Assignment 1

Parameter estimation and predictions based on noisy data from 3-variable drug-response models

2018-09-17

Kristina Benevides

Automated predictive modelling 1MB516, 5.0 c

Contents

Abstract 3
1. Introduction 3
2. Results 5
2.1. Task 2 5
2.2. Task 4 6
2.3. Task 5 8
2.4. Task 6 10
2.5. Task 7 10
3. Discussion and Conclusions 11
4. References 11
5. Self-reflections 12

Abstract

In this report parameter estimation and prediction for drug response levels are studied using three different models; the Hill-based model, a simplified version of the Hill model and a polynomial model. The estimated effect levels were compared to the observed effect level for each model in order to compare predictive power and the relative errors for the estimated parameters in the Hill-based model were calculated. The results indicate that the combined Hill-based model perform better than the other two models, which is consistent with Ning et al. (2014) findings.

1. Introduction

Combining multiple drugs for cancer chemotherapy is of great interest as it can increase drug potency due to synergistic interactions as well as reduce drug resistance. Consequently follows a need to understand the mechanisms and interactions of the combination and the choice of an optimal combination. However, costs and other considerations restricts the number of experiments and drug quantities. It is thus important to do experimental optimal design.

This assignment was based on a paper by Ning et al. (2014) which combined two models; one that is used to study biological response which assume sigmoidal shape, the Hill model, and the other, the polynomial model, which fits a response surface with a full or fractional factorial design, into an improved model to study a drug combination for lung cancer treatment.

The Hill model can be used to model the dose-effect relationship. It is common to use ATP levels as marker for proliferation of cells, which will represent the effect level of the drug combination. For a two-drug system each combination with a fixed ratio of respective drug is assumed to be a 'new' drug, as in the following:

$$Y_r = \frac{1}{1 + \left(\frac{C}{IC_{50,r}}\right)^{\gamma}} + \varepsilon_r,$$

Where index r is the fixed ratio, C is the total concentration of the combination, Y_r is the dose-effect level, γ and $IC_{50,\mathrm{r}}$ are to estimates from the data and ε is random noise. This model can be used for prediction outside the experimental range but is limited to only apply to fixed-ratio combinations which makes it unsuitable when you want to do a global response surface across various concentration ratios.

The polynomial model is however usually used to fit the global response surface together with some experimental design:

$$Y = \beta_0 + \sum_i \beta_i x_i + \sum_{i,j} \beta_{ij} x_i x_j + \varepsilon,$$

Where x is the concentration for drug *i*, *Y* is the effect level, the beta parameters represent the main effects and interactions. Nonetheless, this model's predicted effect level is boundless as the concentration increases to infinity which clearly is not true.

These two models were combined into one which kept the advantages of each model while discarding their respective limitations. Primarily, it is applicable on a global scale and can address different combinations of several drugs. This model is referred to as the combined Hill-based model or Hill-based global model and defined as following:

$$Y = \frac{1}{1 + \left(\frac{C}{IC_{50}(\theta)}\right)^{\gamma(\theta)}} + \varepsilon,$$

where we have *p* drugs, C_i is the dose level for drug *i*, *C* is the sum of the concentrations of all drugs and $\theta_i = C_i/C$ which represents the proportion of drug *i* in the combination. $IC50(\theta)$ and $\gamma(\theta)$ are the original parameters in the Hill model and are now functions of θ. Like before, *Y* represents the effect-level and ε is random noise. The way the polynomial model is included is in the functions for IC_{50} and γ:

$$IC_{50}(\theta) = b_0 + \sum_{i=1}^{p-1} b_i\theta_i + \sum_{1\le i\le j\le p-1}^{p-1} b_{ij}\theta_i\,\theta_j;$$

$$\gamma(\theta) = a_0 + \sum_{i=1}^{p-1} a_i\theta_i + \sum_{1\le i\le j\le p-1}^{p-1} a_{ij}\theta_i\,\theta_j,$$

where b_i, b_{ij}, a_i and a_{ij} are parameters. Notably, only *p-1* components are used to ensure identifiability of the parameters.

2. Results

The results for each of the tasks are summarized in figures and text below.

