Hierarchical and grouped time series forecasting: an application to Emergency Medical Services

Bahman Rostami-Tabar[a,*], Rob J. Hyndman[b]

[a] *Cardiff University, Cardiff Business School, United Kingdom, CF10 3EU*

[b] *Monash University, Department of Econometrics and Business Statistics, Australia, VIC 3800*

Abstract

The primary goal of an emergency service is to minimize the ambulance response time. Accurate forecasts of ambulance demand are crucial inputs into the decisions of planners and policymakers on the resource (e.g. staff and fleet) planning and dynamic deployment. Informing those decisions require detailed forecasting of demand at multiple levels of hierarchy from the national and regional levels to the nature of incidents and their priorities. These forecasts are often generated independently by different teams. However, independent forecasts may not add up to the other levels resulting in mismatched forecasts that lead to conflicting decisions and a lack of coordination in the care service. To address this issue, we explore the inherently hierarchical and grouped structure of Emergency Medcine Services. We apply forecast reconciliation methods to generate both point and probabilistic forecasts that are coherent, for the first time in the healthcare literature.

Keywords: Emergency services, healthcare, hierarchical time series forecasting, forecast reconciliation

1. Introduction

Inability to match the resources with the demand in Emergency Medical Services (EMS) results in an overcrowding care system. This is a serious problem causing challenging situations on patient flow, that ultimately leads to increasing patient's length of stay, higher waiting times, low patient satisfaction, high staff pressure, increasing health care costs and inaccuracy in electronic medical record (Ekström et al., 2015; Rostami-Tabar and Ziel, 2022). Demand forecasting in EMS is a critical element that helps to depict various courses of action to avoid the mismatch, which can result in massive savings in terms of patient safety and lives. An accurate forecasting of the daily demand enables planners and decision makers to manage resources to meet anticipated patients, reconfigure units, redeploy staff and vehicles, where necessary. This will have many advantages for patients, staff and the quality of healthcare services provided.

Demand forecasts are typically required at multiple levels of cross-sectional granularities to inform various planning and decision-making processes across the EMS (Hulshof et al., 2012). There are some planning process at the national level (more strategic or long-term) such as workforce resource planning and budgeting; sub-national and regional level (tactical or medium-term) such as temporary capacity expansions, resource sharing and staff-shift scheduling; and healtboard, hospital or station level (operational or short-term) such as planning rosters for staff and ambulance deployment. Demand forecasts might be also required at different

*Corresponding author

Email addresses: `rostami-tabarb@cardiff.ac.uk` (Bahman Rostami-Tabar), `Rob.Hyndman@monash.edu` (Rob J. Hyndman)

level for a specific group of interest such as nature of demand (e.g. type of incidents) or demand priority. Moreover, there exists an inherent hierarchical and grouped structure relationships that typically occurs in EMS data. Demand for emergency services at the country level can be disaggregated in a geographical hierarchy into sub-national, regions and healthboards or divided into groups such as nature of incidents or demand priority.

Despite a large number of studies dedicated to forecasting for Emergency Medical Services (Shi et al., 2022; Gul and Celik, 2020; Ibrahim et al., 2016; Wargon et al., 2009), the main focus has been on producing independent forecasts at a single level, ignoring the inherent hierarchical or grouped structure relationships of the time series demand. However, forecasting the EMS demand across all hierarchical and grouped levels is paramount for an effective planning and decision making. The hierarchical forecasting approach can not only achieve more accurate forecasts than the independent (base) forecasts, but also creates consistent forecasts (Hyndman et al., 2011). Obtaining consistent forecasts at different levels is important for reliable decision-making to avoid making conflicting decisions. With hierarchical forecasting, plans at any level are based on identical forecasts and therefore can be aligned. Implementing and sustaining improvements in EMS requires alignments and coordination between different stakeholders, without which teams operate in isolation that may lead to conflicts, duplication work, rework, or work that runs counter to overall goals to improve hospital-wide patient flow. Hierarchical forecasting framework can also be used as a tools to improve coordination between teams across the care services at the national, sub-national, regional and hospital levels. To our knowledge, there is not only no research in the EMS forecasting to account for the hierarchical and grouped structure of the system but also in the entire field of forecasting for healthcare management (also referred to as non-clinical healthcare).

