2022 44th Annual International Conference of
the IEEE Engineering in Medicine & Biology Society (EMBC)
Scottish Event Campus, Glasgow, UK, July 11-15, 2022

2022 44th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC) | 978-1-7281-2782-8/22/$31.00 ©2022 IEEE | DOI: 10.1109/EMBC48229.2022.9871010

Forecasting of Continuous Vital Sign Using Multivariate Auto-Regressive Models

Søren M. Rasmussen[1], Jesper Mølgaard[2,3], Camilla Haahr-Raunkjær[2,3], Christian S. Meyhoff[3,4,*], Eske Aasvang[2,4,*], Helge B. D. Sørensen[1,*] *Senior Member, IEEE.*

***Abstract*— This project assessed the use of multivariate auto-regressive (MAR) models to create forecasts of continuous vital signs in hospitalized patients. A total of 20 hours continuous (1/60Hz) heart rate and respiration rate from eight postoperative patients, where used to fit a centered MAR model for forecasting in windows of 15 minutes. The model was fitted using Markov Chain Monte Carlo sampling, and the model was evaluated on data from five additional patients. The results demonstrate an average RMSE in the forecast window of 11.4 (SD: 7.30) beats per minute for heart rate and 3.3 (SD:1.3) breaths per minute for respiration rate. These results indicate potential for forecasting vital signs in a clinical setting.**

I. INTRODUCTION

Complications occurring in postoperative patients remain a significant problem despite the last decades advances within the medical sector. About one in six patient develop postoperative complications after elective surgery during the stay in the hospital [1]. With an increasing number of over 300 million surgical procedures performed globally each year, the need for improved care for inpatients is essential for improved patient outcome [2]. The current standard practice for patient monitoring during hospitalization relies on manual measurements at specified intervals performed on the patients according to the escalation protocol Early Warning Score (EWS). EWS has though been criticized due to incompleteness of records, human influence on the results, and the lack of continuity [3].

The Wireless Assessment of Respiratory and circulatory Distress (WARD) is a collaborate project between the Technical University of Denmark, Bispebjerg Hospital and Rigshospitalet that aim to improve on the standard care by implementing continuous monitoring of patients using small biomedical sensors attached on the patient and autonomously interpreting the signals obtained. This way the physiological state of a patient can be assessed at all times with potentially small effort from staff, providing the possibility to attend to other matters. With continuous recording of vital signs the foundation for data analysis becomes greater with respect to standard care, thus enabling more advanced patterns to be interpreted from the data. One obvious potential of this if to make forecasts, thus not only notifying staff when deterioration occurs, but predicting whether it is likely to occur in the near future.

The use of continuous measurements of vital signs to detect future clinical events have shown increasing interest in research. Forkan et al. proposed using hidden Markov models to predict future clinical events defined as multiple vital signs deviating from the expected range [4]. Colopy et al. proposed using Gaussian processes to create patient specific forecasts of vital signs and that were compared to the real values. Deviations from the trajectory were then used to early detect clinical events [5].

Auto-regressive models are no new introduction within the analysis of time series, and the application and use of the model is well known in analysis of stochastic time series, such as stock market or weather forecasts. This project sought to assess the potential of using Multivariate Auto-Regressive (MAR) models to create a forecast projection of vital signs parameters based on past measurements. Forecasting vital signs could help identify deviations of normal physiology that is likely to occur in the near future.

II. THE MULTIVARIATE AUTO-REGRESSIVE MODEL

Consider a set of variables $\mathbf{y} = y_1, ..., y_N$, where each element $y_t = [y_{t1}, ..., y_{tm}]$ is the response at time t, N is the signal length and m the number of modalities in the signal. The response at time, t, as defined by the MAR model is given by

$$y_t = \alpha + \sum_{k=1}^{K} \beta_k \cdot y_{t-k} + \epsilon \quad , \tag{1}$$

where α is a vector of m elements, β_k is a matrix of size $[m, m]$ from the array $\boldsymbol{\beta} = [\beta_k, ..., \beta_K]$. Thus, in the AR-model the value of y_t is given as a linear combination of the previous K elements of $\mathbf{y}$, the intercept α and the weights in $\boldsymbol{\beta}$.

