Time Series Analysis of Affective Physiological States

Stat 248, Spring 2016

Jamie Palumbo

May 9, 2016

1 Introduction

The question that this study will consider is whether or not a time series model exists that can pick up on the unique fingerprints physiological signals leave behind in sensor readings during different emotional states. Physiological pattern recognition of emotions has several important applications. For example, if emotions such as anger and anxiety can be quantified through physiological signals, psychiatric diseases could potentially be pinpointed more accurately, allowing for higher success rates of treatment. Another important application is improving human-computer interaction. Specifically, changes in physiological signals could help detect aggravation or frustration with a product feature which could then help developers redesign and improve their product. This paper will take a time series approach in studying physiological signals by first performing exploratory data analysis, then looking at different fitted models and forecasting schemes, and finally performing a frequency side analysis.

2 Data Description

The dataset used in this analysis was generated as part of the MIT Affective Computing Group's research. The data was compiled from a single test subject–a healthy female graduate student with two years of acting experience. In the experiment, the researchers had the test subject purposefully elicit emotion in a close to real-world environment with the aid of

a computer prompting system called the "Sentograph" and emotion specific imagery. The order of emotions was as follows: no emotion, anger, hate, grief, platonic love, romantic love, joy, and reverence. This sequence was designed to make it easier for the subject to transition from emotion to emotion. While the test subject was experiencing these emotions, four physiological sensors simultaneously measured electromyogram (EMG) in microvolts, blood volume pulse (BVP) in percent reflectance, skin conductance (GSR) in microSiemens, and respiration (RES) in percent maximum expansion. This experiment was repeated 20 days in a row with each day's session lasting 25 minutes, resulting in 16,000 samples per physiological signal per day or 2,000 samples per emotion per physiological signal per day.

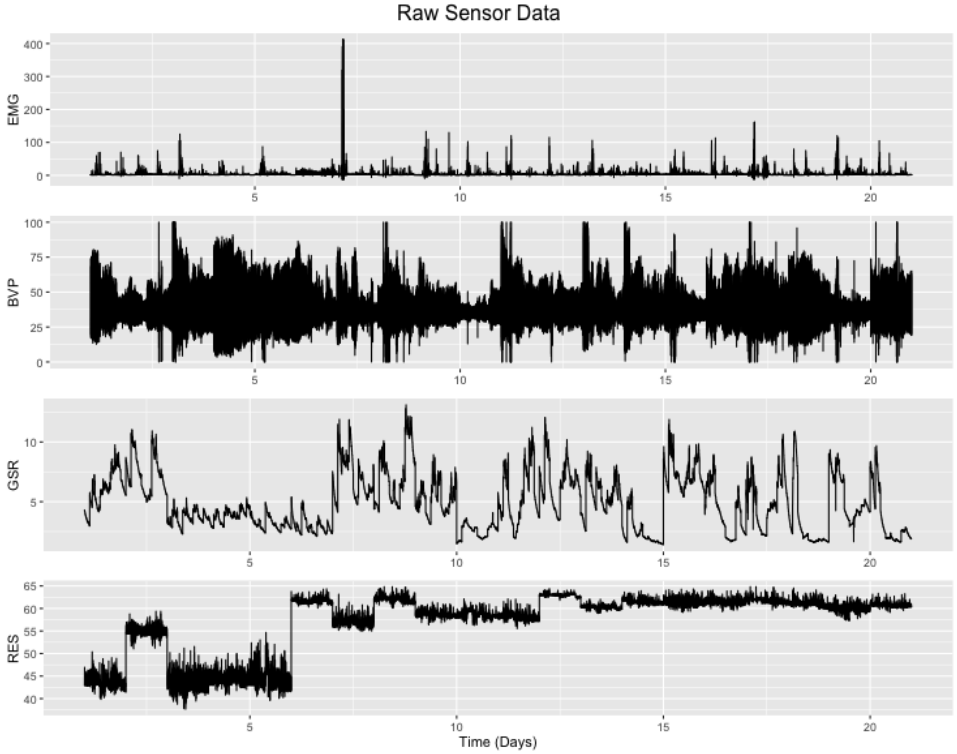

Figure 1: Raw sensor data over all 20 days

3 Exploratory Data Analysis

We begin our exploratory data analysis by plotting the raw sensor data in Figure 1. Due to the heavy outliers expressed in the EMG plot, we perform a Box Cox transformation by

logging the series in order to reduce the occurrence of outliers and to make the data more stationary. In Figure 2, we plot the transformed sensor data for anger, grief, and joy on Day 2. This plot shows that there do appear to be visibly different patterns for different emotions in the physiological sensor readings. However, this is generally not the case. In order to get a better sense of the distributions of these sensors, we plot histograms in the left panel of Figure 3. The BVP, GSR, and logged EMG data appears to be skewed right with the median being less than the mean. While logged EMG and BVP appear to have unimodal distributions, GSR and RES could potentially be specified as bimodal – the RES sensor in particular. This is made more obvious with the overlaid kernel density estimation function which provides a smoothed estimation of the probability density function.

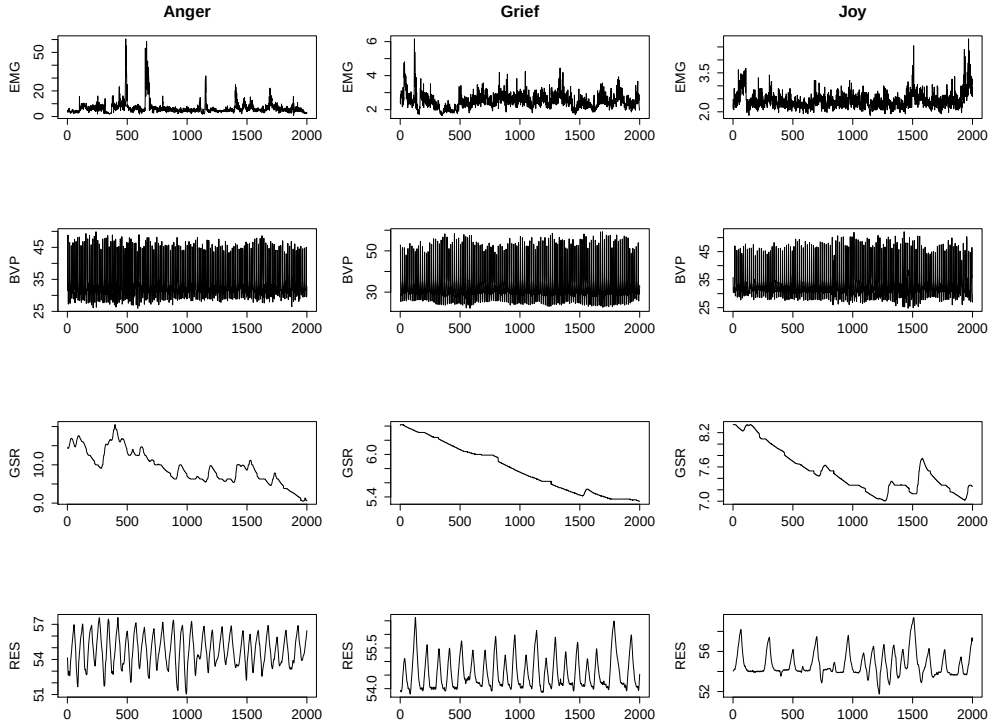

Figure 2: Transformed sensor data for anger, grief, and joy on Day 2

After getting some sense of the raw data that will be analyzed, we next perform a seasonal-trend decomposition of each individual sensor since it would be reasonable for daily periodicity (16,000 observations per sensor per day) to exist in this dataset. We assume an additive

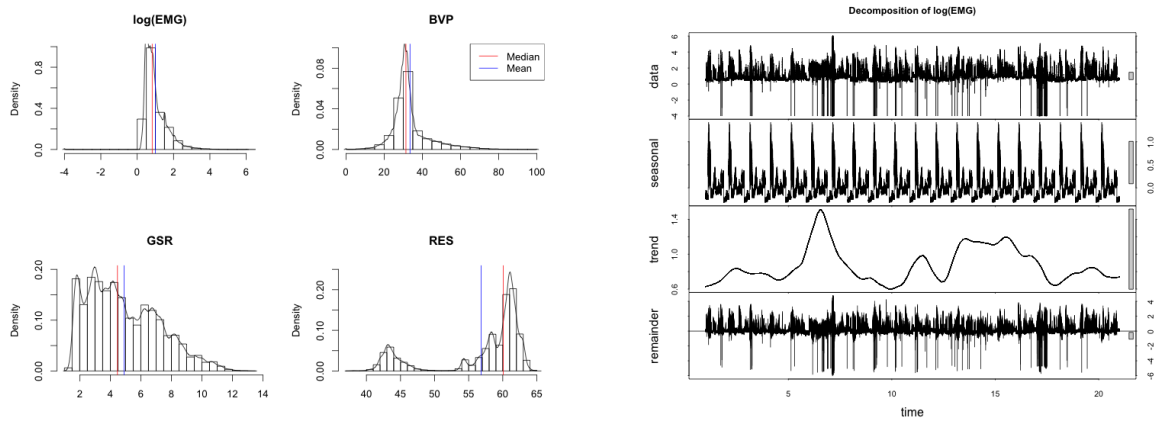

Figure 3: EDA distribution and seasonal decomposition plots

decomposition of $Y_t = S_t + T_t + E_t$ where the trend component T_t is determined using loess regression and the seasonal component S_t is calculated from the differences $Y_t - T_t$. Figure 3's right panel illustrates this decomposition of the logged EMG series. The vertical range bars on the right of the plots indicate equal heights for easier readability. Surprisingly, we find that the trend and seasonal components do not explain much of the variation in the data and that the remainder component is very nearly the original plot. This appears to also be the case in the remaining sensors – though GSR and RES have stronger trends. The remaining decomposition plots can be found in the appendix (Figures A.1 – A.3).

4 Modeling

We next turn to time series modeling to determine whether the physiological signals can be accurately predicted. We focus our attention on ARIMA, GARCH, and harmonic regression.

