You Can't Stop the Beat: Modeling Heartrate in Endurance Running

John Baierl
UCLA – STATS 221

Abstract:

Endurance athletes rely on heart rate tracking during both training and competition to maintain the constant effort level necessary for proper pacing. Modeling the variations in heart rate during this type of constant-effort aerobic activity has both personal and commercial applications, enabling fitness tracking devices to provide more accurate predictions identifying when athletes drift above or below their desired intensity level. During this case study, we show that an ARMA model fits medium-intensity heart rate data well, outperforming vanilla forecasting methods over short-term (60 second) time horizons.

I. Introduction

Race strategy in long-distance endurance sports is highly reliant on proper pacing throughout the early stages of the event. Overexertion in the opening miles can easily spoil a quality race later on. Biomechanically, the optimal strategy in endurance events is to complete the race at a constant effort level (Ely et al 2008). For athletes competing on flat courses in ideal conditions such as track runners or marathoners on pancake-flat courses, this simply equates to a constant pace. However, changing course conditions complicate things. Ultramarathoners, Nordic skiers, or distances runners competing on hilly courses for example must adapt their pace to changing conditions. Athletes must then compensate for hills or changes in temperature by adjusting pace to maintain constant race effort.

Additionally, the popularity of fitness trackers has grown dramatically over the past decade both for personal and professional use, allowing athletes to augment their perception with real-time heart rate data. Developing a model for constant effort low- to medium-intensity running has applications both for personal use as well as incorporation into the ever-growing set of analytic tools built into tracking devices. In this paper, we will use modeling techniques for time-series data to forecast future heart rate for this type of running.

II. Experiment and Data

For this case study, we fit a time series model to heart rate data gathered during a sample long-distance run. The data to be studied was provided by the author, who is a 30-year-old male with competitive endurance sport backgrounds in Nordic skiing and distance running, but is otherwise an amateur athlete. The workout of study was a 12-mile run that took place in the beginning of a marathon training cycle. This was completed on a flat course (50 ft elevation change per Google maps), at a comfortable temperature (71°F at start), and in negligible wind (SE 3.8 mph at start). The goal of this run was to maintain at a constant perceived "breathy conversational" effort, translating to a target heartrate range of approximately 148-168 BPM (74%-84% of maximum heart rate). While a truly blinded experiment was impossible given the author's participation, real-time heartrate or pace were not consulted during the run, instead relying on perceived effort to set pace. Simulating race conditions, mile splits were viewed throughout the run. These are provided in the *Table 1*, which verify an approximately constant effort.

Heart rate was measured using an electrode-based chest strap paired to a Garmin Forerunner 935 watch. Measurements were recorded at one-second intervals. Readings from the first and last mile were omitted from the data to account for warm-up and cool-down periods, resulting in a 10-mile study period.

III. Model-Building

The full heart rate versus time data for the study portion of the run is plotted in *Fig 1*. We apply a 75-25 train-test split, withholding the last quarter of the data for model evaluation and forecasting. The training portion of the data is plotted in *Fig 2*. Most immediately noticeable is the steady, upward trend in heart rate over time. This exemplifies a phenomenon known as cardiovascular drift (Coyle, González-Alonso 2001). For low- to medium-intensity aerobic exercise, athletes typically experience a steady increase in heart rate despite no changes in perceived effort or breathing. Since little has been written about the precise functional form of this phenomenon, we use fit a linear model via ordinary least squares to capture this trend. The fitted line is shown below:

$$\widehat{HR} = 156.412 + 0.002753 * Time$$

Our estimated slope suggests an expected heart rate drift of approximately 9.912 BPM/hr at this intensity level. Note that caution should be taken extrapolating this trend to higher-intensity workouts. Indeed, we may expect a different functional form for this trend as athletes' approach and exceed lactate threshold pace, necessitating further study.

We then detrend these data to remove the effect of cardiovascular drift, considering the residuals from the linear model in subsequent model-fitting. These residuals are plotted as a time series in *Fig. 3*. Visual inspection suggests no obvious changes in variance over time. The augmented Dickey-Fuller test confirms our suspicion of stationarity, rejecting the null hypothesis at the 1% significance level. After detrending, the remaining variability in the data can be attributed to (*i*) measurement error from our device, (*ii*) natural fluctuations in heart rate while maintaining constant effort, and (*iii*) variations in our runner's ability to maintain constant effort in the first place.