A. Centered Multivariate Auto-Regressive Model

Due to the nature of the vital signs signals, the physiological expectation of the temporal evolution in the signals is that homeostasis will cause the value to return to some patient specific baseline value. It can be advantageous to construct a model that includes the 'pull' towards a baseline value. This can be achieved by creating the MAR model centered around the intercept parameter. A popular implementation is center the model around the mean of the signal, μ_y, where

Corresponding author: Søren Møller Rasmussen (smora@dtu.dk)
[1] Technical University of Denmark, Department of Health Technologies, Kongens Lyngby, Denmark
[2] Department of Anesthesiology, Copenhagen University Hospital - Rigshospitalet, Copenhagen, Denmark
[3] Department of Anesthesia and Intensive Care, Copenhagen University Hospital - Bispebjerg Hospital, Copenhagen, Denmark
[4] Department of Clinical Medicine, University of Copenhagen, Copenhagen, Denmark
[*] Shared last authorship

the response y_t, computed in equation (1), instead comes from

$$y_t = \mu_y + \sum_{k=1}^{K} \beta_k \cdot (y_{t-k} - \mu_y) \quad . \tag{2}$$

As the value of μ_y, when computed from the time series available, does not necessarily reflect the true baseline, this can be fixed globally or as a parameter fitted in the model. The implementation for this project will be elaborated more later.

III. Data

In this project, we used data from an observational study with 500 postoperative cancer patients monitored for up to 4 days after major abdominal surgery. The data were obtained at Rigshospitalet and Bispebjerg Hospital in Copenhagen, Denmark from February 2018 to August 2020. The Danish Data Protection Agency approved the study (2012-58-0004) and the trial registered at http://ClinicalTrials.gov (project: NCT03491137) [6].

Patients were monitored with a single lead ECG patch (Lifetouch Blue), a wrist-worn pulse oximeter (Nonin WristOx$_2$), and a cuff-based blood pressure monitor (TM-2441). From the sensors the following modalities were available: Heart Rate (HR) (1/60 Hz), respiration rate (RR) (1/60 Hz), peripheral oxygen saturation (SpO$_2$) (1/60 Hz) and systolic and diastolic blood pressure (measured every 30 minutes). All data were transmitted to a central server by the Isansys Patient Status Engine.

A subset of the measurements were selected from the cohort, to perform the inference of the parameters in the model and evaluate the predictive accuracy. Only measurements of HR and RR values were used. The data extracted was ensured to not have any missing values for HR or RR in the time period. To fit the model, the subset consisted of 150 minutes of simultaneous HR and RR measurements from eight different patients chosen at random. This gave a total of 20 hours of data for inference. The time series used are shown in Fig. 1.

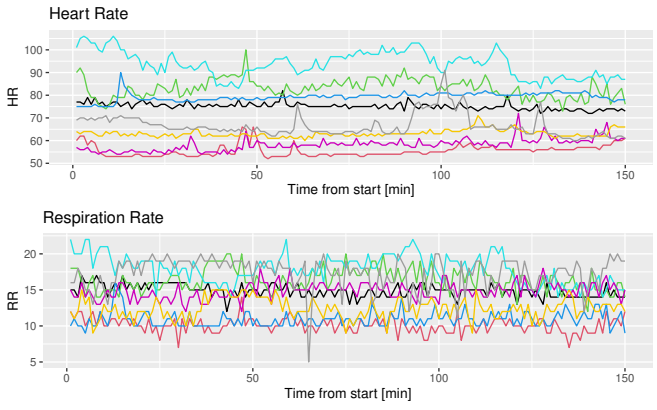

Fig. 1. Time series vital signs for the patients used to fit the MAR model.

To test the predictive accuracy of the model 400 minutes from five different patients were used. The patients in the subset used for evaluation were ensured to be different from the subset used for inference.

IV. Fitting the Multivariate Auto-Regressive Model

The MAR model was constructed as a pooled model. A pooled model defines a model, where the same parameters are fitted across several different data sources, in this case different patients vital signs signals. This results in a single set of model parameters used for all future patients. In the case of the MAR model this means, that the parameters α, $\boldsymbol{\beta}$ and Σ are kept equal for all patients, P. The model is shown in the graph in Fig. 2.

For the model, the priors for the parameters were kept uninformative and given by normal distributions. As the intercept, α, is used as a global baseline, values for this were chosen to reflect common baseline values for the HR and RR. For HR the mean was set to 70 and for RR it was set to 12. All parameters in β had priors set to follow a standard normal distribution.

