Spectral Processing of Signals
Homework 2:
Rational Parametric Methods

Philip Ahl

I have chosen to investigate the least squares approach to the AR and ARMA models. First, looking at the least square AR (LSAR), the model order choice was aided by plotting the white noise variance estimate against the order of the AR polynomial. The plots can be seen for all data sets in Fig.1. The white noise variance estimate gives a picture of how much of the spectrum is given by the signal and how much is given by the noise. When the noise variance estimate decreases with increased order, less of the full signal variance is represented by the error. Here a trade-off is made between model order and variance (not to mix up with the variance of the whole signal). Using this methodology of analysis, the loglynx data set order seems good at 12, the lynx data set at 8 and the sunspot at 9.

Four different information criterion were used separately on each data set. Akaike's Information Criterion (AIC), Akaike's Information Criterion with correction (AICc), Bayesian Information Criterion (BIC), Generalized Information criterion (GIC) were used. Four different were used to get a nuanced picture of where the criterion suggests the order, but mainly the AIC was considered. They were usually pretty coherent, as seen in Fig.2 for the sunspot data set where 9 seemed as a good minimum. This fenomena was seen in all three data sets, where all criterion minimum matched with the variance analysis, hence the above mentioned orders were chosen for each data set.

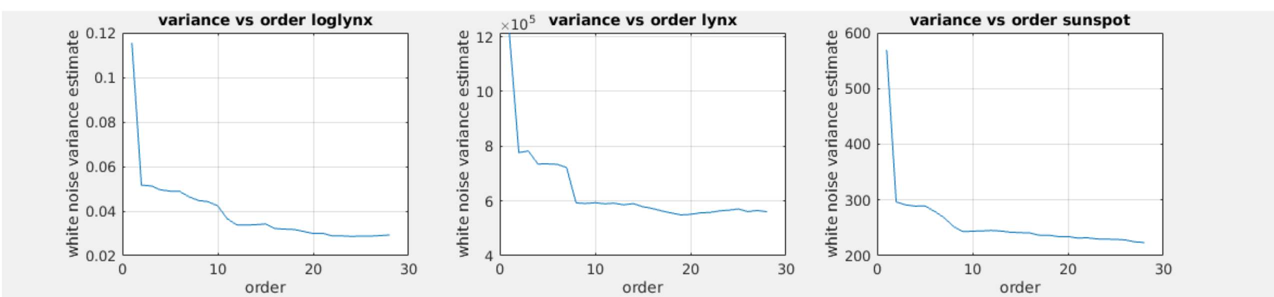

Figure 1

For the ARMA model, similar analyses were made, but slightly more complex since the ARMA model consists of three different orders to choose, AR, MA and the truncated model size K. The methodology of analysis was with the following scheme. A fixed K was chosen, while all permutations of M and N up to the value of K was investigated. The same information criterion was used. The best (lowest) for each criterion and data set was saved for each tried M, together with the corresponding N that yielded that value, were plotted against each other. In Fig.3 the criterion for the lynx data set is shown with K fixed at 20. It is not as clear as with the AR model, but the M (MA) order was set to 11 and the corresponding N (AR) order was found as 8. Similar

