**1st meeting (15th April 10:00)**

Discussion about:

The topic and research question.

* SIR/SEIR model for disease transmission.
* Expansion of SEIR model for better analysis of disease transmission.
* Expansion of the model by introducing additional states

Agent variables that define the conditions of agents

* Variables can define the infection rates of each agent. E.g., an additional variable that defines if an agent is wearing a mask or not.
* Agent-related variables can define the conditions(characteristics) of agents. This may alter infection rates on each agent (some are more vulnerable to diseases than others).

Analysis tools for model assessment and optimisation

* openMole and BehaviourSpace

Discussion about the ethics form.

**2nd meeting (9th May 14:30)**

Things I did and want to do:

* I wrote the introduction and literature reviews and created a model description diagram.
* I currently implemented a basic SIR model with vaccination on Netlogo.
* As a next task, I wish to build an environment using walls as objects.
* Add more variables to define infection rates depending on various conditions of agents.
* Add more states (incubation and death states)
* Alter agents' movement
* Read more about openMOLE and get used to the functionalities it provides.

**3rd meeting (8th June 10:00)**

Questions that I want to ask:

* Show the codes I have built and explain them to the supervisor (ask how to calculate the infection rate). The infection rate depends on the properties of an agent. For example, the infection rate decreases if an agent wears a mask and increases if an agent suffers cardiovascular disease, etc.
* Refine the model description diagram and show it to the supervisor. If things need to be changed, make an amendment, and improve the model description.
* Ask why the openMole is not allowing me to click the files. Plus, what kind of analysis method should I use to analyse the models I have built?

**Supervisor feedback:**

* Instead of using random-normal, use random uniform. When calculating the infection rate, it would be advisable first to calculate the proportion of infected agents on a patch. For example, if there are 20 agents on a patch, we first need to categorise them based on the type of transportation. If there are five agents with transportation as 'train' and out of those, only one agent is infected, the infected proportion is 20 percent. Then the rest of the four agents get infected based on this calculation: 0.2 \* infection\_rate\_train. Then, we further consider if an agent is wearing a mask or suffering from cardiovascular disease and re-calculate the infection rate.
* When assigning agents to each company, they must not be assigned with equal random chance. This is because each company has a different size. Therefore, agents must not be assigned equally.
* The timeline must be fixed. The test of our model must be fixed within a specific timeline—for example, ten days or 20 days. Standardising the testing time is necessary.
* The cumulative total count of agents in each model state must be monitored. Instead of monitoring the current number, we must take note of the cumulative total of agents in each state of the model.
* In the model description ppt, do not put the whole NetLogo code; use pseudo-code instead.
* When using openMole analysis, find the effectiveness of vaccination. Keep in mind that the cost of vaccination is expensive; hence we are looking to minimise the cost of vaccination and maximise the effect of reducing the number of infected agents in the shortest period.
* The flow of the experiment is as follows:
  + Statistical distribution of output
  + Direct sampling
  + Sensitivity analysis
    - This analysis shows the importance of the inputs of the model.

**4th meeting (1st July 15:30)**

1. How should I explain Saltelli's method in my paper?

             \* How detail should I explain the mathematical concepts behind them?

             \* Explain the mathematical equations and the terms in the literature review or methodology?

2. How do I interpret the results of DS?

             \* Present the results of both methods and discuss them with the supervisor.

                            - DS has a sample (model run)

                            - GA has termination and parallelism. What are they exactly?

                            - Saltelli has a sample (model run)

             \* Should I analyse the results with Morris SA as well?

                           - Morris has a sample and level. What is the level referring to exactly?

                            - Also, how is the result interpreted? How should I represent the result?

4. How should I present the statistical distribution of output?

             \* In the previous meeting, he told me to do three things which are:

                            - present statistical distribution of outputs (Ask the supervisor what this means)

                            - data sampling

                            - sensitivity analysis (Saltelli and maybe Morris)

                            - in addition, I am doing optimisation using NSGA2

5. How to visualise the results? In what ways? How many graphs to show?

             \* bar graph, line graph, box & whisker plot etc.

6. How should I determine the probability of infection, and how to explain based on what?

             \* The factors that define infection rate are:

                          - by company, transportation

                          - by mask, cardiovascular disease

                          - the proportion of sick agents out of total agents on a patch

             \* How do I justify the rate of infection calculated by these factors?

7. What should I include in the project scope? How should I describe the scope? In what ways?

             \* Define parameters, timeline, schedule of a day, initially infected agents, initial population etc.

8. Regarding the SIR model, do I have to explain it in detail? I have constant changes in population, but the SIR assumes:

              - that there is no change in population since it does not consider the birth and death of agents in the population.

