are the lowest. Moreover, there appears to be a drop in the number of calls during 2014. An additional finding is the relatively higher number of calls during the weekend compared to the weekdays.

We performed an additional analysis on the daily call volumes for the whole period, as is shown in Figure 3 which depicts a slightly downward trend in the number of emergency service demand in the period between February 2012 and April 2015.

The daily ambulance call volume is not only subject to special day effects and events, it is also subject to usual daily and monthly effects. Figure 4 shows the Box-plots for the average daily demand for the months. The figure demonstrates relatively more volatile demand for the winter months with December demand as the peak. In contrast, the months (April-

Figure 1. Total calls on hourly basis by emergency type. **Figure 2.** Average daily demand by month.

Figure 3. Call volumes (daily).

Figure 4. Box-plot of monthly demands.

July) seem more stable. This finding will inform the multiple regression model for the purpose of forecasting.

In order to further understand the underlying trend and seasonality within the data, decomposition was used to divide the data into its constituent parts. Figures 5-7 show the decomposition plots for the daily call volume and the Category A and C categories respectively.

In order to gain some insight into the inter-relationship of factors, a series of correlation analyses was computed for consecutive days of ambulance demand up to the 7th day. Moreover, we carried out some analyses on the correlation between the daily ambulance demand by emergency type with weather conditions, and average daily temperature in degrees Celsius. Results show that the correlation between ambulance demand on consecutive days up to seven days is significant at 0.01 level. In terms of weather conditions, there was no significant

Point-biserial correlation coefficient existing for fog, heavy rain, sleet, light snow and heavy snow. However, for temperature, a significant Spearman correlation coefficient of -0.69 is found for emergency category (A) and 0.56 for emergency category (C) which means that temperature is weakly positively correlated with non-serious ambulance calls but negatively correlated with serious ambulance calls.

Considering all the analysis performed in this section, the following insights have been gained:

- There has been a slight downward trend in the daily number of calls in the study period
- While the highest demands are observed in December and early January, demands in August and November are relatively lower
- A positive statistically significant Pearson correlation coefficient (0.01 level) is found for consecutive days ambulance demand up to the 7th day

Decomposition of additive time series

Figure 5. Decomposition plot for the daily call volume.

- There is no significant correlation for weather conditions. However, for temperature, significant Spearman correlation coefficients of -0.69 and 0.56 have been found for emergency categories (A) and (C) respectively.
- This means that temperature is weakly positively correlated with not-serious ambulance calls, but weakly negatively correlated with serious calls.

4. Forecasting models

Before introducing each of the forecasting models we discuss the approach used in terms of dividing the data into two subsets commonly known as training and test sets. The training set is used to build the model and derive the necessary parameters for a chosen forecasting approach. The test set is used to test the model and to determine whether the model provides the required level of accuracy for the data set. If the chosen model provides a good fit to both the training and the test set

it can be used to predict the future number of ambulance calls over the specified planning horizon.

4.1. Forecasting approach

All the analysis was performed on the total daily ambulance call volume demand as well as per emergency category demand. The software package R was utilised for all the analyses (the following packages are used: ggplot2, forecast and Rssa). The analysis and forecasting approach are summarised in Figure 8. Each time series model is built on the first 2.5 years of the data to obtain the model parameters for the required forecasting approach (e.g. Equations (4)–(7) for the Holt-Winters method) and tested on the last six months (using the estimated parameters). Within and out of sample validation are assessed through four error statistics: Root Mean Square Error (RMSE), Mean Absolute Percentage Error (MAPE), Mean Error (ME) and Mean Absolute Error (MAE). Several forecasting horizons are considered for each

Figure 6. Decomposition plot for Category A calls.

model: 7-days, 30-days and 90-days. The within sample validation is calculated using the training set and the out of sample validation is determined using the test set.

Forecasting accuracy measures are statistics that allow us to compare the actual time series vs the fitted time series for different forecasting models. Such measures are not very informative on their own. However, they become much more informative upon comparing them for different models on the same time series. For all forecasting accuracy measures used here, the smaller is the better and suggests less errors. The choice of which measure to use is an executive decision designed to fit a certain policy prescribed by the management.

