Time series models for temperature and rest activity of four healthy patients

B.Hadj-Amar, N. Cunningham, S.Ip

March 9, 2016

Abstract

1 Introduction

2 Methodology

2.1 SARMA

Suppose a time series $\{Y_1, Y_2, ..., Y_T\}$ was observed and, without lost of generality, has zero mean. An ARMA(p, q) model can be fitted which has the form

$$Y_t = \sum_{i=1}^{p} \phi_i Y_{t-i} + \epsilon_t - \sum_{j=1}^{q} \theta_j \epsilon_{t-j} \tag{1}$$

where ϵ_t are some i.i.d. distributed noise with zero mean and variance σ^2.

By using the backshift operator $\widehat{B}$ where

$$\widehat{B} Y_t = Y_{t-1}$$
$$\widehat{B} \epsilon_t = \epsilon_{t-1} \ ,$$

Equation (1) can be expressed as

$$Y_t = \sum_{i=1}^{p} \phi_i \widehat{B} Y_t + \epsilon_t - \sum_{j=1}^{q} \theta_j \widehat{B} \epsilon_t$$

which can be rearranged to

$$\theta \left(\widehat{B} \right) Y_t = \phi \left(\widehat{B} \right) \epsilon_t$$

where

$$\phi(x) = 1 - \sum_{i=1}^{p} \phi_i x$$
$$\theta(x) = 1 - \sum_{i=1}^{q} \theta_i x$$

1

are called the AR and MA characteristic equations respectively.

Seasonality, of lag s, can be considered by including the AR and MA seasonal characteristic equation, $\Phi(x)$ and $\Theta(x)$ respectively, where

$$\Phi(x) = 1 - \sum_{i=1}^{P} \Phi_i x^{is} \tag{2}$$

$$\Theta(x) = 1 - \sum_{j=1}^{Q} \Theta_j x^{js} . \tag{3}$$

The SARMA$(p,q) \times (P,Q)_s$ model uses the seasonal characteristic equation where the model has the form

$$\phi\left(\widehat{B}\right)\Phi\left(\widehat{B}\right)Y_t = \theta\left(\widehat{B}\right)\Theta\left(\widehat{B}\right)\epsilon_t \tag{4}$$

Assuming the noise is Normally distributed, the maximum likelihood can be used to estimate the SARMA parameters: $\phi_1, ..., \phi_p, \theta_1, ..., \theta_q, \Phi_1, ...\Phi_P, \Theta_1, ..., \Theta_Q$. The assumption of Normal i.i.d. noise can be checked by investigating the residuals of fitting the SARMA model onto the data.

The objective was to fit SARMA onto activity and skin temperature time series data and see how they perform.

2.2 Spectral Analysis and Harmonic Regression

Let us consider the periodic model
$$X_t = A\cos(2\pi\omega t) + Z_t,$$

where Z_t is white noise, A is the amplitude, ϕ is a phase shift that determine the initial point of the cosine function and ω is a fixed frequency. Using trigonometric identities we rewrite X_t:

$$X_t = \beta_1\cos(2\pi\omega t) + \beta_2\sin(2\pi\omega t) + Z_t, \tag{5}$$

where $\beta_1 = A\cos(\phi)$ and $\beta_2 = -A\sin(\phi)$.

We use the periodogram as a tool to pick the frequencies that drive the data; then we estimate for each frequency, its corresponding regressions coefficients by using least squares estimates; finally, we can easily obtain the estimate of the amplitude A and the phase shift ϕ:

$$A = \sqrt{\hat{\beta}_1^2 + \hat{\beta}_2^2}, \qquad \phi = \tan^{-1}\left(\frac{\hat{\beta}_2}{\hat{\beta}_1}\right) \tag{6}$$

As it discussed in [1], the *Spectral Representation Theorem* states that any weakly stationary time series can be approximated as the sum of sinusoidal waveforms oscillating at different frequencies:

$$X_t = \mu + \sum_{k=1}^{K}\left\{\beta_{k1}\cos(2\pi\omega_k t) + \beta_{k2}\sin(2\pi\omega_k t)\right\}, \tag{7}$$

where $\{\omega_k\}_{k=1}^{K}$ is a collection of distinct fixed frequencies.

In this scenario, least squares estimates from this regression can be effectively obtained by associating the data to the frequency domain via the Discrete Fourier Transform (DFT), and then obtaining a statistic called the *periodogram*, which is an unbiased estimator of the *spectral density function*. Informally, the periodogram measures the squared correlation of the data with sines and cosine oscillating at frequency $\omega_j = j/T$

3 Applications

3.1 SARMA

Selecting p, q, P, Q, s to fit the SARMA model, onto the time series data, is troublesome because it involves investigating 5 parameters.

s was selected and fixed to be 24 hours because it was observed, by inspection, that there may be a cyclic behaviour, with a period of 24 hours, in the time series.

