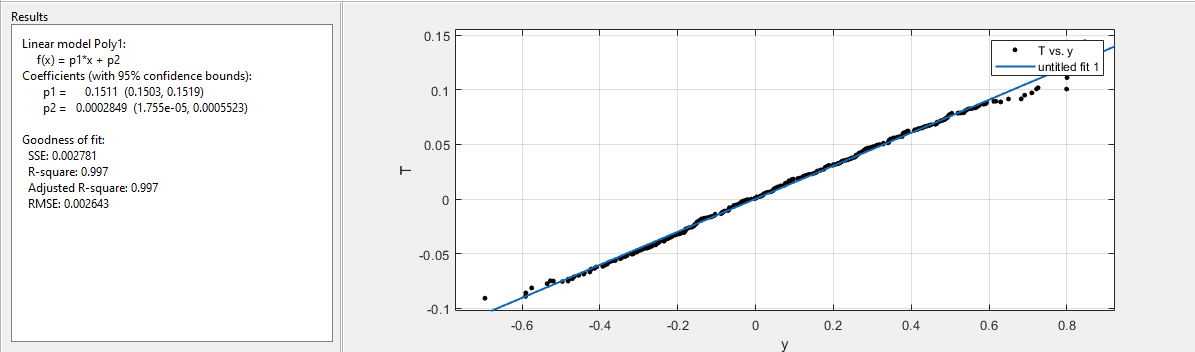
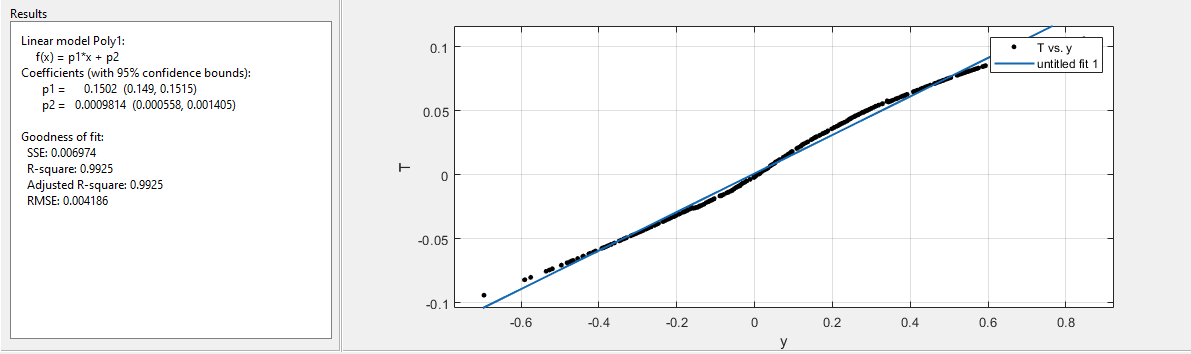
Problem 2

1. For the given synthetic data file, a function was defined in matlab ,ACE(syn, span, smooth\_method), where smooth\_method is the one of the smoothing methods specified in the smooth() function of matlab, and ‘span’ is the the span of the smoothing function and syn is the CSV version of the syn.data file provided. The ACE function can give the outputs for y,x1,x2 and x3 by reading in the csv file, and also give outputs for theta, phi1, phi2, and phi3. The function generates a figure with 4 plots, between Y and Y\_Tr, x1 and phi1, x2 and phi2 and x3 and phi3.





