EE 672 System Identification: Assignment 1, Part 1

Furkan Gül, Student ID: 28795

March 23, 2021

1 Comment on I/O Data

In Matlab's System Identification toolbox, there is an example dataset which is called Dryer Data. This data is collected from a SISO system of hair dryer for 80 seconds with 0.08 sample time. The single input of this hair dryer system/plant is electrical power (denoted by *u2*). As it can be seen from Figure 1, this power input data is a binary sequence of two different input levels (approximately 3.5 and 6.5). The width/duration of each pulse in this input data seems almost random, but it is not. Actually, there is a pseudo algorithm behind generating this kind of data. So, the power input data is a type of pseudo random binary sequence (PRBS in short) which is a deterministic signal not a stochastic. The single output of the hair dryer plant is the temperature (denoted by *y2*) at the outlet of hair dryer. By looking at Figure 1, the output data is changing between 3.5 and 6. After making some mathematical operations on Matlab by using built-in *mean* and *cov* functions, the mean of this temperature output data is 4.89 and its variance is 0.69. On the other hand, the mean and the variance of the input power is 5 and 2.24 respectively. As it is seen from Figure 1, calculated mean and variance values for the input and output data are make sense from the statistical point of view. *i.e.* 68% of the input or output data lies in the one-standard deviation (square root of its variance) interval around its mean value. For the input signal, this interval is [3.5, 6.5] (from [5-1.5, 5+1.5]). 68% interval for the output signal data is [4.06, ,5.72] (from [4.89-0.83, 4.89+0.83]). This calculation can be extended for two-standard deviation (around 95% of data) and three-standard deviation (around 99.7% of data) intervals for the input and output data.

2 Estimation and Validation Data Split

In Dryer data, there are 1000 samples/measurement points. Estimation data set and validation data set for system identification of hair dryer plant will be generated by splitting whole Dryer data into two pieces of 80% and 20% respectively. It is similar to data splitting process in Machine Learning problems but without shuffling. Estimation data set is identified as the first 80% of the original Dryer data by setting Samples as from 1 to 800 in the Select Range window. Then, this data set is called Estimation-Dryer (can be seen in blue

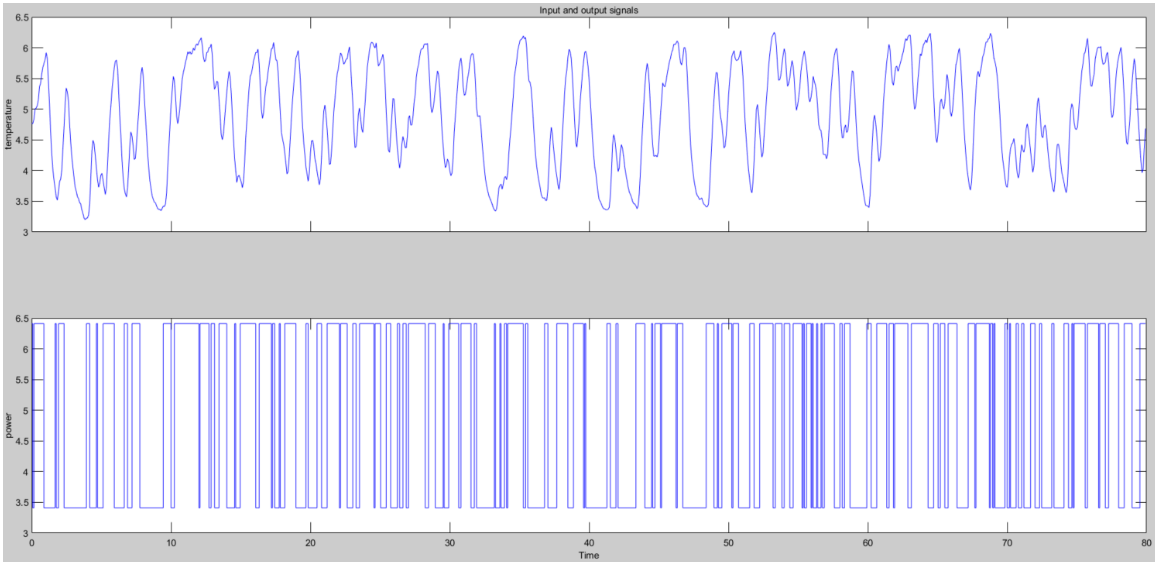

Figure 1: A visual inspection of Dryer Data (An example of SISO System)

color on data window). Validation data set is chosen as the last 20% of the original Dryer data by setting Samples as from 801 to 1000 in the Select Range window. Then, this data set is called Validation-Dryer (can be seen in red color on data window). Both estimation and validation data sets can be seen in Figure 2 and 3 respectively.