4.1 ARIMA

4.1.1 Model Selection

The first step in fitting an ARIMA model is to determine the order of differencing required to stationarize the series. This can be accomplished intuitively by inspecting the ACF and

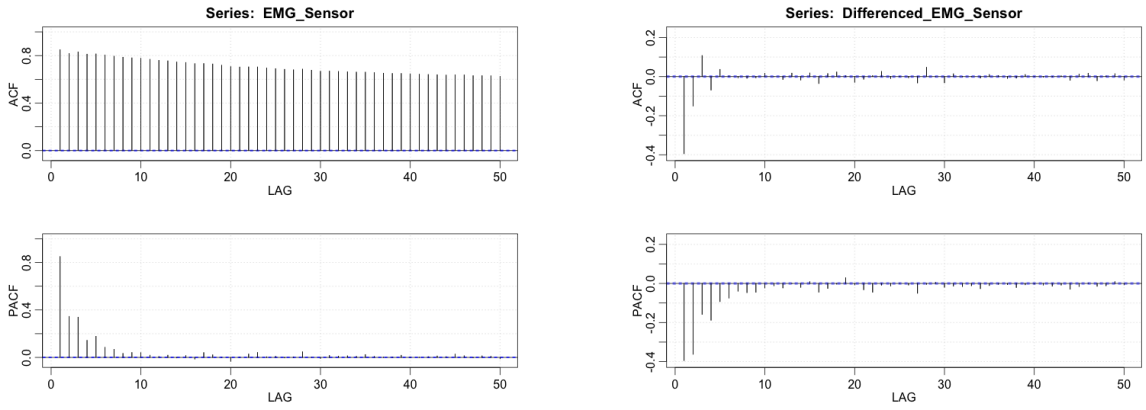

Figure 4: Raw and differenced ACF and PACF of logged EMG sensor

PACF of the series or more analytically by performing an Augmented Dickey-Fuller (ADF) test. Having already taken logs of the EMG sensor data, the ACF and PACF in Figure 4 indicate that an order of differencing might be required to make the series stationary. However, according the the p-value of 0.01 in the ADF test, we reject the null of a unit root and assume stationarity. Indeed, looking at the differenced ACF and PACF plots, we can see that the series appears to be over-differenced. Therefore, we do not include an order of differencing for the EMG sensor. Following a similar approach for the remaining sensors, we find that BVP and GSR (p values = 0.01) appear to be stationary. While the ADF test indicates that the RES sensor is stationary (p value = 0.01), the KPSS test and ACF and PACF plots indicate non-stationarity so we include a differenced term in that model. The ACF and PACF plots for these sensors can be found in the appendix (Figures A.4 – A.6).

After identifying the order of differencing for each sensor, we must next stipulate the number of AR and MA terms to include in the models. Since it is possible for AR and MA terms to cancel each other's effects out, we follow a "forward stepwise" approach to specifying the models, adding either AR or MA terms as indicated by the appearance of the ACF and PACF plots of the residuals and supporting the addition of these terms through the use of the BIC model selection criteria. We use BIC instead of AIC in this case as BIC tends to

	Model Selected	First Runner-up	Second Runner-up
EMG	ARIMA(2,0,1) $BIC = 96095.39$ $MSE = 0.2023135$	ARIMA(1,0,1) $BIC = 97857.57$ $MSE = 0.2026166$	ARIMA(2,0,0) $BIC = 150901.8$ $MSE = 0.2041943$
BVP	ARIMA(2,0,2) $BIC = 1729463$ $MSE = 110.3139$	ARIMA(2,0,1) $BIC = 1772493$ $MSE = 110.3079$	ARIMA(1,0,1) $BIC = 1827365$ $MSE = 110.3087$
GSR	ARIMA(2,0,2) $BIC = -893370.6$ $MSE = 7.539375$	ARIMA(1,0,1) $BIC = -892080.3$ $MSE = 7.564715$	ARIMA(2,0,2) $BIC = -892059.7$ $MSE = 7.539375$
RES	ARIMA(1,1,1) $BIC = -750578.6$ $MSE = 0.5358516$	ARIMA(2,1,1) $BIC = -750567.6$ $MSE = 0.5383754$	ARIMA(2,1,2) $BIC = -750553.5$ $MSE = 0.5363165$

Figure 5: Model Selection

choose a more parsimonious model and penalizes overfitting more so than AIC. The model selected for each sensor is listed in Figure 5 along with the two next lowest BIC models.

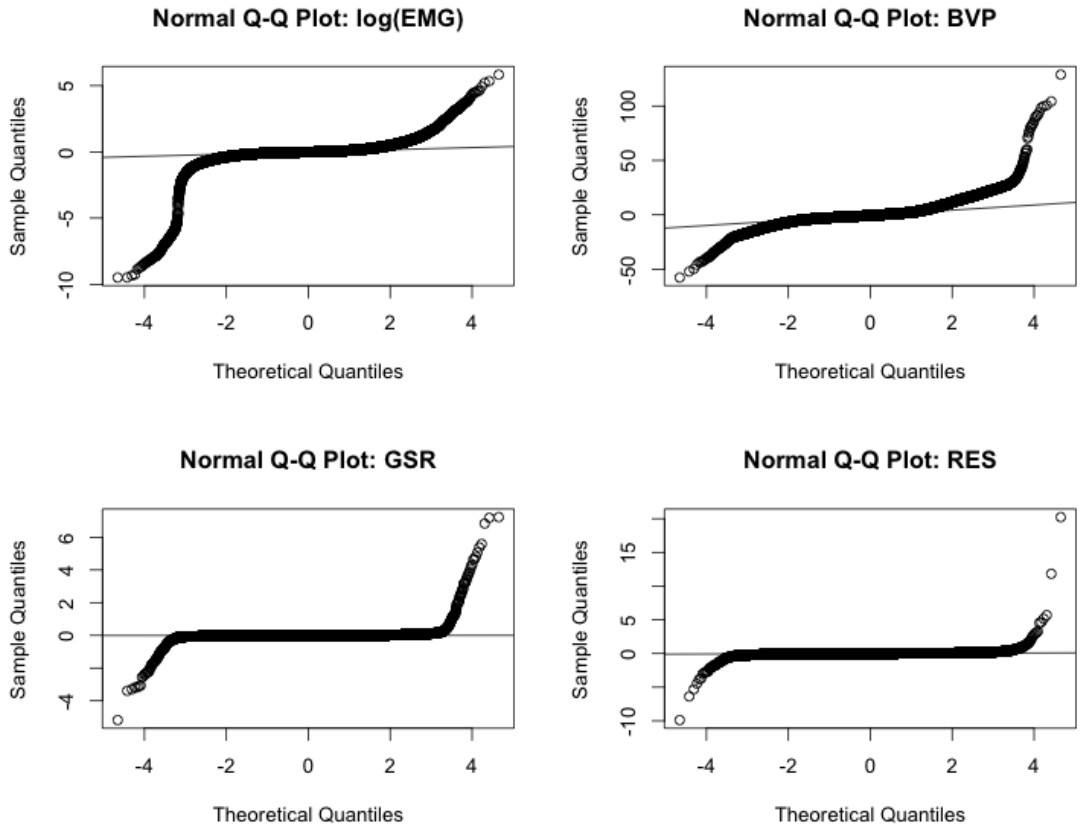

Figure 6: Normal Q-Q plots of the fitted model residuals

Despite stationarizing the series and choosing the lowest BIC models, it appears that the Q-Q plots of the fitted model residuals do not closely follow normal distributions. In fact, the Q-Q plots in Figure 6 suffer from incredibly heavy tails, meaning that the residuals have a higher probability of taking on extreme values compared to the normal distribution.

4.1.2 Forecasting

Having chosen the ARIMA models for each sensor, we can then forecast. Since the data from this experiment is incredibly specialized for a single individual, we choose to instead forecast on Day 20 of the experiment to see how well our model performs, leaving Days 1-19 as the training set and Day 20 as the test set. As shown in Figure 7, the ARIMA models do

