Response Letter

ID: THSS-2022-0054
Title: "Probabilistic forecasting of hourly Emergency Department arrivals"
Editors: Samir Chatterjee, Daniel Gartner, Nelson King and Kathy Kotiadis.
Due: January 15th, 2023.

1 Letter to Editor and summary of changes

2 Response to Area Editors

This paper deals with a very interesting problem, forecasting of hourly Emergency Department arrivals. I received two referee reports, both raising important points. Both referees recommend a major revision based on critical comments on the extend of discussion, introduction and conclusion sections. Furthermore, they are both reasonable in the amount and the nature of revisions they inquire.

Please see the attached reviewer comments for further details about necessary revisions.

In conclusion, I advise major revision and expect authors to address all of the comments raised by the referees, either by modifying the paper or by addressing these in a response. Moreover, the language of the paper should be improved. Though this is a major revision, it is certainly a very doable one. I look forward to the improved version of the paper.

Thank you for giving us the opportunity to revise the manuscript. We have addressed all the issues and provided a detailed answers in this document. Our answers are shown in green, while the corrections in the text are in blue.

We have also included a new section for reproducibility. If the paper is accepted for publication, we will include a url link to the GitHub repository to produce all results and the entire paper.

3 Response Letter to Reviewer 1

Summary and general comment The paper presents an innovative model based on generalized additive models and an advanced dynamic model based on exponential smoothing for hourly probabilistic forecasts of emergency department (ED) arrivals for a prediction window of 48 hours. Paper fits well to the scope of Health Systems journal. It suggests a relatively untouched topic among ED arrival forecast studies. Generally, daily basis is studied in the literature. By this aspect, the current research has some merits both for application and theory. I have the following comments which helps improving the paper from my side.

Response: Thank you for providing valuable feedback to improve the paper. We have addressed all comments and detailed responses below.

- Comment 1: The paper presents good results and comparison among different models but it lacks of discussion on what the managerial implications and practical suggestions for hospitals are. The authors must think mostly on the practical outcomes of the numerical results.

  Response:

  Thank you. We have added the following paragraphs in the revised version of the paper to highlight the managerial implications of the research. Please refer to page 16, just before the conclusion section

  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.

- Comment 2: How the models and their results will benefit to the ED process? More discussion can be given o this. Authors mention this concern on the conclusion part but it is limited.

  Response: We have included the following paragraph in the revised version of the paper. Please refer to page 16. We have also provided a more detailed explanation of the implications in the text provided in Comment 1 above.

  Probabilistic forecasts of hourly ED arrivals can benefit ED process because they 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.

- Comment 3: In Table 1: one of the title "Length" is related to the data set? If so, state it clearly.

  Response: We changed the title of the column to "Length of dataset"

- Comment 4: In figure 1, the "minnight" in y-axis of the figure must be corrected. Language.

  Response: We changed "minnight" to "midnight"

- Comment 5: The language is good but it must be checked in terms of English and some linguistic flaws and writing errors. Language.

  Response: Thank you very much for the comment. We have now checked the article and corrected multiple linguistic errors throughout the paper. Our revised manuscript has also been proofread by a native competent English speaker.

4 Response Letter to Referee 2

This referee report discusses an article about forecasts of Emergency Department arrivals per hour. Different forecasting methods with a focus on probabilistic forecasting are compared using real-world UK data from 2014 to 2019. Based on the comments below, I recommend a revision of the paper. The authors focus a topic which is of importance from an application and practitioners' perspective. There is undoubtedly also a methodological contribution to a certain extent, but in my opinion, it is of secondary importance. For example, standard R functions are mainly used for the predictions. In my opinion, the focus of the literature part and the figures should be subject to revisions. Please also address the minor comments given below.

Response: Thank you very much for sharing your time and valuable comments. We have responded to your concerns regarding the literature, figures, the motivation and discussions in the following sections. We believe that they enhance the quality of our article and hope your expectations are met.

- Comment 1: **Motivation**
  Reading the introduction without having the information of the following sections, the intention of some of the contributions are hard to understand. Please make sure that sufficient motivation is given for all contributions in the introduction. Please refer to the Introduction section for details and changes:

  Response: The motivation for this study comes from the practical implications of probabilistic forecasting in informing decision making and planning in Emergency Departments and managing risks. Also, there are multiple gaps identified in the literature which motivated us to undertake this research. These gaps include lack of research on probabilistic ED forecasting, limited number of studies considering multiple seasonal cycles, low volume hourly time series (e.g. some hours with no arrival, zero values) which cause problems for traditional methods leading to negative forecasts, limited length of datasets, and lack of reproducible research in this area. We have highlighted the motivation of this study for all major contributions in the paper and added the following explanation in the introduction section:

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. 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.