The Akaike information criterion (AIC) and Bayesian information criterion (BIC) was used to select which p, q, P, Q to fit the SARMA model onto the data. The information criterions have the form

$$\text{AIC} = T \ln \widehat{\sigma}^2 + (p + q + P + Q)2 \tag{8}$$

$$\text{BIC} = T \ln \widehat{\sigma}^2 + (p + q + P + Q) \ln T \tag{9}$$

where $\widehat{\sigma}^2$ is the estimated variance of the residuals. Firstly, ARMA(p, q) models were fitted for $p = 0, 1, ..., 5$ and $q = 0, 1, ..., 5$. The p, q pair which minimises an information criterion was selected and fixed. Secondly, SARMA$(p, q) \times (P, Q)_s$ models were fitted for $P = 0, 1, 2$ and $Q = 0, 1, 2$. The P, Q pair was selected which minimises an information criterion. MATLAB *regARMA* was used to fit the SARMA model onto the data.

The model was tested by fitting the SARMA model, using the AIC/BIC procedure, onto the first 3 days of the data. The model was then extrapolated to do a 24 hour forecast. Such a forecast was then compared with the actual data using the mean squared error (and the logarithmic scoring). That is suppose $\{\widehat{Y}_{T+1}, \widehat{Y}_{T+2}, ..., \widehat{Y}_{T+h}\}$ was the forecast of $\{Y_{T+1}, Y_{T+2}, ..., Y_{T+h}\}$, then the mean squared error (MSE) was calculated using

$$\text{MSE} = \frac{\sum_{k=1}^{h} \left(\widehat{Y}_{T+k} - Y_{T+k} \right)^2}{h} \ . \tag{10}$$

The forecast error, $\{\widehat{e}_{T+1}, \widehat{e}_{T+2}, ..., \widehat{e}_{T+h}\}$, was obtained and used to estimate approximately the error in the MSE using the Delta method given as

$$\sigma_{\text{MSE}} = \frac{2}{h} \sqrt{\sum_{k=1}^{h} \left(\widehat{Y}_{T+k} - Y_{T+k} \right)^2 \left(\widehat{e}_{T+k} \right)^2} \ . \tag{11}$$

3.2 Spectral Analysis and Harmonic Regression

We now apply the techniques summarised above on a dataset consisting of two different measurements of 4 healthy patients (whose ID are 2, 8, 24, 26). These observations are temperature and rest activity.

Temperature

We initially focus on the temperature of patient 8, whose time series is given in Figure 1.

The periodicity of this time series suggests applying an harmonic regression to fit the data. The first step for working on spectral analysis is considering a zero mean process. Therefore, we performed our analysis on the residuals, by subtracting the trend.

In order to find the frequencies that drive the data, we explored the periodogram; notice that it is recommended to investigate behaviours of the smoothed periodogram, which is a *consistent* estimator of the spectral density function.

In Figure 2, we show the periodogram for this time series, and two different smoothing version of it, using uniform weights and *Daniell* weights. We can notice how the main frequency which drives the data is around 0.04, i.e 1/24 hours.
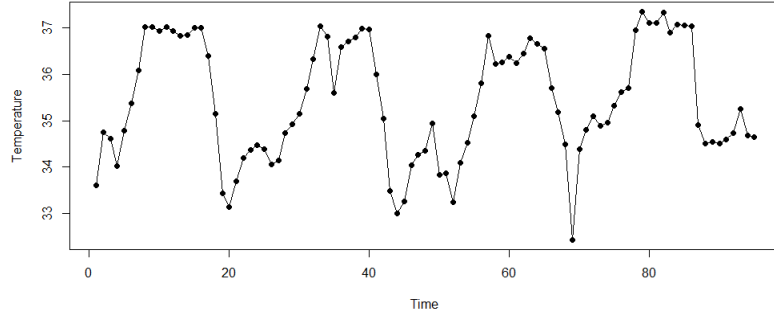
Figure 1: Time series of the temperature of patient 8.

Furthermore, it can be shown that the spectrum of any (weakly) stationary time series can be approximated through the spectrum of an AR(p) model, whose explicit form is known. In order to find the *best* lag parameter p we fit several AR(p) to the data, for increasing values of p. We then calculated two measures of quality of the model, BIC and AIC; the smallest value for BIC or AIC is a good suggestion for choosing the value of p. We give in Figure 3 an idea of this procedure to find the value of p; hence, we show the spectrum of an AR(2) and AR(9).