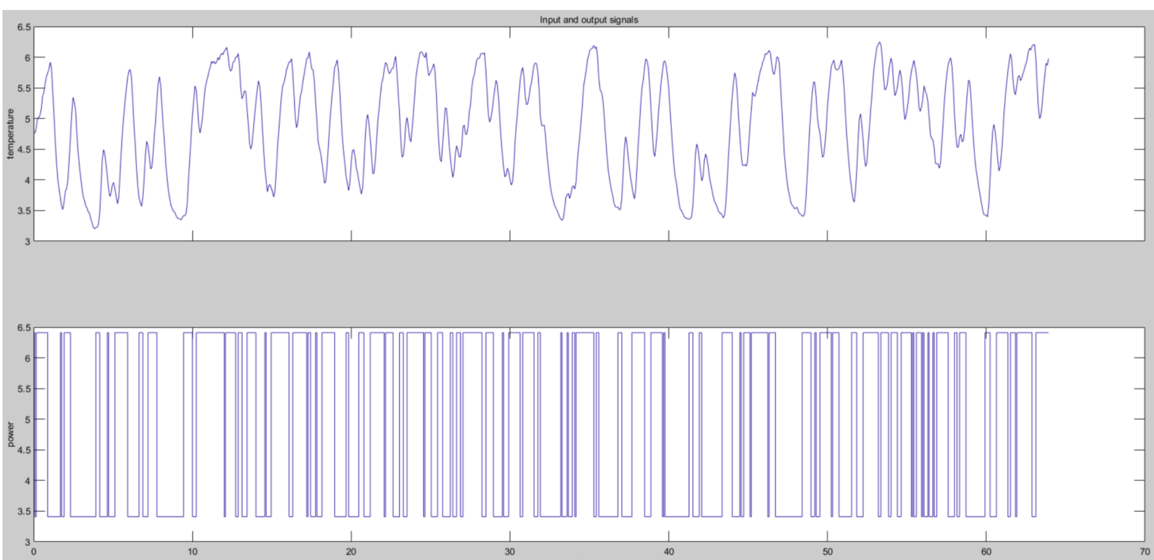

Figure 2: Estimation Data Set for Hair Dryer Plant

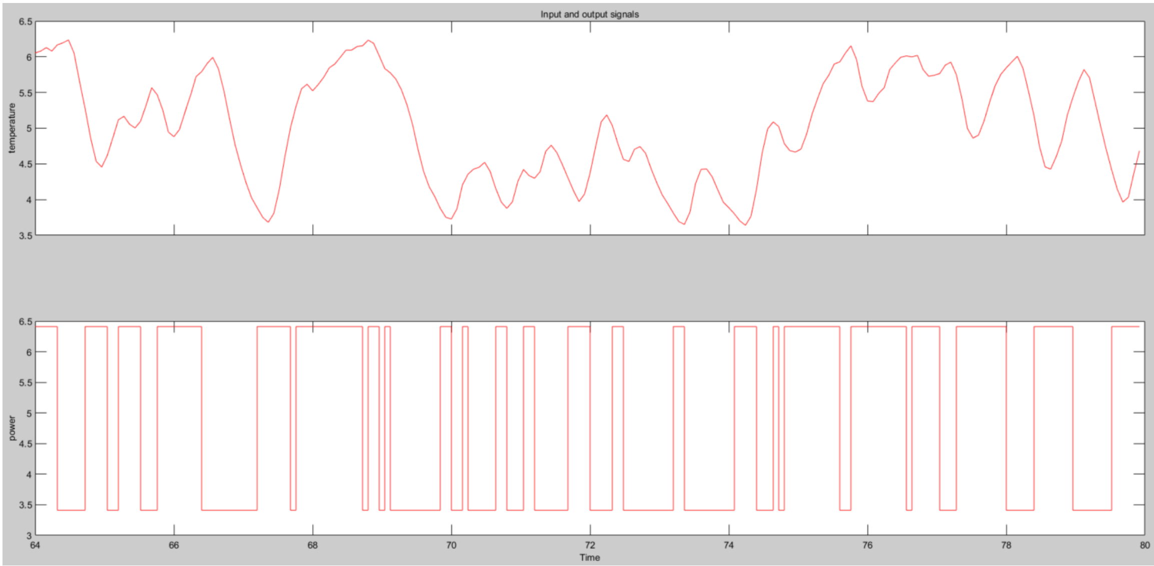

Figure 3: Validation Data Set for Hair Dryer Plant

3 Detrend Operations

In Preprocess section (in the middle of SID toolbox), there are "Remove means" and "Remove trends" options. They both are originated from the built-in detrend function in Matlab. "Remove means" option will basically remove the mean value of the selected channel from itself. Here, since taking the mean is just returning a scalar value, it is called *0-th* order detrend operation. However, when "Remove trends" is applied on the selected channel, it will remove the linear trend in this channel (it is like *1-st* order detrend operation). Therefore, both the mean and linear trends will be removed from estimation and validation data set separately. Then, newly updated data sets for estimation and validation are called Detrended-Estimation-Dryer (can be seen in magenta color on data window) and Detrended-Validation-Dryer (can be seen in brown color on data window). In my opinion, there exist some limitations for detrending signals in SID toolbox. For example, it lacks removing quadratic trend option for signals. Detrended versions of estimation and validation data sets can be seen in Figures 4 and 5 respectively. From Figures 6 and 7, details of how these two signals are created in terms of functional operations on Matlab can be seen inside the rectangular shapes of magenta and brown color for detrended estimation and detrended validation data set respectively.

4 ARX Model and its Parameters

ARX model can be mathematically represented as $Ay = Bu + e$ where e is white-noise, u represents the input signal, and y stands for the output of the system. In our hair dryer plant, y is temperature of dryer's outlet and u is the electrical power supplied to the hair drier. In ARX model, there exist three parameters: n_k for the number of pure delay in

