**Script to go with presentation**

**Slide One:**

Hello everyone.

My name is Jack Collins and for the past few weeks I’ve been working on optimisation in parameter fitting for pandemic modelling under the supervision of Niall Madden.

**Overview:**

Today I’ll be talking about the models we had to choose from, what the coefficients represent, issues when working with real world data, the optimisation of the coefficients, and what the next steps should be.

**Resources:**

The main tools we used in this project were MATLAB, and MATLAB add-ons such as CHEBFUN, the optimisation toolbox for MATLAB and the MATLAB app designer which was used for demonstration purposes. All images were created using the above unless otherwise stated like I have here in the bottom corner of the slide.

**Available Models:**

There is a vast wealth of options to choose from when you consider modelling a disease, ranging from Compartmental to Individual to Spatial and so on. Each have their own advantages and disadvantages associated with them, but for the purpose of this project I choose to stick primarily to compartmental models as they are the most widely used.

**SIR Model:**

That brings us on to the SIR model. The SIR model is one of the most well recognised compartment models and offers a good foundation upon which to further build. It consists of three differential equations. Each equation represents the rate of flow of people from one compartment to another. They are as follows:

S-prime = - Beta\*I\*S/N

I-prime = Beta\*I\*S/N – Gamma\*I

R-prime = Gamma\*I

Here S,I and R represent the compartments ‘Susceptible, Infected and Recovered’ while Beta and Gamma are the coefficients we’ll be investigating. It is also worth noting that S + I + R is equal to the total population N, which remains constant.

**The coefficients and what they mean:**

The reason these coefficients are so interesting and worth exploring is that they relate to physical attributes. Beta, the infection coefficient is the number of people one person is likely to infect should they contract the disease, whilst gamma, the recovery coefficient is how fast an infected individual will recover. Together they give us the famous reproduction number, R naught that we’ve heard so much about on the news. More importantly, because they have physical translations and are not purely mathematical, they can be used to compare and contrast the viral outbreak in different parts of the world as well as evaluate and critique government restrictions and their effectiveness.

**The importance of the values:**

In order to be able to achieve this we need to understand just how sensitive the model is to the value of beta and gamma. To show this I have prepared a short demonstration using MATLAB app designer which I’m going to open up now.

**Open App:**

Everything is already set up and running as you can see. Here we are fitting the SIR model to real world Irish data. The x-axis represents the days since the first case appeared and the y-axis is the number of daily new cases. Beta and gamma are set to predetermined values of 9.6 and 9.4 which give a reasonable fit for this data. To emphasise how important the accuracy of the value is, I’m going to show you what happens if I change either of the values ever so slightly. If I move the beta slider to slightly less than 9.6, you’ll see despite a small change in the value there is a dramatic change in the peak of the curve and when it occurs. Increasing beta slightly above 9.6 makes the peak happen sooner as well as increasing it. You’ll also see changing the gamma slider has just as dramatic an effect on the fit. If I move it slightly below 9.4, it happens sooner and is larger, slightly above 9.4 and it happens later and is smaller

**Apply to real world data:**

Now that we’re aware of how sensitive the model is to the coefficients, we must look at applying our model to real world data. On the right you’ll see a bar chart of the reported daily new cases in Ireland. It’s clear to see that there is a lot of noise from things such as over and under reporting and if you look closely enough, you’ll notice it follows a weekly trend. On top of this there was a lack of testing in the initial weeks of the outbreak which means some cases went undetected and in Ireland in the beginning. Also the recovered cases were not reported daily. These will all have an effect on the values we’re able to attain for our coefficients.

**Noise Reduction:**

So the next step is naturally to try and minimise the affect this noise has. To do this we created an SIR model with arbitrary Initial values and population figures and sampled the data. We then applied artificially created noise to the sample data which you can see on the right-hand side. To best replicate our problem the artificial noise followed a seven-day trend.

**Noise Reduction (2):**

Next, we explored various noise reduction techniques and found that the *smoothdata* function in MATLAB was the most accurate for what we needed. It takes a seven-day moving gaussian average of the noisy data and returns a rather close representation of the sample data we started with. The difference between the noisy and smooth data can clearly be seen in the figure on the right.

**Noise Reduction (3):**

Applying this technique to the Irish data set has a profound affect and manages to reduce a lot of the issues we faced when working with real world data

**Coefficient Optimisation:**

But why do we need to optimise the coefficients? Well as you’ve seen in the app, small changes in the values of either of the coefficients can lead to rather dramatic changes in the fit. On top of this, it allows us to be more precise in our comparisons between regions and gives us a more accurate evaluation of restrictions and policies that were implemented. In order to optimise them, we used the MATLAB optimization toolbox.

**Optimisation Process:**

How does it work? First you choose a solver. After some trial and error, we decided to run with LSQNONLIN, which is a multi-variant nonlinear solver. It had the huge advantage of being able to set upper and lower bounds. This allowed us to prevent negative values for the coefficients which would have not made sense in our model as well as increasing the efficiency of the process by narrowing down the values it had to search through.

Once that was decided, we had to define our SIR model in chebfun along with the parameters and initial guesses. The initial guess were particularly important as the better they were the easier and more efficiently the solver was able to optimize our problem.

Then we ran our solver which works by minimising the least square difference in a vector which is outlined below, and it returned new values for Beta and Gamma.

One could attain more accurate values be repeating the process but using the newly found values as initial guesses.

**Optimisation Process (2):**

Here’s a quick display of the before and after of the process on the Irish data set. On the left is our model with the initial guess, which aren’t terrible but the fit itself is. On the right is a much more accurate fit and the new optimised values which were found. As you can see, it’s quite affective.

**Comparisons:**

Now that we’re able to attain reliable figures for the coefficients, we’re able to use them to compare regions. Here you will see the optimised values for both Ireland and Italy as well as the R\_0 value. You can see that Italy has a larger rate of infection which mean more people are contracting the virus on a daily basis, but it also has a faster recovery rate then Ireland which means more people recovering as well. As a result Italy actually has a lower R\_0 value despite having more overall cases then Ireland.

**What’s Next?**:

What’s next? Well taking inspiration from the two papers listed here on the screen, the next goal would to be to develop a neural network which could not only attain more accurate values, but would also be able to use a clustering algorithm to group countries of similar beta, gamma and R\_0 values together. By doing this and applying to real world data we could potentially tell if a country which is only at the beginning of the outbreak is going to follow a trend like China and have only one major peak in daily new cases, or will experience something like Israel and encounter a second substantial rise in daily cases. Furthermore, by allowing the Coefficients to vary in time we could more accurately recreate the trends seen as a result of preventive measures and establish what had the largest impact on the figures.

Thanks for listening, that is all from me, and if you’ve any questions feel free to ask.