



EINDHOVEN UNIVERSITY OF TECHNOLOGY

COMPUTATIONAL BIOLOGY GROUP

Bachelor Eind Project SBMD Report Logs

Student

J.H. Cao, **studentid:** 1258540, *Department of Biomedical Engineering*

Supervised by

prof.dr.ir. N.A.W. van Riel, *Department of Biomedical Engineering*

dr. D. Bosnacki, *Department of Biomedical Engineering*

Contents

1	Preface	4
2	Introductory meeting 19-march-2020	5
2.1	Notulen	5
2.2	SSA	5
2.3	Results	6
3	"Startbijeenkomst" 08-april-2020	7
3.1	Notulen	7
3.2	SSA	7
3.3	Results	8
4	Meeting on 16-april-2020	9
4.1	Notulen	9
4.2	SSA	9
4.3	Results	10
5	Meeting on 23-april-2020	11
5.1	Notulen	11
5.2	SSA	11
5.3	Results	12
5.3.1	Logistic Growth	12
5.3.2	Exponential growth trajectory	13
6	Meeting on 30-april-2020	14
6.1	Notulen	14
6.2	SSA	15
6.3	Results	16
6.3.1	Kalman Filter	16
6.3.2	SIR/SEIR model	16
6.3.3	Logistic Growth Curve	17
7	Meeting on 07-may-2020	19
7.1	Notulen	19
7.1.1	Logistic Growth Curve	19

7.1.2	SIR/SEIR model	19
7.1.3	Kalman Filter	20
7.2	SSA	22
7.3	Results	23

1 Preface

The purpose of this **Report Log** is to keep track of the notes, tasks and every other information provided during the Bachelor Eind Project (BEP) meetings. The table of contents contains a list of many different dates. For each of the given date, with the name ***Notulen***, there is a description, notes or information received during the meeting. The dates named ***SSA***, contain a summary or listings of all the findings/information and tasks performed on that day as a self-study assignment. Lastly, the section ***Results*** contain the results of the performed SSA of the given date.

2 Introductory meeting 19-march-2020

This section contains information regarding the introductory meeting of this BEP.

2.1 Notulen

The introductory meeting was mainly focused on the different kind of topics regarding the **BEP SBMD**.

The different **topics** were:

1. **Hospital patient's data/clinical data regarding:**

- (a) Obesitas after gastric bypass surgery
- (b) Heart failure, monitoring Troponin
- (c) Liver transplantation template model

2. **Dynamic, biological data:**

- (a) Diabetic patients
- (b) Metabolism/reaction on different meals
- (c) Parameter estimation

3. **Sars-Cov-2, aka COVID-19:**

- (a) Epidemic models
- (b) estimation of the infected/exposed

2.2 SSA

The tasks to be completed as self-study assignment before the next meeting are:

- Fill and sign BEP form
- Read the articles and give preferred topic

2.3 Results

After the given tasks, it was important to read and analyse the different articles and information, given by prof.dr.ir N.A.W. van Riel. All topics were interesting, but the topic of **COVID-19** is the most "recent and ongoing" topic and is a major concern for the global health.

The obtained results or the completed tasks from this SSA are:

1. The BEP form is signed and sent
2. Read all the given articles
3. The preferred topic is **Sars-Cov-2 aka COVID-19**

3 "Startbijeenkomst" 08-april-2020

The very first meeting after choosing the preferred subject.

3.1 Notulen

The chosen subject for this BEP is **COVID-19**.

Interesting topics to research regarding the subject of COVID-19:

1. Kalman filter to predict the number of daily infection
(<https://towardsdatascience.com/using-kalman-filter-to-predict-corona-virus-spread-72d91b74cc8>)
2. The main programming language to be used is Python
3. Public Data as resource for model training
4. John Hopkins map and data is useful
(<https://coronavirus.jhu.edu/map.html>)
5. Using the SIR/SEIR model
6. Making a blogpost and Github as "final report"
7. Should focus on the Netherlands
8. Would be good to compare 2 or 3 models (Kalman vs Logistic fit vs SEIR model)

Furhtermore, there is an outreach of the TU/e regarding research of the COVID-19.