In order to test the performance of the four procedures, we use the two commonly used measures, called the root mean squared error (RMSE) and mean absolute percentage error (MAPE). Each method was used for three different forecasting periods of seven days, 30 days, and 90 days to examine the method's strength

in forecasting short-term to long-term demands. Assuming n to be the total number of observations in the complete data set and a sample size of $m < n$, these two error statistics are defined as follows:

$$RMSE = \sqrt{\frac{1}{n-m} \sum_{t=m+1}^n (y_t - \hat{y}_t)^2} \quad (1)$$

$$MAPE = \left(\frac{1}{m} \sum_{t=1}^m \frac{y_t - \hat{y}_t}{y_t} \right) * 100 \quad (2)$$

where y_i and $\hat{y}_i$ are the actual observation and the forecast value respectively for time t . The parameter m is the size of the training or test set when the error statistics are determined. The error statistics are calculated for both the training and test sets as can be seen in Table 3. Time series cross validation as a method of determining the forecasting accuracy of each approach was not used in this study. Instead, the model parameters estimated using the training set

Decomposition of additive time series

Figure 7. Decomposition plot for Category C calls.

were applied to the test data and the forecasts and associated error statistics calculated for the second sample.

As well as the error statistics, residual diagnostic testing was conducted. The ACFs of the residuals in both the training and test sets were analysed for each of the four forecasting approaches conducted on the three data sets (Category A calls, Category C calls, all calls).

In the following sections, we will shortly introduce the forecasting models investigated and present their performance for the WAST dataset with different call types and time frames.

4.2. Forecasting methods

In the following sections, we present and apply three conventional parametric forecasting models on the WAST dataset and compare their performance in terms of their errors. These methods include Autoregressive Integrated Moving Averages (ARIMA), Holt-Winters (HW) and a Multiple Linear

Regression (MLR). Then, the Singular Spectrum Analysis (SSA) method will be applied as a non-parametric forecasting procedure.

4.2.1. Autoregressive integrated moving average (ARIMA)

ARIMA models are a popular choice for prediction purposes considering their flexibility. Generally, an ARIMA model is represented as $ARIMA(p, d, q)$ as a mixture of an autoregressive component of order p , a moving average component of order q , and a non-seasonal difference denoted as d . ARIMA models are a general class of models; however, they include random walk, random trend, and exponential smoothing models as special cases (Sen et al., 2016).

Residual diagnostics is an important step in assessment of the ARIMA model validity to make sure that the residuals are independent. In order to test the independence of the model residuals, we performed a Ljung-Box Portmanteau test which detects the

Figure 8. Schematic representation of the approach taken.

Table 3. ARIMA results for the training and test sets.

	RMSE		MAPE (%)	
	Train set	Test set	Train set	Test set
Total demand	17.54	19.83	7.29	8.14
Emergency (A) demand	10.45	13.36	10.82	13.87
Emergency (C) demand	12.95	12.85	9.11	8.84

autocorrelations r_k at lag k using the Q statistics which is defined as:

$$Q = n(n+2) \sum_{k=1}^h \frac{r_k^2}{n-k} \quad (3)$$

Under the null hypothesis that the data is independent, the statistic Q approximately follows a Chi-square distribution with $(h-m)$ degrees of freedom where h is the number of lags the test is computed for, m refers to the number of parameters used in fitting the model, and n is the sample size. The p-values of the Portmanteau test for lags up to ten are plotted in Figure 9. Further details about Portmanteau tests can be found in Makridakis et al. (1998).

The Autocorrelation Function (ACF) of residuals shows no significant spikes which further affirms

independence. The same is true for demand types (A) and (C). Another test we carried out was the out-of-sample validation by splitting the data into training and test sets which led to the conclusion that the training set and the test set capture the general level of the series while consistently miss out the outliers.

