Fitting SARMA Models Onto Activity and Skin Temperature Time Series Data

Sherman Ip   March 10, 2016

Abstract

Abstract

1 Introduction

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

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

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

$$\widehat{B} Y_t = Y_{t-1} \qquad (2)$$
$$\widehat{B} \epsilon_t = \epsilon_{t-1} \ , \qquad (3)$$

Equation (??) can be expressed as

$$Y_t = \sum_{i=1}^{p} \phi_i \widehat{B} Y_t + \epsilon_t - \sum_{j=1}^{q} \theta_j \widehat{B} \epsilon_t \qquad (4)$$

which can be rearranged to

$$\theta\left(\widehat{B}\right) Y_t = \phi\left(\widehat{B}\right) \epsilon_t \qquad (5)$$

where

$$\phi(x) = 1 - \sum_{i=1}^{p} \phi_i x \qquad (6)$$

$$\theta(x) = 1 - \sum_{i=1}^{q} \theta_i x \qquad (7)$$

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

$$\Phi(x) = 1 - \sum_{i=1}^{P} \Phi_i x^{is} \qquad (8)$$

$$\Theta(x) = 1 - \sum_{j=1}^{Q} \Theta_j x^{js} \ . \qquad (9)$$

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

$$\phi\left(\widehat{B}\right) \Phi\left(\widehat{B}\right) Y_t = \theta\left(\widehat{B}\right) \Theta\left(\widehat{B}\right) \epsilon_t \qquad (10)$$

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 Methods

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

$$\text{AIC} = T \ln \widehat{\sigma}^2 + (p + q + P + Q)2 \qquad (11)$$

$$\text{BIC} = T \ln \widehat{\sigma}^2 + (p + q + P + Q) \ln T \qquad (12)$$

where $\widehat{\sigma}^2$ is the estimated variance of the residuals. Firstly, $\text{ARMA}(p, q)$ 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, $\text{SARMA}(p, q) \times (P, Q)_s$ 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

$$\text{MSE} = \frac{\sum_{k=1}^{h} \left(\widehat{Y}_{T+k} - Y_{T+k} \right)^2}{h} \ . \qquad (13)$$

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

$$\sigma_{\text{MSE}} = \frac{2}{h} \sqrt{\sum_{k=1}^{h} \left(\widehat{Y}_{T+k} - Y_{T+k} \right)^2 \left(\widehat{e}_{T+k} \right)^2} \ . \quad (14)$$

3 Results

$\text{SARMA}(p, q) \times (P, Q)_{24}$ models were fitted onto the 8 time series data independently. As an example, Figure **??** shows the selected SARMA model, using AIC, being fitted onto the time series data of subject 8. The model was extrapolated to do a 24 hour forecast.

It was observed that the forecast error on the activity time series blew up in the order of $\sim 10^2$. A forecast with such an error is practically useless thus it was considered that the SARMA model did not fit well with the activity time series.
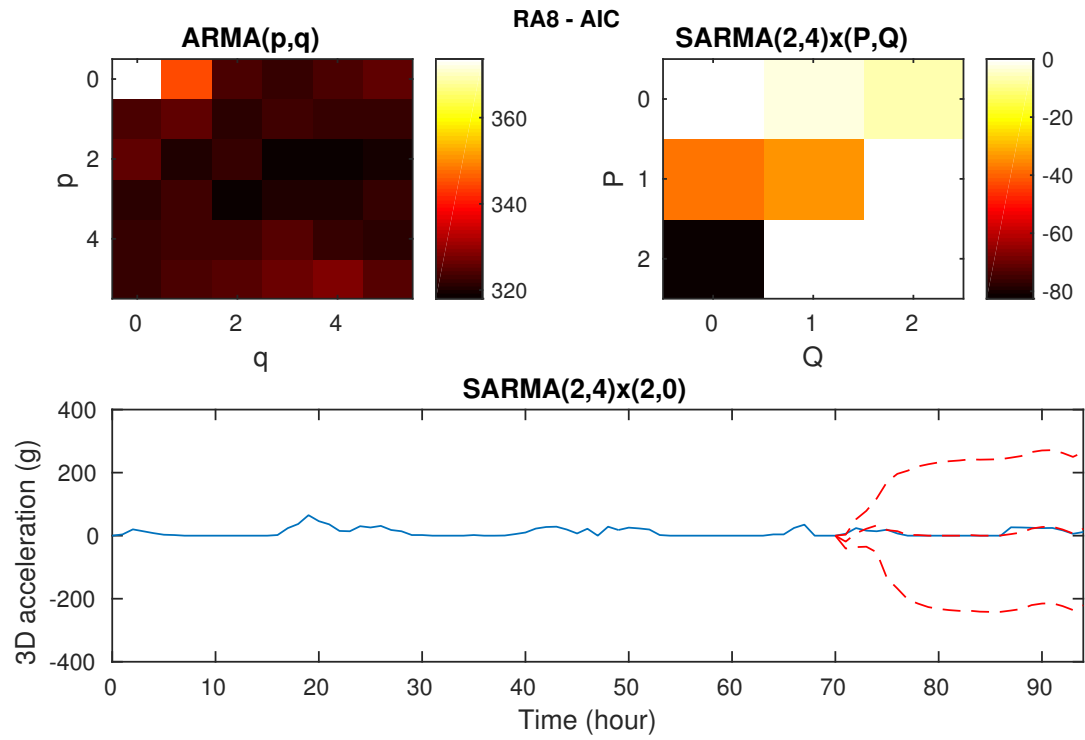
The forecast error on the temperature time series was sensible, for example in Figure **??**.

