Over the summer, I was able to work with my Pembroke tutor Raphael Hauser in order to generalise an algorithm for handling data dimension-reduction he and his collaborators (Armin Eftekhari and Andreas Grammenos) had developed. Retaining my room in the GAB, I was able to meet with Raphael regularly to discuss the work, as well as take full advantage of the many libraries and academic resources that can be found in Oxford.

The original algorithm in question is called MOSES, standing for Memory-limited Online Sub-space Estimation Scheme. I began my time on the internship by reading though Raphael's paper on MOSES, helping me to gain a better understanding of the algorithm for myself.

It can often be useful to have a lower-dimensional representation of a set of data. This representation is typically found by manipulating the dataset in its entirety. Rather than doing this, MOSES considers data in small 'blocks' and uses these to reduce the dimension of the data. This provides a running estimate of lower dimensional representations of the data. For very large sets of data – as are common in the real world – using MOSES saves a significant amount of time. In fact, Raphael's paper provides proofs of bounds on the computational cost of the algorithm, reflecting its speed and efficiency.

However, the MOSES algorithm relies on the assumption that the distribution from which the data is sampled from doesn't change over time. This is an unrealistic assumption to make in the real world. For instance, in financial data it's very clear that the distributions governing markets changes rather quickly.

This observation led me to seek a generalisation of Raphael's MOSES algorithm, which I called Time-dependent MOSES. The idea behind this generalisation is to introduce weights to the data, whereby chronologically older data entries have a smaller weight and less significance. I did this by geometrically scaling down older data by multiplying by a fixed constant between 0 and 1. This is useful for making predictions of the distribution of data at the time of inspection. For instance, weather information from 10 years ago is less useful for predicting today's forecast than yesterday's weather is.

I spent the next period of the internship trying to analyse my Time-independent MOSES algorithm. At first, I studied the proofs and methods in the original MOSES paper, hoping that I would be able to adapt the work done there and apply it to my own problem. This was proved to not be very fruitful. After some time, I realised that by applying a matrix transformation to the data, I could effectively transform my problem into the previous MOSES problem, which had already been explored!

Using the online program Overleaf, I was able to write up my method and findings in a report. This was a good learning experience for me, as it helped me to improve my Latex writing skills, something that will be crucial for me in fourth year, as I write my statistics dissertation.

This internship was a very enjoyable and educational experience for me. I enjoyed learning about the MOSES algorithm and coming up with a variation of it. Having never done any form of mathematical research before, the internship tested a new range of skills for me. The work forced me to persevere when I was unsure how to analyse the Time-Dependent MOSES, and I was encouraged to think creatively to come up with a solution. I am grateful for this welcome introduction to the world of mathematical research.