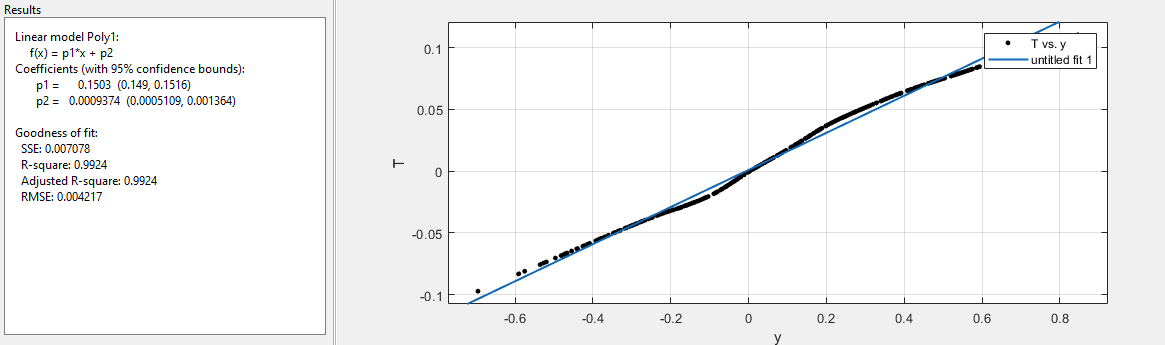
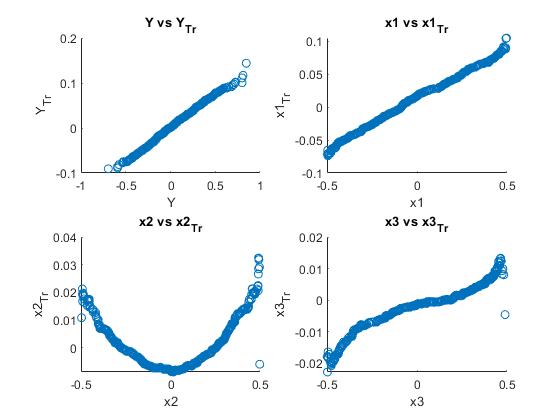
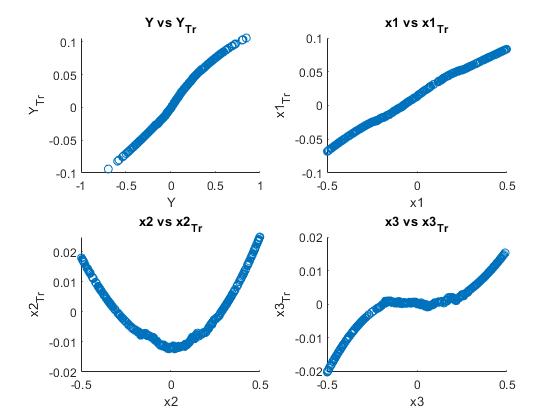


Figure : (a) Moving average (b)Savitzky-Golay (c)LOESS

The ACE function was run using 3 different smoother – moving average filter, Savitzky-Filter and LOESS filter. The moving average filter is A lowpass filter with filter coefficients equal to the reciprocal of the span. The Savitzky-Golay filter is a generalized moving average with filter coefficients determined by an unweighted linear least-squares regression and a polynomial model of specified degree. The LOESS filter is where local regression is done using weighted linear least squares and a 2nd degree polynomial model. **Fig 1** demonstrates the fits into a single degree polynomial for each of the case for Y and Y\_Tr, where the span used was 0.5. This was generated using ‘cftool’.**Fig 2** shows the plots between the original and transformed variables for each of these smoothers, which are generated by the ACE function.. As we can observe in **Fig 1(a)**,**1(b)** and **1(c)**, the curve fit for the ‘LOESS’ was noticeably smoother than the ‘moving average’ case , and slightly smoother than the ‘Savitzky-Golay case’.

Hence, the LOESS filter is considered for further optimal transformations.





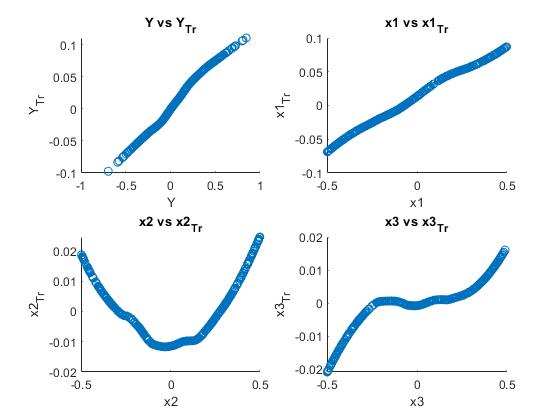


Figure : Plots for original and transformations for y, x1, x2 and x3 using (a) Moving-Average (b) Savitzky-Golay (c) LOESS filters (span : 0.5)

The plots shown in **Fig. 2** further illustrate the effect of smoothers on the optimal transformation curves.

* *Y vs Y\_Tr plots* : Upon Comparison of Y and Y\_Tr case plots for all the filters as shown in **Fig. 2(a), Fig. 2(b) and Fig. 2(c)**, the plots for the moving average case(Fig. 2(a)) is noticeable less accurate , and less smoother than the plots in Fig. 2(b) and Fig. 2(c). The plots in the savitzky-golay and LOESS are considerably more similar to each other. We then go on to compare the plots of optimal transformations for the other variables.
* *X2 vs X2\_Tr :* The plots for the cases of *X2 vs X2\_Tr*  compared for the cases of Savitzky-Golay filter(Fig. 2(b)) and the LOESS filter case(Fig. 2(c)). The optimal transformation curve is more smoother in the case of the LOESS filter.
* *X3 vs X2\_Tr :* The plots for the cases of *X3 vs X3\_Tr*  compared for the cases of Savitzky-Golay filter(Fig. 2(b)) and the LOESS filter case(Fig. 2(c)). The optimal transformation curve is more smoother in the case of the LOESS filter.