3.2 SSA

The tasks to be completed as self-study assignment before the next meeting are:

- Study the Kalman filter and its applications
- Take a look at the outreach of TU/e regarding research of the COVID-19
- Make a Github repositories for the codes of this BEP

3.3 Results

It was important to read the medium article regarding the modelling of COVID-19, using the **Kalman Filter**. The Kalman filter is combines system dynamics, (real-time) data and statistics/stochastics. Furthermore, since the spread of COVID-19 in the Netherlands can be seen as a system that is continuously changing and as a real-time problem, using the Kalman filter might be an ideal approach.

The obtained results or the completed tasks from this SSA are:

1. Kalman filter and its application is studied. Background information is obtained
2. Wrote a proposal for the application regarding the TU/e - outreach for the research of COVID-19.
3. Public repository, containing the codes and progress of the work, is available on Github at : <https://github.com/jiehaocao/BEP-SBMD-COVID19>.

4 Meeting on 16-april-2020

This time, dr. D. Bosnacki joined the team as a supervisor.

4.1 Notulen

It is difficult to use the SEIR model, since there are currently no available data regarding the **Recovered** patients of COVID-19 in the Netherlands. dr. Bosnacki suggested to use a model without the recovered, since the recovery at the beginning is almost negligible. Furthermore, it is difficult to know the **Exposed**, since it is difficult to track the amount of people who are actually exposed to this virus, due to lack of test kits.

It is still advisable to study the Kalman filter, and applying this to the data of RIVM, made available to the public. This way it is possible to check the performance of the Kalman filter on the data of the Netherlands. Another suggestion from dr. Bosnacki, was to use the data regarding the amount of patients taken in the hospital or ICU. This is more trustworthy and this data is also available. Furthermore, it would be possible to use the Kalman filter to predict the daily hospitalized patients due to COVID-19. Lastly, it is advisable to write a proposal for the TU/e innovation space, regarding the outreach and offering help to the research of COVID-19.

4.2 SSA

The tasks to be completed as self-study assignment before the next meeting are:

- Try to model and find a parameter for **Exposed**
- Model the hospital intake
- Try to implement the codes of Kalman filter, specifically on the data of the Netherlands (predicting the daily infected people)
- Proposal for TU/e innovation space (Offer help)
- Study the potential SEIR model

4.3 Results

The number of **Exposed** can be estimated from the SEIR model, if and only if the estimated parameters are accurate. Since the SEIR model, is a dynamic model and consists of differential equations, it is really sensitive to outliers, thus the different parameters. By changing of the parameters, the graph and model would have a complete different output. It would also be possible to add "Hospital" as an extra compartment to the SEIR model, since these models are compartmental models.

The implementation of the Kalman filter was unsuccessful. There were many errors when converting the R script to Python. Therefore, it was better to focus on the SEIR model and implement this in Python for to try to model the epidemic in the Netherlands.

The obtained results or the completed tasks from this SSA are:

1. **Exposed** can be estimated with the SEIR model, but is highly dependent on the estimated parameters (for the differential equations of the different compartments).
2. Extra compartment can be added to model the hospital intake.
3. Proposal is written and sent.
4. The SEIR model is implemented and available on Github repository.
5. Attempt was made to implement the Kalman filter, but it was not successful. z

5 Meeting on 23-april-2020

5.1 Notulen

It is quite difficult to estimate the parameters using the SIR or SEIR models, since this is a dynamic model. Therefore, one would have to estimate the parameters and solving the differential equations, simultaneously.

It is interesting to apply the SEIR model on the hospital admissions or ICU admissions. Furthermore, it is also possible to add extra compartments to this model, taking hospital and ICU admissions into account.

Instead of estimating the parameters, since these are sensitive (sensitivity can be tested by taking the derivative of these parameters function as time). One can also use **parameter sweep**. However, it might be better to use values obtained from literature or the ongoing research on COVID-19.

It is recommended to focus on the **Logistic Growth Curve (Verhulst)** and using this function to estimate the parameters (less complicated), with the help of *Maximum Likelihood Estimation*. With these estimations, the SEIR model can be tweaked and a simulation can be performed on the SEIR model. Finally, the combination of the information obtained from the SEIR and Verhulst model, can be applied on the Kalman Filter.