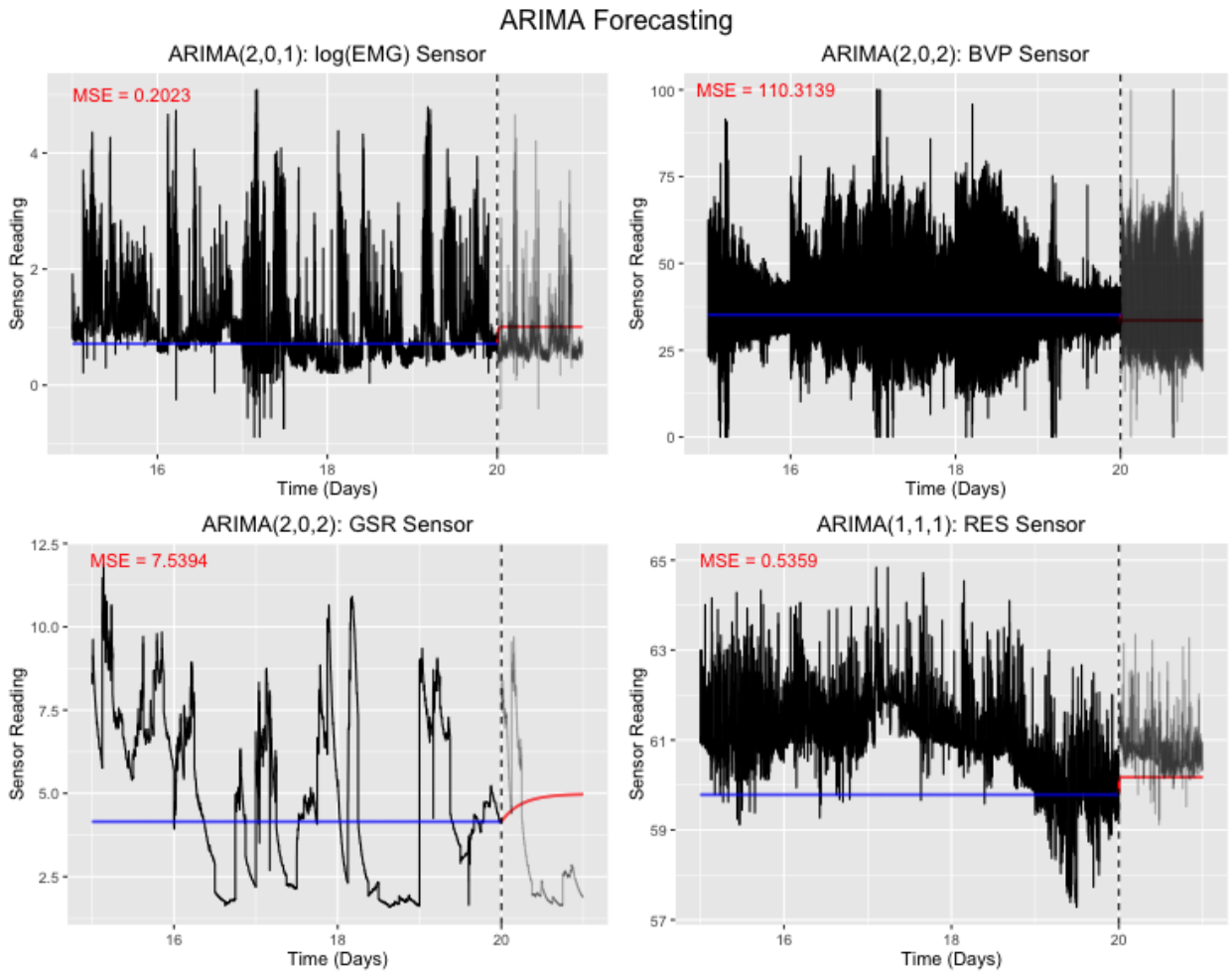

Figure 7: ARIMA Forecasting

	mu	omega	alpha1	beta1
EMG	0.6883442***	0.0016981***	0.2769400***	0.7150818***
BVP	3.365e+01***	1.162e-02	1.843e-02***	9.818e-01***
GSR	4.527e+00***	1.187e-05	2.229e-01***	7.241e-01***
RES	6.095e+01	4.233e-04	6.308e-01***	3.636e-01***

Levels of Significance: 0***, 0.001**, 0.01*, 0.05.

Figure 8: Model Selection

not pick up hardly any of the variability in the sensor data. While the MSEs seem to be low, the models appear to be predicting the mean of previous time stamps and do not seem to be viable models for predicting emotions.

4.2 GARCH

As illustrated in the previous section, ARIMA models did not appear to capture the variability in the sensor readings. This is likely due to the fact that ARIMA models assume constant variance. Due to the volatile nature of the data, we attempt to fit a GARCH(1,1) model. According to Figure 8, the majority of the coefficients in the GARCH models are highly significant at the 0.001 significance level, lending some credence to the use of this model in prediction. However, the Jarque-Bera test suggests that the residuals of every sensor contain some level of non-normal skewness and kurtosis. Other tests including the Shapiro-Wilk test point to non-normality in the residuals as well. To better understand the GARCH volatility predictions, we plot the data in Figure 9 for Day 2 along with the one-step-ahead predictions of the corresponding volatility $\pm 2\hat{\sigma_t^2}$. While it seems that GARCH is capturing more of the volatility in the data than ARIMA did, the model overpredicts volatility since it responds slowly to large isolated jumps in the sensor readings.

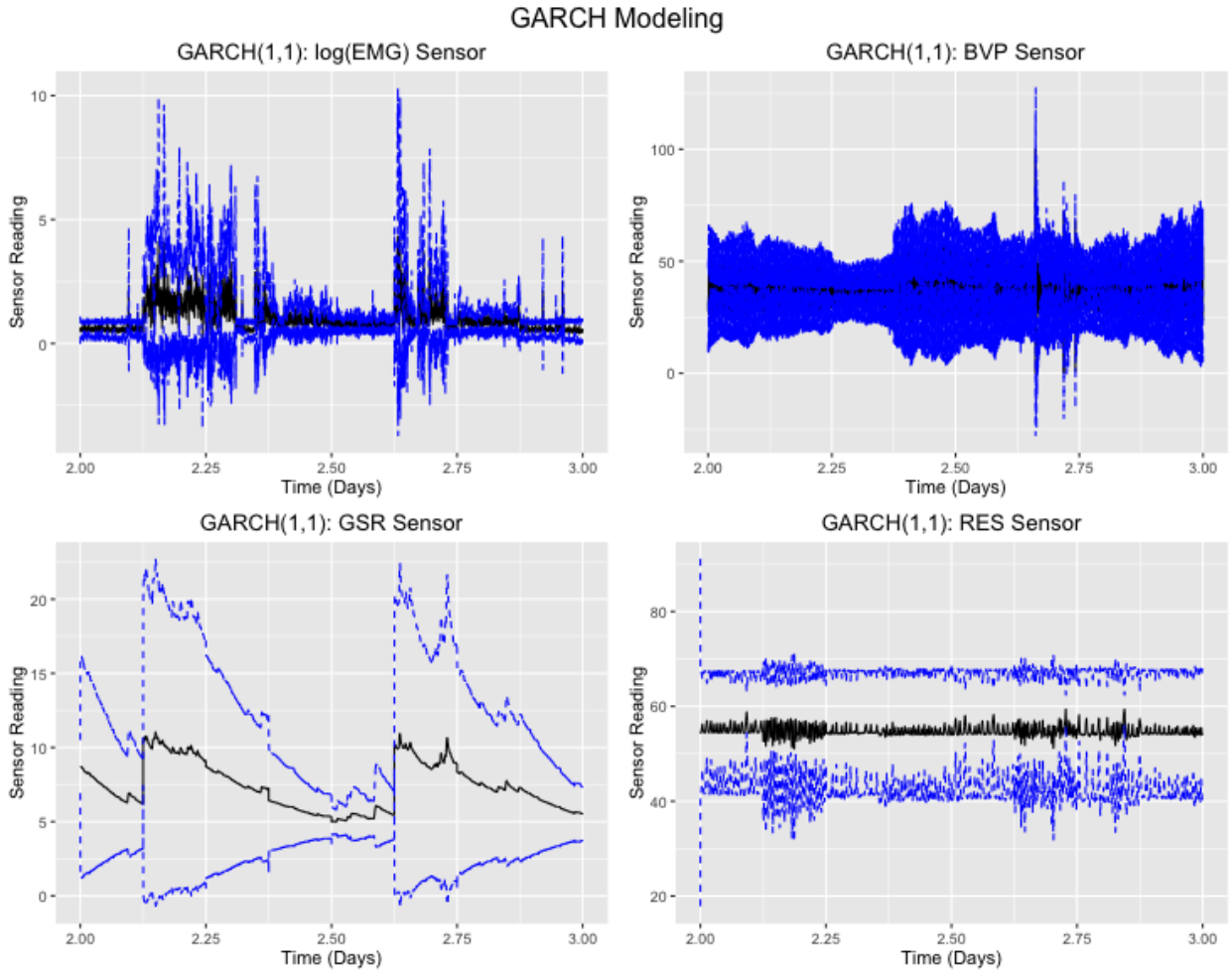

Figure 9: GARCH Models

4.3 Harmonic Regression

After attempting to fit ARIMA and GARCH models with varying levels of success, we turn to harmonic regression to determine whether our series can be modeled from a combination of sine and cosine waves. We regress each of the sensor readings on sine and cosine waves for frequencies that represent the following periods: 1/8 day, 1/2 day, 1 day, 2 days, 3.5 days, 7 days, 10 days, and 20 days. The period 1/8 day was selected to represent each emotion, 1/2 day grouped positive and negative emotions separately, 1 day represented a full emotional cycle, and so forth. The EMG and GSR series coefficient estimates were all highly significant at the 0.001 level while the BVP coefficient estimates were not significant

at all. The RES coefficients were all highly significant at the 0.001 level except for the sine wave indicating 1/8 of a day. As illustrated in Figure 10, the combination of sine and cosine waves appears to capture some of the variation in the EMG, GSR, and RES sensors but does not seem to model the BVP sensor well. In comparison to ARIMA, the harmonic regression performs better for BVP and GSR in terms of MSE. Harmonic regression is a viable option for predicting emotions, but the choice of parameters requires more fine-tuning which is outside the scope of this paper.