Hence, after careful consideration, the LOESS filter was chosen due to it’s ability to smooth the variables and obtain more accurate optimal transformation curves.

3.

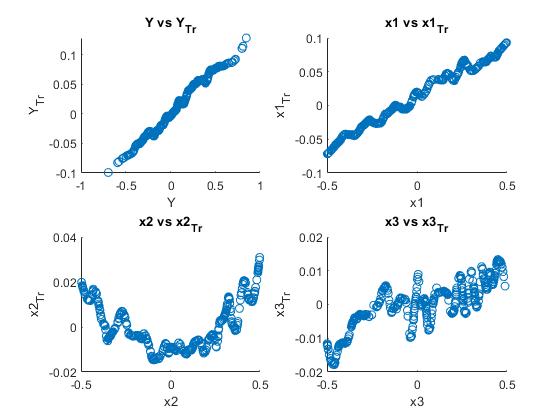


Figure : LOESS with Span 0.1

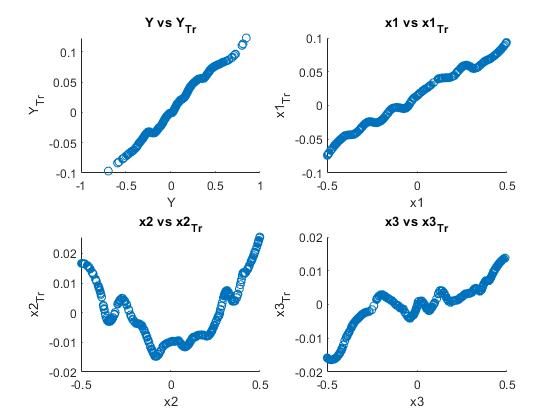


Figure : LOESS with Span 0.2

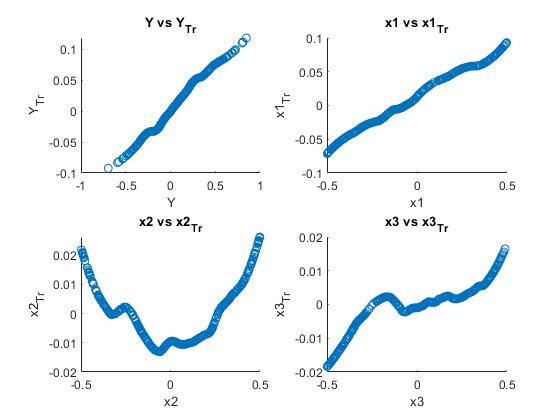


Figure : LOESS with Span 0.3

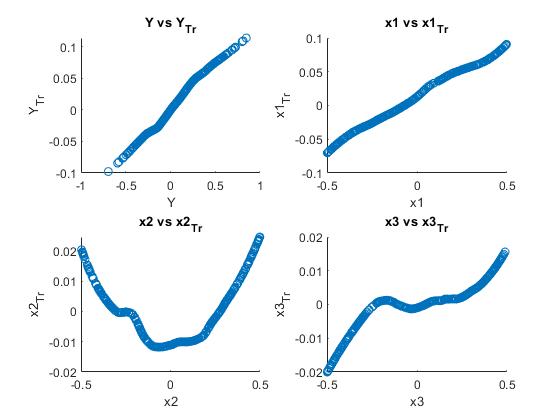


Figure : LOESS with Span 0.4(best case chosen)