From now on, prepare the work and questions on a Powerpoint Presentation, same as OGO computational biology (recap, work done, upcoming, questions for supervisors)

5.2 SSA

The tasks to be completed as self-study assignment before the next meeting are:

- Implement Logistic Growth Curve on Python
- Use Logistic Growth Curve to estimate the parameters using MLE
- Tweak and improve the Logistic Growth Curve
- (Optional) Implement these estimated parameters in the SEIR model

5.3 Results

The **logistic growth curves** are successfully implemented in Python and on the data of COVID-19 in the Netherlands. This logistic growth is applied on:

1. Number of confirmed cases (daily) in the Netherlands
2. Hospital Admissions (daily) in the Netherlands
3. ICU Admissions (daily) in the Netherlands

Furthermore, a **exponential growth trajectory** is also implemented in Python, regarding the data of COVID-19 in the Netherlands.

The detailed results and explanations are shown in the README.md on Github.

5.3.1 Logistic Growth

After estimating the parameters and fitting the model, it is possible to get an idea of the predicted maximum number of cases in the Netherlands. These models can also be used to simulate and check whether a lockdown had any effect, by fitting a model with data before the strict measurements and a model with data during these strict measurements. A difference in the **predicted maximum** can be a sign that there the measurements are effective.

Furthermore, this is important, because it can be predicted what the **maximum intake** is and when is "peak" is achieved. With this information, it is possible to take preventive measurements (such as necessary lockdowns, making sure that there are enough medical supplies and/or that the hospitals and ICU are not overloaded).

Lastly, it is possible that the actual data passes the predicted maximum, but these models are still valid, meaning that there exist a chance that the actual admissions can pass the predicted maximum. Therefore, it is important use these models to achieve insights when taking preventive measurements.

Note: detailed explanation and results are available on Github

5.3.2 Exponential growth trajectory

Eventhough exponential growth function are quite abstract and difficult to tell when the peak is reached, by plotting the **log-scale**, a straight line is obtained. This represents the trajectory of confirmed COVID-19 cases. By plotting the actual data (in the Netherlands), it can be seen whether the actual data deviates from the trajectory.

A downward deviation, would mean that things are getting better, but a upward deviation is a sign that more preventive measurements should be taken.

6 Meeting on 30-april-2020

6.1 Notulen

Today everything went as planned. A Powerpoint presentation with the work done and the results were given. Throughout the presentation, some important remarks and feedbacks were mentioned. One of the important feedback is to make sure that the source of the data that were used for the research, are mentioned. This is mentioned on **Github**.

The data of the daily confirmed cases and the hospital admissions are obtained from the website of **RIVM (National Institute for Public Health and the Environment)**, whereas the data of ICU admission is obtained from **Stichting Nice : National Intensive Care Evaluatie**: <https://www.stichting-nice.nl/>.

Furthermore, most other important remarks regarding the logistic growth model were:

1. Logistic growth is not 'suitable' for hospital admissions, because the logistic growth curve is only taking the admissions into account, but not the discharge of the patients (recovered or deceased). Thus, the section of **Hospital Admission** under Logistic Growth, is *deprecated (on Github)*.
2. Moreover, it is not 'suitable' for ICU admissions either, because the logistic growth curve is only taking the admissions into account, but not the discharge of the patients (recovered or deceased). Thus, the section of **ICU Admissions** under Logistic Growth, is also *deprecated (on Github)*.

For the above applications (hospital & ICU admissions), a SEIR model would be much better (since this deals with both admissions and discharge). Lastly, more the SEIR and the Parameters of the logistic growth (as well as the caveats of a logistic growth curve) should be studied.

6.2 SSA

The tasks to be completed as self-study assignment before the next meeting are:

1. Implementation of Kalman filter (on Python) and:
 - Applied on the daily confirmed cases
 - The hospital admissions
 - The ICU admissions
2. Improve SEIR model
 - Study the article and see what extra compartments and parameters are used in for the model. Furthermore, these parameters are not a constant, but rather a function of time.
 - Ultimately, it is interesting to add extra compartments to this model and try to mimic the course of the spread of COVID-19 in the Netherlands (our own parameters for each compartment).
3. Improve or analyse the Logistic growth curve carefully
 - Try to use less data (first two weeks), so the model does not use all available data for the training, but rather using the first two weeks and making a prediction. This predicted value is then compared to the data *"that have never been seen"*. Therefore, it is possible to calculate the error margin/deviation.
 - By applying the above method, more insights regarding the parameters are obtained. The parameters are not constants, as the model estimates and returns, but are rather a function of time.
 - The predicted maximum will be different than the predicted maximum when using all available data.

