

# Time-Series Prediction Of The Electrocardiogram (ECG) Signal

Madhur Behl  
Dept. of Electrical and Systems Engineering  
University of Pennsylvania  
{mbehl}@seas.upenn.edu

---

---

## 1. Autoregressive Moving Average Model (ARMA)

Many observed time series exhibit serial autocorrelation; that is, linear association between lagged observations. This suggests past observations might predict current observations. The autoregressive (AR) process models the conditional mean of a signal  $y_t$  as a function of past observations,  $y_{t-1}, y_{t-2}, \dots, y_{t-p}$ . An autoregressive (AR) process that depends on  $p$  past observations is called an AR model of degree  $p$ , denoted by  $AR(p)$ . The  $AR(p)$  model is written as:

$$y_t = c + \psi_1 y_{t-1} + \dots + \psi_p y_{t-p} + \epsilon_t \quad (1)$$

where  $\epsilon_t$  is a zero mean uncorrelated white noise term (also referred to as the innovation process term).

The moving average (MA) model captures serial autocorrelation in a time series  $y_t$  by expressing the conditional mean of  $y_t$  as a function of past innovations,  $\epsilon_{t-1}, \epsilon_{t-2}, \dots, \epsilon_{t-q}$ . An MA model that depends on  $q$  past innovations is called an MA model of degree  $q$ , denoted by  $MA(q)$ . The  $MA(q)$  model is written as:

$$y_t = c + \theta_1 \epsilon_{t-1} + \dots + \theta_q \epsilon_{t-q} \quad (2)$$

For some observed time series, a very high-order AR or MA model is needed to model the underlying process well. In this case, a combined autoregressive moving average (ARMA) model can sometimes be a more parsimonious choice. The electrocardiogram is one such example of a time-series. An ARMA model expresses the conditional mean of  $y_t$  as a function of both past observations,  $y_{t1}, y_{t2}, \dots, y_{tp}$ , and past innovations,  $\epsilon_{t1}, \epsilon_{t2}, \dots, \epsilon_{tq}$ . The number of past observations that  $y_t$  depends on,  $p$ , is the AR degree. The number of past innovations that  $y_t$  depends on,  $q$ , is the MA degree. In general, these models are denoted by  $ARMA(p, q)$  and written as:

$$y_t = c + \psi_1 y_{t-1} + \dots + \psi_p y_{t-p} + \epsilon_t + \theta_1 \epsilon_{t-1} + \dots + \theta_q \epsilon_{t-q} \quad (3)$$

We use MATLAB to implement the  $ARMA(p, q)$  models based on [2, 1].

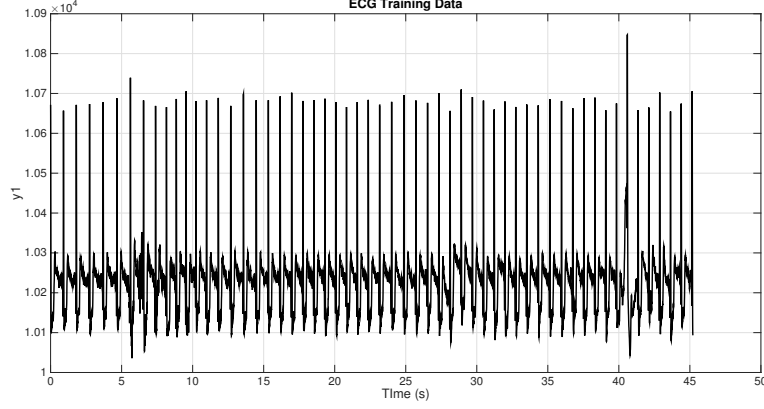


Figure 1: Training data for the ARMA(4,4) model

## 2. Normal Sinus Rhythm Prediction

We use ECG data, sampled at 2.8ms, from a healthy subject and logged for one minute to train and validate the ARMA model. 80% of the data (Figure 1) is used for training and learning the coefficients of an  $ARMA(4, 4)$  model (Eq. 3). The trained model is then used for predicting the ECG signal from the remaining 20% of the data-set.

For predicting the ECG signal, we can vary the prediction horizon. In this study, we evaluate two different prediction horizon lengths- 2 step prediction ( $\sim 6$ ms) and 5 step prediction ( $\sim 15$ ms). The longer the prediction horizon the less accurate is the prediction. The prediction accuracy obtained for the 2-step

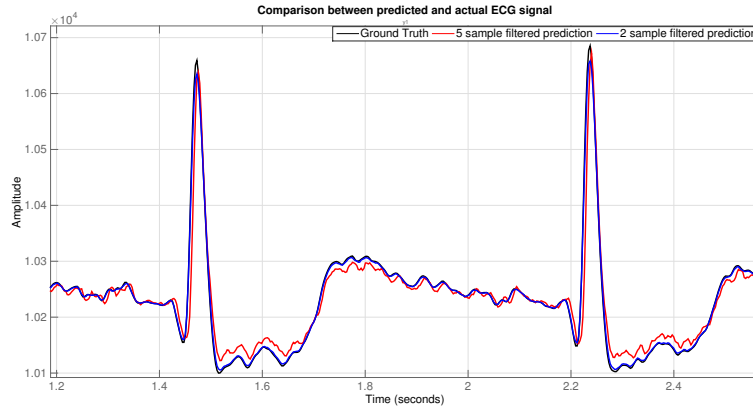


Figure 2: Comparison between predicted ECG response and actual ECG signal for different prediction horizons.