The optimal ARIMA model for the data has been found as a seasonal ARIMA (1,1,1) (2,0,0) for the emergency (C) demand, and a non-seasonal ARIMA (1,0,2) for the emergency (A) demand. We divided the data into a training set and a test set and ran the model for both types to see if the ARIMA performs similarly well on both, based on the RMSE and MAPE indices. Table 3 reports on the performance of the model for the training and test sets. Results show that the accuracy of the ARIMA model on both datasets is reasonably close, affirming the out of sample validity. In absolute terms, ARIMA predictions are within 15 calls from the actual demand which is a reasonable error according to the WAST experts. Moreover, results show that on average, while ARIMA overestimates demand for category (C), it underestimates the category (A) demands.

Figure 9. ARIMA total daily demand residuals diagnostic.

4.2.2. Holt-winters

Exponential smoothing methods such as Holt-Winters make use of past values and recent observations with different weights to predict future events. Holt-Winters methods perform well where the historical data exhibits a linear trend or a seasonal pattern and are modelled as additive or multiplicative forms. The choice of the optimal model for a dataset depends on the seasonality of the data. While the additive model is useful when the seasonal variation is relatively constant over time, the multiplicative model is a better choice whenever the seasonal variation increases over time. Holt-Winters models rely on three smoothing parameters, namely the level (L), trend effects (b), and seasonal patterns (S) which are normally determined according to the historical data.

In order to find the better choice for our study, we tested both the additive and multiplicative models on the dataset for the test and training datasets and concluded that using the additive model for the dataset in hand brings about slightly smaller error measures. The fundamental equations of the additive method are given as Equations (4)-(7).

$$L_t = \alpha(Y_t - S_{t-s}) + (1 - \alpha)(L_{t-1} + b_{t-1}) \quad (4)$$

$$b_t = \beta(L_t - L_{t-1}) + (1 - \beta)b_{t-1} \quad (5)$$

$$S_t = \gamma(Y_t - L_t) + (1 - \gamma)S_{t-s} \quad (6)$$

$$F_{t+m} = L_t + mb_t + S_{t-s+m} \quad (7)$$

Here, the parameters α , β , and γ are associated with the level, trend, and seasonality in the data, L_t , b_t , and F_t are the level, trend, and forecast at time t respectively. Moreover, Y_t represents the actual value of the data. Studies showed that the method used to designate the initial vector has very little effect on the accuracy of the predictions obtained when smoothing (Tratar and Strmcnik^ˆ (2016)).

The results of applying the Holt-Winters method are given in Table 4 showing that the accuracy measures of the Holt-Winters method are reasonably close to each other which affirms the out of sample validity. Moreover, from the results, it is observed that on average, Holt-Winters predict demands within 14 calls from the actual demand which is slightly less than the ARIMA model. An additional observation is that, similar to the ARIMA model, the Holt-Winters method underestimates the demand for category (A) calls and overestimates the category (C) calls. As with the ARIMA model, the Autocorrelation Function (ACF) of residuals for the Holt-Winters method shows no significant spikes which further affirms independence.

Table 4. Holt-Winters results for the training and test sets.

	RMSE		MAPE (%)	
	Train set	Test set	Train set	Test set
Total demand	18.02	18.46	7.42	8.00
Emergency (A) demand	11.00	11.38	11.18	12.42
Emergency (C) demand	13.37	12.20	9.31	9.21

4.2.3. Multiple linear regression

The preliminary analysis indicated that the total daily number of ambulance calls is related to the day of the week, year, and special events in Wales. A multiple regression model is an explanatory model that linearly relates the variable of interest to a set of explanatory variables. In our case, we may explain ambulance calls per day by daily and monthly variations. We also considered school and public holidays as possible variables that might be entered into the model. We estimated the regression model as a function of the explanatory variables as Equation (8).

$$y_t = c + \beta_i T + \beta_i PH_t + \beta_i SH_t + \sum_{i=1}^{12} m_i M_{it} + \sum_{j=1}^7 w_j W_{jt} + e_t \quad (8)$$

Where y_t is the estimated value of the daily call volume, c is a constant term, β is the Ordinary Least Square estimate. T is the average day temperature in degrees Celsius. PH is public holidays; SH is specific school holidays while excluding those which coincide with public holidays to avoid multicollinearity. M_{it} , is i th month of the year at time t , W_{jt} , is the j th day of the week at time t , m_i is the i th month Ordinary Least Square (OLS) estimate, w_j is the j th day OLS estimate, e_t is the errors. By assumption, the errors are normally and identically distributed. All our variables except temperature (continuous variable) are binary, such that = 1 if condition is satisfied, 0 if not. As temperature is difficult to predict in the long-term it could affect the accuracy of the forecast model. In analysing the significance of each variable in the regression equation, Temperature was found to be insignificant.