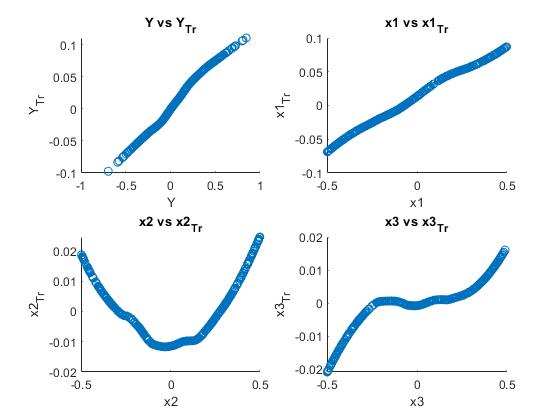


Figure : LOESS with Span 0.5

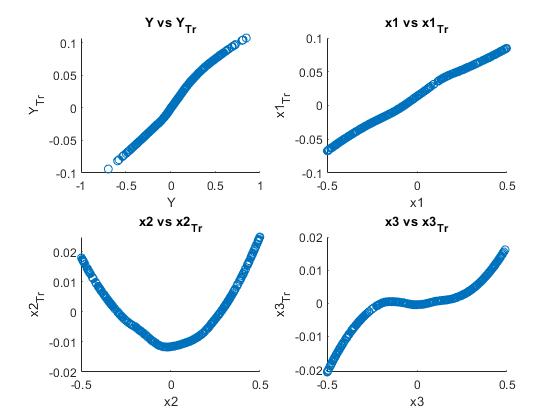


Figure : LOESS with Span 0.6

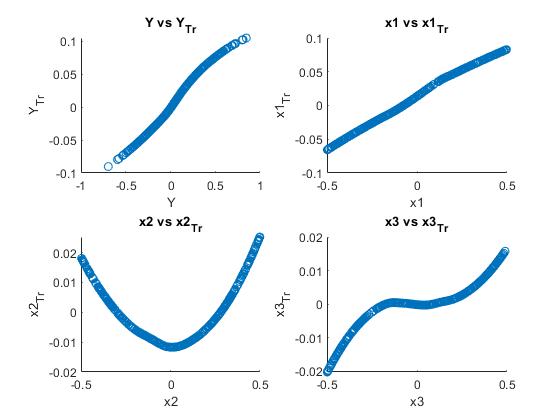


Figure : LOESS with Span 0.7

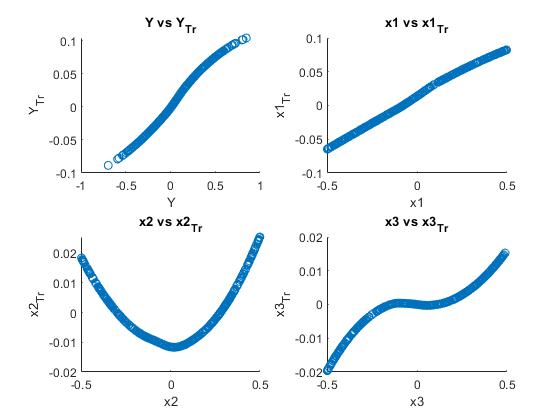


Figure : LOESS with Span 0.8

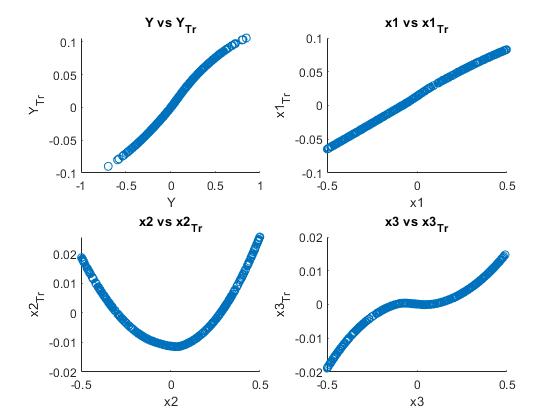


Figure : LOESS with Span 0.9

* The optimal span selection for the LOESS filter is investigated into, paying close attention to the bias-variance trade-off.
* The span of the LOESS filter was changed from 0.1 to 0.9, with a span difference of 0.1 with each gradual step.
* The plots shown in the **Figs. 3 to 11** help illustrate the effect of the change in span.
* As the span is changed by 0.1, we notice that the optimal transformation curves become smoother due to the increase in span. As the span increases, more values are considered for the mean value that is used to smooth the curve at each step. Hence, the bias gradually increases as the span increases as more values are averaged out at each step, thereby reducing the variability of the data as well.
* The bias increases as we increase the span, while the variance in the data will decrease and the curves gradually smoothen out.
* The curves generated by plotting the measured versus transformed variables are rough at low spans, as the bias is low and the spontaneous variability in data is high.
* The optimal span selection is considered to be the span at which the curves become smooth enough to generate a consistent curve without spikes in the data, while also retaining enough variability(or roughness) in the data curve to accurately model the properties and variance in the data set.
* As we go from span of 0.3 to 0.4, we notice a significant decrease in data spikes in the optimal transformation curve and the curve is smoother.
* As we move ahead from span of 0.4 to 0.5, we notice that the curve become smoother, but a lot of data that is critical to model the variability of the data is not graphed accurately.
* Looking at Figs. 5, 6 and 7, we can compare the curves for all the optimal transformations for all the variables, and notice that the span of 0.4 manages to retain vital curve roughness, while also managing to filter out the data that caused the high curve roughness in span of 0.3.
* The span of 0.5 generated a slightly smoother curve compared to span of 0.4, but in the process loses out on modeling the inherent variability of the data(curve roughness) accurately due to larger span, while the span of 0.4 seems to retain a lot of significant data roughness.
* In conclusion, the span of 0.4 is optimal as it gives us smoother optimal transformation curves while also ensuring a good balance between bias and variance.