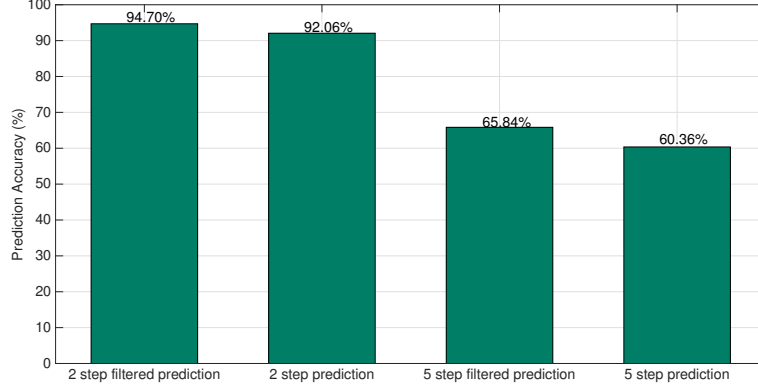


Figure 3: Prediction accuracy comparison for different prediction horizon lengths. Also shown is the improvement obtained by filtering the predicted response for each prediction horizon length.

prediction is 92.06% while a prediction accuracy of 60.36% is obtained for the 5-step prediction. The accuracy is calculated (in percentage) as:

$$\text{Accuracy} = 100 \left( 1 - \frac{\|y - \hat{y}\|}{\|y - \text{mean}(y)\|} \right) \quad (4)$$

The predicted response from each is further passed through a low-pass moving average filter. This helps to smooth out the predicted ECG signal and further improve the model accuracy. Figure 2 shows the comparison between the predicted ECG and the actual ECG signal for the filtered 2-step and 5-step prediction horizons.

Figure 3 shows how the prediction accuracy varies as the prediction horizon is changed from 2 to 5. We can also see that filtering the predicted response improves the prediction accuracy to 94.7% from 92.06% for the 2-step prediction and to 65.84% from 60.36% for the 5-step prediction.

Based on these preliminary results, ARMA models seem like a good choice for short term ECG prediction. One should note that the training phase in the case-study presented in this report was completely offline. For a real-time implementation, the training phase can be executed after a certain number of heartbeats has been observed and stored in a buffer. The model can be continuously re-tuned and used for prediction in an online fashion.

### 3. Future Work and Improvements

So far, we have only considered NSR prediction. In the future, we will evaluate this method for prediction of ECG signals in the case of an arrhythmia present in ECG signal. We will also investigate better methods of tuning and choosing the  $p, q$  parameters of the ARMA model.

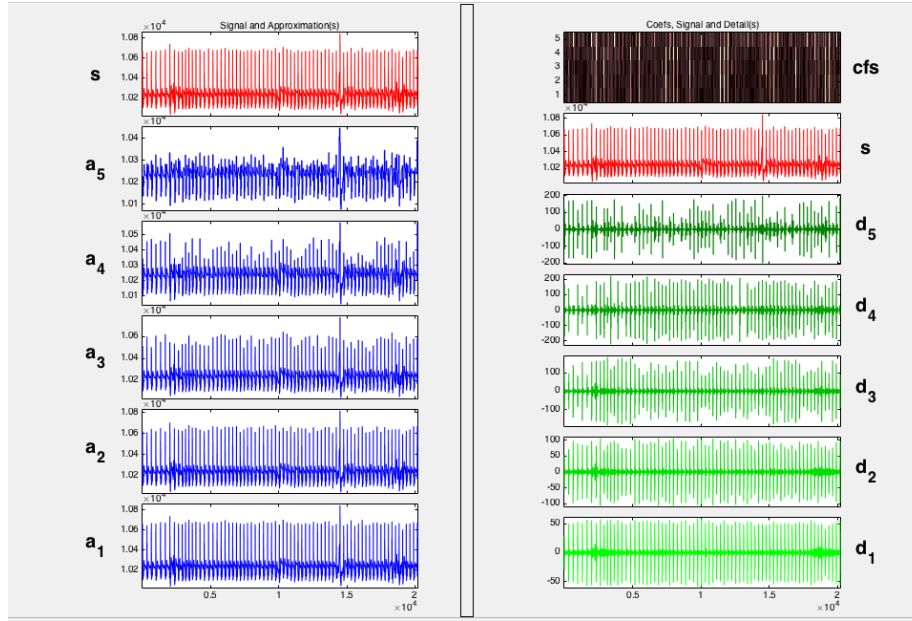


Figure 4: The ECG signal (s) is broken down into its approximation and detail components using Daubechies wavelets decomposition.

ECG signal data is non-stationary and contains transitory characteristics such as: drift, trends, abrupt changes, and beginnings and ends of events. These characteristics are often the most important part of the signal, and a time-series prediction is not completely suited for detecting and predicting ECG characteristics. Using wavelet transforms, we can decompose the ECG signal into a combination of high and low frequency signals. We can then make predictions for each component and recombine the wavelets to obtain a predicted ECG signal.

Lastly, using MATLAB's armax implementation it takes 2.76s to train the  $ARMA(4, 4)$  model. This is very slow for a real-time implementation, therefore, evaluation of how fast this can be done without using MATLAB code will be conducted.

## References

- [1] James Douglas Hamilton. *Time series analysis*, volume 2. Princeton university press Princeton, 1994.
- [2] Herman OA Wold. *The analysis of stationary time series*. Almqvist & Wiksell, 1954.