In this paper, we address this gap by investigating the application of forecast reconciliation methods in the Emergency Medical Services using the daily time series of verified incidents in the Great Britain from 2015 to 2020. The data has a hierarchical and grouped structures with hierarchies at the national, control (sub-national), healthboard (corresponding to regions) and groups by priority and nature of incidents. We compare the point and probabilistic forecast accuracy of the independent forecasts, bottom-up and optimal combination approaches. We use Exponential smoothing, ARIMA and Poisson regression to generated the independent based forecast followed by applying bottom-up and optimal reconciliation approaches. Forecast performance are assessed by the Root Mean Squared Scaled Error (RMSSE) for point forecasting and the Continuous Ranked Probability Score (CRPS) for the probabilistic forecast. This paper complies with the principles of reproducibility (Stodden and Miguez, 2013; Boylan et al., 2015) and utilising fable package (O'Hara-Wild et al., 2022) in R software. Therefore, this study could equally be applied to any healthcare service (e.g. emergency department, primary or social care) subject to the time series demand having a hierarchical and/or grouped structure, which is generally the case in the healthcare sector.

The remainder of this article is structured as follows: In section 2, we provide a brief review of the literature and discuss its limitation to position our work; in section 3, we present the experiment design describing the dataset, forecasting methods and forecast evaluation metrics. In Section 4, we describe the hierarchical and grouped time series forecasting methods for producing point and probabilistic forecasts. In section 5, we present and discuss our results; in section 7, we summarize our findings and present ideas for future research.

2. Research background

Emergency medical services (EMS), also referred to as ambulance, paramedic or pre-hospital emergency services, are a critical component in the delivery of urgent medical care to communities. An effective service delivery requires accurate resource planning that are generally replying on demand forecasts at operational, tactical and strategic levels.

There is a substantial literature on the application of forecasting in the Emergency Medical Services. Various areas have been the main focus. One of the major research topics has been forecasting call volume arrivals. Ibrahim et al. (2016) provide an extensive review of the forecasting models in this context. Another important focus of the literature in EMS is related to forecasting ambulance demand. Although sometimes the definition of the demand is not cleanly stated, however this is generally the demand where a physical resource has been deployed to respond to an incident, which is also referred to attended incidents. Another demand related variable is the verified incidents. These are all incidents that require an action: either send a physical vehicle, deal with via the Clinical Support Desk, get an external (private) provider to respond to it, or send it through to other channels such as police, firefighters or general practitioners. Another domain that has been extensively studied in the literature and it is also close to ours is forecasting Emergency Department attendance. We refer interested readers to Shi et al. (2022), Gul and Celik (2020) and Wargon et al. (2009) for extensive reviews of the literature in these areas. Although crucial to EMS performance, Aringhieri et al. (2017) state that demand forecasting has received limited research attention in the EMS context. In this section, we provide a brief review of studies on forecasting ambulance demand in EMS.

There are generally two main streams of research in this area: i) the first stream is focusing on the application of time series methods and regression approaches on forecasting aggregate ambulance demand as the most widely used approaches (Vile et al., 2012; Sasaki et al., 2010); ii) another stream considers forecasting EMS demand in more finer temporal and geographical granularitties by employing temporal-spatial prediction methods (Zhou and Matteson, 2016; Zhou, 2016). Papers studying the second stream has not used classical regression and time-series approaches like the first one. Our study belongs to the first category, but we briefly review research in both streams.

Sasaki et al. (2010) develop a multivariable regression model to estimate future EMS demands. In addition to the historical demand, the population census for different age groups and counts of the number of companies employing more than five people are included in the regression. The census variables describe groups who may be more likely to need an ambulance. A stepwise ordinary least squares regression analysis with SPSS is used for estimating the parameter and generating forecast. The only performance measure reported in this study is R^2. The research design of this study is not rigorous and the study is not reproducible. Vile et al. (2012) explore using a Singular Spectrum Analysis (SSA) method to generate forecasts of the EMS demand at the national level for 7-day, 14-day, 21-day and 28-day forecast horizons using data provided by the Welsh Ambulance Service Trust (WAST). The performance of this approach is compared with Auto-Regressive Moving Average (ARIMA) and Holt-winter time series methods using Root Mean Squared Error (RMSE). They concluded that point forecasts generated by SSA are more accurate for longer-term, but ARIMA and Holt-winter performance is superior for shorter-term horizons. Vile et al. (2016) further develop a decision support system to integrate forecasts generated by SSA. However, the study does not compare and contrast of the performance of forecasting methods based on some utility measures such as cost, resource utilization or response time. The tool contains options that allow generating forecasts at various levels of granularity, however, it ignore the hierarchical and grouped relationships structure, preventing aligned decision making and coordination.