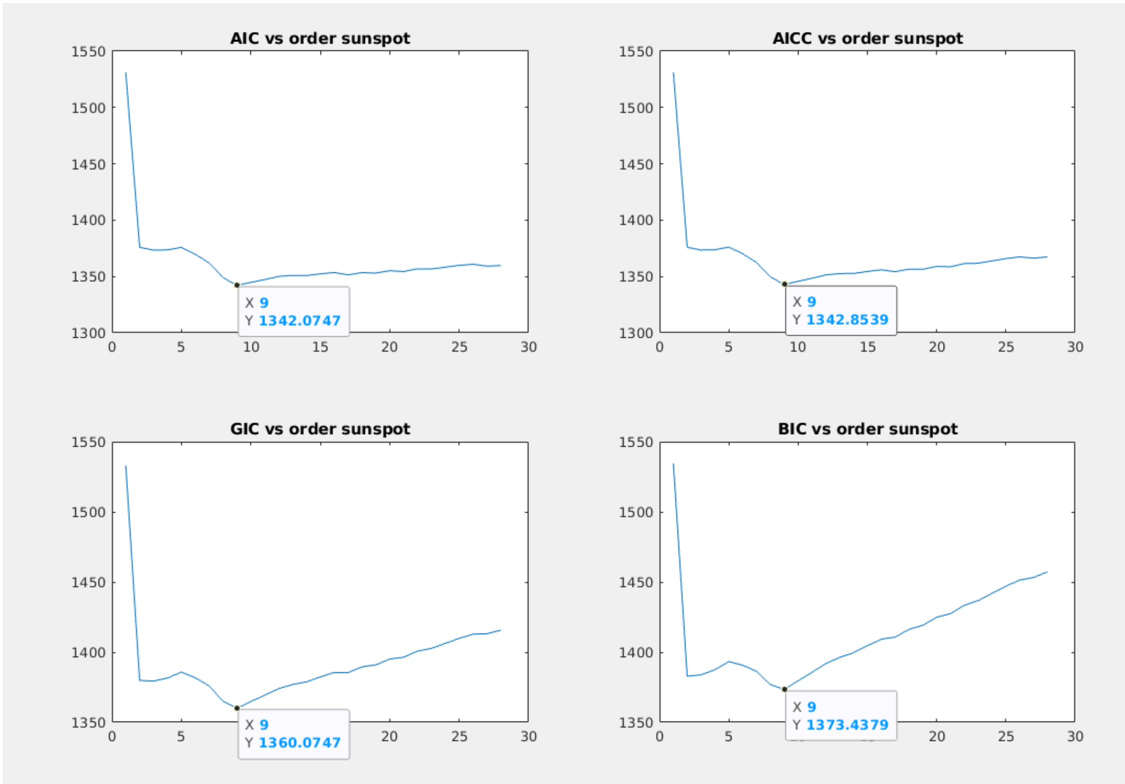

Figure 2

approach was used for the other data sets as well and got M = 5 and N = 2 for the loglynx and M = 9, N = 3 for the sunspot data set. At some point different local minima were tried out, steering towards lower orders due to unstable poles. The steering towards less complex models with lower orders is due to the mind set of Occam's razor.

The spectrum and pole-zero plots of the final models are shown in Fig.4 and Fig.5. Where all poles on the periodogram from zero towards positive pi corresponds on rotating anti clockwise on the PZ-plot starting from "3 o'clock".

The cons of parametric methods such as the ones looked at in this paper is that parameters = freedom of choice. Since parameters can be correlated, all feasible permutations should be considered. For larger data sets, this can be a computationally heavy process. The analysis can grow immensely and the common feel for the goodness of the methods may disappear. The non-parametric method works with less tweaks, henceforth also with less accuracy. A trade-off depending on the signal needs to be done if a full parametric analysis is worth the time and computational burden. A non-parametric method could here be used as a first step as guidance in a rough estimation on how many peaks the spectrum should have.

Having too large order size with respect to the length of the signal can lead to overfitting. This is when we are fitting the model on the training data, giving it a high variance. Having too small order size will lead to underfitting, which will decrease the models variance while increase its bias. Here, I'd say the plotted LSARMA on the loglynx data set was underfitted, biased to a few amount of parameters, loosing information about frequency peaks compared to the others.