Once the correct frequencies are selected, we can apply the harmonics regression explained in Section 1.1. The resulting model is given in Figure 4, where we used a total of 5 harmonics. As we stated before, the main frequency that drives the data results in a period of 24 hours

Finally, we give a 24 hour forecast in Figure 5

Rest Activity

The second time series we analyse is the rest activity, which is measured through an accelerometer placed on the chest of the patient. This data takes a particular shape as it take value zero when the patient is resting, and a continuous range of positive values when he is not resting. Therefore, fitting an harmonic regression appears to be difficult and probably not the best choice. However, we give in Figure 6 the harmonic fitting using 5 harmonics; in the same figure we show an alternative model where we put an indicator function to the fitted values: negative values are taken to be zero.

Testing models

We now turn on testing the models we fit for both temperature and rest activity. In order to do so, we split our data in 3 days of training, and 1 day of testing; to measure the accuracy of our prediction we use standard mean square error (MSE). We summarise this in Table 1 and Table 1
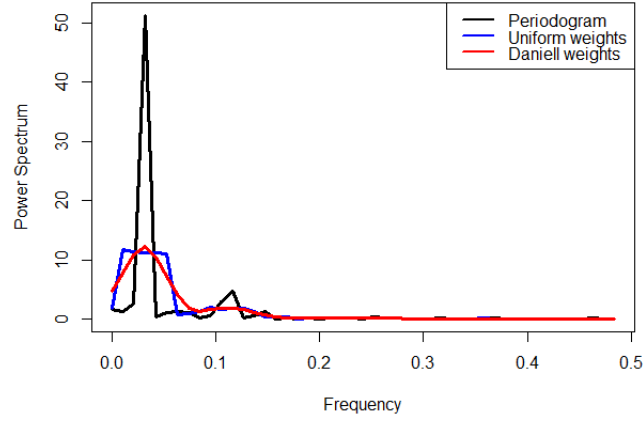
4   Conclusions

Figure 2: Periodogram and smoothed periodograms (uniform and Daniell weights) for temperature of patient 8. This figure is best viewed in colours.
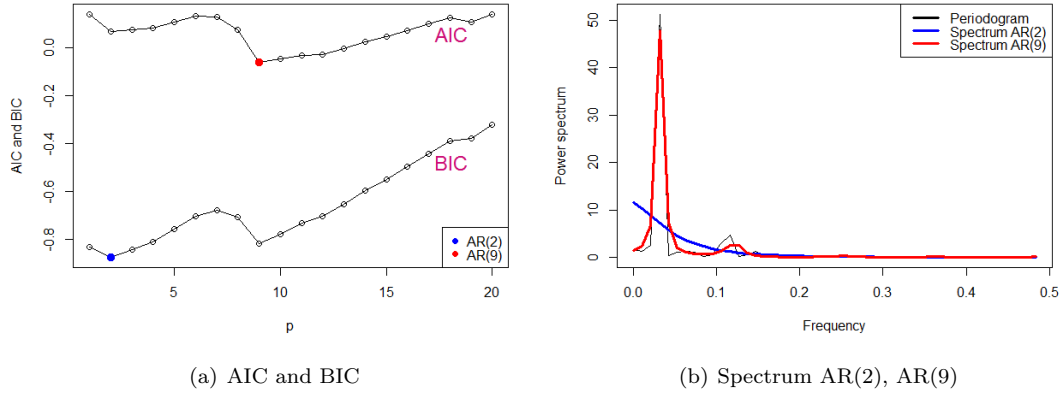

(a) AIC and BIC

(b) Spectrum AR(2), AR(9)

Figure 3: AR(p) approach to obtain the correct frequency representation. Main frequencies that drive the data are around 1/24, and 1/8