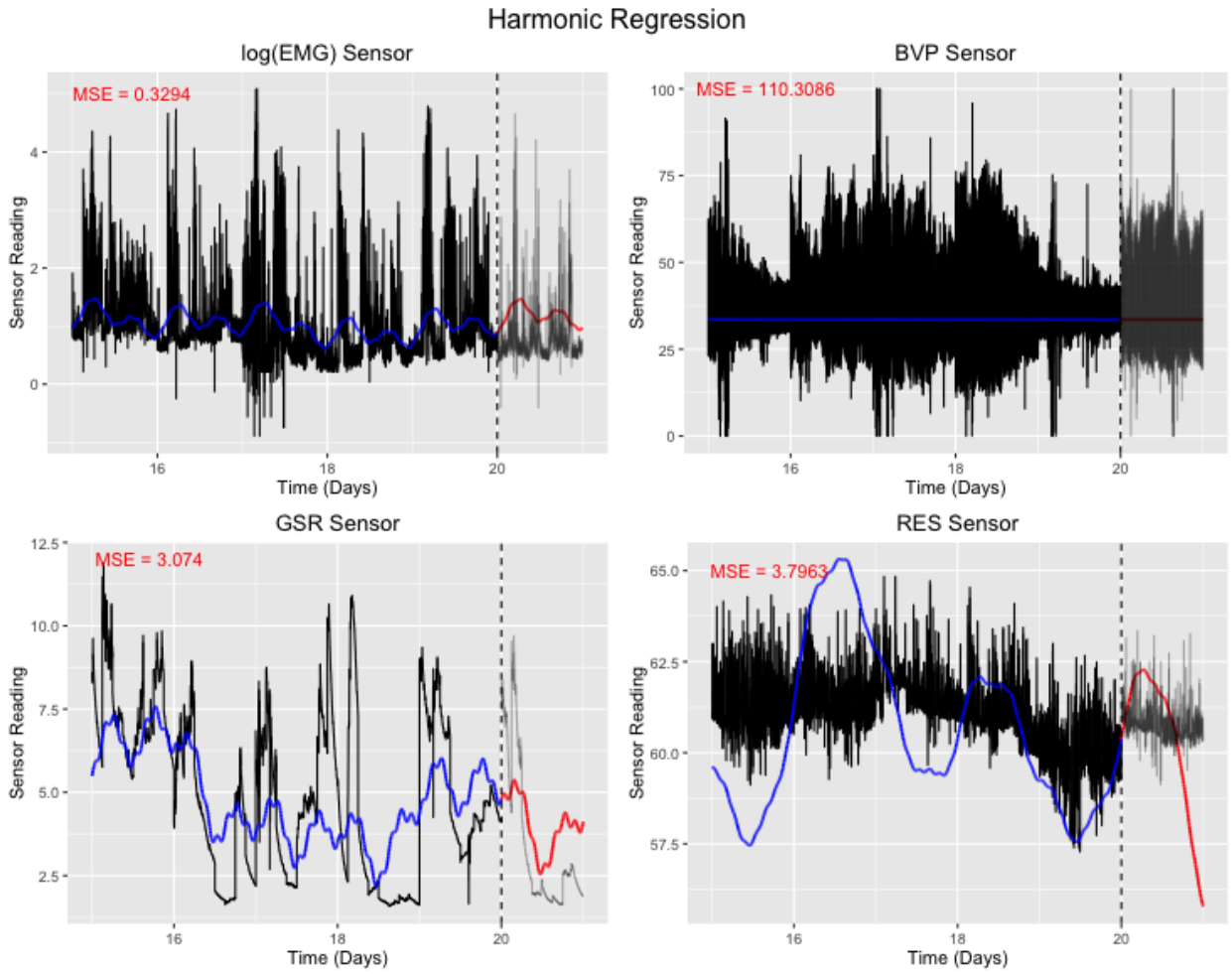

Figure 10: Harmonic Regression

5 Spectral Analysis

Having completed the time side analysis, we briefly turn to spectral analysis. We begin by plotting the logged periodograms for each sensor which can be found in the appendix (Figures A.7 – A.10). Unlike harmonic regression, the periodogram fits sine and cosine waves for all n data points, resulting in a regression that perfectly predicts the data. The periodogram contains peaks at certain frequencies which correspond to significant periodicities in the data. We can see that the raw EMG periodogram indicates that 3/4 day (halfway through the positive emotions), 10 days, and 20 days are significant periodicities. BVP flags 1 day and 2 days as significant. GSR and RES indicate significant periodicities after 2 years with RES also flagging 15 days. In order to make these peaks more obvious, we perform kernel smoothing in Figures A.11 – A.14. In each of the cases, it appears that the Daniell(9,9) kernel which is a $m = 9$ moving average of a $m = 9$ moving average retains sufficient information from the periodogram and results in similar significant periodicities as before.

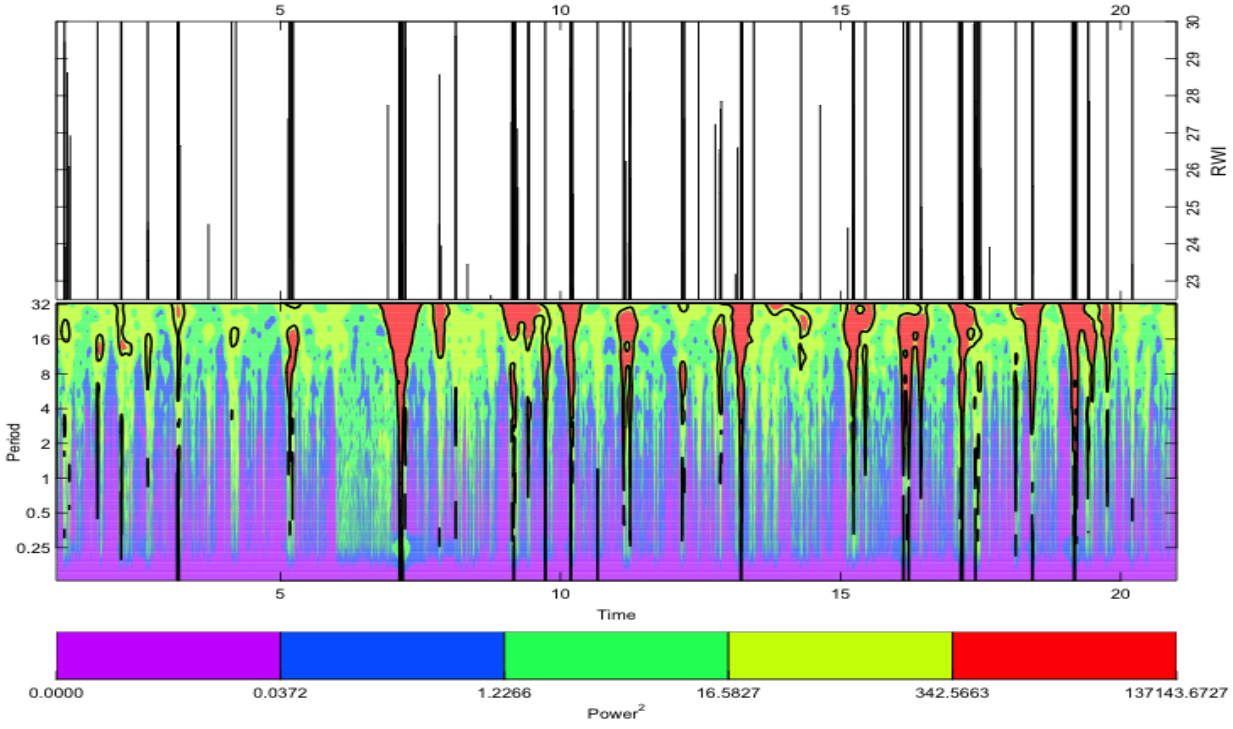

Figure 11: Power spectrum of EMG sensor

Additionally, tapering in performed in Figures A.15 – A.18, but the smoothed periodograms remain essentially the same. Finally, we plot the spectrogram in Figure 11 and Figures A.19 – A.21 after estimating the periodogram through a "morlet" wavelet. We find that EMG and GSR have strong components at 8 to 32 days at nearly all time periods while RES has similar strong components that are more clustered near the beginning time periods. Meanwhile, BVP has strong components at 0 to 1 day at nearly all time periods. A more thorough analysis is required in order to determine whether these strong components are associated more strongly with certain emotions.

6 Conclusion

The answer to my question is yes – certain time series models appear to have the potential to pick up on the unique fingerprints physiological signals leave behind in sensor readings during different emotional states. While ARIMA models performed poorly in picking up the variability in the sensor readings, GARCH seemed to do a better job but overpredicted the volatility in the readings and could be strengthened by more fine-tuning. Despite the performance of GARCH, it seemed that the frequency side analysis had the most promise. Harmonic regression was a move in the right direction. With a more accurate sinusoidal model specification, it would seem that this model could fairly accurately predict physiological signals in different emotional states. Future studies should focus on tuning the model parameters and incorporate the findings from the periodogram into harmonic regression. A machine learning approach could also be interesting in comparing predictive capabilities of these models. Physiological signal processing applied to emotional states is an evolving field and can be greatly strengthened when paired with time series analysis.

References

[1] Picard, Rosalind, Elias Vyzas, and Jennifer Healey. "Toward Machine Emotional Intelligence: Analysis of Affective Physiological State." *IEEE Transactions on Pattern Analysis and Machine Intelligence.* 23, no. 10 (2001): 1175-191. Accessed May 8, 2016.

[2] "Spectral Analysis of Time Series." Rstudio Pubs. Accessed May 8, 2016. http://rstudio-pubs-static.s3.amazonaws.com/9428_1197bd003ebd43 c49b429f22ea4f36e5.html