2.1. Task 2

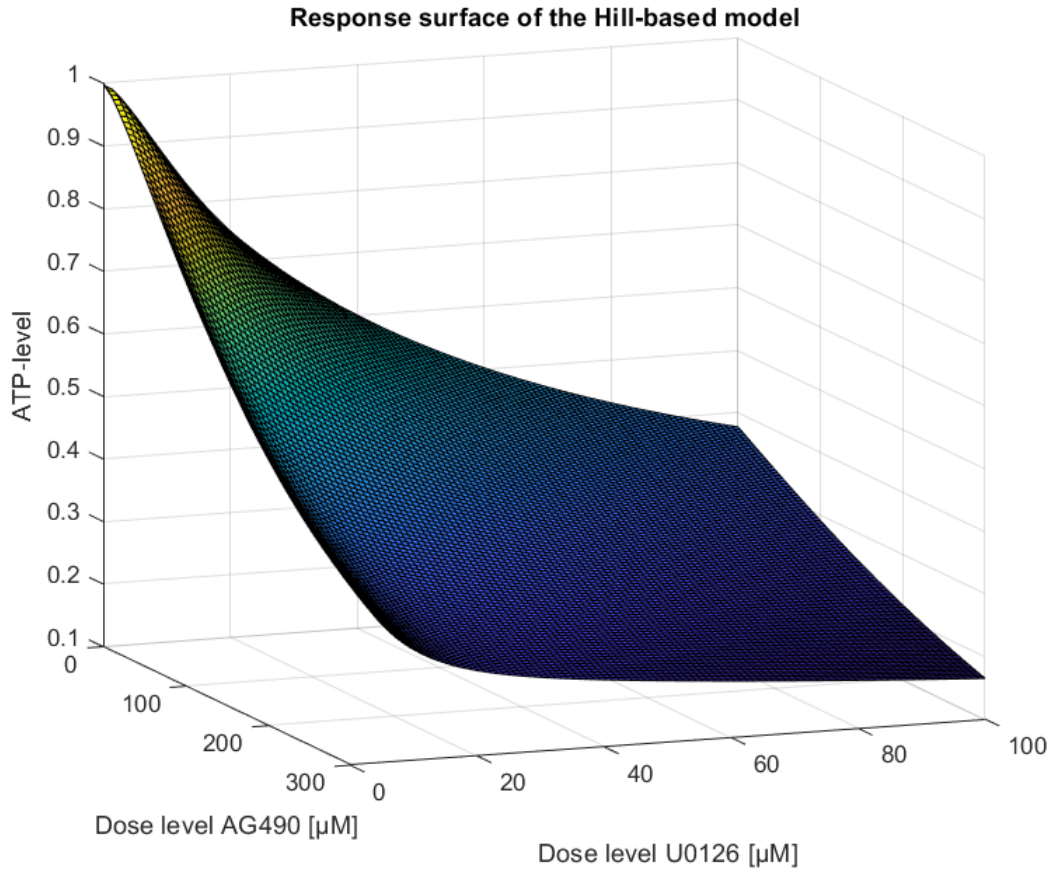

Figure 1. 3D-plot of the response surface of the global Hill model.

2.2. Task 4

a.