Data	MSE	N. Harmonics
Temp2	1.92	5
Temp8	0.58	5
Temp24	0.27	7
Temp26	0.43	3

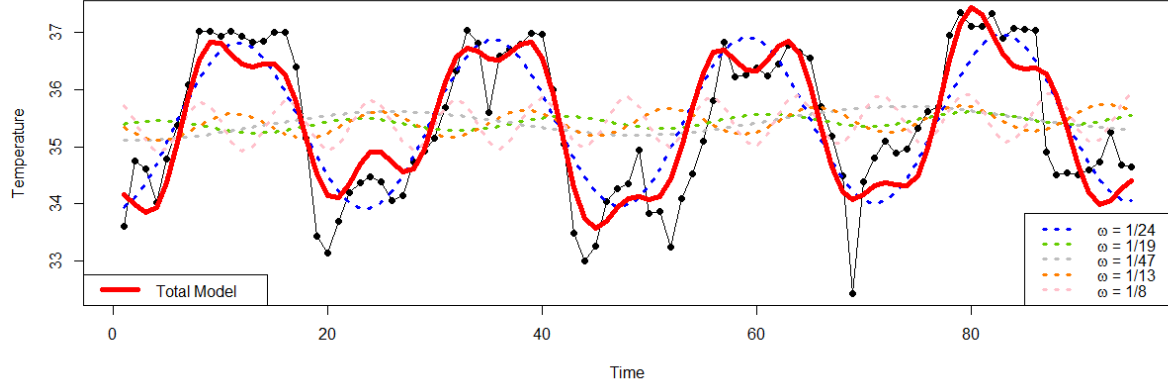Table 1: MSE for temperature. 3 days training, 1 day testing.

Figure 4: Harmonic regression for temperature of patient 8, using 5 different harmonics. Dotted lines represent the fitting for single harmonics; thick red line is the superposition of the dotted harmonics, which is the final fit.
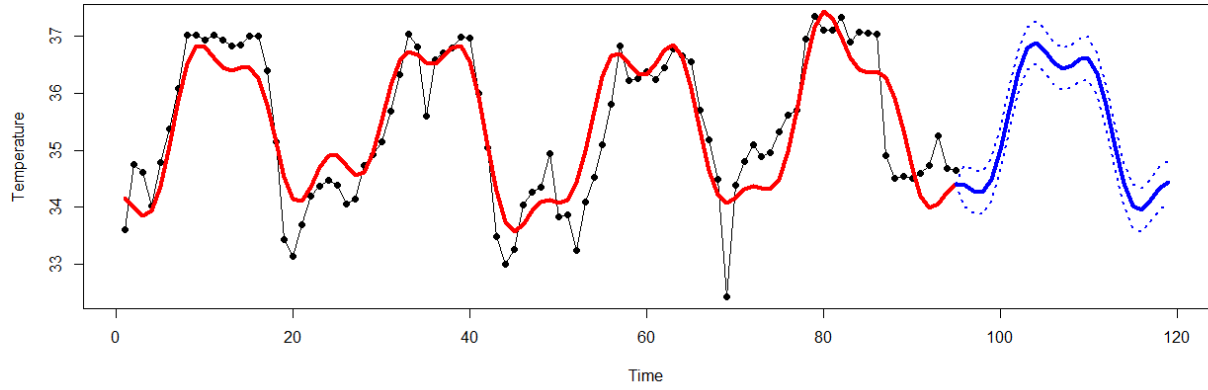

Figure 5: 24 hour forecast for temperature of patient 8.

Data	MSE	N.Harmonics
RA2	430.64	5
RA8	308.60	5
RA24	398.23	6
RA26	577.91	5

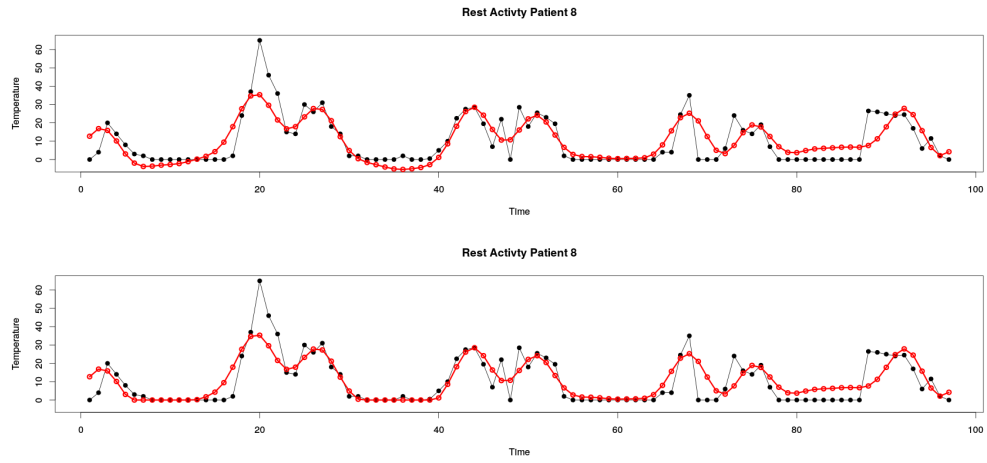Table 2: MSE for rest activity. 3 days training, 1 day testing.

Figure 6: Harmonic regression for rest activity of patient 8, using 5 harmonics. On top it is shown the classic harmonic fitting, whereas on the bottom the alternative model.