Appendix

Figure A.1: Decomposition of BVP Sensor

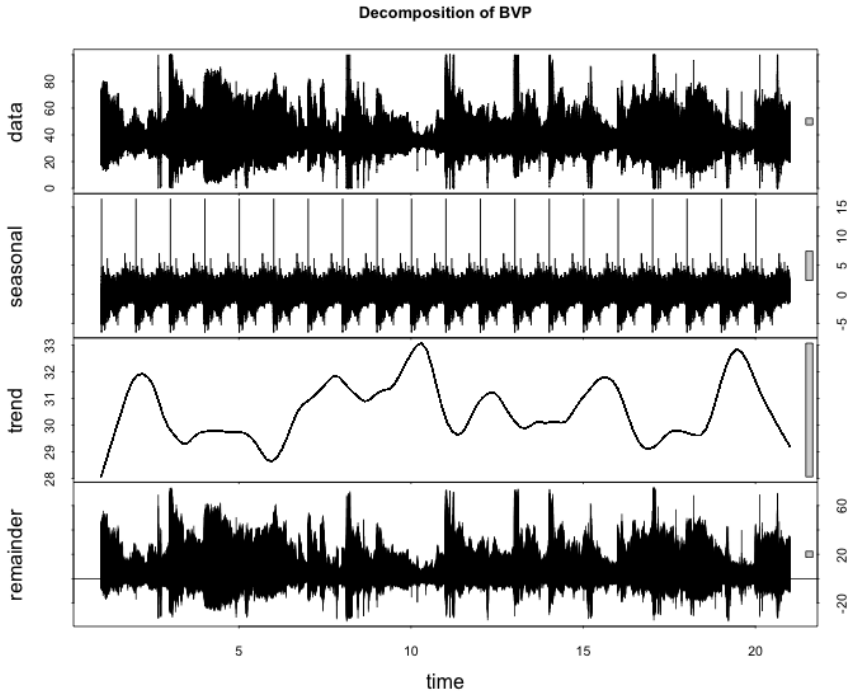

Figure A.2: Decomposition of GSR Sensor

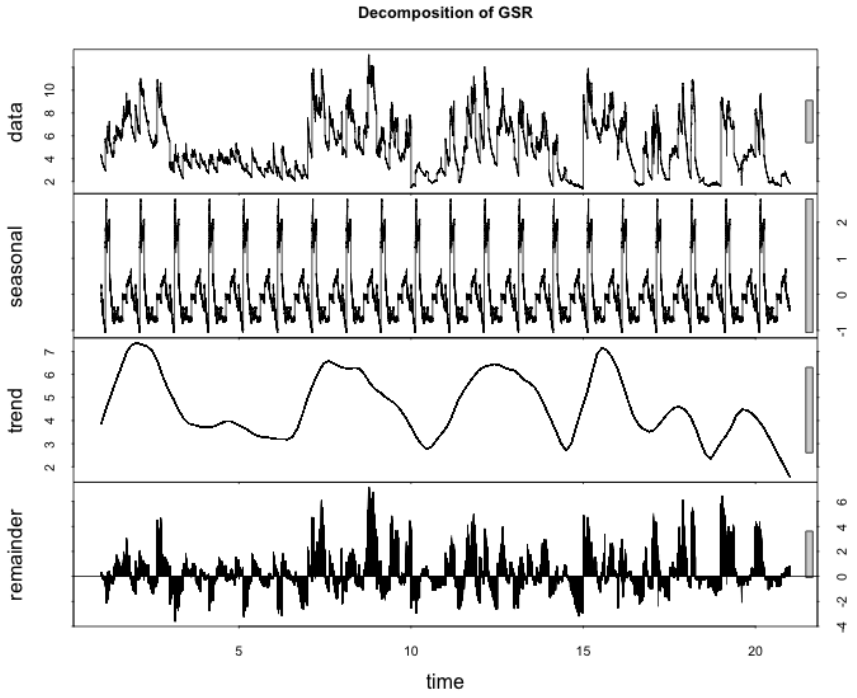

Figure A.3: Decomposition of RES Sensor

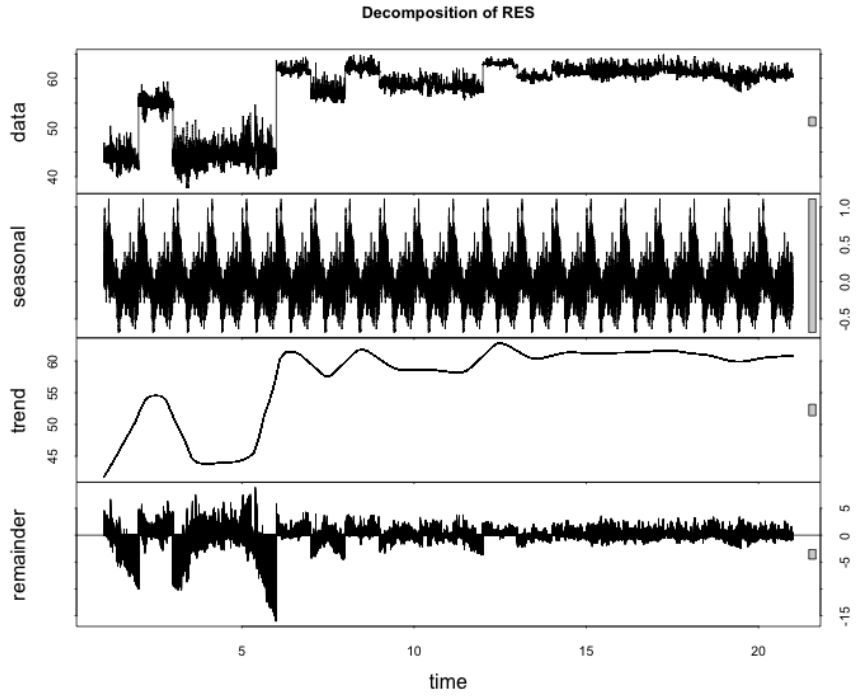

Figure A.4: Raw and differenced ACF and PACF of BVP sensor

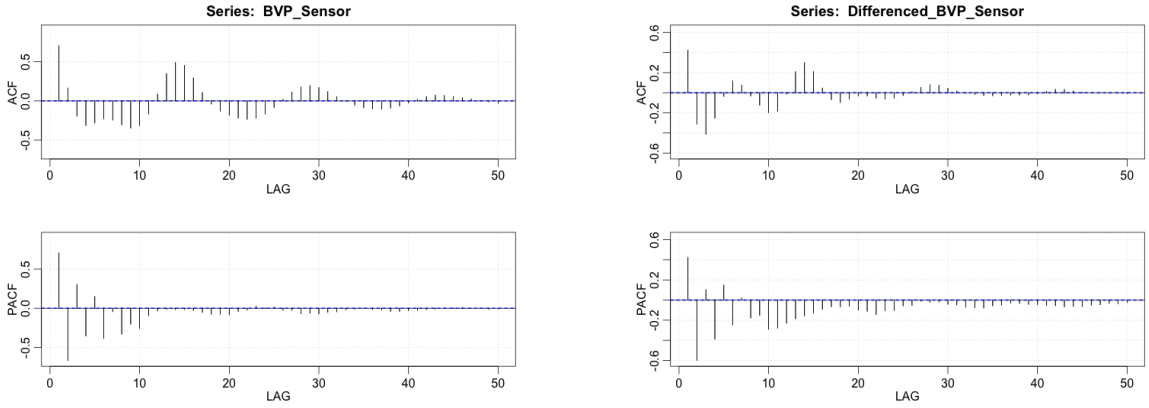

Figure A.5: Raw and differenced ACF and PACF of GSR sensor