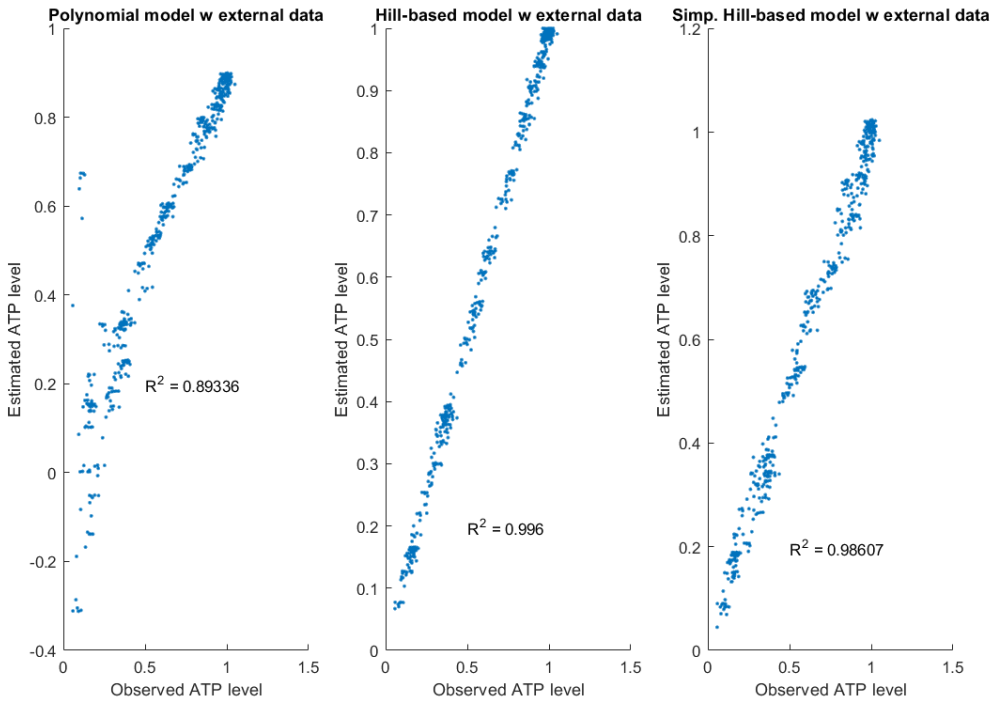

Figure 2. Scatter plots of the estimated effect- level and observed effect level for the three models.

b.

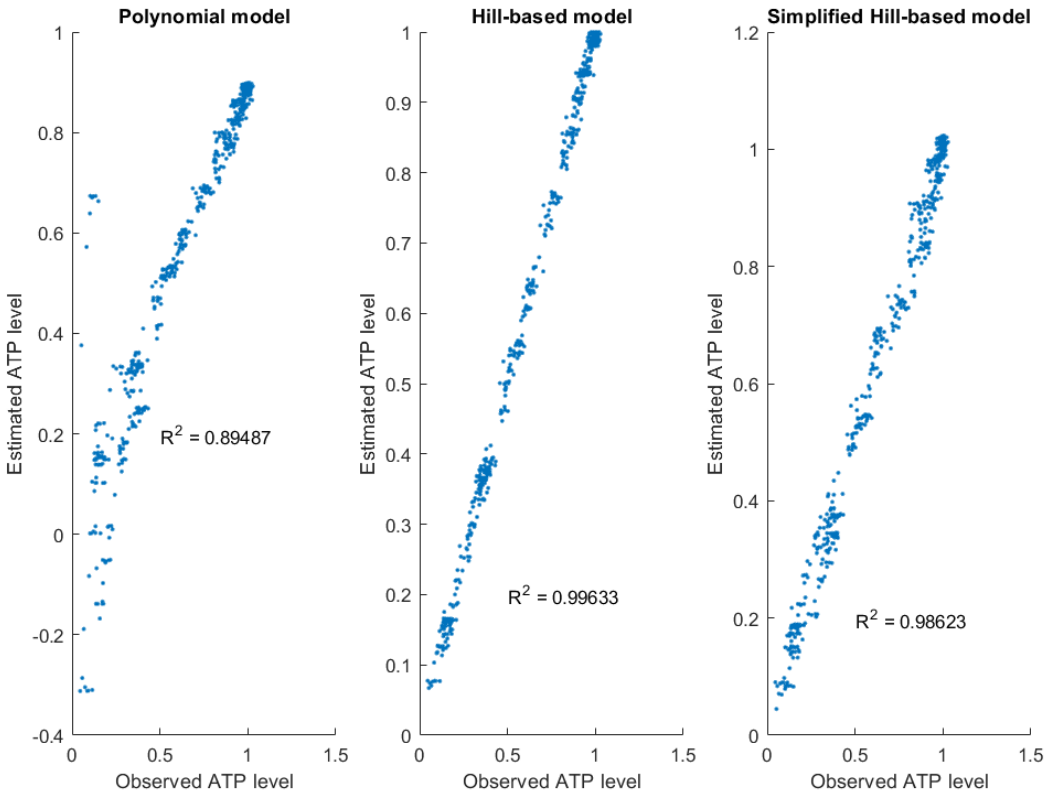

Figure 3. Scatter plots of the estimated effect- level and observed effect level for the three models using the new external dataset Y_observed.

2.3. Task 5

b.