Haugsbø Hermansen and Mengshoel (2021) investigate forecasting EMS demand in a high spatio-temporal resolution of $1{\times}1$km spatial regions and 1-hr time intervals using total incidents in Oslo, Norway from January 1st, 2015 up to and including February 11th, 2019. They used multi-layer perceptron (MLP) and long short-term memory (LSTM) models to forecast the EMS demand, and compare them to simple aggregation methods and baselines. The point forecast accuracy is evaluated using Mean Absolute Error (MAE) and Mean Squared Error (MSE) and the forecast distribution is measures by Categorical Cross-Entropy. They shows that while Neural Network models perform better in producing point forecast, a distribution baseline method based on spatial distribution of the incidents across all time steps provide more accurate forecast distribution. Zhou (2016) propose three methods based on Gaussian mixture models, kernel density estimation, and kernel warping to predict 4 weeks into future for a 1-km2 spatial region over an hour. Two years of incidents attended from Toronto, Canada (years 2007 and 2008 with 391,296 events) and Melbourne, Australia (years 2011 and 2012 with 696,975 events) are used to build the model and examine the performance on test data using mean negative log likelihood. They show that forecasts generated by the proposed methods are significantly more accurate than the current industry practice, a simple averaging formula. Grekousis and Liu (2019) investigate the combination of spatial analysis methods with data mining techniques based on an improved Hungarian algorithm and MLP neural network to identify the most likely locations of future emergency events. The proposed approach is tested using data of 2851 events attended by the EMS in Athens, Greece over 24 weeks. They show that 23.24% of real emergency events lie within 50 m of the predicted ones and nearly 70% of the real emergency events lie no further away than 150 m, which is rather accurate given the granularity of the problem at the city level.

One of the major limitations of the current forecasting research in healthcare and EMS is the lack of a forecasting methodology that accounts for the inherent hierarchical and grouped structure in data to produce coherent forecasts, enabling aligned planning and decision making and possibly forecast accuracy improvement by exploring the demand relationships between different levels. While this methodology has been developed over the past 10 years (Panagiotelis et al., 2022), it has never been explored in this area. Additionally, we found that most research only deals with point forecast estimation and also reproducibility is still a major concern in this area. Also, given the count nature of Considering these limitations from the literature, we develop a forecasting experiment to fill these gaps.

3. Experiment setup

We are interested in generating forecasts to inform the planning horizon of $ph = 7 to 42$ days, interested by planners in the ambulance services trust. The forecast horizon in this study is $fh = 2 \times ph$ days ahead (2*42 days planning horizon). This is because the planning is generally freezed for ph days and considering a forecast horizon of ph days might not be helpful for planning. While forecasts are generated for $2 \times ph$ days ahead, performance evaluation is only assessed based on the last ph days and not the $2 \times ph$ days. The forecasts are produced for the holdout of 365 days using time series cross-validation (Hyndman and Athanasopoulos, 2021). In the following section, we discuss the dataset, describe the forecasting methods used to generate base forecasts and present the point and probabilistic accuracy measures.

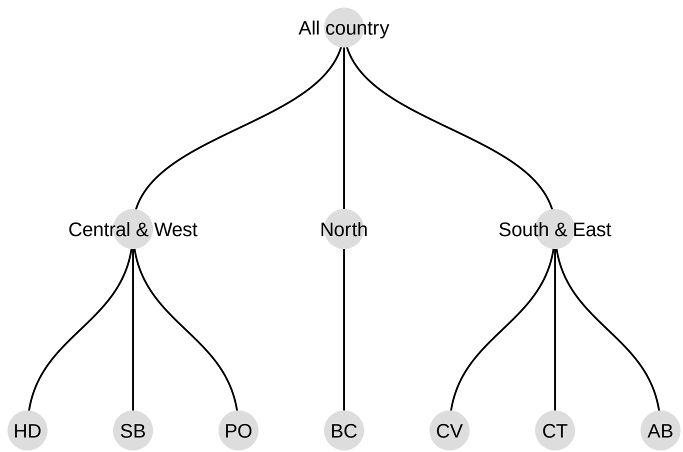
3.1. Data

The dataset used in this study is from a major ambulance service trust in the United Kingdom. It contains information relating to the daily number of attended incidents from 10 October 2015 to 31 July 2019 by nature of incidents, priority, the health board managing the service and the control area (or region). Figure Figure 1 depicts both the hierarchical and grouped structure of the data. Figure Figure 1a illustrates the nested hierarchical structure based on control area and health board and Figure Figure 1b shows the grouped structure by priority and the nature of incident.