6.3 Results

The Kalman filter is a mathematical model that can be used as a predictor or filter. This means that this Kalman Filter can also be applied on the SEIR models or the Logistic Growth Curve, as a filter. But in this section, Kalman Filter is used as a predictor, to forecast the spread of COVID-19 in the Netherlands. Since, the Kalman Filter is ideal for systems that are continuously changing, the spread of COVID-9, which is time-dependent can also be seen as a continuously changing system.

The advantage is that Kalman Filter uses the observed measurements, which are measured over time (confirmed daily cases). These measurements can contain noise and inaccuracies (for example in this case: lack of testing, social distancing etc.), which the Kalman Filter will take into account and therefore produce values that are closer to the actual data (the prediction would be closer to the actual confirmed cases of that day).

6.3.1 Kalman Filter

The **Kalman Filter** is applied on:

1. The daily confirmed cases (**After further evaluation, this approach might be incorrect**)
2. It is chosen to apply it on the SIR/SEIR model, as a filter but also an estimator or predictor. This would mean, that the "**state-space**" would consist of the differential equations from the corresponding models.

Confirmed daily cases, can be used to update the Kalman Filter, which would enhance the next prediction.

6.3.2 SIR/SEIR model

After reading different articles, it is clear that these compartment models can have a high complexity. More compartments, such as **quarantine** or **hospital admission** can also be added. These models would not be dependent on the available data, but the available data is necessary to determine the different parameters in the differential equations found in these models. By applying the **Kalman Filter**, real-time data can be used to provide a more accurate estimation (as a filter on the SIR/SEIR models) and also for the forecast of new cases in the future.

6.3.3 Logistic Growth Curve

By using less data (20 days of the actual data) on the **Logistic Growth Curve**, it was remarkable that the model did **not** perform well, which was not unexpected.

In Figure 1, it can be seen what would happen if only 20 days of data was used to fit a model and using this (without updating the model with new available data) to predict the number of daily cases over a longer period of time. It is clear that the model's predictions are off-track.