The following summarizes the model.

$$y_t \sim \mathcal{N}(\mu_t, \Sigma) \tag{3}$$

$$\mu_t = \alpha + \sum_{k=1}^{K} (\beta_k \cdot (y_{t-k} - \alpha)) \quad , \tag{4}$$

with the priors for α and β being

$$\alpha_{HR} \sim \mathcal{N}(70, 10) \tag{5}$$

$$\alpha_{RR} \sim \mathcal{N}(12, 4) \tag{6}$$

$$\beta_* \sim \mathcal{N}(0, 1) \quad . \tag{7}$$

The lag-parameter, K, was set to $K = 20$, reflecting the past 20 minutes of vital signs data. The model was implemented in Stan [8], the data analysis was done in RStudio [9] using the RStan interface [10] to join the two languages.

A. Inference of Model Parameters

The objective for fitting a model is to establish the parameters of the model, θ, to fit the target distribution $p(\theta|y)$. This is done by either exact or approximate inference, depending on the dimensionality of the problem at hand. Due to the computational complexity of exact inference the current problem would be intractable in an exact approach.

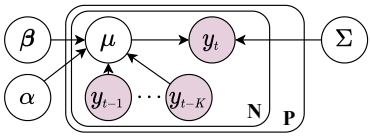

Fig. 2. Probabilistic graphical model of the implemented pooled MAR model. N denotes number of time instances, t, for each patient, P. K is the auto-regressive lag parameter.

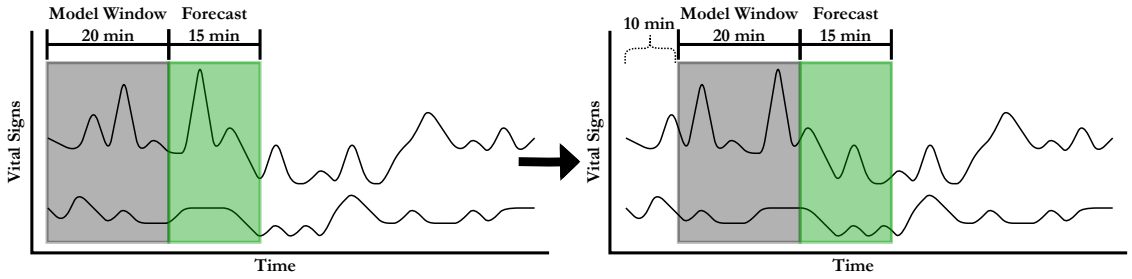

Fig. 3. Illustration of the setup used for evaluating the model on new patients. For each step a forecast (green box) is performed based on the data available in the model window (grey box). The windows are then moved 10 minutes forward and the process is repeated.

Instead approximate inference in the form of Markov Chain Monte Carlo (MCMC)-sampling is used.

MCMC-sampling is a general method based on iteratively drawing samples of θ from approximate distributions and updating these to continuously improve the approximation of the target distribution. The idea is as in Bayesian simulation that the collection of the simulated draws from $p(\theta|y)$ will summarize the posterior density. Hence MCMC-sampling is useful for sampling from Bayesian posterior distributions, where it is intractable to infer θ exactly from $p(\theta|y)$. Due to the random initialization of the sampling algorithm, the samples will have a transition period from initialization to the posterior distribution. To account for this, a warm-up period is defined and the samples from this are rejected.

To ensure that the sampling is stabilized at the posterior distribution, sampling from multiple independent chains were done such that convergence could be quantified by used of the diagnostic measure, $\widehat{R}$. $\widehat{R}$ compares the within-chain variance and the between-chains variance. The idea is that while the individual chains have not mixed and thus not approached the target distribution, the variance of all chains mixed should be larger than that of the chains individually. As the individual chains converge, $\widehat{R} \to 1$ and Vehtari et al. recommends $\widehat{R} < 1.01$ before using the sample [11].

In the used setup, each model was fitted using 4 chains with 2000 iterations in each. Each chain was given a warm-up period of 1000 iterations, thus leaving 1000 for sampling per chain. This provided 4000 posterior samples of the parameters.