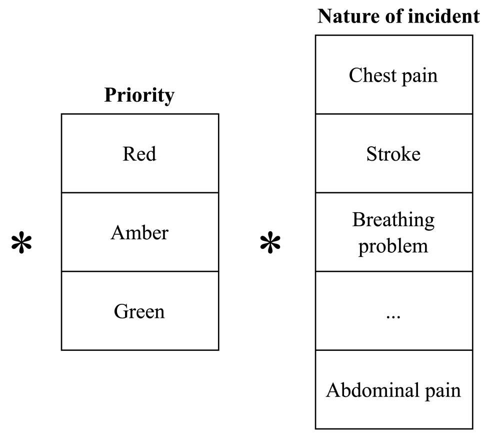
Table Table 1 also displays the structure of data with the total number of series at each level. At the top level, we have the total attended incidents for the country. We can split these total attended incidents by control area, by health board, by priority or by nature of incident. There are 3 control areas breakdown by 7 local health boards. Attended incident data are categorized into 3 priority classes of red, amber and green. There are also 35 different nature of incidents such as chest pain, stroke, breathing problem, etc. In total, across all levels of disaggregation, there are 1530 time series.

Given the total number of time series, direct visual analysis is infeasible. Therefore, we first compute features of all 1530 time series and display the strength of trend and weekly seasonality strength in Figure Figure 2. Each point represents one time series with strength of trend in x-axis and strength of seasonality in y-axis. It is clear that there are some series showing strong trends and/or seasonality, corresponding to series at the higher levels of the hierarchy. The majority show low trend and seasonality with . These are time series belonging to the bottom of the structure, series related to the nature of incidents for each given control, health board and priority.

In addition to displaying the trend and seasonality features, we also highlight few time series at various levels of the aggregation. Figure Figure 3 reveals different information such as trend, seasonality and noise.



(a) Hierarchical structure: Attended incciddents in all country is disaggregated into 3 control areas /regions and then into 7 different healrhboard at the bottom



(b) Grouped structure: Incidents could be grouped into priority (i.e. Red, Amber & Green) and the nature of attenedd incident (i.e. there are 35 different nature of incidents including chest pain , breahtign problems, heart attack, stroke, and son on). The symbol * refers to the crossed attributes between hierarchical and grouped levels.

Figure 1: The hierarchical and grouped structure of attended inciddents (ambulance demand)

Table 1: Number of time series in each level for the hierarchical & grouped structure of attended incidents

Level	Number of series
All country	1
Control	3
Health board	7
Priority	3
Priority * Control	9
Priority * Health board	21
Nature of incident	35
Nature of incident * Control	105
Nature of incident * Health board	245
Priority * Nature of incident	104
Control * Priority * Nature of incident	306
Control * Health board * Priority * Nature of incident (Bottom level)	691
Total	1530

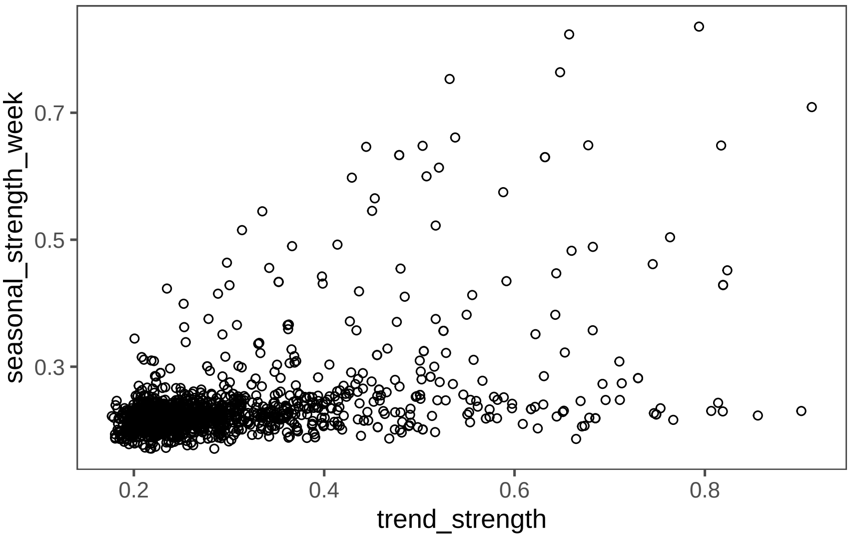


Figure 2: Time series features of attended incidents across all levels (1530 series)

For example, some series depict seasonality and trend, whereas some other series report low volume of attended incidents and entropy, making them more volatile and difficult to forecast. At the level on nature of incidents combined with categories of other levels, there are many series that contain zeros with low counts. As such, the data set represents a diverse set of daily time series patterns.

