A GENERAL APPROACH FOR DYNAMIC MODELING OF PHYSIOLOGICAL TIME SERIES

Pfeifer M, Lenis G, Dössel O

Institute of Biomedical Engineering, Karlsruhe Institute of Technology (KIT), Germany

publications@ibt.kit.edu

Abstract: *Dynamic modeling of physiological time series represents an auspicious approach in the arena of biomedical signal processing. This study illustrates a new methodology for identifying dynamic models that is based on stationary stochastic processes. The method is applied to time series extracted from the ECG. Simulations of the gained models yield physiologically plausible results .*
Keywords: Time Series, Modeling, State Space, Heart Rate, ARMA

Introduction

Time series (TS) analysis represents a fundamental discipline of scientific and clinical practice in biomedical signal processing. A very sophisticated subarea is the investigation of TS including cardio-electric information such as heart rate dynamics and ventricular repolarization [1].

The majority of methods analyzing electrophysiological information have been primarily directed towards static parameters or dynamic models considering only brief TS extracts (<500 elements). In contrast, this study presents a holistic dynamic approach of modeling physiological TS. Thereby, the methods can be applied to a wide context of biomedical applications.

Methods

Autoregressive-moving average (ARMA) processes are capable of modeling any stationary TS at arbitrary accuracy. However, regarding physiological TS, the hypothesis of stationarity usually has to be rejected by means of statistical tests. Applying a finite number of suitable filter operations provides modified TS which reasonably can be expected to be stationary. The filtered TS are modeled using ARMA processes which are subsequently transferred into a state space representation. A subtle extension of the space state model enables an elimination of the filtering influence and therefore provides a dynamic model for the original TS.

Stationarity and Filtering: Evaluating stationarity for measured TS is no trivial procedure. However, in many cases a visual analysis of the TS plot permits initial tendencies towards stationarity properties. As an example, a linear trend provides clear evidence on instationarity. A more tangible approach is based on the analysis of the sample autocorrelation function (ACF), denoted $r(k)$ [2]. A necessary condition for stationarity is $r(k) \rightarrow 0$ for $k \rightarrow N$ where N represents the number of elements of the TS. In physiological TS the decay of $r(k)$ is typically performed very slowly with respect to k, which is a distinct indication for instationarity. A differencing filter embodies an appropriate approach for eliminating this type of instationarity. Each two consecutive elements are subtracted yielding a filtered TS with improved stationarity properties. Differentiating d times consequently supplies a TS that is sufficiently stationary to be modeled by an ARMA process.

ARMA-Modeling: ARMA(p,q) models provide a parsimonious description of a stationary stochastic process. They are characterized by a Gaussian white noise process $\epsilon_k \sim N(0, \sigma_\epsilon^2)$ and two summations, one for the AR and the second for the MA part:

$$Y_k = \sum_{m=1}^{p} a_m Y_{k-m} + \epsilon_k + \sum_{n=1}^{q} b_n \epsilon_{k-n}. \quad (1)$$

In order to identify a model for a given TS the orders (p, q), the parameters (a_m, b_n) and the noise variance (σ_ϵ^2) must be investigated.

Determining the model orders faces a trade off between a strong adaption to statistical properties of the TS and the principle of parsimony. For an initial appraisal the ACF and the partial ACF (PACF) are evaluated and maximal orders p_{max} and q_{max} are estimated. Conciliating the above mentioned trade off, a minimization of the Akaike Information Criterion (AIC) then yields the orders p and q [2]:

$$\text{AIC} = \log \sigma_r^2 + \frac{2(p+q)}{N}. \quad (2)$$

σ_r^2 is the residual variance between TS and an estimated ARMA(p,q) model and therefore a measure of adaption. The second addend penalizes high model orders.

Deploying the specified orders, the parameters a_m and b_n are estimated based on the prediction-error method. In a last step, the noise variance σ_ϵ^2 must be determined. During a literature research, no quantitative relation between σ_ϵ^2 and the variance of Y_k, denoted σ_Y^2, could be found. However, empirical tests suggested a linear dependence between the two variances. Based on system-theoretical fundamentals, a formula was derived quantifying the linear dependence:

$$\sigma_Y^2 = \int_{-f/2}^{f/2} \left| \frac{1 + \sum_{n=1}^{q} b_n e^{-j2\pi fn}}{1 - \sum_{m=1}^{p} a_m e^{-j2\pi fm}} \right|^2 df \cdot \sigma_\epsilon^2. \quad (3)$$

ARMA Model Diagnostics: Diagnostics represent a crucial step of TS modeling [2]. The quality of adaption of a model is assessed based on a comparison between statistical properties of the TS and the model:

- Visual comparison between the theoretical ACF of the ARMA model and the empirical ACF of the TS
- Assessing whether the residuals represent a realization of a white noise process

If the adaption is appraised to be not sufficient, a new model with increased orders has to be identified.