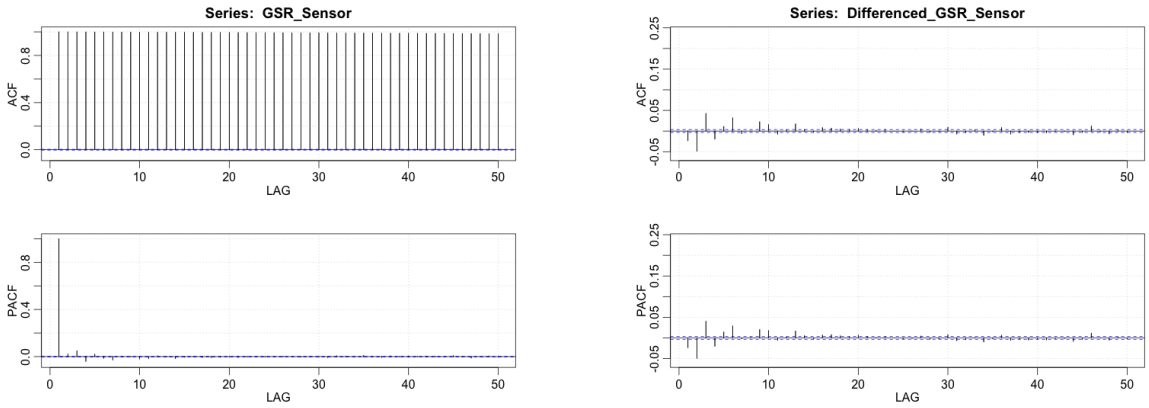

Figure A.6: Raw and differenced ACF and PACF of RES sensor

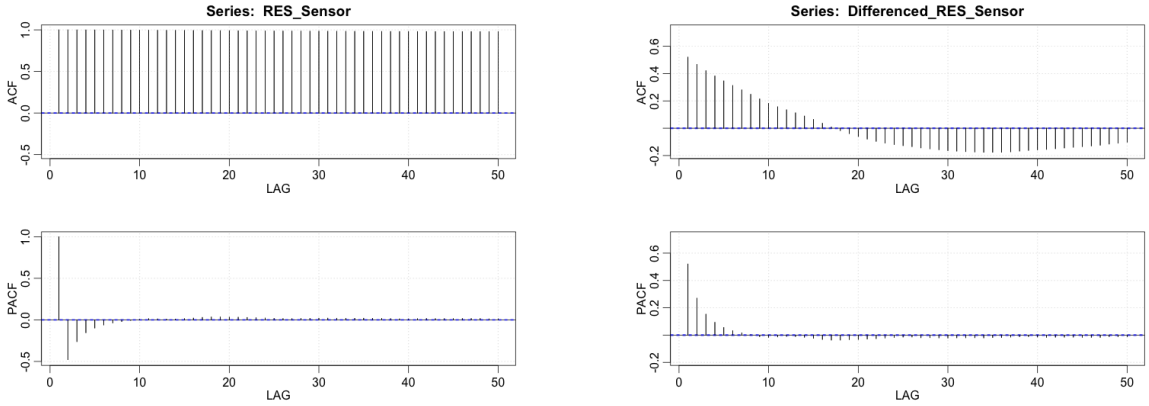

Figure A.7: Logged periodogram of EMG sensor

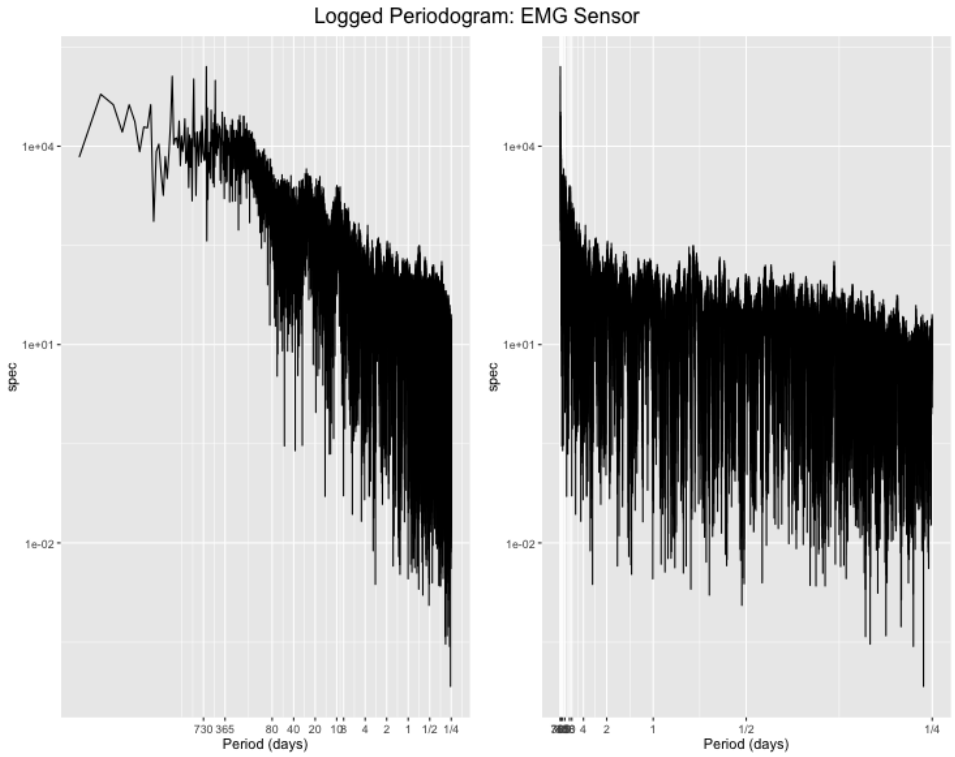

Figure A.8: Logged periodogram of BVP sensor

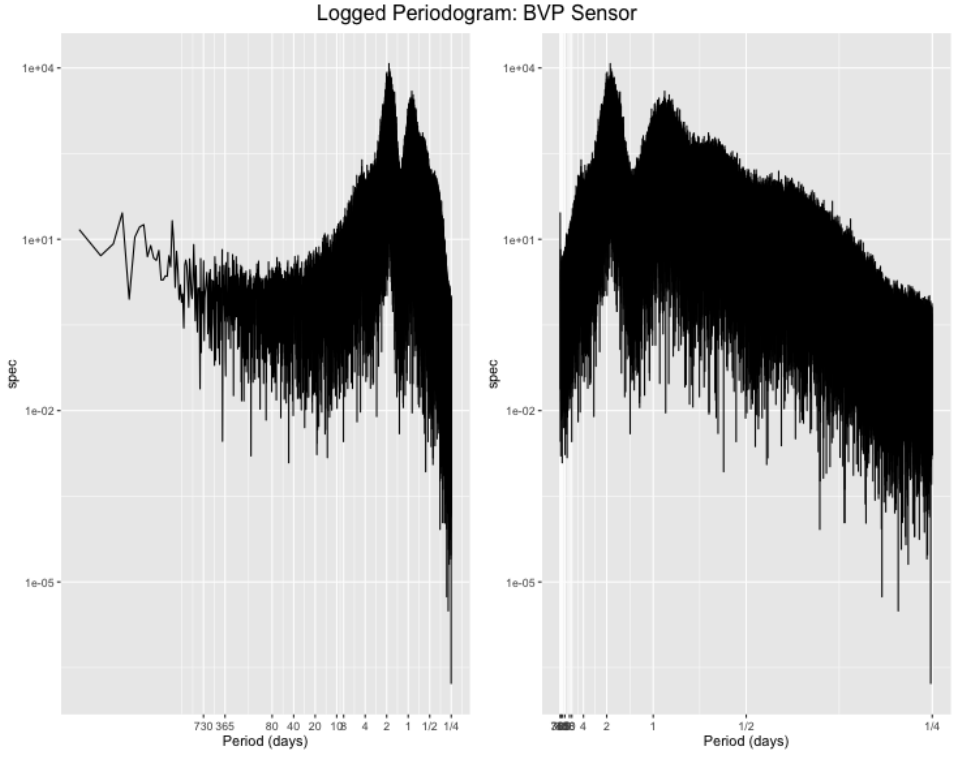

Figure A.9: Logged periodogram of GSR sensor

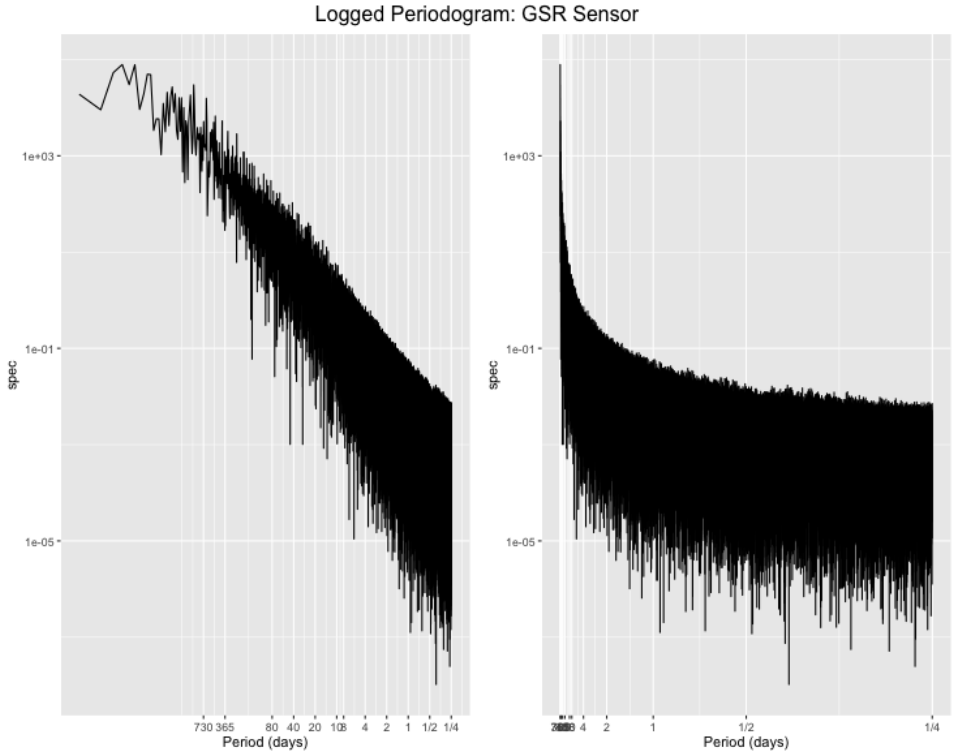

Figure A.10: Logged periodogram of RES sensor