Running the Ljung-Box Portmanteau test for the data results in a p-value of less than 0.01 which rejects the null hypothesis of random errors and calls for an autocorrelation analysis. If autocorrelation is present in the error as in this case, there are three possible approaches that can be used to eliminate the autocorrelation: Cochrane-Orcutt, Hildreth-Lu or first differencing. The first two approaches involve adjusting the regression equation by adding a lagged error term (Cochrane-Orcutt) or performing a transformation (Hildreth-Lu). The final approach requires first differences to be calculated and the regression performed on the differenced data. We used the Cochrane-Orcutt to correct for autocorrelation and this changes Equation (8) to the following one:

Table 5. MLR results for the training and test sets.

	RMSE		MAPE (%)	
	Train set	Test set	Train set	Test set
Total demand	16.88	20.49	6.95	11.97
Emergency (A) demand	10.11	14.54	10.56	15.78
Emergency (C) demand	12.42	15.49	8.61	12.98

$$y_t = c + \beta_i T + \beta_i PH_t + \beta_i SH_t + \sum_{i=1}^{12} m_i M_{it} + \sum_{j=1}^7 w_j W_{jt} + \beta_i e_{t-1} + e_t \quad (9)$$

where e_{t-1} is the lagged error term. The inclusion of the added term to the equation is to extract the systematic information of the non-random errors. Running the Ljung-Box Portmanteau for the residuals of the modified model will result in a p-value greater than 0.1 which confirms the null hypothesis of independent errors. Results of the MLR model is given in Table 5.

4.2.4. Singular spectrum analysis (SSA)

The basic assumptions for using the conventional parametric models do not hold in real life. That is why nonparametric methods have emerged and their application has been increasing recently. Single Spectrum Analysis (SSA) is closely related to a seminal paper by Broomhead and King (1986). It is composed of two stages, called decomposition and reconstruction. In the decomposition stage, the main series is separated into components of individual time series. Such series signify a trend, oscillatory component or random error. In the reconstruction stage, diagnostic plots as well as paired plots are used to select an appropriate number of these time series which are summed up to form the forecast time series. In the course of implementing SSA, the modeller selects two parameters namely, the window length L and the number of time series opted for in the reconstruction phase G . The window length L is commonly selected to be half of the time series length and G is selected based on the number of principal components in the singular values graph. Moreover, the window length refers to the number of lags of Y needed to include in the subsequent SSA steps. For the sake of brevity, we will not cover more theoretical concepts of SSA in this paper. However, interested readers can refer to Golyandina and Zhigljavsky (2013) and references therein for a deeper understanding of the SSA concepts.

Having selected a window length equal to half our sample size, we plotted the singular values of daily ambulance call volumes as well as those for emergency (A) and (C) calls to identify the number of pairs of eigenvectors to retain in the model. In order to avoid over-fitting, the point in the

singular value plot from which there is not a significant variation should be found and the remaining ones should be left out from the model. Following on from Figure 10, retaining 20 eigenvalues for the total ambulance calls is selected (15 and 10 for type (A) and type (C) calls respectively). In this case study, G is 20 for SSA applied to the total ambulance call data and 15 for category (A) calls and 10 for category (C) calls. L is 455 days (half the training data set). Then, the eigenvalues were plotted pairwise to identify the eigenvalue groups associated with seasonality as shown in Figure 11.

It should be noted that although beyond the first principal component, there is no contribution to explaining the series, they might be able to account for the frequencies in the dataset which is needed in the reconstruction stage. Hence, we carried out an additional analysis on the pairwise plots as given in Figures 12–14 which led to the conclusion that there is some seasonality in the dataset. The polygons generated are a sign of seasonality in the demand with their number of sides equal to the seasonal frequencies identified. For example, in the sixth pairwise plot (2nd plot in the 2nd row of Figures 12 and 14, and the 2nd plot in the first row of Figure 13 there are 7 sides which represents the weekly seasonality in each of the three data sets (Category A, total demand and Category C). These components are generally referred to as harmonic components.