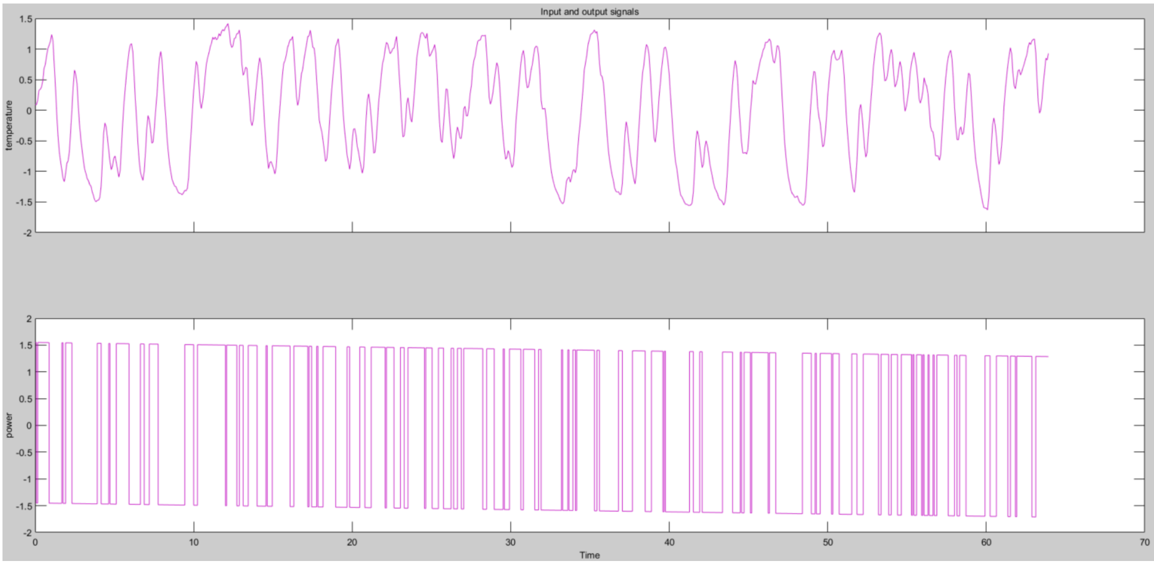

Figure 4: Detrended Estimation Data Set for Hair Dryer Plant

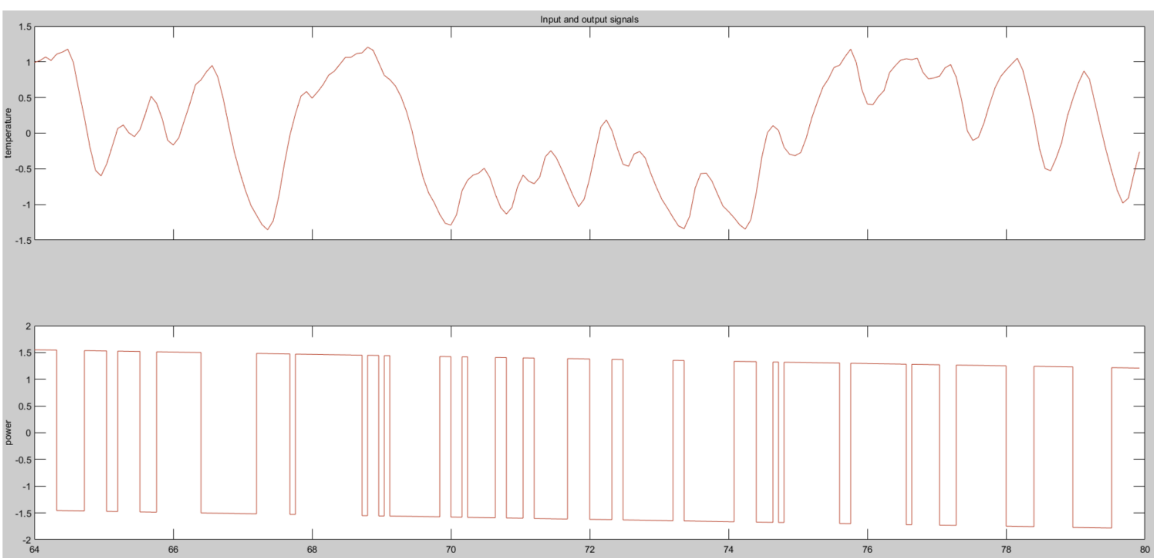

Figure 5: Detrended Validation Data Set for Hair Dryer Plant

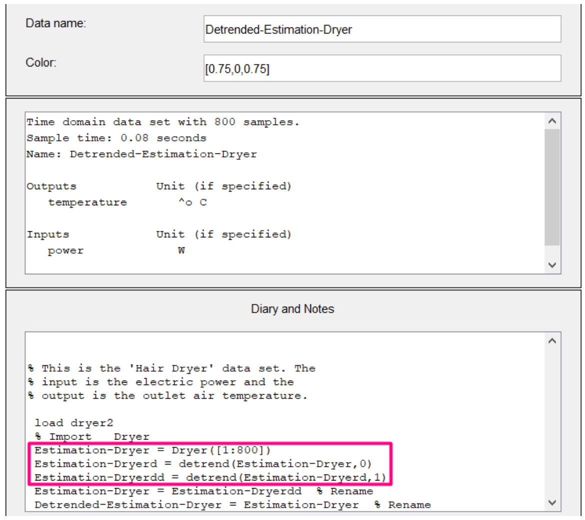

Figure 6: Details of Detrended Estimation Data Set Generation

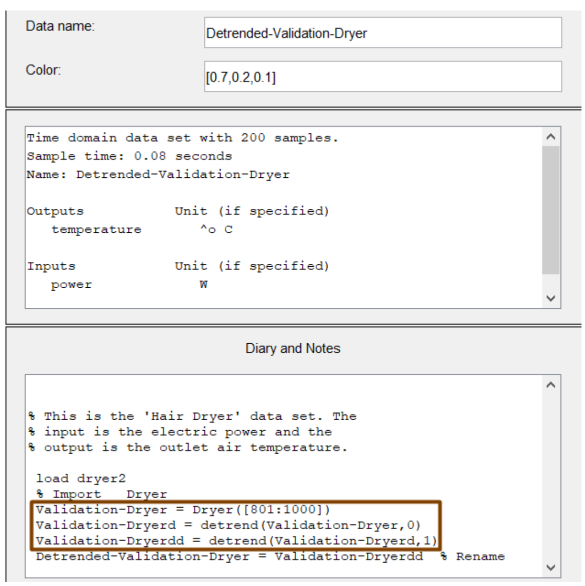

Figure 7: Details of Detrended Validation Data Set Generation

the system, n_a for the degree of polynomial A, and n_b for the degree of polynomial B. Since it is not known in advance how many steps the ARX model should go to the past in terms of the input and output, it is beneficial to use order selection property on Polynomial Models window. By default, the order for each three parameter on ARX model will be optimized by adjusting each of three parameters from 1 to 10. In Figure 8, there are two optimized suggestions for three parameters. The first one is called MDL choice (blue bar on the figure) which recommends 3 for n_a, 4 for n_b, and 2 for n_k. The other one is called Best Fit optimization (red bar on the figure) which offers 6 for n_a, 9 for n_b, and 2 for n_k. Two different polynomial models are created according to these two model structures. The name of the model with MDL choice parameters is MDL_arx342. The name of the model with Best Fit choice parameters is BestFit_arx692.