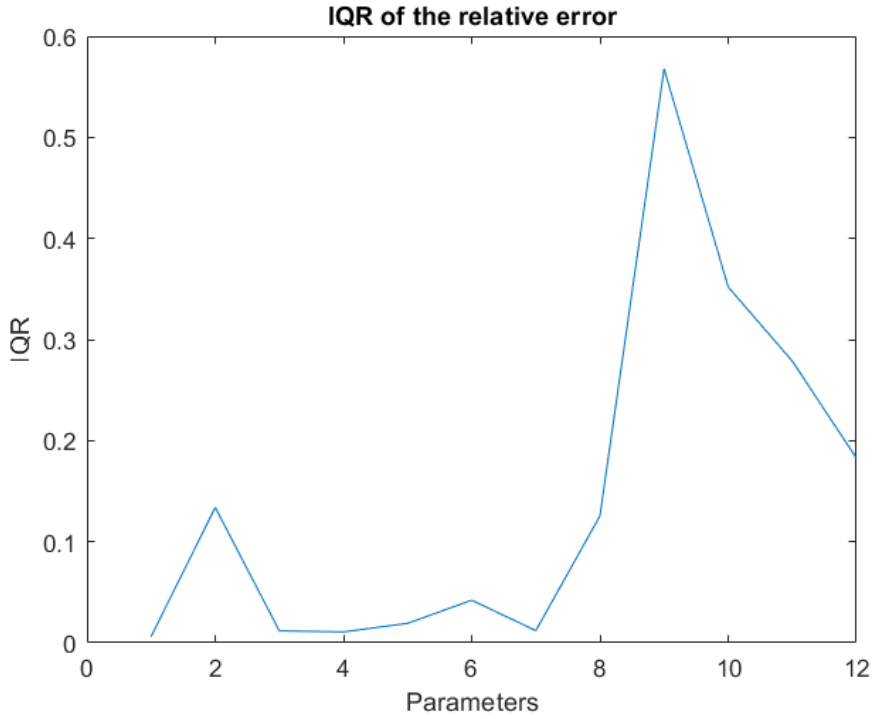

Figure 4. The interquartile range (IQR) for each of the parameters of the combined Hill-based model.

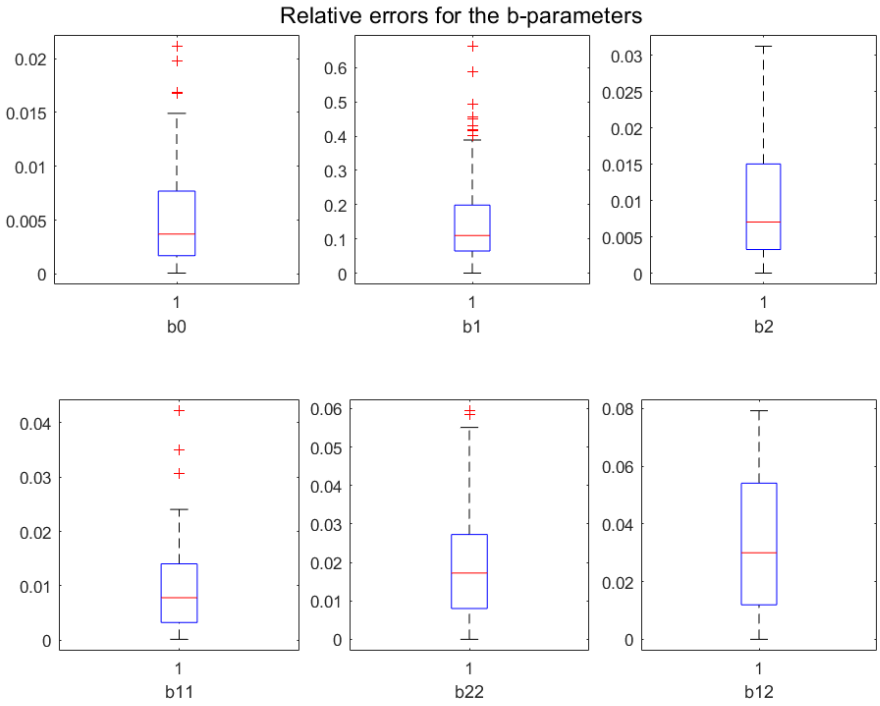

Figure 5. Boxplots of the relative errors for each of the IC_{50} – parameters for the combined Hill-based model.