- Comment 2: **Literature review**

  In the literature section, the authors focus on existing work regarding hourly Emergency Department forecasts and systematize the papers in an interesting table. I suggest extending the literature part to other forecasting tasks in hospitals and similar forecasting contributions in other areas. For example, in the methods part, the authors introduce methodology from climatology and Facebook without classifying the streams in the literature section.

  Response: We agree with the reviewer that the application of forecasting in the healthcare domain is vast and could be applied in emergency care, primary care, secondary care, etc and also on any possible temporal level such as daily, weekly, monthly etc that corresponds to the decision making. However, given the focus of the current study on Emergency Departments and hourly forecasting, we had to limit the scope of the study to the respective literature to make sure that we do not diverge from the main topic. We have now included one paragraph in the literature to highlight a broader applicability of forecasting models in the healthcare. Regarding using Prophet, we have included this method as a benchmark because we are aware that this method is currently used in many Emergency Departments in the UK., so we thought that it would be a suitable benchmark in that regard. Finally, it seems that the term climatology is confusing because this is just a simple naive approach that uses the empirical distribution to produce quantiles. In the new version of the paper, we replaced the term "climatology" with "Naive" for simplicity.

  We have included the following paragraph at the beginning of section 2.

  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.

Considering Prophet, please refer to section 4.4: 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.

- Comment 3: **Figures**
  In my opinion, some figures are confusing for the reader, because an abundance of data is summarized. For example, the authors could simplify figure 2 (seasonal plot: day of week arrivals) by a focus on maxima/minima per respective color. In figure 4, the dates of outliers can be deleted. The information on the outliers is given in the text. In figure 8 (Speed: Fast), it is difficult to differentiate in between the models.

  Response: We have modified some figures following your suggestion. We believed that these figures collectively help us to better understand the data and illustrate its features from different angles. We used the seasonal plot to highlight weekly seasonality. We agree that the seasonal plot in Figure 2 seems to be busy but this is due to the data being too noisy. In order to address this concern, we replaced the seasonal plot with the subseries plot that can provide similar insights, while being less noisy. We have also removed some dates in Figure 4, to make it less crowded, but we believe that having some dates in the figure could still be helpful to provide quick insights. We also increased the space between points in Figure 8 to improve its readability. Readers can also refer to Table 2 for details on running time and accuracy. Please refer to Figures 2, 4, and 8 in the revised version of the paper for details. We have also included the following explanation in the paper, right after Figure 2:

  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.

- Comment 4: **Discussion**
  The data base of the authors (2014-2019) does not include the COVID-19 pandemic. Please discuss the influence of the pandemic for your results.

  Response: Thank you for the suggestion. We have now included a paragraph in the conclusions to highlight the issue of COVID-19 pandemic.

  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.

- Comment 5: **Minor comments**

  The style of writing is readable, but I did spot some typos, orthographical, or grammatical errors. Please find my minor comments below:

  - P. 2/4: Please provide references
    Response: We have added Ramos et al. (2013) Wright et al. (2006), and Rostami-Tabar and Ziel (2020)
  - P. 3/19: "...for all the 168..."
    Response: corrected
  - P. 3/47: "summarizes"
    Response: corrected
  - Table 1: NO vs. No, 4 h vs. 1h
    Response: These changes are made
  - Table 1: Please implement a new row describing your study
    Response: We have added our study to Table 1.
  - P. 5/17: "... is at an acceptable..."
    Response: corrected
  - Figure 1: Midnight
    Response: this is fixed now
  - Figure 2 caption: Seasonal plot: day of weekly arrivals
    Response: corrected
  - P.9/20: ..."function . ..."
    Response: corrected
  - P. 9/29-37: Please differentiate in between formulae (1), (2), ... and variables 1., 2., 3., ...
    Response: We now distinguish variables from equations. We use () only for equations.
  - P. 10/15: "... simple distribution, of the distribution..." 10/19: Please provide references
    Response: we have corrected the typo and included a new reference, Ridgeway (2007)
  - P. 12/22: "...results...."
    Response: corrected

- P. 12/51: "The Pinball..."
  Response: corrected
- P. 13/11: "... per cent..."
  Response: corrected
- P. 13/33-34: This sentence is confusing
  Response: we have modified the sentence to: "Notably, both benchmarks exhibit negative 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."
- P.17/26: Do you mean "...discrete event simulation..."?
  Response: We changed this to discrete event simulation

- Comment 6: **Conclusion**
  I hope this feedback enables the authors to improve the manuscript. If the editor opts for giving the authors the opportunity to revise, then I will be happy to review a future version of the article.

  Response: Thank you for your comments, we hope that our answers have addressed them adequately.

References

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

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.

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.

Rostami-Tabar, B., Ziel, F., 2020. Anticipating special events in emergency department forecasting. International Journal of Forecasting .

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

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.