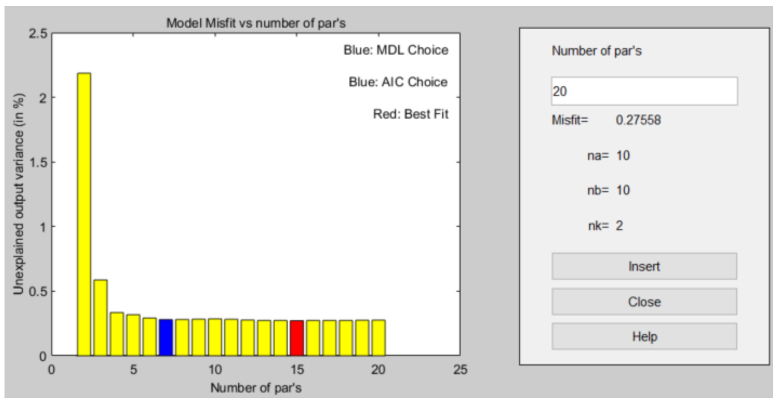

Figure 8: Optimizing Structure of ARX Model

5 Discussion & Comparison about Generated ARX Models

In Figure 9, there are three plots. The black plot is just an output in the detrended validation data set which is also called measurement data or measured model output. The blue and green plots are the output of BestFit_arx692 and MDL_arx342 ARX models respectively. The color convention will be kept the same for the remaining Figures. These two are called simulated model outputs in SID toolbox. On the right panel in Figure 9, there exist Best Fits values for each of two simulated model outputs with respect to the measured model output. It is just an accuracy (in terms of percentage) metric like an identification criteria in System Identification Process. BestFit_arx692 ARX model resulted in slightly better performance around 0.85 % than MDL_arx342 in terms of Best Fits criteria.

In Figure 10, there are two graphs. The first one is auto-correlation of residuals for output temperature. Residual simply means errors in the model. The expected auto-correlation of residuals coming from ARX models will be that of white noise. If the disturbance of the hair drier system is really a white noise, then ideally this top graph in Figure 10 would be that expectation which is auto-correlation of white noise. However, that is not the case for actual

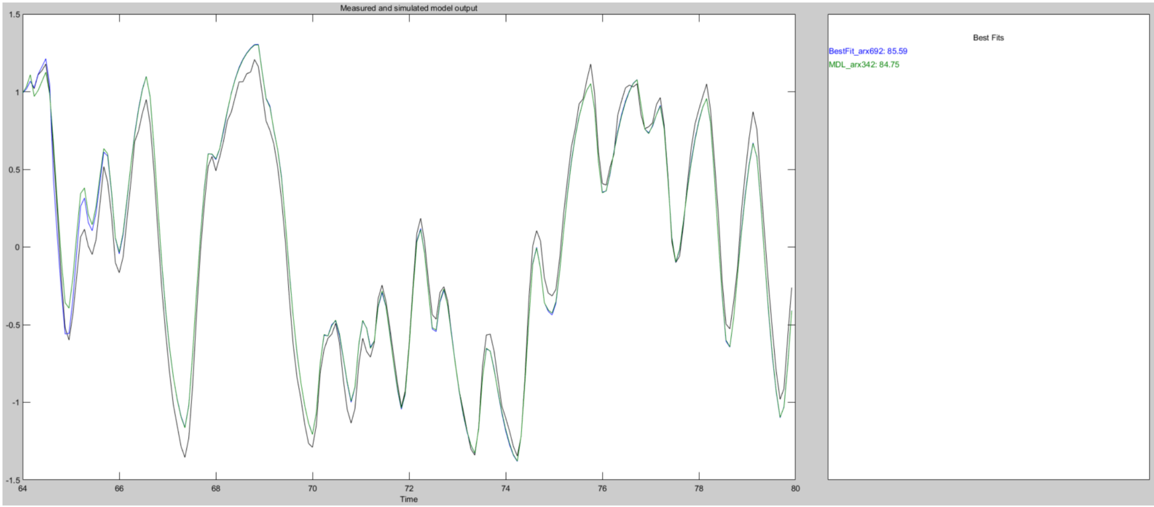

Figure 9: Model Output: Temperature Comparison of both ARX Models and Detrended Validation Output