Data	p	q	P	Q	MSE
Temp2	4	2	2	2	1.1 ± 0.2
Temp8	2	5	2	0	1.5 ± 0.1
Temp24	5	4	2	0	0.54 ± 0.06
Temp26	2	4	2	0	0.69 ± 0.06

Table 1: Hello.

Data	p	q	P	Q	MSE
Temp2	1	0	2	2	1.1 ± 0.2
Temp8	2	1	2	0	1.4 ± 0.1
Temp24	2	4	1	1	0.28 ± 0.03
Temp26	2	0	2	0	0.72 ± 0.07

Table 2: Hello.

(a) Activity
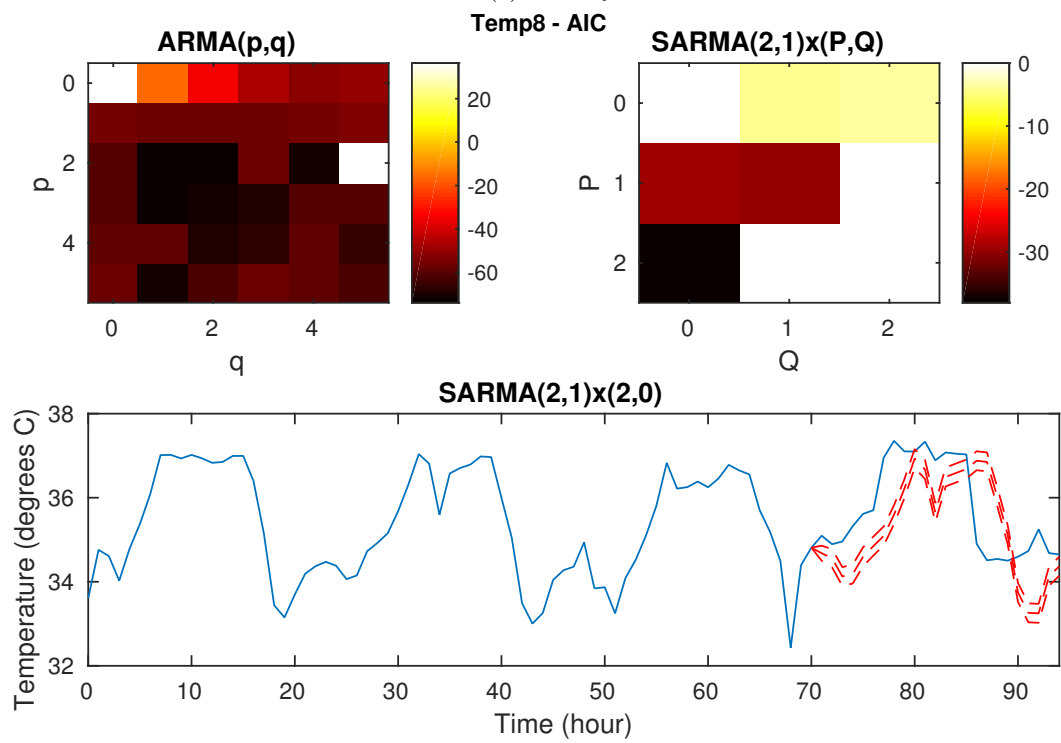
(b) Skin Temperature

Figure 1: hello