Cycles within heart rate data may be possible when considering (*iii*) due to the nature of attempting to maintain a constant effort. Such a task may result in periodic cycles of overexerting, followed by under exerting to compensate. Moreover, we might expect such an effect to be more pronounced for amateur athletes like the test subject as compared to professionals. Identifying such a pattern would be a useful insight for both racing and training. However, the smoothed periodogram (*Fig. 4*) shows a steady decline in magnitude as frequency increases with no noticeable spikes. This is indicative of acyclic, smooth data. We proceed under this assumption moving forward.

The sample ACF and PACF for the detrended data are shown in *Fig. 5*. Note that the slow tail-off of the ACF suggests that measurement error is not overwhelming the heartrate signal, since we should expect a high degree of autocorrelation at the one second level for heart rate data under most conditions. The PACF suggests an AR(3) or AR(4) may be appropriate for these data, due to the cutoff at around lag 3 or 4.

We next conduct a grid search, fitting ARMA(*p, q*) models for $p \in [0{,}9], q \in [0.9]$. AIC values for each model are shown in *Table 2*. The ARMA(3, 2) model minimizes AIC, consistent with our visual inspection of the PACF plot. The final model for heart rate at each time step (x_t) is shown below:

$$x_t = 2.13x_{t-1} - 1.33x_{t-2} + 0.201x_{t-3} - 0.687w_{t-1} - 0.103w_{t-2} + 0.0326$$

These coefficients give some insight into the structure of the data. Despite the general smoothness of the data indicated by the ACF plot, the inclusion of two moving average (MA) terms further smooths the time series over the two seconds preceding each time step. Since we would expect other sources of variability due to biomechanical fluctuations to take place more slowly, this is most likely accounting for measurement error in the form of white noise. The autoregressive (AR) coefficients are quite large for the two preceding time steps, with both having a magnitude greater than one. The large positive coefficient at lag-1 is counteracted somewhat by a smaller negative coefficient at lag-2.

Fig. 6 gives a snapshot of typical model behavior for a small segment of test data using one-step-ahead forecasting. Note that the model tends to fit the steady and gradual increases and decreases in heart rate quite closely. During periods of great variability or sharp changes in direction, the model tends to oscillate due to the push-pull nature of the lag-1 and lag-2 AR coefficients, fitting the data most poorly during these periods. The fitting of these particular coefficients suggests that the heartrate data here is most generally characterized by these gradual upward and downward trends.

IV. Forecasting and Model Evaluation

We consider four models in our performance evaluation: ARMA(3, 2), AR(3), previous-second prediction, and the cardiac drift linear model. Note that the ARMA(3, 2) and AR(3) models also include the linear effect from cardiac drift, further modeling the residuals on top of it. While the ARMA(3, 2) outperformed the AR(3) in terms of AIC during the initial grid search, we include it to get some sense of how much value the inclusion of the additional MA terms adds to predictive performance.

Performing one-step-ahead forecasting on the testing portion of the data yields RMSE values shown in *Table 1*. In part due to the smoothness of the data, all models perform reasonably well, with both included ARMA models falling well within 1 BPM in terms of RMSE. Moreover, these significantly outperform the vanilla linear model prediction. However, forecasting simply based on the heart rate reading at the previous second remains the best predictor in this short time horizon. This suggests that the smoothness of the data is dominating any larger trends at the one-second level. This is again visible in *Fig. 6* where we can see the poor model fit during rapid changes in the overall trend. Here, simply relying on the reading from the previous second is clearly the superior strategy.

Longer-term forecasting is likely the more interesting application for most use-cases. We carry out a forecast at the 60 second time horizon, testing on the first 60 seconds of test data. The same four models are used as in the one-second-ahead case. Note that the previous-second forecast uses the most recent heart rate reading as a constant prediction over the full 60 second test interval. The resulting RMSE values are shown in *Table 2*, with the forecasted ARMA(3, 2) values and corresponding 1-SE bounds plotted in *Fig. 7*.

In this context, the ARMA(3, 2) model shows its advantage over the other three models, offering about 22% improvement over the previous-second model. *Fig. 7* highlights some of its advantages. The ARMA(3, 2) captures the shape of the overall trend during this period remarkably well during this period, albeit to a less extreme degree. This again aligns with previous discussion that the AR coefficients are well-fit to the more gradual overall trends in the data, while the MA coefficients are smoothing the more short-term measurement error.

V. Discussion and Conclusions

These results show strong potential for improvement in predictive performance over a time horizon up to 60 seconds for heart rate data. The ARMA(3, 2) model both outperforms vanilla approaches and offers a reasonable forecast for the overall shape of the trend moving forward. Fitness trackers frequently offer functions that alert users when their heartrate passes above a certain threshold. For example, the Garmin watch used in this test allows the user to pick a specific heart rate reading before the run begins at which to send this alert. These results suggest that such features could be improved by using time series forecasts to predict when athletes are trending towards exceeding that threshold, increasing their utility by preventing overexertion before it actually occurs.