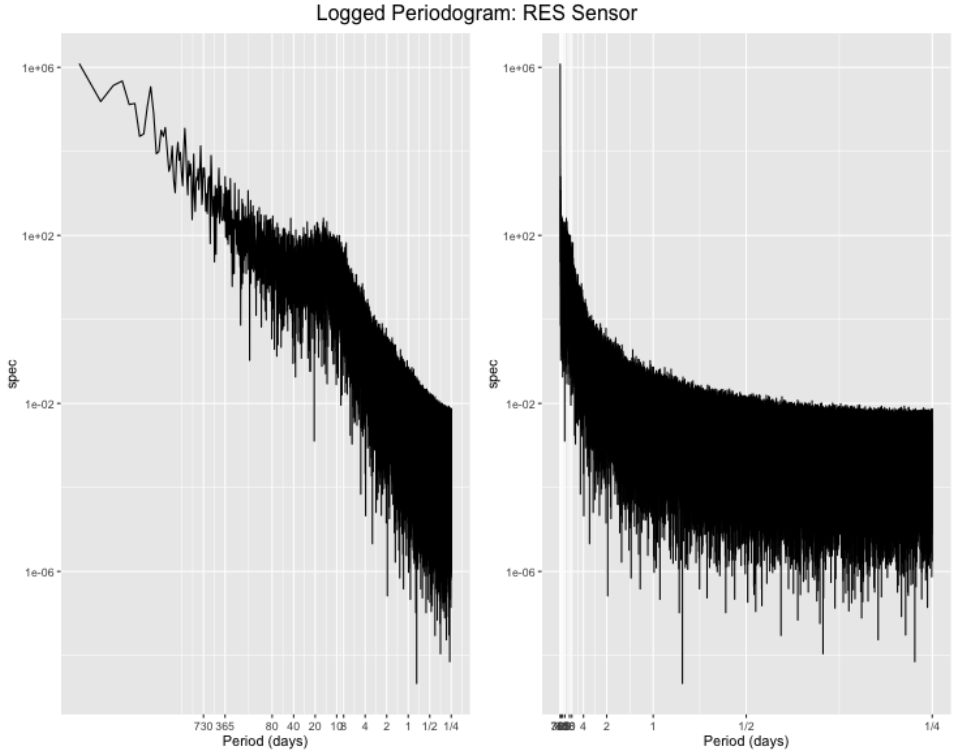

Figure A.11: Kernel smoothing of EMG sensor

Figure A.12: Kernel smoothing of BVP sensor

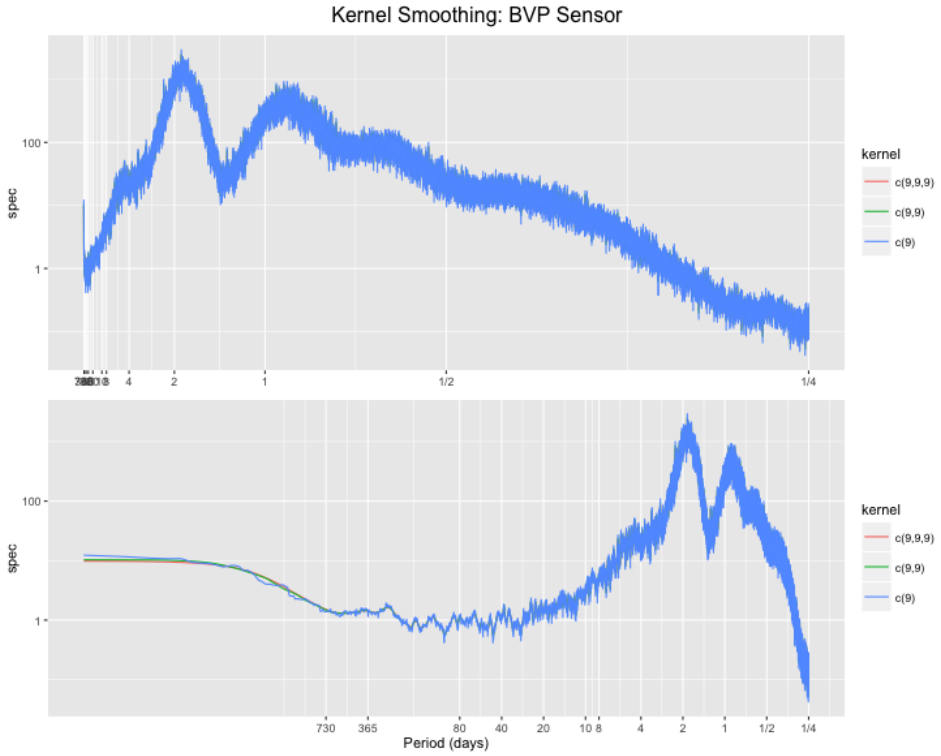

Figure A.13: Kernel smoothing of GSR sensor

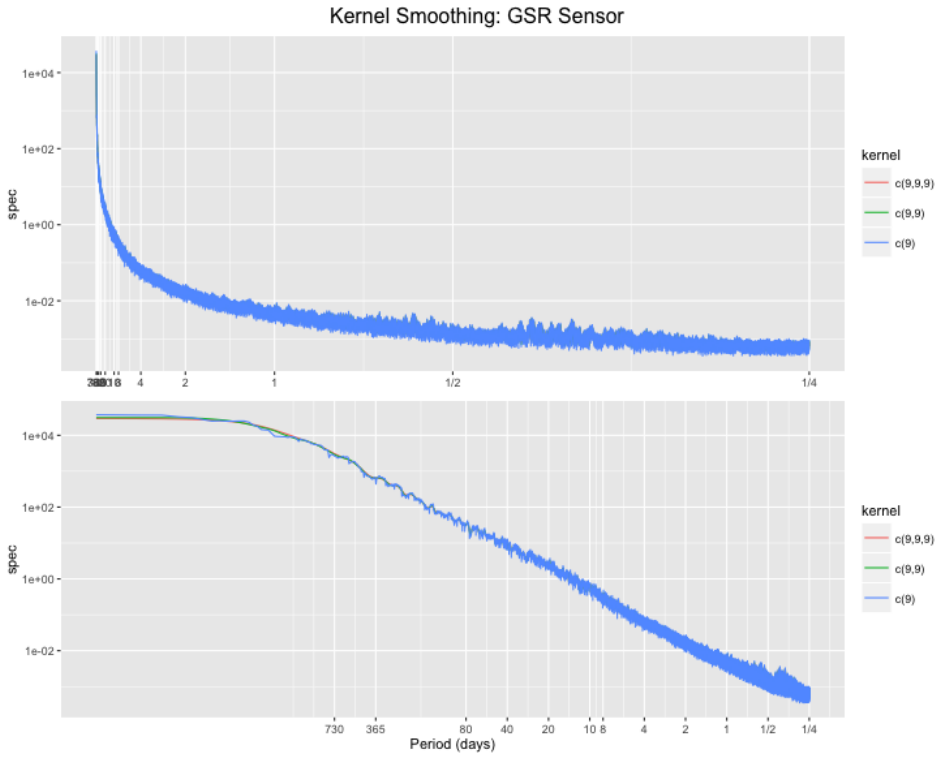

Figure A.14: Kernel smoothing of RES sensor

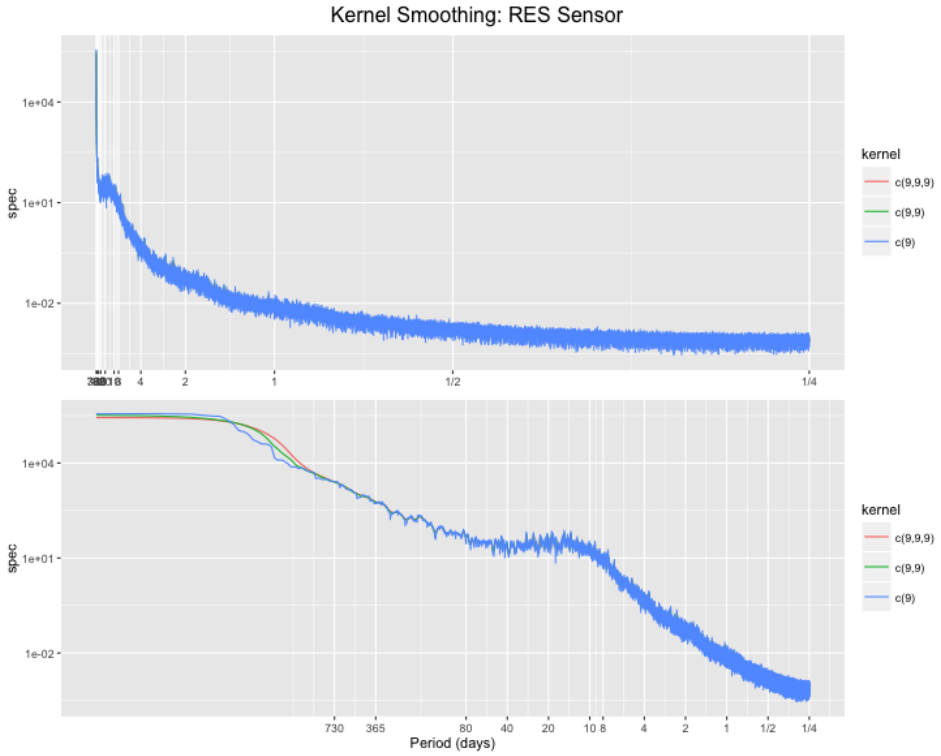

Figure A.15: Tapering of EMG sensor

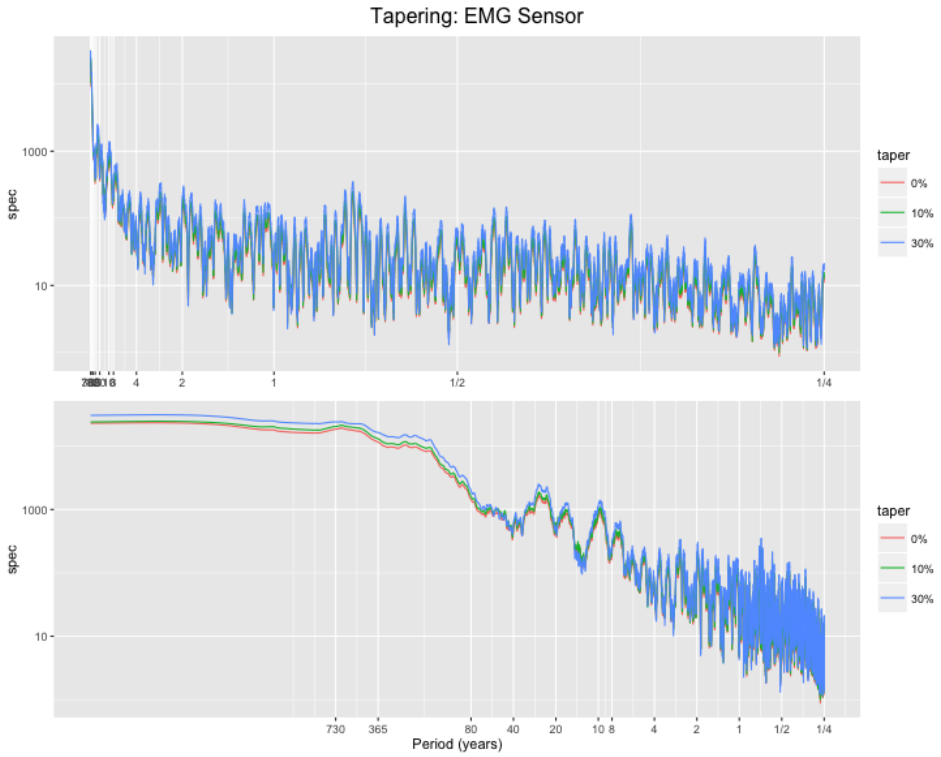