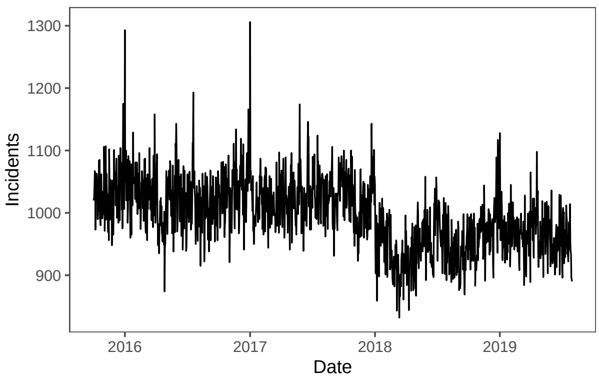
3.2. Forecasting methods

Given the presence of various significant patterns in the past attended incidents, we consider three different forecasting models to generate the base forecasts. Once the base forecasts are produced, hierarchical and grouped time series methods are used to reconcile them across the all levels. We briefly discussed forecasting models in the following sections, and the hierarchical forecasting methods are discussed in Section 4.

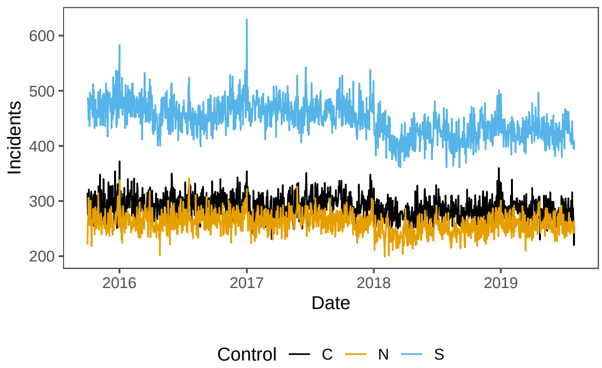
Exponential Smoothing State Space model (ETS): ETS models (Hyndman and Athanasopoulos, 2021) can combine trend, seasonality and error components in a time series through various forms such as additive, multiplicative or mixed. The trend component can be none ("N"), Additive ("A"), damped ("Ad") or multiplicative ("M"). The seasonality can be none ("N"), Additive ("A"), or multiplicative ("M"). The error term can also be additive ("A") or multiplicative ("M"). To forecast the attended incidents at each level, we use the `ETS()` function in the `fable` package (O'Hara-Wild et al., 2022) in R. To identify the best model for a given time series, the ETS function uses the corrected Akaike's Information Criterion (AICc).

Auto-Regressive Integrated Moving Average (ARIMA): ARIMA (Hyndman and Athanasopoulos, 2021) is another effective family of forecasting models that capture autocorrelation in time series. They can model various range of time series considering the stationarity of series, the Auto-Regressive (AR) and moving average (MA) orders. In this study, an automatic ARIMA algorithm implemented using `ARIMA()` function of fable package in R is used to forecast the attended incidents. To select the best model for a given time series, the algorithm uses unit root test, Maximum Likelihood Estimation (MLE) and the AICc.

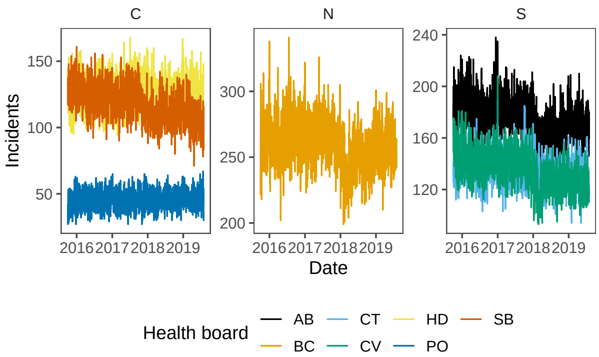
Poisson Regression: Despite the popularity and the relevant of automatic `ETS` and `ARIMA` in this study, but they produce non-integer attended incidents and forecasts might also be negative. However, the number of attended incidents is an integer and non-negative. When using `ETS` and `ARIMA`, a time series transformation approach such as `sqrt()` could be used to generate strictly positive forecasts, however forecasts are still not integer. Another alternative is to use forecasting models that produce integer, non-negative forecasts. One