The most obvious extension of these results is to explore how the model responds to changes in biomechanical or environmental conditions. This model was fit to a single, medium-intensity run in nearly optimal conditions with ideal levels of rest and hydration. Repeating similar tests on a treadmill would control for natural variations in pace, effectively isolation heart rate variability due to natural fluctuations.

Additionally, factors like rest and hydration are likely to vary substantially over the course of a training program. The stability of this type of model in the face of changing levels of hydration or rest remains to be seen, since these factors can have substantial impacts on performance. As heartrate-based indicators of overtraining have become built into fitness tracking products, predictable changes in heart rate behavior from run to run could serve as an additional indicator for detecting and alerting users to deficiencies in the quality of their rest and level of recovery.

Appendix:

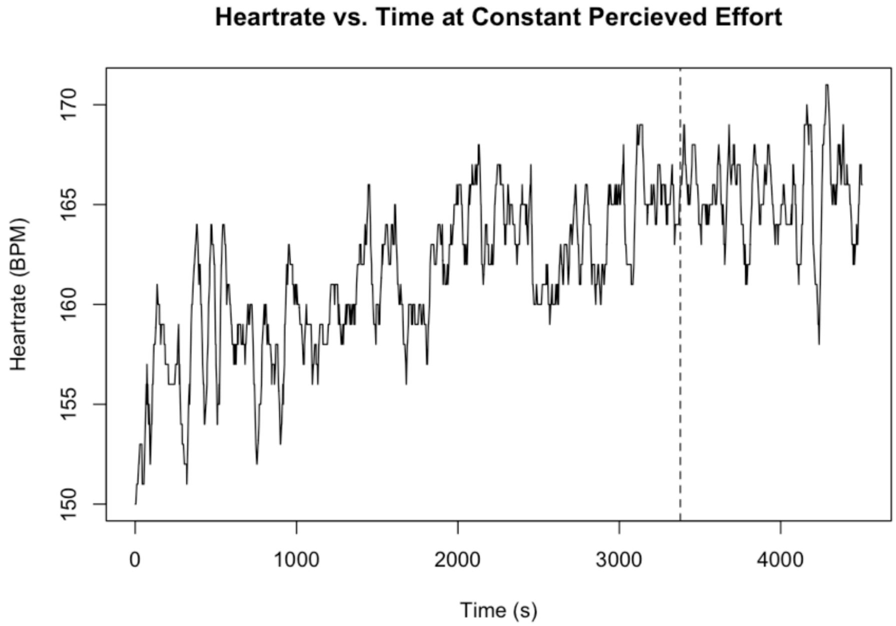


Fig. 1: 75-25 train-test split