Figure A.16: Tapering of BVP sensor

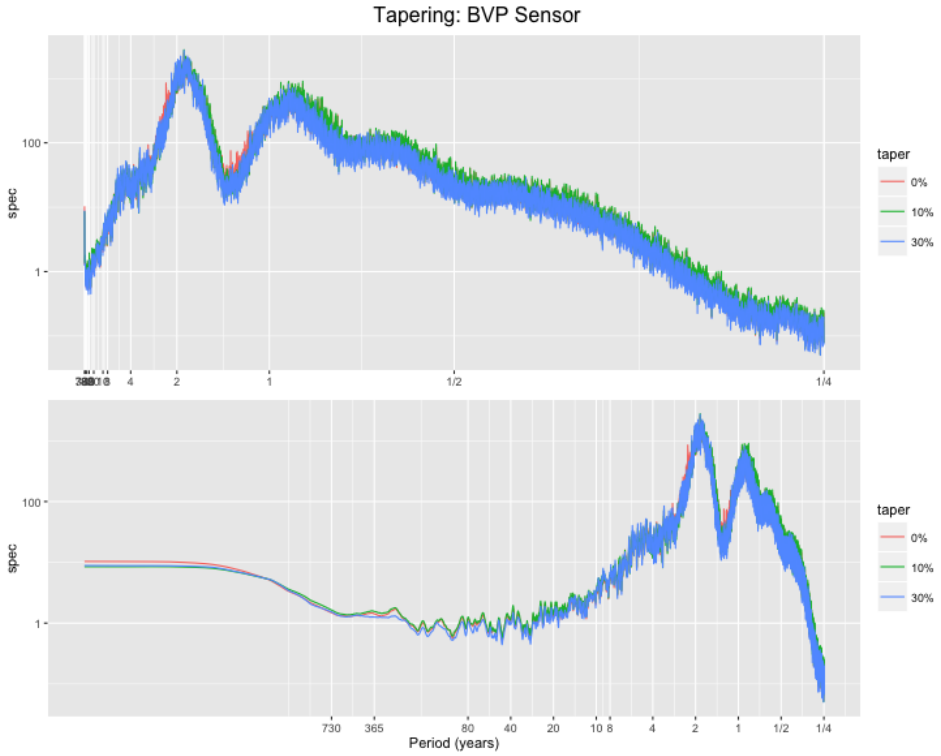

Figure A.17: Tapering of GSR sensor

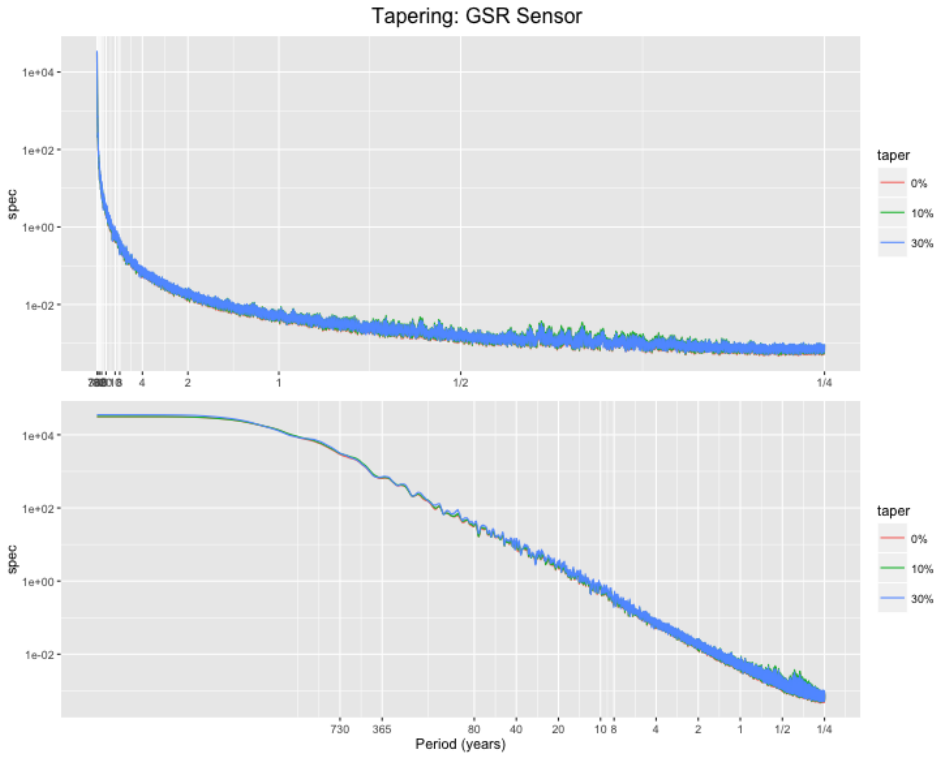

Figure A.18: Tapering of RES sensor

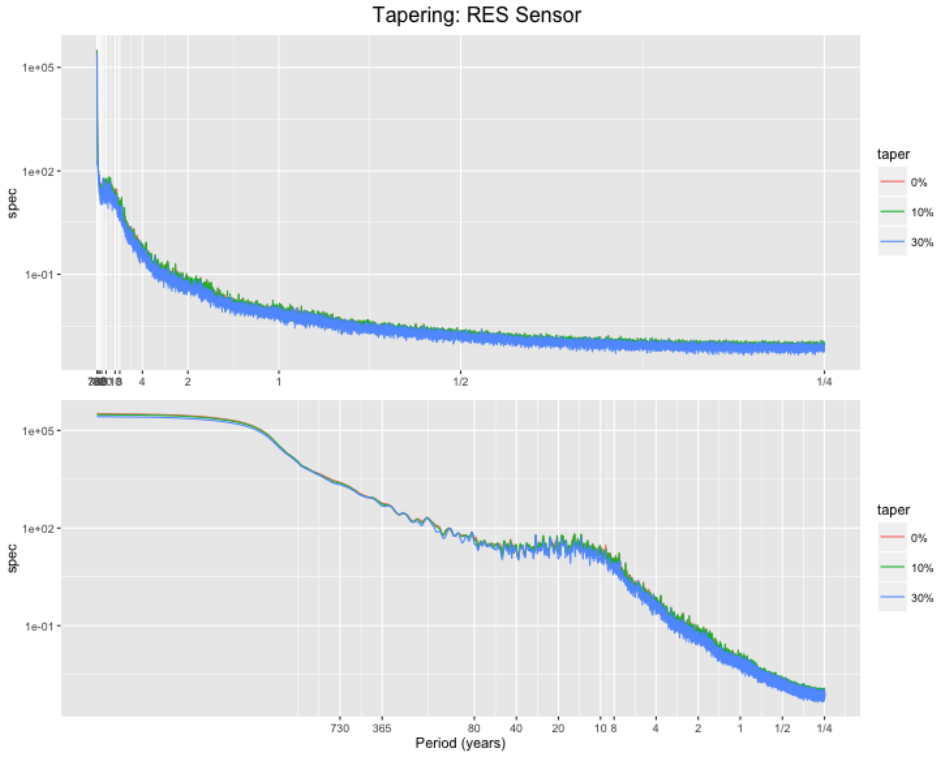

Figure A.19: Power spectrum of BVP sensor

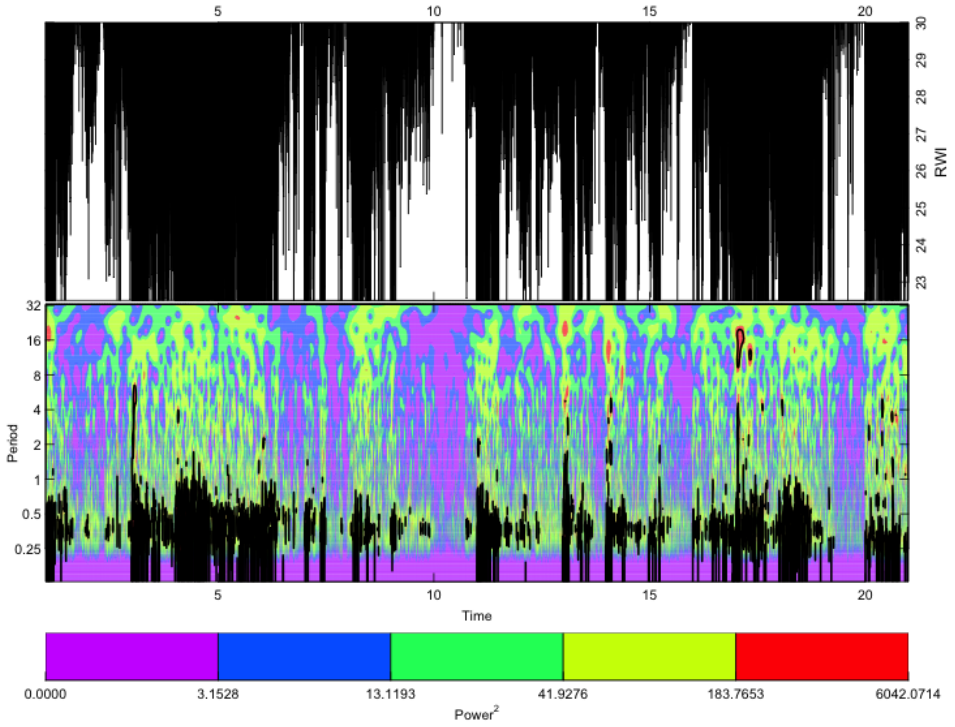

Figure A.20: Power spectrum of GSR sensor

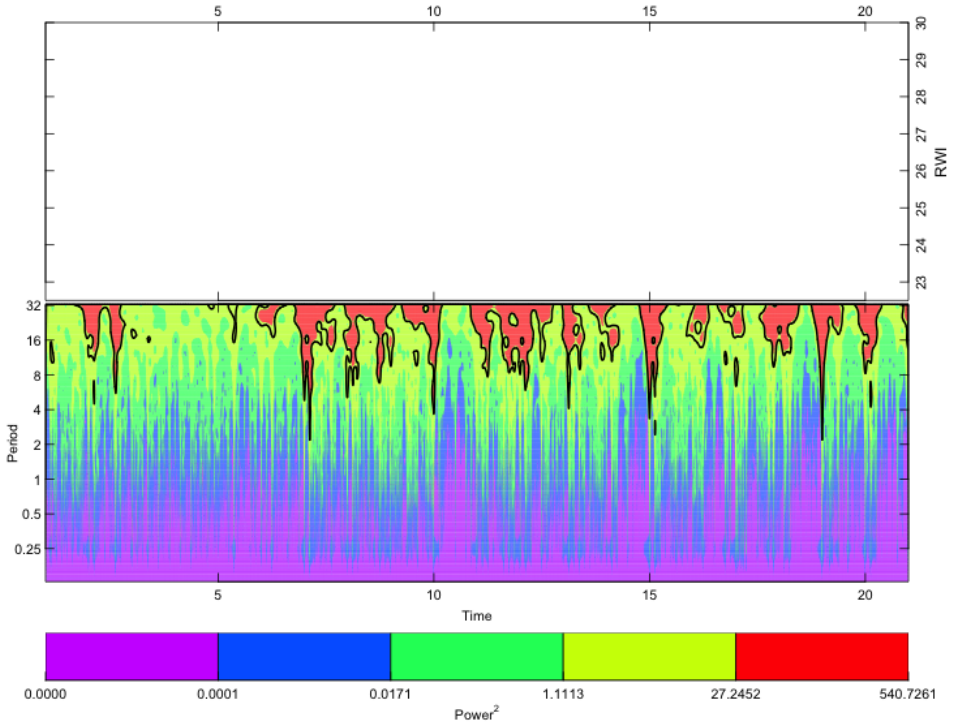

Figure A.21: Power spectrum of RES sensor