ARMA and ARIMA Models in State Space: By introducing $r=\max(p,q)$ and setting a_m and c_n to zero for $m=p+1,...,r$ and $n=q+1,...,r$ a transformation into a companion state space realization is viable. Applying a slight extension eliminates the differencing influence of the initial filtering by contributing d integrating operations. For $d=1$ the model is given by:

$$\vec{x}_{k+1} = \begin{bmatrix} 0 & 0 & \ldots & 0 & -a_r & 0 \\ 1 & 0 & \ldots & 0 & -a_{r-1} & 0 \\ 0 & 1 & \ldots & 0 & -a_{r-2} & 0 \\ \vdots & \vdots & & \vdots & \vdots & \vdots \\ 0 & 0 & \ldots & 1 & -a_1 & 0 \\ 0 & 0 & \ldots & 0 & 1 & 1 \end{bmatrix} \vec{x}_k + \begin{bmatrix} b_r - a_r \\ b_{r-1} - a_{r-1} \\ b_{r-2} - a_{r-2} \\ \vdots \\ b_1 - a_1 \\ 0 \end{bmatrix} \epsilon_k \quad (4)$$

$$Y_k = \begin{bmatrix} 0 & 0 & 0 & \ldots & 0 & 1 & 1 \end{bmatrix} \vec{x}_k + \epsilon_k. \quad (5)$$

Formulas (4) and (5) represent a so-called autoregressive integrated moving average (ARIMA) process.

Results

Based on a measured ECG, a RR interval TS containing 2954 elements was constructed. Using this TS the presented methodology is now illustrated. Since the ACF of the TS shows a slow decay, the TS needs to be filtered. The resultant TS represents values of ΔRR and is sufficiently stationary to be modeled with an ARMA process ($d=1$). Analyzing the ACF and PACF yields the maximum orders $(p_{max}, q_{max}) = (5,5)$. Applying this, the AIC is minimized for $(p,q) = (1,4)$ which yields the model:

$$Y_k = 0.9Y_{k-1} + \epsilon_k - 1.1\epsilon_{k-1} + 0.2\epsilon_{k-2} - 0.1\epsilon_{k-4}. \quad (6)$$

For descriptive representation, parameter values are rounded to the first decimal place. Inverse application of (3) yields an input noise variance of $\sigma_\epsilon^2 = 433.5(\text{ms})^2$. Conducting diagnostics raises no doubts on the validity of the model. By using $r=4$ and $d=1$ the ARMA process is transformed into ARIMA state space representation. For validating the gained model, the subsequent illustrations depict model simulation results (red). A simulation is accomplished by simply applying a realization of the input noise process to the model. For comparison, the measured TS (blue) and the simulation results of an identified $(p,q,d)=(20,20,1)$ model (green) are also presented. It is important to note that the objective is not to approximate the measured TS, but to share its dynamics as exact as possible. Fig. 1 suggests that both models provide a suitable reproduction of the original ΔRR dynamics. Recalling formula (3), a correct reproduction of the ΔRR variance by the models is ensured. Fig. 2 confirms the ARIMA(1,4,1) model to
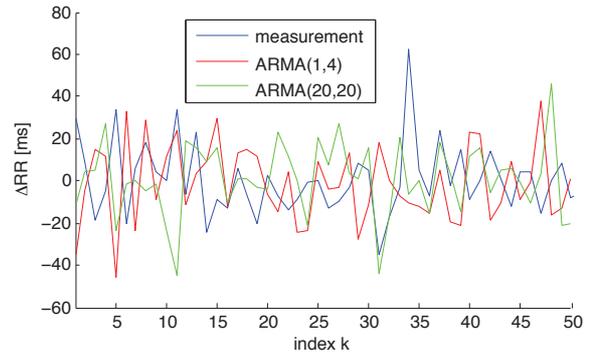

Figure 1: TS of 50 measured and simulated ΔRR intervals.
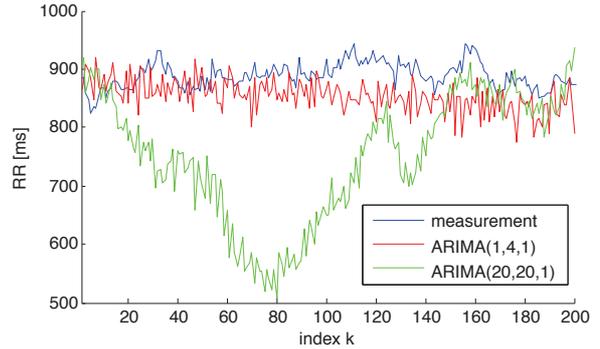

Figure 2: TS of 200 measured and simulated RR intervals.

represent a suitable description for the RR interval dynamics. However, the ARIMA(20,20,1) model tries to assimilate inaccuracies originating from the empirical ACF/PACF during the modeling procedure. Due to the integrating behavior of the ARIMA models these minor deficiencies will lead to physiologically implausible results.

Discussion

The presented methodology allows a versatile dynamic modeling of physiological TS only under the restriction of inaccuracies when regarding finite TS. Simulations yielded plausible results for parsimonious models. However, it is necessary to be cautious when simulating extensive models or long periods of time. Due to the integration of a power signal, the ARIMA models generate TS with variances increasing over time. The introduced models provide an enormous amount of system characteristics to be analyzed. As an example, the eigenvalues of the system matrix in formula (4) involve direct significance to dynamic properties of the generating physiological process. This information could be applied on medical classification studies concerning cardiac disease or drug safety.

Bibliography

[1] K. Kotani *et al.*, "Model for complex heart rate dynamics in health and diseases," *Physical Review E*, vol. 72, no. 4, 2005.

[2] J. D. Hamilton, *Time Series Analysis*, vol. 2. Cambridge University Press, 1994.