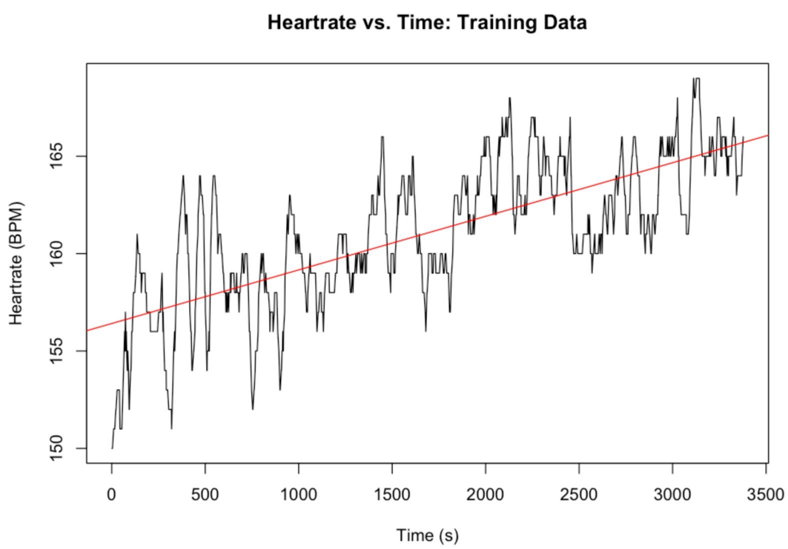


Fig. 2: Training data with cardiovascular drift model shown

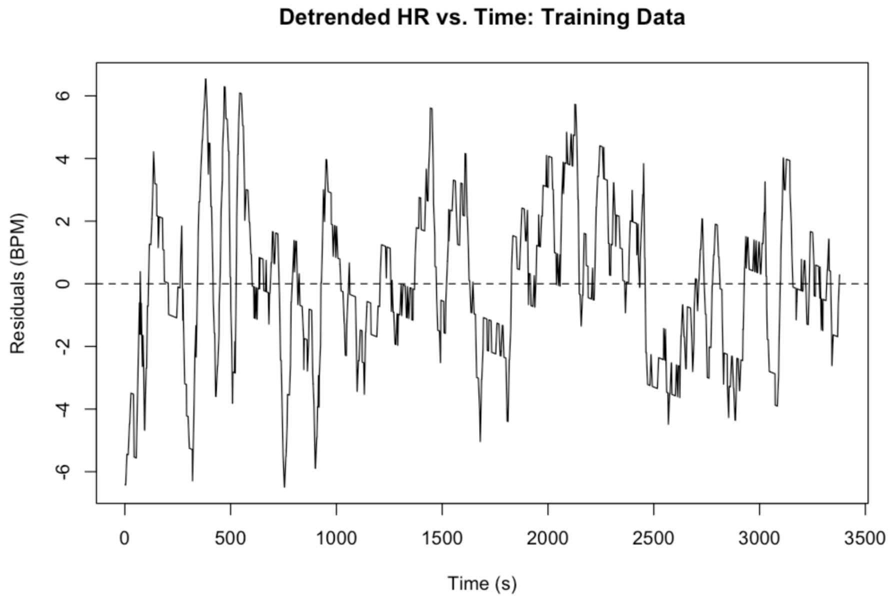


Fig. 3: Stationary residuals after detrending

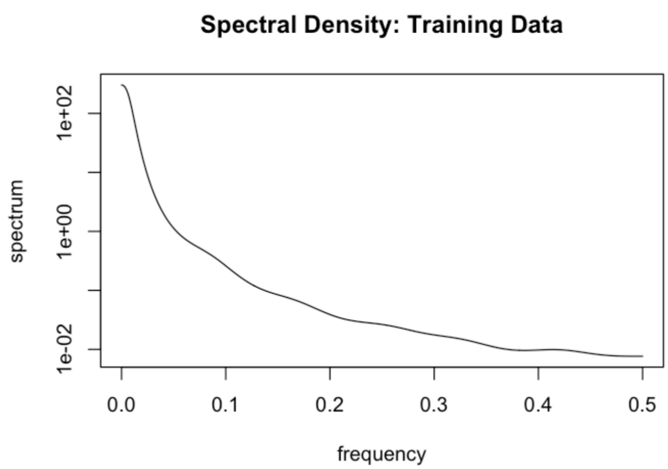


Fig. 4

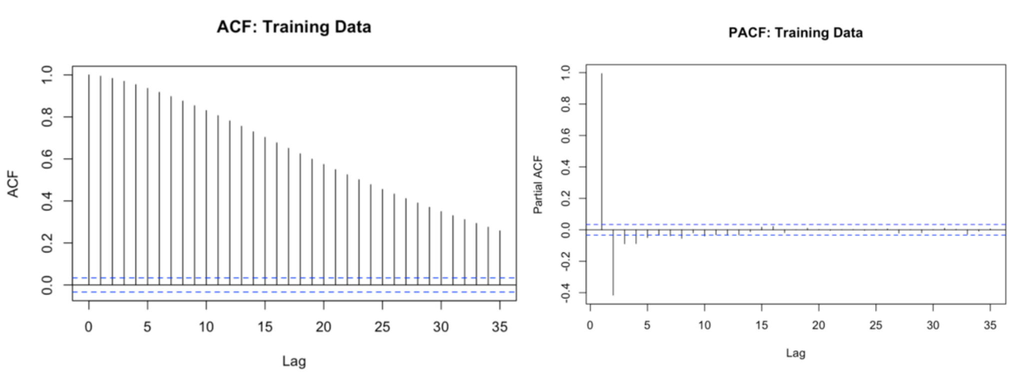


Fig. 5

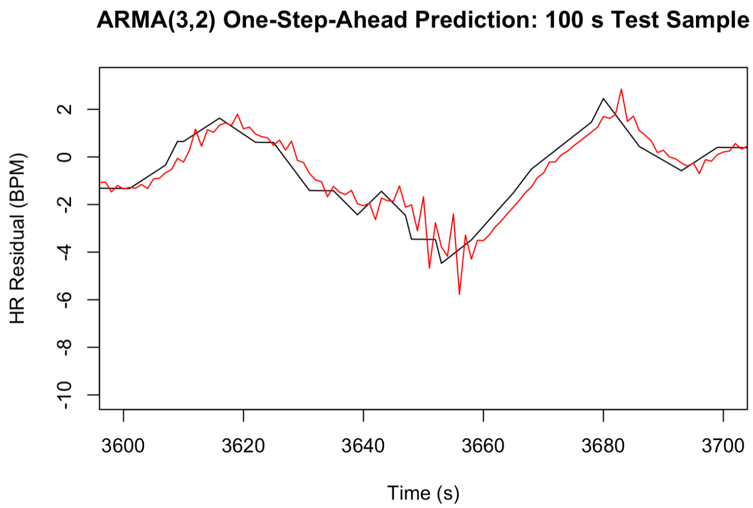
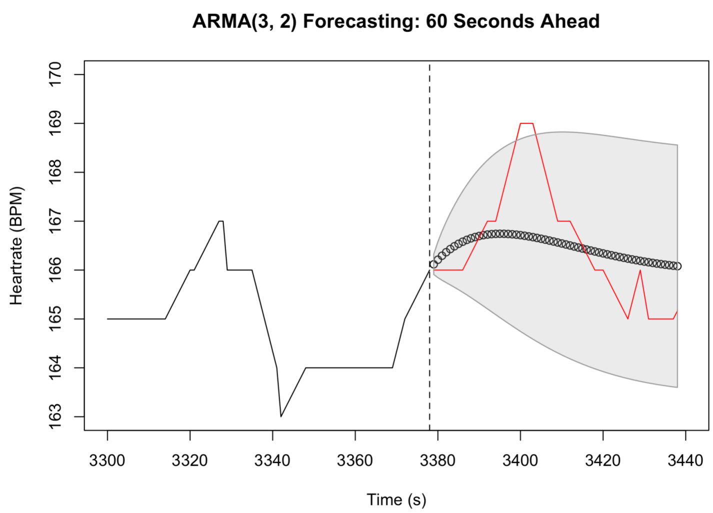


Fig. 6



*Note that confidence bounds show ±1-se interval

Fig. 7

Mile:	1	2	3	4	5	6	7	8	9	10
Pace (min/mi):	7:41	7:48	7:46	7:39	7:30	7:35	7:36	7:29	7:42	7:47

Table 1: Mile splits during study portion of run

	$q = 0$	1	2	3	4	5	6	7	8	9
$p = 0$	15662.09985	11258.3705	7846.7703	5428.9211	3914.5009	2751.5058	1954.3973	1379.9610	977.4814	645.2821
1	56.73862	-699.7150	-818.6965	-873.4013	-908.6979	-923.8749	-925.1763	-940.9218	-942.7782	-946.3252
2	-911.99188	-930.2301	-981.3150	-983.8812	-983.6118	-983.0203	-920.0667	-984.2050	-982.6207	-982.0244
3	-922.16935	-983.5437	-985.0997	-977.3199	-981.2060	-979.8789	-981.4795	-983.0880	-980.2403	-871.0766
4	-952.41528	-984.5798	-983.9854	-982.3128	-976.5519	-984.0574	*NA*	*NA*	*NA*	*NA*
