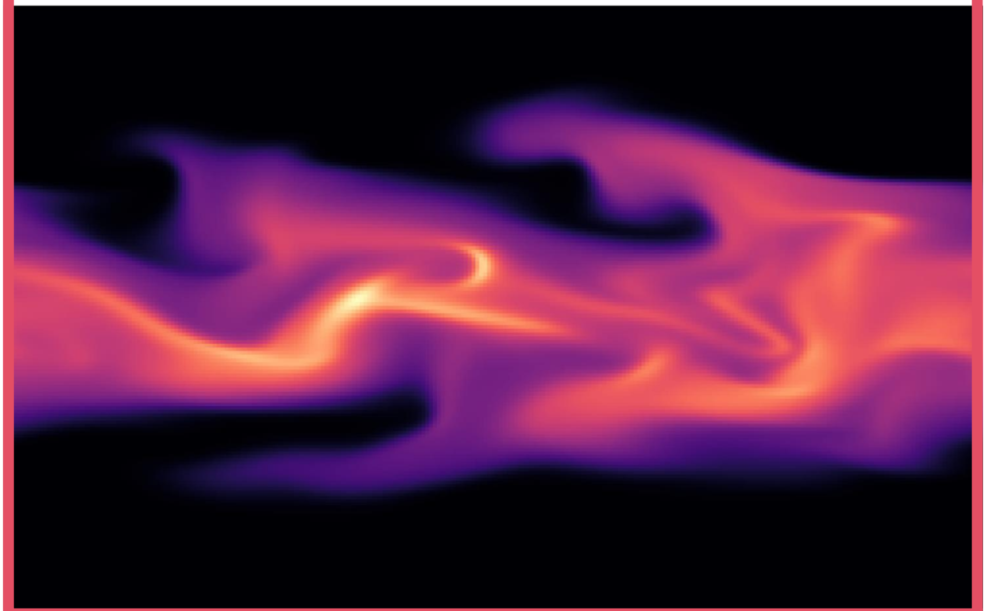
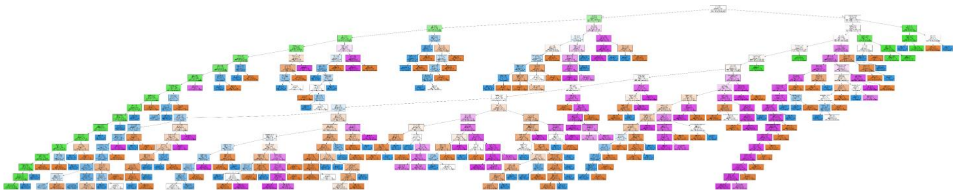
Data Science Portfolio

William Booth-Clibborn

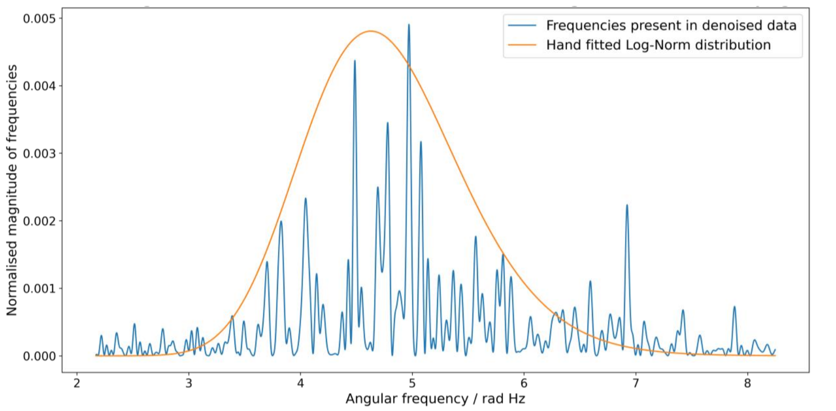
Fluid dynamics simulation of early protoplanetary formation.



Racism in the UK Rental Market



Analysis of accelerometer data from Royal Marines Landing Craft



Contents

Introduction

This is a document containing and explaining my choice of samples of work. These projects are either personal projects or are in-progress pieces of research. Therefore there is no write-up as such, but instead I have presented here some conclusions and Jupyter notebooks associated with the samples of work that I've lead. This document is formatted for easy reading and is not in the form of a report. Text that is blue in colour are hyperlinks to repositories, click on them to be redirected to GitHub. In case hyperlinks stop working (because of PDF processing issues) a list of URLs can be found as the final chapter of this document.