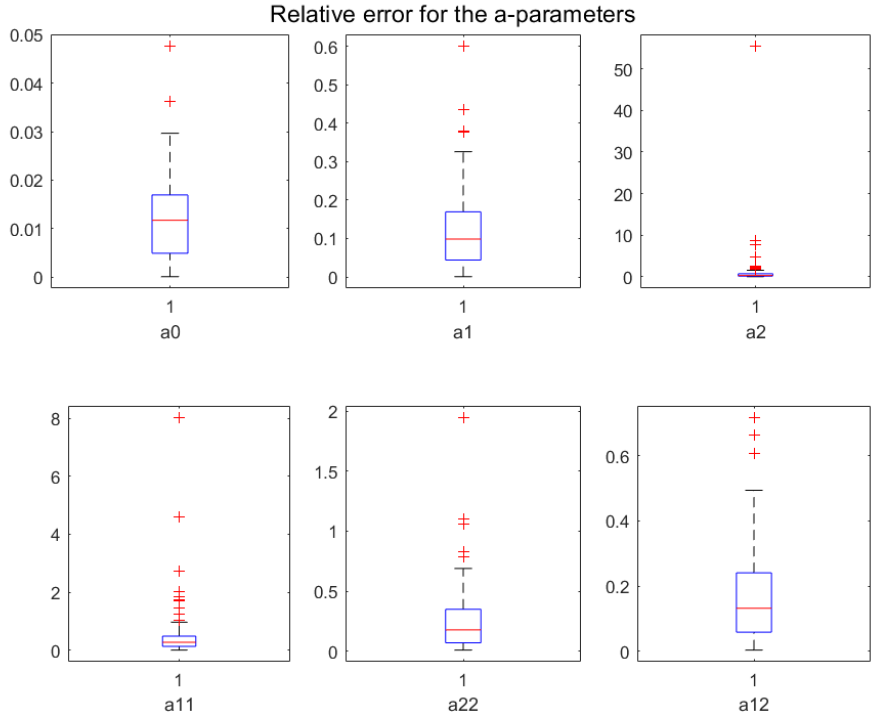

Figure 6. Boxplots of the relative errors for each of the γ – parameters for the combined Hill-based model.

2.4. Task 6

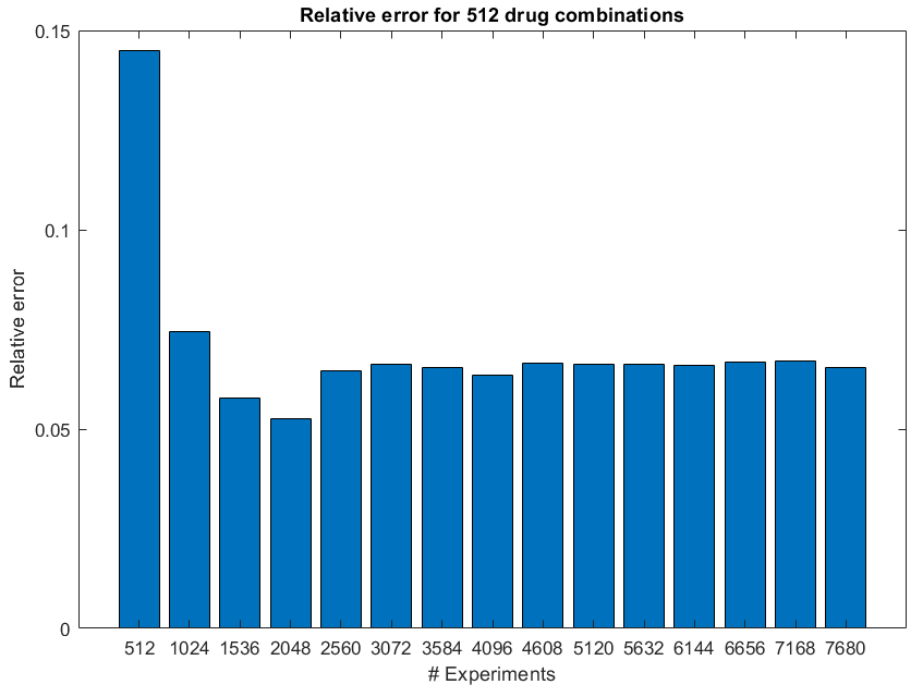

Figure 7. Relative error for 512 drug combinations for increasing batch size. Generated with zero mean normal distribution noise term, $\sigma = 0.15$