B. Evaluating the Predictive Accuracy

To evaluate the model's predictive accuracy, the model was applied to data from 5 unseen patients. A window matching the lag parameter, $K = 20$, was provided to the model to create a forecast of 15 minutes. The forecast segment was compared to the true values within the window. For this, the root mean squared error (RMSE) was used to quantify the accuracy of the expected value in the forecast window with respect to the original signal. The window was moved 10 minutes forward and the process was repeated for the entirety of the time series. The setup for the first to steps is shown in Fig. 3

V. Results

The results of the evaluation of the predictive accuracy of the forecasts are presented in table I. The parameters, α, β and Σ, of the model showed proper convergence with all values of $\widehat{R} < 1.01$. The average RMSE for HR across all patients was 11.4 bpm with the lowest and highest being 0.4 bpm and 32.1 bpm, respectively. For RR the average RMSE was 3.3 brpm with the lowest and highest being 0.9 brpm and 7.4 brpm, respectively. For HR the results in table I show a large difference between patients, where the lowest average RMSE for one patient was 4.7 bpm and the highest 20.5 bpm.

The resulting responses of the MAR model are visualized in Fig. 4. For visual purposes, the plots show the last 50 minutes leading up to the forecasting window and the 15 minutes within the forecasting window.

VI. Discussion

Predicting upcoming deviations in HR and RR is challenging, as the nature of the signals implies rapid changes, that are not known in advance. Sudden physical activation of the patients will lead to changes in their vital signs, that will not be possible to predict before the activation occur. The difficulty to capture this can be seen in Fig. 4, where rapid changes within the forecast window is not captured by the prediction.

The proposed model demonstrates promising results when applied to different patients. The range in the subset used for evaluation shows that both in low and high values of HR and RR the model still provides a good forecast. Though, the variation occurring over multiple days and under different

TABLE I

RMSE between the forecast and the true values for the 5 patients used in the evaluation setup.

RMSE – mean ± std		
Patient	**Heart Rate** [bpm]	**Respiration Rate** [brpm]
#1	20.5 ± 5.78	3.6 ± 1.29
#2	4.7 ± 3.98	3.2 ± 1.37
#3	6.7 ± 3.68	3.6 ± 1.45
#4	14.0 ± 4.97	3.1 ± 1.38
#5	11.1 ± 5.55	2.9 ± 0.94
Average	11.4 ± 7.30	3.3 ± 1.30

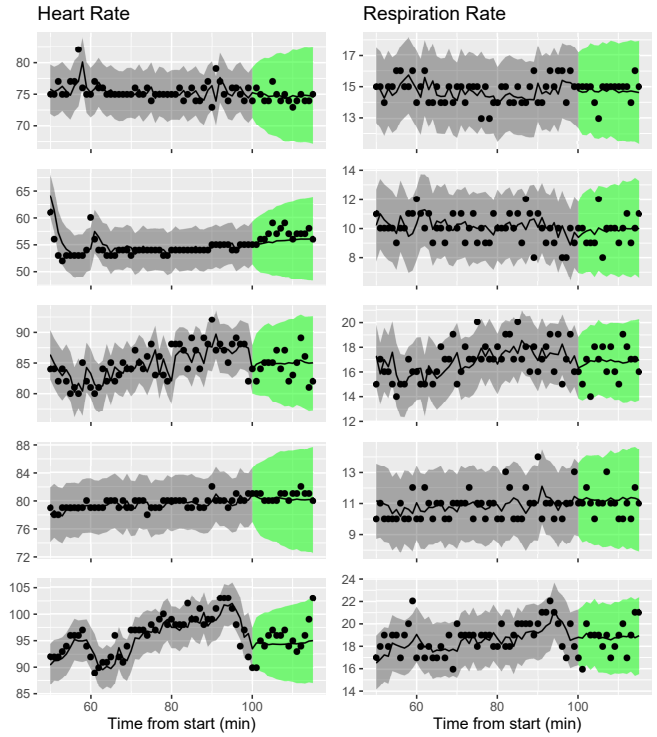

Fig. 4. Visualisation of the response of the hierarchical AR-model fitted to the heart rate and respiration rate data. (●): Time series of the original signal. (–): The expected value of the AR-model. (Grey area): Predictive interval (95%) of the AR-model. (Green area): Predictive interval (95%) of the forecasts from the AR-model.

circumstances has only barely been assessed and there will most likely be rare events, that has not been represented in the evaluation.