Before reading this document please visit my Portfolio Repository and watch the animation seen in the README.md for some context on my protoplanets project. This animation is from the protoplanets research project from over 100GB of simulation data that I generated, and is the culmination of the most involved data science that I have undertaken.

The three projects I am displaying for my work sample are the following:

1. **Fluid mechanics of protoplanetary disks**, supervised by Dr Pablo Loren-Aguilar. This started as my dissertation and was extended into a research project over the summer. We received £1500 of funding and an unlimited HPC budget from the University of Exeter.

2. **Prevalence of racism in the UK rental market**, lead by Dr Amy Binner. With the pilot study completed, we're planning to expand to the whole UK rental market with a funding application planning to be submitted February.

3. **Predictive Shock Absorbing Seats for the Royal Marines Landing Craft**, for a personal project assisting a friend's engineering project.

1 Fluid mechanics of protoplanetary disks

This project started as a group dissertation project investigating the streaming instability in protoplanetary disks, which is a kind of fluid instability involving a streaking motion of density. We discovered that this effect is likely not physical, and is the result of the boundary conditions of the simulation, rather than expected physical behaviour. This project was submitted as a poster and achieved the highest dissertation grade in the Year in the Physics and Astronomy department. A digital copy of this poster can be found in Portfolio Repository: Protoplanets Poster.

The main challenge of this project was the computational scale of the simulation. When this was a dissertation project, we were running these simulations using WSL on the laptops of 7 undergraduate students overnight. I had to coordinate the parameters everyone would use, share over 80GB of data over OneDrive, and chase people up when they inevitably forgot to run them before bed. We received £1500 of funding and an unlimited budget on the HPC at Exeter to continue the research over the summer. It is a large two-fluid computational simulation run in python using a spectral partial differential equation solver called Dedalus. To generate a dispersion relation we used linear instability analysis on the set of equations that govern our behaviour, and plotted this using Mathematica.

Protoplanetary disks are disks of matter surrounding a young star. They are dispersed clouds of gas and dust, and eventually condense to form planetary systems such as our own solar system. The mechanics of how $\sim 1\mu m$ sized particles can become $\sim 100km$ sized planets is unknown. The large difference in size has been suggested to be influenced by fluid mechanics, because of turbulence's ability to scale across many orders of magnitude of size. One mechanism this may be able to occur through is the gas in the protoplanetary disk, which may exert a fluid drag on the dust. Over short time spans this effect is negligible, but across 100s of orbits this effect becomes substantial.

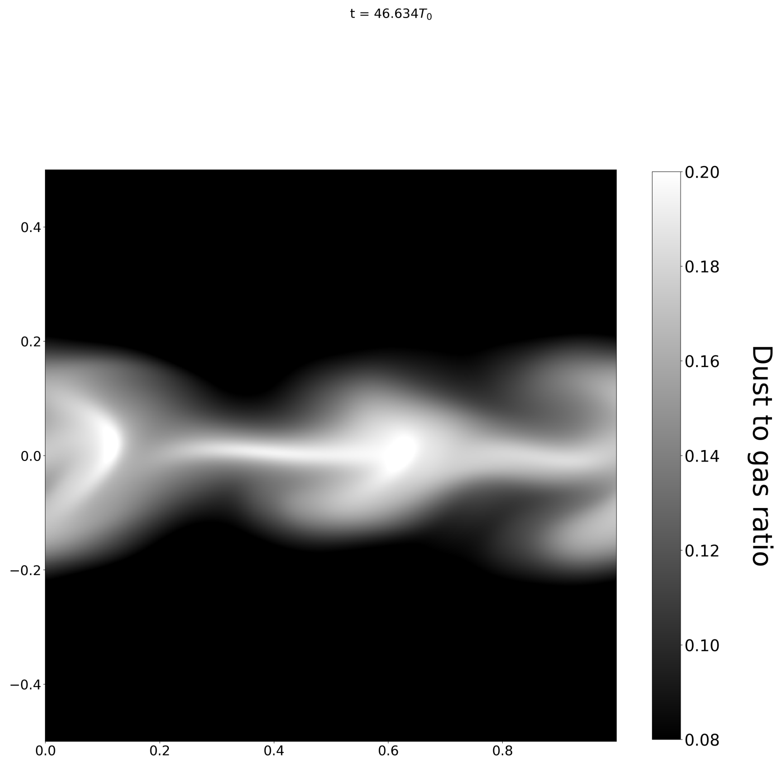


Figure 1: This is an image we generated from a 100GB simulation. This is a 2D map of density and is being viewed as a co rotating sheering box viewed as a cross section of the protoplanetary disk. As we can see here behaviour consistent with the shape of the Kelvin-Hemlholtz instability.