systems. Auto-correlation of white noise is the variance of it for lag=0 and 0 for elsewhere. In Figure 10, auto-correlation of both ARX models is very close to 0 except lag=0, but not exactly 0. If we manage a successful modeling, then auto-correlation of residuals should be very close to auto-correlation of white-noise. Because auto-correlation function is an even function, symmetric behaviour can be observed at the top graph in Figure 10. Since the blue line (BestFit_arx692) is more close to 0 auto-correlation, it can be inferred that it is more accurate than green in terms of ARX model perspective. That deduction also agrees with the fact that BestFit_arx692 ARX model has a higher Best Fits value than MDL_arx342 model. Ideally, it is desired to have 0 as cross-correlation for the input and output residuals. If there is a correlation between the input and output, then there may be some unexplained component of the input in residuals. That behaviour is not desired. Thus, it is so natural to see both cross-correlation plots are very close to 0 value at the bottom graph in Figure 10. This means that these two ARX typed models have a good performance. Moreover, since cross-correlation is not an even function, it is not symmetric.

Step response of both ARX models for hair drier plant can be seen in Figure 11. This is an ideal behaviour for the transient response of a system. There is one=time step (step response plot starts increasing at t=0.08) delay in the output due to dynamics/inertia of the system.

In Figure 12, there is frequency response function (FRF) which is a type of non-parametric models. It is basically a bode plot. FRF is obtained by substituting s=jw into input transfer function (G(s)) to get G(jw)=FRF. G(jw) is a complex number. That's why it has the plot of both the magnitude and phase for the representation of FRF. Both models almost have the same frequency response in terms of magnitude and angle/phase.

The graph in Figure 13 shows zeros and poles of two hair dryer models separately. BestFit_arx692 model has 8 zeros and only 3 of them are unstable zeros. Also, BestFit_arx692 model has 6 poles in total and only 2 of them are stable. However, MDL_arx342 model has

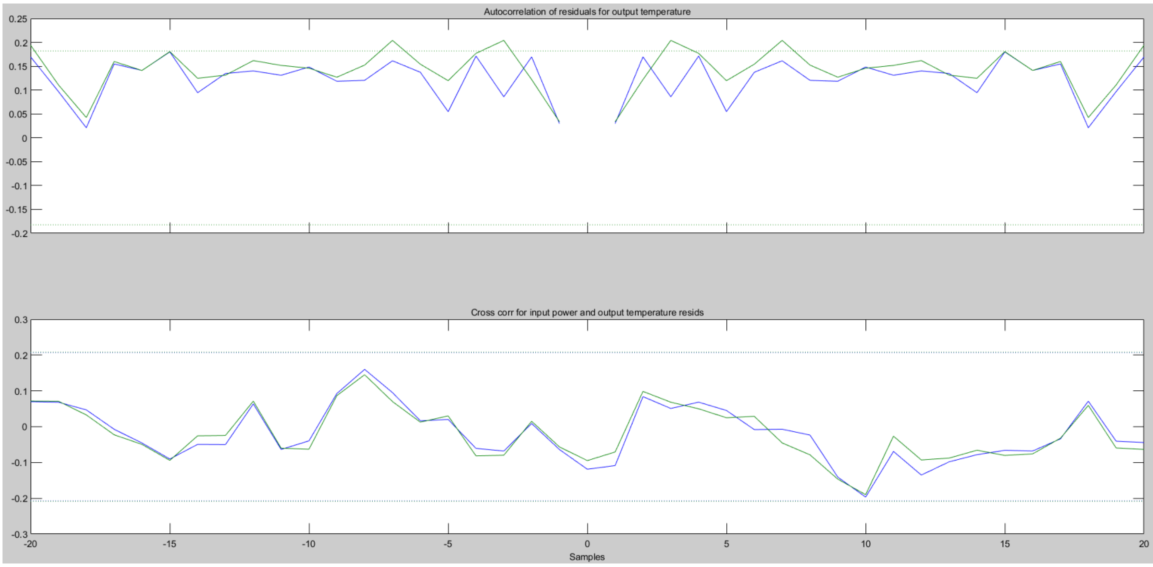

Figure 10: Model Residual Analysis with Auto-Correlation and Cross-Correlation Functions