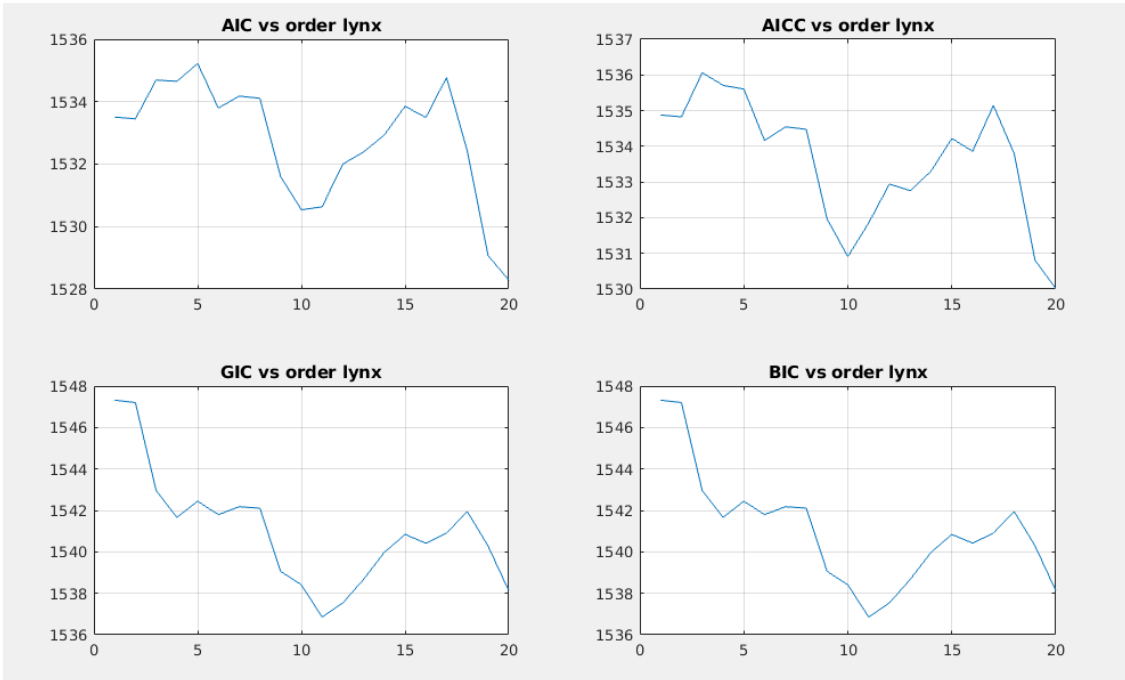

Figure 3

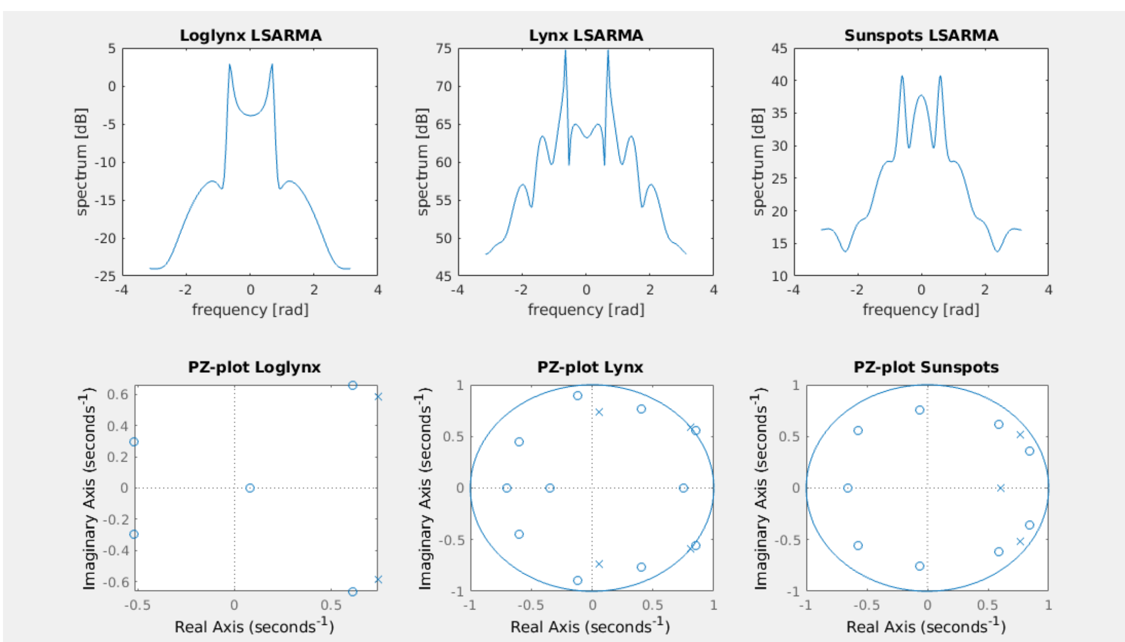

Figure 4

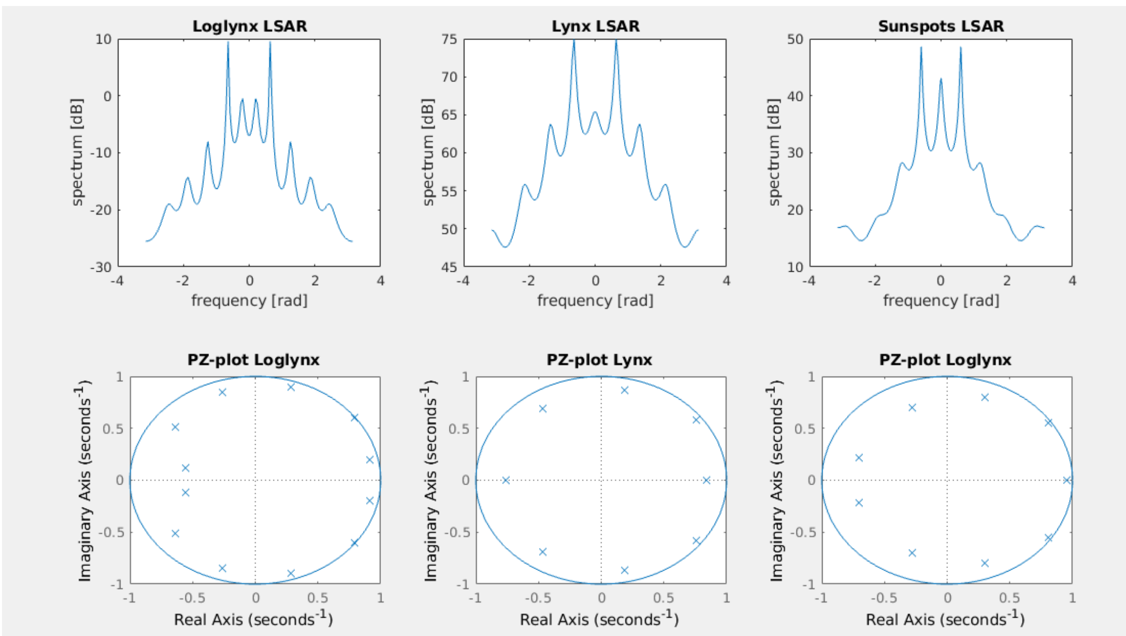
Loglynx LSAR
spectrum [dB]
frequency [rad]
Lynx LSAR
spectrum [dB]
frequency [rad]
Sunspots LSAR
spectrum [dB]
frequency [rad]
PZ-plot Loglynx
Imaginary Axis (seconds-1)
Real Axis (seconds-1)
PZ-plot Lynx
Imaginary Axis (seconds-1)
Real Axis (seconds-1)
PZ-plot Loglynx
Imaginary Axis (seconds-1)
Real Axis (seconds-1)

Figure 5