              - Plus, I have a vaccination state. I know I am simply using ABM based on the concept of the SIR model, but to what degree should I explain the SIR model and how?

9. I have three input parameters, including vaccination rate, infectious and immunity period of agents. Are they enough?

10. Should people's immunity period differ based on recovery and vaccination? Which one should be longer?

**[Justifying the parameter values]**

             - In justifying the values of contact rates and all other variables about infection rate calculation, it would be advisable to relate to the literature review

             - But this is usually the hard case, so try carrying out with various values and find reasonable outcomes.

             - We can explain that specific rates are fixed with some values because we came up with a stylised scenario like that... etc.

             - Because the combinations of these values are almost infinite, there should indeed be fixed values and explained why.

             - Try to justify the parameters of the research topic. For example, I am interested in these input parameters in particular...

             - Can change the number of companies and transportations even.

             - We can also introduce other input parameters eg. alter the variables that deal with contact rates, and find out if the result changes dramatically.

**[The hook to display the results]**

             - Change the format of the hook section of the .osm file to the standard format so a result can be shown in a matrix format.

**[Saltelli's Sensitivity Analysis]**

             - The analysis does not need to be explained in detail, but the concept can be explained in some mathematical forms to backup.

             - Morris's sensitivity analysis does not need to be explained since Saltelli is used.

             - The interpretation of the result, the values are in percentage values which can range from 0 to 1 only. Plus, round them up to 2~3 decimal points.

**[Input parameters]**

             - The input parameters can be considered up to 6 at max but keep it 2 to 3 parameters.

             - The objective parameters should only be considered for upto 2.

**[Statistical Distribution of Outputs]**

             - Using direct sampling

             - With every possible combination of the input parameters, show the overall distribution of the outputs by altering seed values.

             - https://next.openmole.org/Uniform+Sampling.html#Samplewithinauniformdistribution

**[Progress of Work]**

             - Statistical Distribution of Outputs

             - DS (Data Sampling)

             - SA (Sensitivity Analysis)

             - GA (Genetic Algorithm)

**[Project Scope]**

             - Project scope should be more broad and generalised.

             - The things I have written are suitable for methodology.

**[The SIR model]**

             - The SIR does not need to be explained in detail since ABM is the central part of the dissertation.

             - Site more ABM-related papers rather than the SIR model.

**[The waves of the infected agents]**

             - Currently, the outbreak's waves are not recorded and cannot be visualised since they are represented by numbers only.

             - It would be good if there was a way to count the number of peaks (number of waves)

             - We can use the 'peak detection algorithm' to calculate the number of peaks in our data. (Use Netlogo to implement the code)

             - If this is too difficult to do, try to export the graphs from Netlogo and visualise the number of waves.

**[UCL computer environment service]**

             - The supervisor has sent the email to the head of the dissertation to allow access to the UCL research computing service.

**[Sending draft of dissertation]**

             - I must send a draft of my dissertation to the supervisor during the weekend.

**5th meeting (13th July 15:30)**

1. Should the input parameters be Double or Int when conducting analysis?

   Can we also display the values of inputs used to generate specific outputs in the table?

             We have 4 analyses to carry out:

                            - Uniform Sampling

                            - Direct Sampling (Grid or complete sampling)

                            - NSGA2

                            - Saltelli SA

                            \* VIF & Multi-Linear Regression

2. Finding the peak (wave) is inaccurate using z-score and threshold.

   Python code built in .ipnyb which can be shown to the supervisor in the meeting.

3. Ask if VIF and multi-linear regression can be used to determine the impact of the

   input parameters or if this is not necessary since SA is conducted.

4. Why is an essential statistical distribution of output using a uniform sampling method of inputs?

   They are similar to the second analysis, which uses Direct Sampling of inputs that are evenly spaced.

5. For the first and second analysis (statistical distribution of the outputs and the DS of input parameters)

   What graphs can you recommend me to use to visualise the results? What should be the output parameters?

   Should it be just cumulative numbers or proportional to the total population?

6. For the statistical distribution of outputs, we need to identify

   the types of distribution it represents after we get the results. Like whether it is Poisson distribution or...

   If so, can we check the chi-squared goodness-of-fit?

7. The total combination of the input parameters is 81. Should each combination be executed

   10,000 times? Wouldn't it cost too much computational power and time?

8. OpenMole for my workspace does not work. (Bad Gateway)

9. Confidence interval calculation

CI = 2 sigma z / sqrt(n)

n = ((2 sigma z) / CI) ^ 2

ci = sigma / 2

n = (4 sigma z)^2

z = 1.96 for 95 confience interval

n = 64