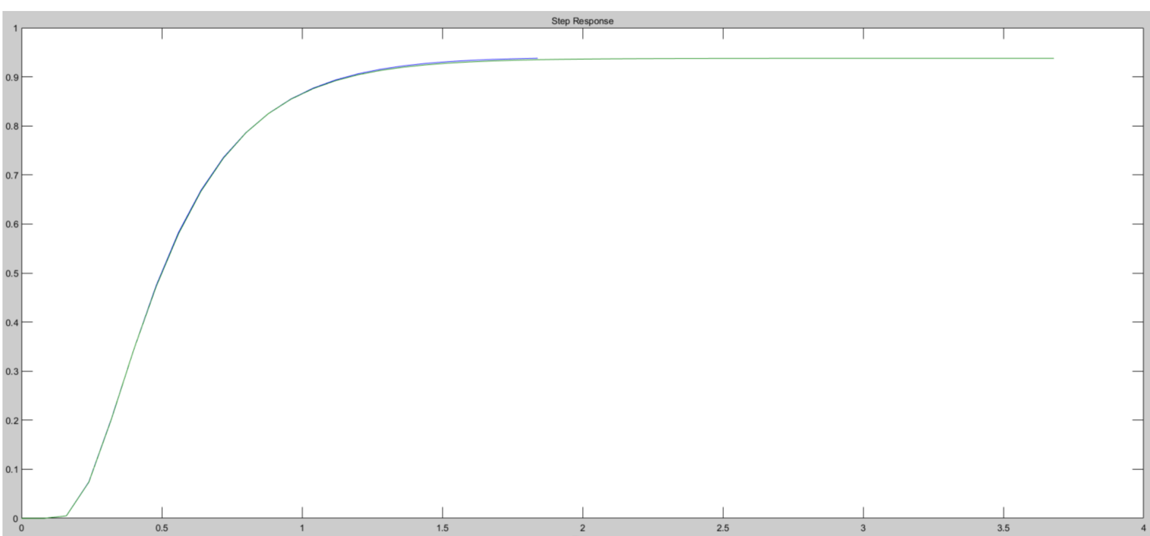

Figure 11: Step Response for ARX Models

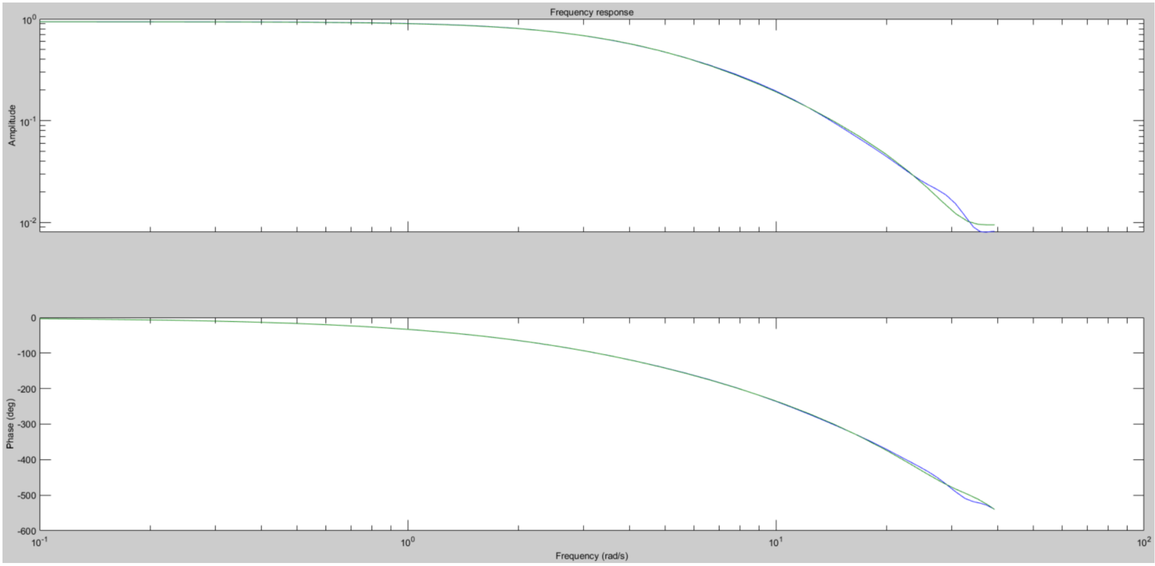

Figure 12: FRF of ARX Models

3 zeros and all of them are stable zeros. Also, MDL_arx342 model has 4 poles in total and only one of them are stable.

Power spectral density of white noise is constant at its variance value for all its' frequency content which is infinite. White noise can not be applied to a system since it is random and all the frequency of the system can not be excited. In practice, systems will have a certain frequency range, one should not excite the system beyond it's frequency range. So, it not practical to apply white noise to actual systems. Instead, PRBS signals are applied as in the case of electrical power input of the hair drier plant. The system should be excited with input signals which demonstrate good capability of pumping energy into the system almost at all frequencies we are interested in. In Figure 14, the power spectral density of noise can be observed.