This still image in Figure 1 is taken from an animation hosted on Portfolio Repository which is generated separately each time the simulation is run. This allows for a quick preview of the simulation, but also allows for the flexibility to change plotting settings without needing to redo the simulation step.

The source code that I designed for the HPC at University of Exeter called ISCA can be found in my Two-Fluid Repository but it is very specialised for use on our HPC ISCA. My code is currently running simulations, and may be used in the future for more dissertation lab projects.

The next steps of the project will involve:

- Run far more simulations changing each parameter to observe the stability of the Kelvin-Helmoltz Effect in different conditions.
- Explore further methods for taking 2D fourier transforms of 1D waves
- Attempt to recreate some conditions of the streaming instability

2 Prevalence of racism in the UK housing market

I was the data lead the for a research project analysing racism in the UK rental market at University of Exeter with Dr Amy Binner. We were researching the impacts of different races being disproportionately denied housing in certain areas. We correlated this with data on pollution and access to recreational spaces to prove that the racial discrimination in the rental market can push people of certain races into less safe areas. I started on the project only responsible for setting up data redundancy on the HPC, but as I worked within the team I took more responsibility until I was managing the data gathering software, managing 5 interns from Duke University and University of Illinois. This year I hired 3 more interns to categorise the data we gathered in Qualtrics.

I am working on my own extension of the project, to determine whether training a ML model with geolocal data can identify the differences between accents. This requires a sentiment analysis model that is categorising the sentiment ordinarily, which I have been developing by testing decision trees and support vector machines. The output of the decision tree can be seen below, with more detail found in my repository.

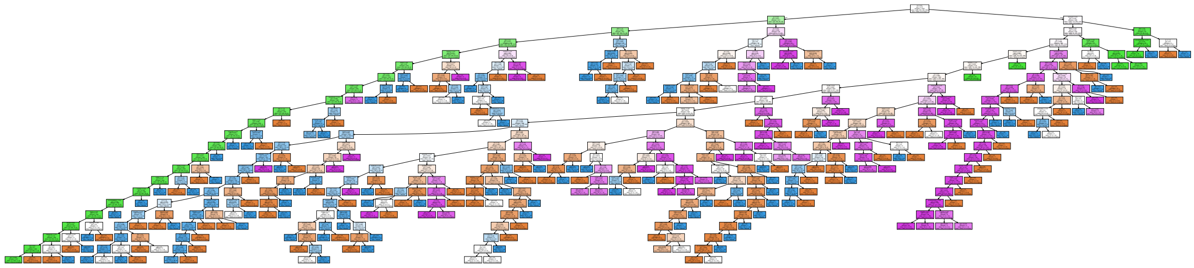


Figure 2: This is a plot of a decision tree with the maximum predictive ability to do sentiment analysis with ordinal classification. A higher resolution image can be found in my Portfolio Repository: Decision Tree.

This project is currently looking for funding to expand the scope to the whole UK rental market. Hence, the code and many outputs from this I cannot shared as it is active research. We are in the process of writing up the pilot study research paper for publication once we receive the funding for the larger scale project.

3 Analysing Accelerometer data for Predictive Shock Absorbing Seats for the Royal Marines Landing Craft

The new specification for updating Landing Craft Vehicle Personnel (LCVP) MK5 involves the landing craft reaching higher speeds and further distances to allow for over the horizon deployment at sea. The craft is physically capable of reaching these speeds, however the impact of hitting waves is very hazardous to the commandos inside. The impact is enough to break bones and slip spinal disks and has been measured as high as an instantaneous 18g, therefore shock absorption is necessary to protect the commandos. A G-force of 9g is high enough to make most people blackout after seconds. Active shock absorption allows for lower forces on the commandos, protecting their health while maximising their effectiveness.

My analysis was a part of a larger engineering project run by a friend of mine. This analysis was designed to demonstrate that there were signals present in the data we gathered that may have predictive power. The data was gathered by attaching a smartphone to the floor of the inside of the landing craft and reading accelerometer data from the smartphone. We were approached by CGI to continue this research after the completion of our degrees, however my friend is now a Commissioned Officer in the Royal Marines, hence our project is on hold.