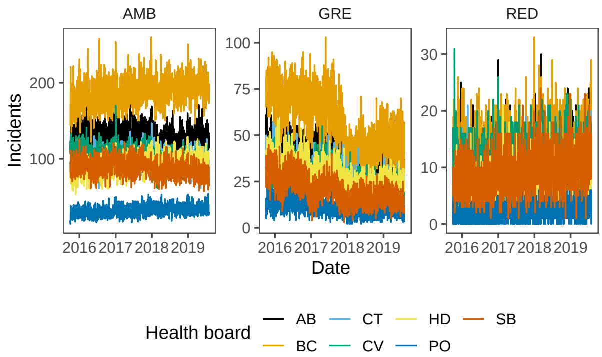
(a) all country



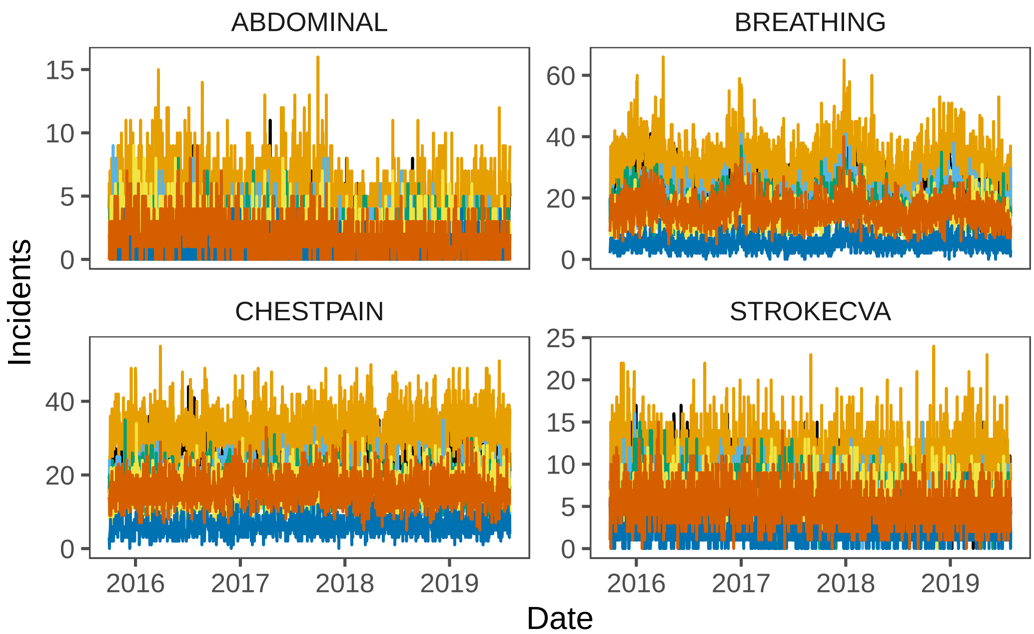
(b) Control area



(c) Health board and control area



(d) Healthboard and priority



(e) Nature of incident and health board

Figure 3: Time series of attended incidents at various levels

of the models that is frequently used in practice is the Poisson regression (see, for example, McCarthy et al., 2008), which can be summarized as

$$\begin{aligned} y_t &\sim \text{Poisson}\,(\lambda_t) \\ \log \lambda_t &= \mathbf{x}_t'\beta \end{aligned} \tag{1}$$

The logarithm in Equation 1 is needed to ensure that the parameter of Poisson distribution is always positive. This model can be estimated via maximization of the likelihood function based on Poisson mass function. For taking into account serial dependence we include a regression on the previous observation. We try to capture the short range serial dependence by a first order autoregressive term and the weekly seasonality by a 7th order autoregressive term. We use the `tscount()` function in the fable package in R, which is a wrapper function written based on `tscount` package. We fit a model to each time series using the function tsglm, with the identity link function, defined by the argument link.

3.3. Performance evaluation

Forecasting performance is evaluated using both point and probabilistic error measures.

The point forecast accuracy is measured via Root Mean Squared Scaled Error (RMSSE):

$$\text{RMSSE} = \sqrt{\text{mean}(q_j^2)}, \tag{2}$$

where,

$$q_j^2 = \frac{e_j^2}{\frac{1}{T-m}\sum_{t=m+1}^{T}(y_t - y_{t-m})^2},$$

e_j is the point forecast error j and $m = 1$ for non-seasonal series and $m = 7$ for daily seasonal series, y_t is the observation for period t and T is the sample size (observations used for training the forecasting model). Smaller RMSSE values suggest more accurate forecasts. Note that the measure is scale-independent, thus allowing us to average the results across series.