5	*NA*	*NA*	*NA*	*NA*	*NA*	*NA*	*NA*	*NA*	*NA*	*NA*
6	*NA*	*NA*	*NA*	*NA*	*NA*	*NA*	*NA*	*NA*	*NA*	*NA*
7	*NA*	*NA*	*NA*	*NA*	*NA*	*NA*	*NA*	*NA*	*NA*	*NA*
8	*NA*	*NA*	*NA*	*NA*	*NA*	*NA*	*NA*	*NA*	*NA*	*NA*
9	*NA*	*NA*	*NA*	*NA*	*NA*	*NA*	*NA*	*NA*	*NA*	*NA*

Table 2: AIC values for ARMA(p, q) grid search

Model:	**RMSE (BPM):**
ARMA(3, 2)	0.597
AR(3)	0.694
Linear model only	2.993
Previous second	0.249

Table 3: RMSE for One-step-ahead forecasting on test data

Model:	**RMSE (BPM):**
ARMA(3, 2)	1.003
AR(3)	1.231
Linear model only	1.287
Previous second	1.287

Table 4: RMSE for 60-second forecast

Works cited:

Coyle, E. F., and J. González-Alonso. "Cardiovascular Drift during Prolonged Exercise: New Perspectives." *Exercise and Sport Sciences Reviews*, vol. 29, no. 2, 2001, pp. 88–92., https://doi.org/10.1249/00003677-200104000-00009.

Ely, Matthew R., et al. "Effect of Ambient Temperature on Marathon Pacing Is Dependent on Runner Ability." *Medicine & Science in Sports & Exercise*, vol. 40, no. 9, 2008, pp. 1675–1680., https://doi.org/10.1249/mss.0b013e3181788da9.