6 Appendix

Overall view of SID toolbox with corresponding data and models can be seen below.

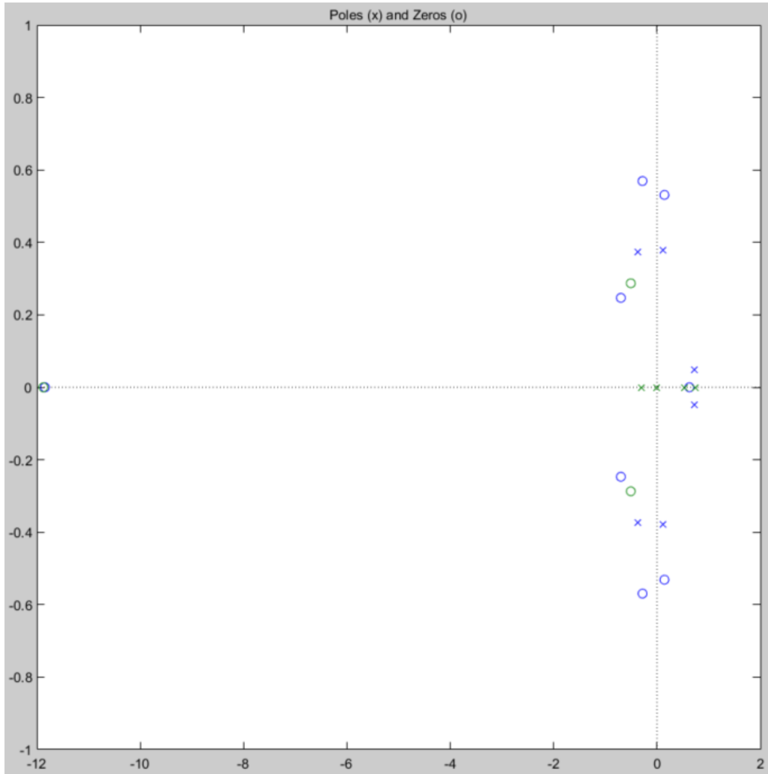

Figure 13: Zeros & Poles Configurations for ARX Models

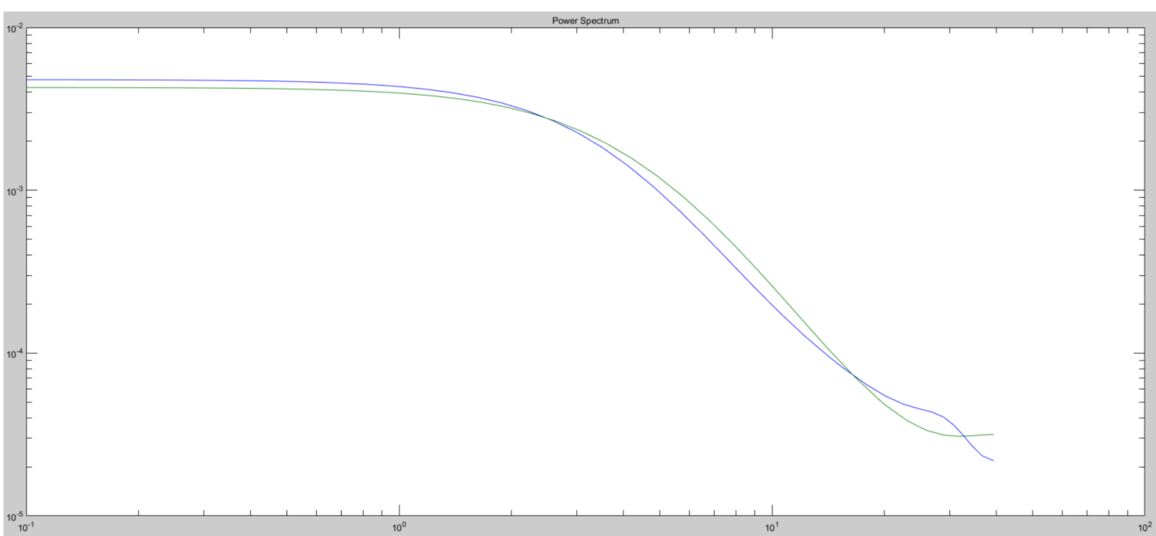

Figure 14: Power Spectral Density of Noise

Figure 15: SID Toolbox with Corresponding Data Sets and Models