Results show that the principal components six and seven account for the weekly cycle in both total demand as well emergency (C) demand. In the emergency type (A), eigenvectors two and three are responsible for the weekly cycles. In all the time series, the eigenvectors one and two account for trend in our time series (the first plot of top row in each of Figures 12–14. After an out-of-sample validation step of the SSA, we reported the results of the SSA forecasting as given in Table 6. The Autocorrelation Function (ACF) of residuals in the training and test set shows no significant spikes which further affirms independence.

From the SSA results, it can be concluded that SSA predicts demands within 15 calls from the actual demand and it overestimated the demand regardless of its type.

5. Results and analysis

As stated earlier, in order to test the performance of the four procedures, we use the two commonly used measures, called the Root Mean Squared Error (RMSE) and Mean Absolute Percentage Error (MAPE). A common problem in any time series analysis study is that using different measures can lead to different choices. Moreover, whenever there is a zero value in the data, these indices can be undefined or infinite. In this study and to have a level of consistency, we gave a priority to the MAPE error statistic.

Figure 10. Singular values of ambulance demand.

Figure 11. Eigenvector plots of the total ambulance demand with their percentage of variation.

Figure 12. Pairwise Eigenvector plots of the type (A) ambulance calls.

Figure 13. Pairwise Eigenvector plots of the type (C) ambulance calls.

Figure 14. Pairwise Eigenvector plots of the total ambulance calls.

Table 6. SSA results for the training and test sets.

	RMSE		MAPE (%)	
	Train set	Test set	Train set	Test set
Total demand	15.66	19.77	6.49	8.05
Emergency (A) demand	9.48	11.86	9.75	12.3
Emergency (C) demand	11.55	12.53	8.12	8.61

Figures 15–17 depict the performance of the four methods for the weekly, monthly and 3-monthly predictions. The actual observations are in grey and the predictions in black. The figures provide a good illustration of how well each method represents the data and predicts across the required planning horizon.

Tables 7–9 summarise the performance of different methods to forecast type (A), type (C) and overall demands. For example, with the 7-day planning horizon, the Holt-Winters approach provides the most accurate forecasts for emergency Category A calls with a MAPE of 10%.

6. Conclusions and future research

This paper compared the performance of three parametric and one non-parametric forecasting method. In terms of the prediction of the total ambulance calls, results show that ARIMA captures the series level with caution ignoring the random peaks or troughs in the data. The performance of SSA was very close to ARIMA where it performed

relatively better for the long-term predictions and falls a little bit behind ARIMA for weekly and monthly predictions. Among the methods applied, Holt-Winters mimics the fluctuations of the data in a cyclical manner. Multiple linear regression was not as successful as the others in terms of capturing the fluctuations in the data and performed the worst in comparison to the other algorithms we used. This may be due in part to the variables that were included in the regression model, for example, daily temperature which will also need to be predicted before its inclusion.

The results are different for prediction of emergency calls of types on their own though. Results of the RMSE measure show while Holt-Winters is a better method to forecast category (A) calls on a weekly and monthly basis, SSA should be preferred when it comes to forecast for the longer period of three months. Applying the MAPE measure, Holt-Winters, ARIMA, and SSA should be used for weekly, monthly, and three-monthly predictions respectively.

Results of the category (C) calls are relatively more conclusive as they suggest ARIMA as the best tool among the others to be used regardless of the time frame for prediction. Results show that the errors of SSA are not considerably different than the ARIMA results and it performs even better for weekly predictions using the RMSE measure.

Figure 15. Comparing the performance of algorithms for the weekly dataset (predictions are drawn as black).

Figure 16. Comparing the performance of algorithms for the monthly dataset (predictions are drawn as **black**).

Figure 17. Comparing the performance of algorithms for the three-monthly dataset (predictions are drawn as **black**).

Table 7. Comparison of methods for emergency (A) calls.