To measure the forecast distribution performance, we calculate the Continuous Rank Probability Score (Gneiting and Katzfuss, 2014). It rewards sharpness and penalizes miscalibration, so it measures overall performance.

$$CRPS = \frac{1}{h}\sum_{j=1}^{h}\int_{-\infty}^{\infty}\left(F_j^f(x) - F_j^0(x)\right)^2 dx \tag{3}$$

where $F_j^f(x)$ is the forecasted Cumulative Density Function (CDF) of period j and $F_j^0(x)$ is the true CDF of period j.

4. Hierachcail and grouped time series forecasting techniques

There are many applications in the healthcare and in particular in EMS where a collection of time series is available. These series are generally hierarchically organized based on multiple levels such as area/region, health board and/or are aggregated at different levels in groups based on nature of demand, priority of demand or some other attributes. While series could be strictly hierarchical or only grouped bases on some attributes, in many situation a more complex structures arise when attributes of interest are both nested and crossed, having hierarchical and grouped structure. This is also the case as discussed in Section 3.1

4.1. Independent (base forecast)

A common practice in healthcare (and EMS) to predict hierarchical and grouped series relies on producing independent forecast, also refereed to as base forecast, typically by different teams as the need for such forecasts arise. This is also known as base forecast. We observe n time series at time t, across the entire hierarchical and grouped structure, written as y_t. The base forecasts of y_{T+h} given data $y_1, \dots, y_T$ are denoted by $\hat{y}_h$ for h steps-ahead for all n series ($n = 1530$ in this study). Forecasts generated in this way are nor coherent.

4.2. Reconciliation methos

Traditionally, alternative approaches to produce coherent forecasts for hierarchical and grouped time series involves using bottom-up and top-down methods by generating forecasts at a single level and then aggregate or disaggregate. Top-down requires having a unique structure to disaggregated forecasts generated at top level by proportions. However, given that we have multiple grouped attributes combined with the hierarchical structure, there is no unique way to disaggregate top forecast. Hence the top-down can not be used in this case, so either we can do some kind of reconciliation or must define our own top-down method for each hierarchy. The recommended approach is to use reconciliation. In the following sections, we first discuss some notations and then present bottom-up and forecast reconciliation approach used in this study to generate coherent forecasts.

4.2.1. Notations

Let b_t be a vector of n_b “bottom-level” time series at time t, and let a_t be a corresponding vector of $n_a = n - n_b$ aggregated time series, where

$$a_t = Ab_t$$

and A is the $n_a \times n_b$ “aggregation” matrix specifying how the bottom-level series b_t are to be aggregated to form a_t. The full vector of time series is given by

$$y_t = \begin{bmatrix} a_t \\ b_t \end{bmatrix}.$$

This leads to the $n \times n_b$ “summing” or “structural” matrix given by

$$S = \begin{bmatrix} A \\ I_{n_b} \end{bmatrix}$$

such that $y_t = Sb_t$.

4.2.2. Bottom-up (BU) and linear reconciliation methods

Bottom-Up is a simple approach to generate coherent forecasts. It first involves creating the base forecasts for the bottom level series. These forecasts are then aggregated to the upper levels which results in generating coherent forecasts. BU approach can capture the dynamics of the series at the bottom level, but they may be noisy and difficult to forecast as well. BU approach is limited on using only forecasts at the bottom level and does not utilize all the information available across the hierarchical and grouped structure. Forecast reconciliation approaches fill this gap and combine and reconcile all the base forecasts in order to produce coherent forecasts.

Given the summing matrix and base forecasts, bottom-up and linear reconciliation methods can be written as $\tilde{y}_h = SG\hat{y}_h$ for different matrices G.

Optimal reconciled forecasts are obtained with $G = (S'W^{-1}S)^{-1}W^{-1}$, or $\tilde{y}_h = M\hat{y}_h$, where the $n \times n$ "mapping" matrix is given by $M = S(S'W^{-1}S)^{-1}W^{-1}$, where $\hat{y}_h$ are the h-step forecasts of y_{T+h} given data to time T, and W is an $n \times n$ positive definite matrix. Different choices for W lead to different solutions such as Ordinary Least Square (OLS) , Weighted Least Square (WLS) and Minimum Trace (MinT) (Wickramasuriya et al., 2019). We use the implementation of these methods in the fable package in R in the experiment.