In this study, the model was implemented in a pooled construction which has advantages in a clinical setting. As the pooled model relies on a single set of parameters to span all patients, there is no requirement of perform inference of the parameters for each patient, which is resource demanding in computational power when done in an iterative Bayesian approach. This can also be a disadvantage of the pooled model compared to other constructions such as the separate or hierarchical model, where patient specific variations can be built into the model. It could be advantageous if the model has difficulties in fitting to the diversity in data that different patients will present. However, as there is no clear patient specific deviation, use of these models must be held against the increased computational requirements.

Another aspect of the natural representation of vital signs, not included in this project, is heteroscedasticity. The current model assumes the data to be homoscedastic within each modality, i.e. the data has the same variance across patients and temporal location. It becomes clear from the plots in Fig. 4, where plots one, two and four from above have very little variance in the data, and the third and fifth show a large variance, that the naive model assumption of homoscedasticity does not hold in reality. Two solutions to achieve heteroscedasticity could be to model the variance in 1) an auto-regressive manner likewise to the current modelling of the mean, or 2) to model the variance in a hierarchical way, where the parameters α, β and Σ relies on the variance in the data. Future work should explore this.

The construction of a model that creates a forecast will lead to the question of how to use the forecast. As the nature of the signals entails rapid changes, the conception that it will be possible to predict far into the future does not resemble reality. Instead, it could be advantageous to use the forecasts as baseline prediction and evaluate deviation from this based on the true values in a novelty detection setup. This concept has been utilized by Colopy et al. [5] were changes from forecasts created by Gaussian processed were used to detect events of deterioration in hospitalized patients. Quantifying rapid changes from the forecast values could be a way to use the model to detect deviations in a real time setting.

VII. Conclusion

We were able to forecast HR and RR time series based on previous measurement, by implementation of a pooled MAR model. Though there were large deviations in the predictive accuracy in the forecast window between patients, an fairly low RMSE of 11.4 bpm for HR and 3.3 brpm for RR was achieved on average. Event though the work is based on a small subset of patient data, it demonstrates promising results for forecasting vital signs in a clinical setting.

References

[1] International Surgical Outcomes Study group. (2016). Global patient outcomes after elective surgery: prospective cohort study in 27 low-, middle- and high-income countries. British journal of anaesthesia, 117(5),

[2] Weiser, T. G., Haynes, A. B., Molina, G., Lipsitz, S. R., Esquivel, M. M., Uribe-Leitz, T., Fu, R., Azad, T., Chao, T. E., Berry, W. R., Gawande, A. A. (2016). Size and distribution of the global volume of surgery in 2012. Bulletin of the World Health Organization, 94(3).

[3] Pedersen, N. E., Rasmussen, L. S., Petersen, J. A., Gerds, T. A., Østergaard, D., Lippert, A. (2018). A critical assessment of early warning score records in 168,000 patients. J Clin Monit Comput. 32(1).

[4] Forkan, A. R. M.,Khalil, I. (2016). A probabilistic model for early prediction of abnormal clinical events using vital sign correlations in home-based monitoring. IEEE International Conference on Pervasive Computing and Communications

[5] Colopy, G. W., Roberts, S. J., Clifton, D. A. (2019). Gaussian Processes for Personalized Interpretable Volatility Metrics in the Step-Down Ward. IEEE Journal of Biomedical and Health Informatics, 23(3).

[6] Haahr-Raunkjaer C, Mølgaard J, Elvekjaer M, Rasmussen S.M., Achiam M.P., Jorgensen L.N., Søgaard M.I.V., Grønbaek K.K., Oxbøll AB, Sørensen H.B.D., Meyhoff C.S, Aasvang E.K. (2022) Continuous monitoring of vital sign abnormalities; association to clinical complications in 500 postoperative patients. Acta Anaesthesiol Scand.

[7] Gelman, A., Carlin, J.B., Stern, H.S., Dunson, D.B., Vehtari, A., Rubin, D.B. (2013). Bayesian Data Analysis (3rd ed.). Chapman and Hall/CRC.

[8] Stan Development Team. (2019). Stan Modeling Language Users Guide and Reference Manual, 2.28. https://mc-stan.org

[9] R Core Team (2021). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/

[10] Stan Development Team (2020). RStan: the R interface to Stan. R package version 2.21.2. http://mc-stan.org/.

[11] Vehtari, A., Gelman, A., Simpson, D., Carpenter, B. Bürkner, P-C. (2021). Rank-normalization, folding, and localization: An improved $\widehat{R}$ for assessing convergence of MCMC. Bayesian Analysis.