	RMSE			MAPE (%)		
	7 Days	30 Days	90 Days	7 Days	30 Days	90 Days
Current	99	105.4	97.71	118	126	118
ARIMA	10.25	10.85	12.5	10.5	10.9	12.9
HW	9.14	10.7	11.2	10	11.5	11.9
MLR	16.4	17.6	20	16.6	15.7	19.5
SSA	12.2	10.9	10.7	12.5	11.5	11.3
RPI	91%	90%	89%	92%	91%	90.4%
Best	HW	HW	SSA	HW	ARIMA	SSA

Table 8. Comparison of methods for emergency (C) calls.

	RMSE			MAPE (%)		
	7 Days	30 Days	90 Days	7 Days	30 Days	90 Days
Current	66.7	72	63.8	55.7	60	52.7
ARIMA	13.7	9.2	13.26	8.7	6.1	8.9
HW	15	10	15	9.2	6.7	10.1
MLR	20.4	18.33	25.3	15.7	12.6	18.3
SSA	13.2	9.3	13.9	9.2	6.4	9.4
RPI	80%	87%	80%	84%	90%	83%
Best	SSA	ARIMA	ARIMA	ARIMA	ARIMA	ARIMA

Table 9. Comparison of methods for total calls.

	RMSE			MAPE (%)		
	7 Days	30 Days	90 Days	7 Days	30 Days	90 Days
Current	28.86	23.31	29.92	12/67	9.07	11.95
ARIMA	15.77	13.39	19.92	4.42	5.44	7.97
HW	16.56	13.65	20.43	7.10	5.98	8.41
MLR	25.17	19.48	23.08	11.93	8.02	9.30
SSA	16.90	14.21	19.23	6.60	5.67	7.83
Best	ARIMA	ARIMA	SSA	ARIMA	ARIMA	SSA

Although all four offered reasonably small errors and performed much better than the current practice, there have been some differences which can be summarised as follows:

- Forecast accuracies are higher when emergency categories are separated
- None of the methods outperform the others for the three planning horizons considered

The results of this study are of utmost value to WAST as it provides the best techniques to be used for different time frames. The current forecasting models used in WAST is based on averaging the three previous annual figures while matching the day of the week. Matching the day for this year is achieved by taking away today's date from 364. This is repeated for two more years and the average is computed in Excel.

Results showed that ARIMA is the best forecasting method for weekly and monthly prediction of demand compared to the other three algorithms and the long-term demand is best predicted using the SSA method.

Future research could be directed towards adding other forecasting methods such as artificial neural

networks for the sake of comparison. In terms of measuring the accuracy of each forecasting approach used in this study, time series cross validation could be considered. Another research which would be worth pursuing is the analysis using a larger dataset in order to shed light on the performance of these four algorithms for larger datasets.

Disclosure statement

No potential conflict of interest was reported by the authors.