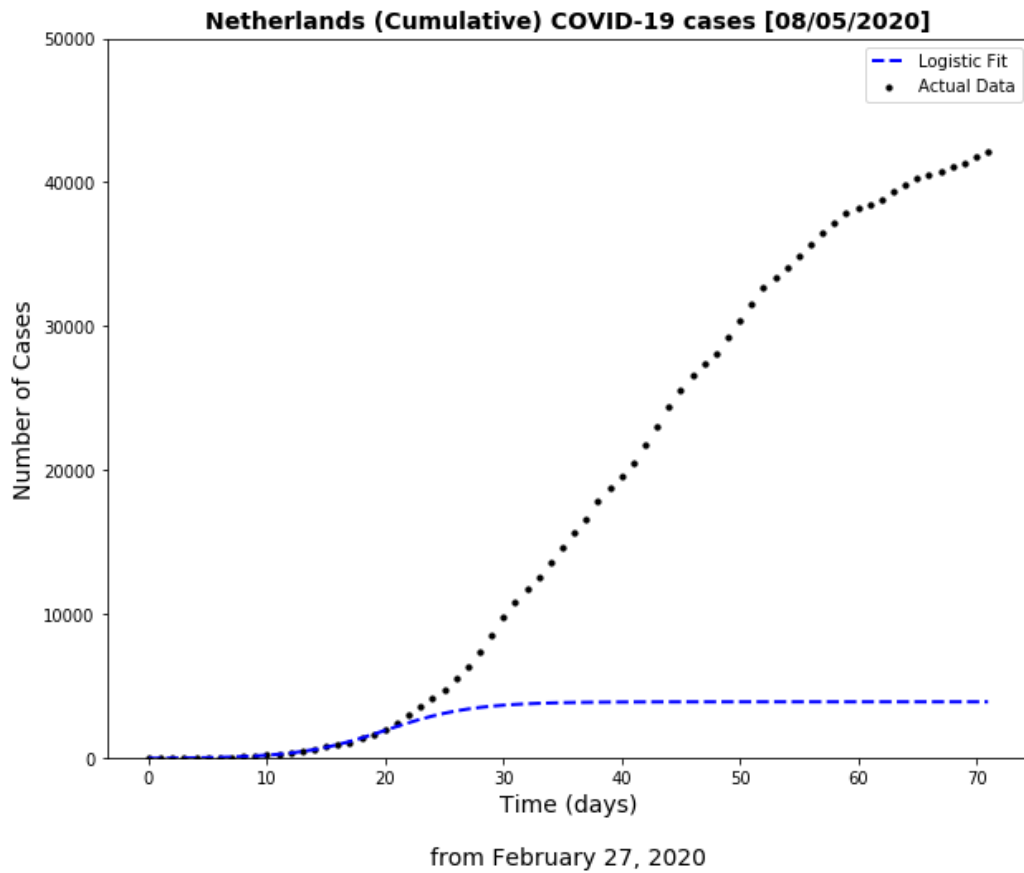


Figure 1: *Logistic Growth Curve model (only 20 days of available data and no further feeding of new data)*

Thus, the **Logistic Growth Curve** model, is a powerful model on predicting the number of cases accurately, but only for a short period of time. Furthermore, this model is highly dependent on the available data and has to be fed with new data continuously for an accurate model. Only with large data and continuously updating this model with new data, the predictions are accurate. However, the latter one is only accurate for prediction of short period of time (such as the cases in the upcoming 2-5 days).

Lastly, depending on the purpose or what question has to be answered, different models should be used. These models are only as good as the data they rely on. Such **Logistic Growth Curve** can be used to accurately predict the cases of the coming days (when sufficient data is available), but for a longer period of time, the SIR/SEIR (compartmental models) should be used.

7 Meeting on 07-may-2020

7.1 Notulen

7.1.1 Logistic Growth Curve

As mentioned before, the **Logistic Growth Curve** does have caveats. These caveats were mentioned in the previous results and also on the Github page. While this type of model does a really good job in predicting the new cases in the upcoming days, it still remains a challenge to have a "*constant*" **predicted maximum**. Each time, this model is fed with new data, the predicted maximum changes, thus it is really difficult to have a definite estimated **predicted maximum**. One challenge would be to use the available data and statistics to achieve a rather accurate estimation of the **predicted maximum** (in the coming weeks or days). An approach for this latter problem is to vary the predicted maximum with time, this way the predicted maximum would be in a certain range.

At last, it depends on the type of question that has to be answered in order to choose the proper model. No models are "good" or "wrong", but it depends on its applications and are only as good as the data they rely on.

7.1.2 SIR/SEIR model

These model, as mentioned before, can get really complex. For the Netherlands, it would be interesting to add "**Verzorgingshuizen**" also known as **Nursing home** as an extra compartment to this model. The reason is because there are many elderly who passed away during this COVID-19 pandemic, but is not certain that the cause was COVID-19, since these elderly at the nursing homes do not get tested (whether the death is indeed caused by COVID-19). This new compartment model can be seen in Figure 2 below:

- Adjusting the Kalman Filter to predict new cases over a longer period of time, since it is accurate for predicting the cases of the upcoming day, but can have high uncertainty when predicting the upcoming week (just like weather forecasting).
- This Kalman Filter approach might even be more accurate than the Logistic Growth Curve.
- Kalman Filter applied on the SIR/SEIR might actually be a discretization of the SIR/SEIR models' differential equations.

Thus, there are many applications where one can make use of the Kalman Filter, whether it is for estimation/prediction, smoothing or filtering. It can even be used for estimating the parameters, thus constantly updating these parameters over time. Depending on the type of application, one can consider using a Kalman Filter.

7.2 SSA

The tasks to be completed or interesting to look into as self-study assignment before the next meeting are:

1. Try to find an algorithm or method to have an accurate estimation of the so called ***predicted maximum*** for the Logistic Growth Curve, since this remains a big challenge with such model.
2. Add Nursing Home compartment to SEIR model and play around with the parameters.
3. Play around with the parameters, by making them time-dependent, to check whether lockdown actually has any effect on the spread of COVID-19.
4. Read articles or information regarding the update of parameters by using a Kalman Filter. Thus, using the Kalman Filter to estimate and update the parameters over time.
5. Read the blogpost and try to figure out which "**state-space**" the author used, or was it just using ***time*** as a "state".
6. **Logistic Growth VS Kalman Filter:** Which is best for forecasting and when should it be used? Furthermore, which is better for long period or short period forecasting?
7. Kalman Filter applied on the SIR/SEIR might actually be a discretization of the SIR/SEIR models' differential equations. Take a look at this phenomena.

7.3 Results