5. Results and discussion

5.1. Point forecast

5.2. Probabilistic forecast

6. Conclusion

References

Aringhieri, R., Bruni, M.E., Khodaparasti, S., van Essen, J.T., 2017. Emergency medical services and beyond: Addressing new challenges through a wide literature review. Computers & Operations Research 78, 349–368.

Boylan, J.E., Goodwin, P., Mohammadipour, M., Syntetos, A.A., 2015. Reproducibility in forecasting research. International Journal of Forecasting 31, 79–90.

Ekström, A., Kurland, L., Farrokhnia, N., Castrén, M., Nordberg, M., 2015. Forecasting emergency department visits using internet data. Annals of emergency medicine 65, 436–442.

Gneiting, T., Katzfuss, M., 2014. Probabilistic forecasting. Annual Review of Statistics and Its Application 1, 125–151.

Grekousis, G., Liu, Y., 2019. Where will the next emergency event occur? predicting ambulance demand in emergency medical services using artificial intelligence. Computers, Environment and Urban Systems 76, 110–122.

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

Haugsbø Hermansen, A., Mengshoel, O.J., 2021. Forecasting ambulance demand using machine learning: A case study from oslo, norway, in: 2021 IEEE Symposium Series on Computational Intelligence (SSCI), pp. 01–10. doi:10.1109/SSCI50451.2021.9659837.

Hulshof, P.J., Kortbeek, N., Boucherie, R.J., Hans, E.W., Bakker, P.J., 2012. Taxonomic classification of planning decisions in health care: a structured review of the state of the art in or/ms. Health systems 1, 129–175.

Hyndman, R.J., Ahmed, R.A., Athanasopoulos, G., Shang, H.L., 2011. Optimal combination forecasts for hierarchical time series. Computational statistics & data analysis 55, 2579–2589.

Hyndman, R.J., Athanasopoulos, G., 2021. Forecasting: principles and practice. OTexts. URL: https://otexts.com/fpp3.

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.

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.

O'Hara-Wild, M., Hyndman, R., Wang, E., Caceres, G., 2022. fable: Forecasting models for tidy time serie. URL: https://fable.tidyverts.org/. r package version 0.3.2.

Panagiotelis, A., Gamakumara, P., Athanasopoulos, G., Hyndman, R.J., 2022. Probabilistic forecast reconciliation: Properties, evaluation and score optimisation. European Journal of Operational Research .

Rostami-Tabar, B., Ziel, F., 2022. Anticipating special events in emergency department forecasting. International Journal of Forecasting 38, 1197–1213. URL: https://www.sciencedirect.com/science/article/pii/S0169207020300017, doi:https://doi.org/10.1016/j.ijforecast.2020.01.001.

Sasaki, S., Comber, A.J., Suzuki, H., Brunsdon, C., 2010. Using genetic algorithms to optimise current and future health planning-the example of ambulance locations. International journal of health geographics 9, 1–10.

Shi, M., Rostami-Tabar, B., Gartner, D., 2022. Forecasting for unplanned care services: A literature review. Working Paper.

Stodden, V., Miguez, S., 2013. Best practices for computational science: Software infrastructure and environments for reproducible and extensible research. Available at SSRN 2322276 .

Vile, J.L., Gillard, J.W., Harper, P.R., Knight, V.A., 2012. Predicting ambulance demand using singular spectrum analysis. Journal of the Operational Research Society 63, 1556–1565.

Vile, J.L., Gillard, J.W., Harper, P.R., Knight, V.A., 2016. Time-dependent stochastic methods for managing and scheduling Emergency Medical Services. Operations Research for Health Care 8, 42–52.

Wargon, M., Guidet, B., Hoang, T., Hejblum, G., 2009. A systematic review of models for forecasting the number of emergency department visits. Emergency Medicine Journal 26, 395–399.

Wickramasuriya, S.L., Athanasopoulos, G., Hyndman, R.J., 2019. Optimal forecast reconciliation for hierarchical and grouped time series through trace minimization. Journal of the American Statistical Association 114, 804–819. URL: https://doi.org/10.1080/01621459.2018.1448825, doi:10.1080/01621459.2018.1448825.

Zhou, Z., 2016. Predicting ambulance demand: Challenges and methods. arXiv preprint arXiv:1606.05363 .

Zhou, Z., Matteson, D.S., 2016. Predicting melbourne ambulance demand using kernel warping. The Annals of Applied Statistics 10, 1977–1996.