References

- Arar, O. F., & Ayan, K. (2015). Software defect prediction using cost-sensitive neural network. *Applied Soft Computing*, 33, 263–277. <https://doi.org/10.1016/j.asoc.2015.04.045>
- Beneki, C., & Yarmohammadi, M. (2014). Forecasting exchange rates: An optimal approach. *Journal of Systems Science and Complexity*, 27(1), 21–28. <https://doi.org/10.1007/s11424-014-3304-5>
- Broomhead, D. S., & King, G. P. (1986). Extracting qualitative dynamics from experimental data. *Physica D: Nonlinear Phenomena*, 20(2–3), 217–236. [https://doi.org/10.1016/0167-2789\(86\)90031-X](https://doi.org/10.1016/0167-2789(86)90031-X)
- Channouf, N., L'Ecuyer, P., Ingolfsson, A., & Avramidis, A. N. (2007). The application of forecasting techniques to modelling emergency medical system calls in Calgary, Alberta. *Health Care Management Science*, 10(1), 25–45. <https://doi.org/10.1007/s10729-006-9006-3>
- Constantino, H., Fernandes, P., & Teixeira, J. P. (2016). Tourism demand modelling and forecasting with artificial neural network models: The Mozambique case study. *T'ekhne*, 14(2), 113–124. <https://doi.org/10.1016/j.tekhne.2016.04.006>
- Gillard, J., & Knight, V. (2014). Using singular spectrum analysis to obtain staffing level requirements in emergency units. *Journal of the Operational Research Society*, 65(5), 735–746. <https://doi.org/10.1057/jors.2013.41>
- Golyandina, N., & Zhigljavsky, A. (2013). *Singular spectrum analysis for time series*. Springer Science & Business Media.
- Hassani, H., Heravi, S., & Zhigljavsky, A. (2009). Forecasting European industrial production with singular spectrum analysis. *International Journal of Forecasting*, 25(1), 103–118. <https://doi.org/10.1016/j.ijforecast.2008.09.007>
- Hassani, H., Heravi, S., Zhigljavsky, A., & Alexandrovich, A. (2013). Forecasting UK industrial production with multivariate singular spectrum analysis. *Journal of Forecasting*, 32(5), 395–408. <https://doi.org/10.1002/for.2244>
- Hassani, H., Webster, A., Silva, E. S., & Heravi, S. (2015). Forecasting US tourist arrivals using optimal singular spectrum analysis. *Tourism Management*, 46, 322–335. <https://doi.org/10.1016/j.tourman.2014.07.004>
- Hyndman, R. J., & Ullah, M. S. (2007). Robust forecasting of mortality and fertility rates: A functional data approach. *Computational Statistics & Data Analysis*, 51(10), 4942–4956. <https://doi.org/10.1016/j.csda.2006.07.028>
- Ibrahim, R., Ye, H., L'Ecuyer, P., & Shen, H. (2016). Modelling and forecasting call center arrivals: A literature survey and a case study. *International*