2.5. Task 7

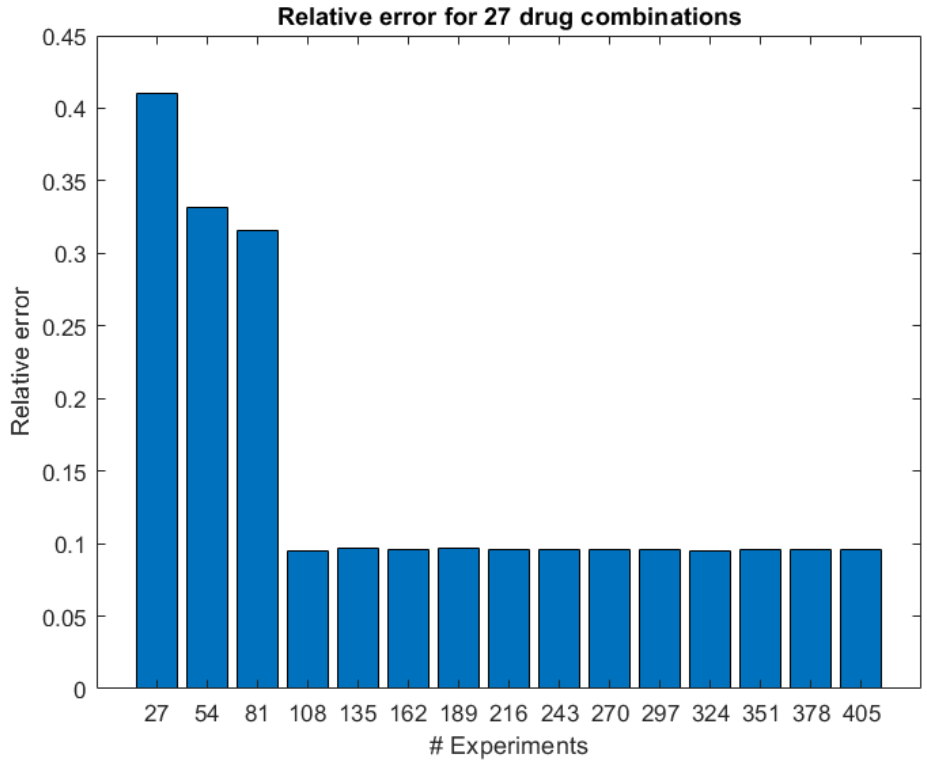

Figure 8. Relative error for 27 drug combinations for increasing batch size. Generated with zero mean normal distribution noise term, $\sigma = 0.08$

3. Discussion and Conclusions

The response surface derived in Task 2 seems consistent with theory seeing that the ATP level is 1 when the dose levels are 0 and decreases as dose levels increase, which means that the cancer cells have 100 % viability when no drugs are added, and they are progressively inhibited or killed as the drug concentrations increase.

The Hill-based model and the simplified Hill model both fit the data very well with $R^2 > 0.98$, although the Hill-based model is a little bit better. The polynomial model does not perform as well as the other two but is still quite good with $R^2 \approx 0.89$. There is no great difference between the predictions made with the original data and the ones made with the external data, all methods perform basically the same as in the previous task. From this I would conclude that the combined Hill model is better at predicting its parameters than the polynomial and simple Hill model are, which is expected since it is supposed to be an improvement of the other two models.

Overall, the estimated parameters for the combined Hill-based model appears to be close to the "true" ones, i.e. the estimates in Table II in the paper by Ning et al. (2014). Something noticeable it the very large IQR for parameters a_2 and a_{11}. However, I have not found an explanation to this.

For Task 6 and 7 the errors decrease when batch size increases but they never seem to reduce towards zero. It is difficult to draw conclusion about how many experiments are needed for a relative error below 5% or 1%. For 512 drug combinations after around 2560 experiments the relative error seems to stop decreasing. For 27 drug combinations the number of experiments where relative error stops reducing is 108, albeit the final error rate is twice as high as for the run with 512 drug combination (0.1 and 0.05 respectively). In both cases I had to increase the standard deviation (originally $\sigma = 0.02$) in the noise term so see a decrease in the error.

4. References

Ning S, Xu H, Al-Shyoukh I, Feng J, and Sun R. An application of a Hill-based response surface model for a drug combination experiment in lung cancer. *Statistics in Medicine*, 33:4227–4236, 2014.

5. Self-reflections

In this assignment I found it useful to practice reading and understanding a scientific paper. In addition, I felt like I got a better understanding of the main findings of the article by testing their model myself in the different tasks and comparing it to other models. I also appreciated the subject of the article/assignment and found it very relevant to my education. It was good to include tasks where you got to tackle problems hands-on and write your own code. This assignment's tasks that focused on the parameter estimation and prediction part of the paper felt familiar from the Multivariate Data Analysis course, which I thought was nice to review some important concepts from that course. Yet, I felt like this was a more in depth practical than we have had before so I did learn some new things. The time it took to complete the assignment was a little bit more than expected from a course which is 5.0 c, but still I would not say the work-load was extremely high. I think I struggled a bit because I haven't used Matlab in a while, therefore I thought the lectures where we discussed the assignment were of great help. One suggestion for improvement is to clarify the instructions, for example to write in the instructions that one can solve for the parameters in the polynomial model analytically and then using those parameter values as starting guesses when trying to estimate them numerically.