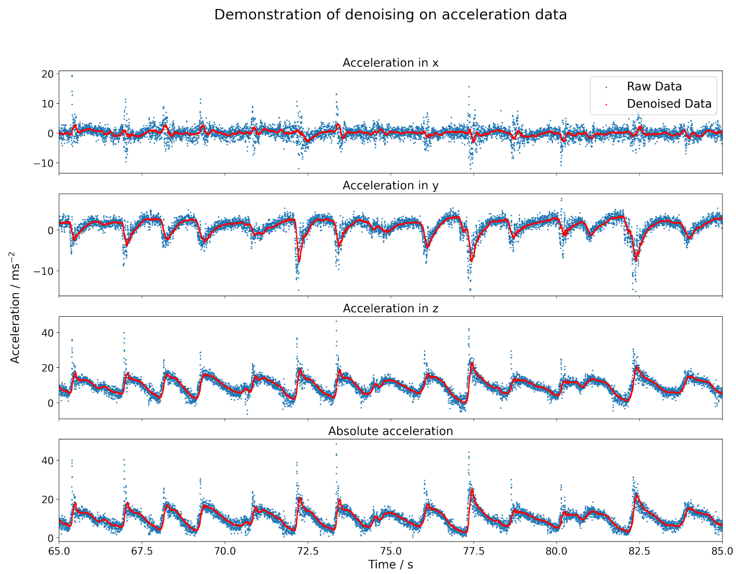


Figure 3: Rolling window denoising algorithm used because of the low precision of the accelerometer in the smartphone. This was generated using an inbuilt function in pandas.

This gave us a smoothed out view of the impacts. This removed their more dramatic peaks, which is ideal for signal processing, but not so much for predicting the magnitude of the impact.

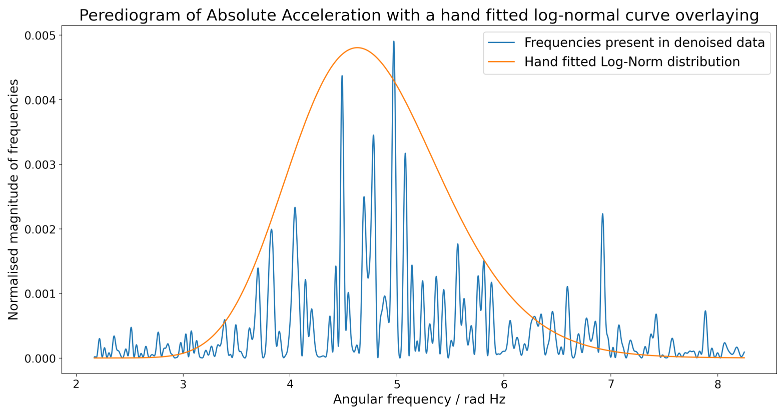


Figure 4: Perediogram of the wave data, with a log-normal curve hand fitted. This was generated using least-squares spectral analysis in astropy, as FFT would not work on our data because of inconsistent sampling.

This is a view of the frequencies present in the dataset. I cropped the periodogram around these frequencies as this is the correct order for frequencies that we see in our data. I manually fitted a log-normal curve to the graph, in the suspicion that the impacts we observe would share the same frequency profile as ocean waves. I believe this was a good fit and the spikes in frequency are expected with our smaller sample size.

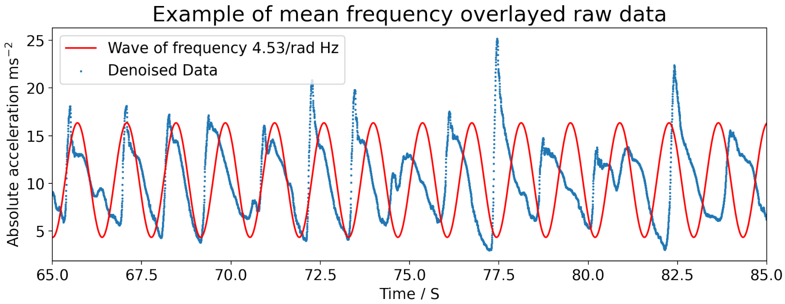


Figure 5: Overlay of a sinewave with the mean frequency we found in our log-normal fit

To demonstrate that our log-normal produces fit sensible results, this is a sinewave with the same mean frequency as found in our log-normal fit, and same height as our wave data. A jupiter notebook of the exact step by step can be found in my Portfolio Repository: Royal Marines Analysis.

The next steps for the project would involve the following:

- More data in different conditions to generalise our findings better, in different weather conditions, and different speeds.
- Predict the magnitude of the impacts
- Predict the direction of the impact to be used with gimbaled seats to make the impact pass straight through the commandos spine
- Try fitting saw waves to the data

URLs

Link Name	**URL**
Portfolio repository	https://github.com/w2l12am/Portfolio
Protoplanets HPC code repository	https://github.com/w2l12am/Two-Fluid-ISCA