- Journal of Forecasting*, 32(3), 865–874. <https://doi.org/10.1016/j.ijforecast.2015.11.012>
- Jalal, M. E., Hosseini, M., & Karlsson, S. (2016). Forecasting incoming call volumes in call centers with recurrent neural networks. *Journal of Business Research*, 69(11), 4811–4814. <https://doi.org/10.1016/j.jbusres.2016.04.035>
- Kamenetzky, R. D., Shuman, L. J., & Wolfe, H. (1982). Estimating need and demand for prehospital care. *Operations Research*, 30(6), 1148–1167. <https://doi.org/10.1287/opre.30.6.1148>
- Kumar, U., & Jain, V. (2010). Arima forecasting of ambient air pollutants (o₃, no, no₂ and co). *Stochastic Environmental Research and Risk Assessment*, 24(5), 751–760. <https://doi.org/10.1007/s00477-009-0361-8>
- Leknes, H., Aartun, E. S., Andersson, H., Christiansen, M., & Granberg, T. A. (2016). Strategic ambulance location for heterogeneous regions. *European Journal of Operational Research*, 260(1), 122–133. <https://doi.org/10.1016/j.ejor.2016.12.020>
- Lowthian, J. A., Jolley, D. J., Curtis, A. J., Currell, A., Cameron, P. A., Stoelwinder, J. U., & McNeil, J. J. (2011). The challenges of population ageing: Accelerating demand for emergency ambulance services by older patients, 1995–2015. *Medical Journal of Australia*, 194(11), 574. <https://doi.org/10.5694/j.1326-5377.2011.tb03107.x>
- Mahmoudvand, R., Alehosseini, F., & Zokaei, M. (2013). Feasibility of singular spectrum analysis in the field of forecasting mortality rate. *Journal of Data Science*, 11, 851–866.
- Makridakis, S., Wheelwright, S. C., & Hyndman, R. J. (1998). *Forecasting Methods and Applications*. John Wiley & Sons.
- Matteson, D. S., McLean, M. W., Woodard, D. B., & Henderson, S. G. (2011). Forecasting Emergency Medical Service Call Arrival Rates. *The Annals of Applied Statistics*, 5(2B)(2B), 1379–1406. <https://doi.org/10.1214/10-AOAS442>
- Nickel, S., Reuter-Oppermann, M., & Saldanha-da Gama, F. (2016). Ambulance location under stochastic demand: A sampling approach. *Operations Research for Health Care*, 8, 24–32. <https://doi.org/10.1016/j.orhc.2015.06.006>
- O’Keeffe, C., Nicholl, J., Turner, J., & Goodacre, S. (2011). Role of ambulance response times in the survival of patients with out-of-hospital cardiac arrest. *Emergency Medicine Journal*, 28(8), 703–706. <https://doi.org/10.1136/emj.2009.086363>
- Pakravan, M. R., Kelashemi, M. K., Alipour, H. R. (2011). Forecasting Iran’s rice imports trend during 2009–2013. *International Journal of Agricultural Management and Development*, 1(1), 39–44. doi: 10.22004/ag.econ.143492
- Ramos, P., Santos, N., & Rebelo, R. (2015). Performance of state space and ARIMA models for consumer retail sales forecasting. *Robotics and Computer-integrated Manufacturing*, 34, 151–163. <https://doi.org/10.1016/j.rcim.2014.12.015>
- Reeves, C. E. *Integrated scheduling for ambulances and ambulance crews*, PhD thesis, Queensland University of Technology, 2015.
- Rukhin, A. L. (2002). *Analysis of time series structure SSA and related techniques*. Taylor and Francis.
- Salimi, F., Henderson, S. B., Morgan, G. G., Jalaludin, B., & Johnston, F. H. (2016). Ambient particulate matter, landscape fire smoke, and emergency ambulance dispatches in Sydney, Australia. *Environment International*, 99, 208–212. <https://doi.org/10.1016/j.envint.2016.11.018>
- Sen, P., Roy, M., & Pal, P. (2016). Application of ARIMA for forecasting energy consumption and GHG emission: A case study of an indian pig iron manufacturing organization. *Energy*, 116(1), 1031–1038. <https://doi.org/10.1016/j.energy.2016.10.068>
- Sheela, K. G., & Deepa, S. (2013). Neural network-based hybrid computing model for wind speed prediction. *Neurocomputing*, 122, 425–429. <https://doi.org/10.1016/j.neucom.2013.06.008>
- Shen, H., & Huang, J. Z. (2008). Inter-day forecasting and intraday updating of call center arrivals. *Manufacturing & Service Operations Management*, 10(3), 391–410. <https://doi.org/10.1287/msom.1070.0179>
- Silva, E. S., & Hassani, H. (2015). On the use of singular spectrum analysis for forecasting us trade before, during and after the 2008 recession. *International Economics*, 141, 34–49. <https://doi.org/10.1016/j.inteco.2014.11.003>
- Tratar, L. F., & Strmcnik, E. (2016). The comparison of holt–winters method and multiple regression method: A case study. *Energy*, 109, 266–276. <https://doi.org/10.1016/j.energy.2016.04.115>
- Vile, J. L., Gillard, J., Harper, P. R., & Knight, V. A. (2012). Predicting ambulance demand using singular spectrum analysis. *Journal of the Operational Research Society*, 63(11), 1556–1565. <https://doi.org/10.1057/jors.2011.160>
- WAST. (2016). *Welsh ambulance services NHS trust annual report 2015/16*.
- Wong, H.-T., & Lai, P.-C. (2014). Weather factors in the short-term forecasting of daily ambulance calls. *International Journal of Biometeorology*, 58(5), 669–678. <https://doi.org/10.1007/s00484-013-0647-x>
- Xiao, Y., Liu, J. J., Hu, Y., Wang, Y., Lai, K. K., & Wang, S. (2014). A neuro-fuzzy combination model based on singular spectrum analysis for air transport demand forecasting. *Journal of Air Transport Management*, 39, 1–11. <https://doi.org/10.1016/j.jairtraman.2014.03.004>
- Zhang, G. P. (2003). Time series forecasting using a hybrid ARIMA and neural network model. *Neurocomputing*, 50, 159–175. [https://doi.org/10.1016/S0925-2312\(01\)00702-0](https://doi.org/10.1016/S0925-2312(01)00702-0)
- Zhou, Z., Matteson, M. S. (2017). Predicting Melbourne ambulance demand using kernel warping. *The Annals of Applied Statistics*, 10(4), 1977–1996. <https://doi.org/10.1214/16-AOAS961>
- Zuidhof, G. M. (2010). *Capacity planning of ambulance services: Statistical analysis, forecasting and staffing* MSc Dissertation, University of Amsterdam.