A computer model for investigating the biomechanical effects of radiation exposure on pathological and non-pathological living human cells

Gary James Johnston

A thesis submitted in partial fulfilment of the requirements of Liverpool John Moores University for the degree of Doctor of Philosophy

February 2017

TABLE OF CONTENTS

I. Abstract

The cellular response to radiation insult and studies have been carried out to investigate aspects of the cytoskeleton and the force response of the cell when probed by an AFM. Confirmed for the first time that there was a statistically significant difference for the PNT2 and PC3 cell lines in response to probing with the AFM tip, and that time was eliminated a possible influencing factor in the short term (1 hour) for the force response. Showed that the Hertz model is not sufficient for distances greater than 500nm due to the strain hardening effect for biological cells and that the biological cells non-linear force response becomes marked after the 500nm region.

The orientation of actin was investigated and a bimodal variation was statistical significant, although the larger tendency was for a 90 degree separation there was indications that earlier theoretical work by Pollard, 2008 was present.

The importance of the contact point when considering the cell lines PNT2, DU145 and PC3 and greater than 500nm indentation is shown and four different methods are tested and the most robust of these chosen as the method for the distance and cell lines involved. That being 'line projection' method created by the author.

A method that normalises the data for AFM force curves is presented, the method minimises the contact point error at the same time and therefore provides biologists with a way to test cell lines using standard normal population tests.

II. Acknowledgements

I would like to thank my Parents for their love and support. Hafthor who did his best to keep me sane, and my Supervisors for putting up with my insanity. To Fran for all his help in proofing and patience with my wood for trees problems. To Mark for his help and cheerfulness. To Dave for his wisdom and good ideas throughout the PhD. To Annette for her help with the yeast strains, and for the Tayto crisps. To Jonno and Steve for all the lunches and helpful advice. To Fred for all his help with my problems. To Critical Role and Day 9 for keeping me laughing in spite of myself.

III. List of abbreviations and symbols.

AFM	Atomic Force Microscope
AIC	Akaike Information Criterion
ATP	Adenosine Triphosphate
DU145	Prostate Medium metastatic cells
Eb	Young's modulus or x^2 co-efficent of Extended Hertz formula.
ECACC	European Collection of Animal Cell Cultures
F-actin	Filamentous Actin.
GUI	Graphical User Interface
Gy	Gray unit of radiological energy.
IBW	Igor Binary wave proprietary format
K^+	Potassium
K1	X^3 component of the extended hertz formula
K2	X^4 component of the extended hertz formula
Kstest2	Kolmogorov-Smirnov test 2 sample variation.
Mg_{2+}	Magnesium,
Na^+	Sodium

ODBC	Open DataBase Connectivity driver.
PC3	Prostate high metastatic cells
PNT2	Prostate normal tissue cells (Immortalised)
RGB	Red Green Blue
SQL	Set Query Language

IV. List of publications arising from this work.

Johnston G, Burton DR, Lilley F, Doyle A, Murphy MF, Madden G, Gdeisat MA, Moore CJ, Marchant T, Matuszewski B. (2011). Analysis of microscopy and reconstructive images for applications in medicine and biology *Proceedings - International Conference on Image Processing*, ICIP, :3069-3072

Ahtaiba A, Gdeisat MA, Burton DR, Lilley F, Murphy MF, Johnston G. (2011). A novel technique for the restoration of atomic force microscope images enabling an approximation of AFM impulse response *Proceedings - International Conference on Image Processing*, ICIP, :3065-3068

Doyle A, Crosby SR, Burton DR, Lilley F, Johnston G, Perez WB, Kinzy TG, Murphy MF. (2013). Deletion of the TPM1 and MDM20 genes affect the mechanical and structural properties of yeast cells MCB *Molecular and Cellular Biomechanics*, **10** :275-288

Murphy MF, Lilley F, Lalor MJ, Crosby SR, Madden G, Johnston G, Burton DR. (2013). Evaluation of a nonlinear Hertzian-based model reveals prostate cancer cells respond differently to force than normal prostate cells. *Microscopy Research and Technique*, **76** :36-41

V. Table of Contents

Table of Contents

VI. List of figures

VII. List of Tables

Chapter 1 : Introduction

This thesis reports upon a research programme into the development of novel software tools and statistical methods for life scientists to analyse and explore atomic force microscope (AFM) data. The AFM can be used not only to image live cells, but also as a microscopic force probe. In this latter mode the instrument mechanically probes living cells, testing the cells' force response and in this study this process is performed on both cancerous and non-cancerous cells. The output from the AFM is complex and the data numerous and therefore extracting meaningful data is best handled using automated techniques rather than manual methods and it is in this area that this research programme mainly seeks to make a contribution.

This research programme is linked to previous work that has been carried out at Liverpool John Moores University into cell mechanics and particularly into the mechanical response of cells to ionizing radiation insult, in a manner that is similar to that seen in radiotherapy treatment of cancer (Gabryś et al. 2007; Murphy, 2007).

The software tools that have been developed as part of this research programme have also enabled the author to make a contribution in population-based cell mechanics for cells exposed to ionizing radiation insult.

The novelty of the work performed as part of this research programme therefore lies in the following aspects:

- The automation of the process of AFM force-distance curve data capture and analysis.
- Prior to this work no data transformation approach existed to convert the AFM cell force response data from a non-normal population to a normal population.
- The investigation of larger cell indentation depths of 1.75μm, greater than the 500 nm typically used in previous studies of AFM force probing on the live human cell types (PNT2, DU145, PC3).
- The cell types (PNT2, DU145, PC3) have been tested here in order to ascertain for the first time whether these cell populations respond differently to ionising radiation insult, that is whether, or not, the cell mechanics of prostate cancerous cells (PC3) and normal prostate tissue cells (PNT2) have differing responses when exposed to therapeutic radiation dosages. DU145 is an intermediate cell line being carcinoid, but having a low metastatic potential (less invasive to other tissue).

The thesis was deemed to be episodic in nature and its' structure reflects this, each of the chapters has a conclusion and discussion section and therefore there is not a separate discussion section at the end of the thesis.

1.1 Background

1.1.1 The Atomic Force Microscope (AFM)

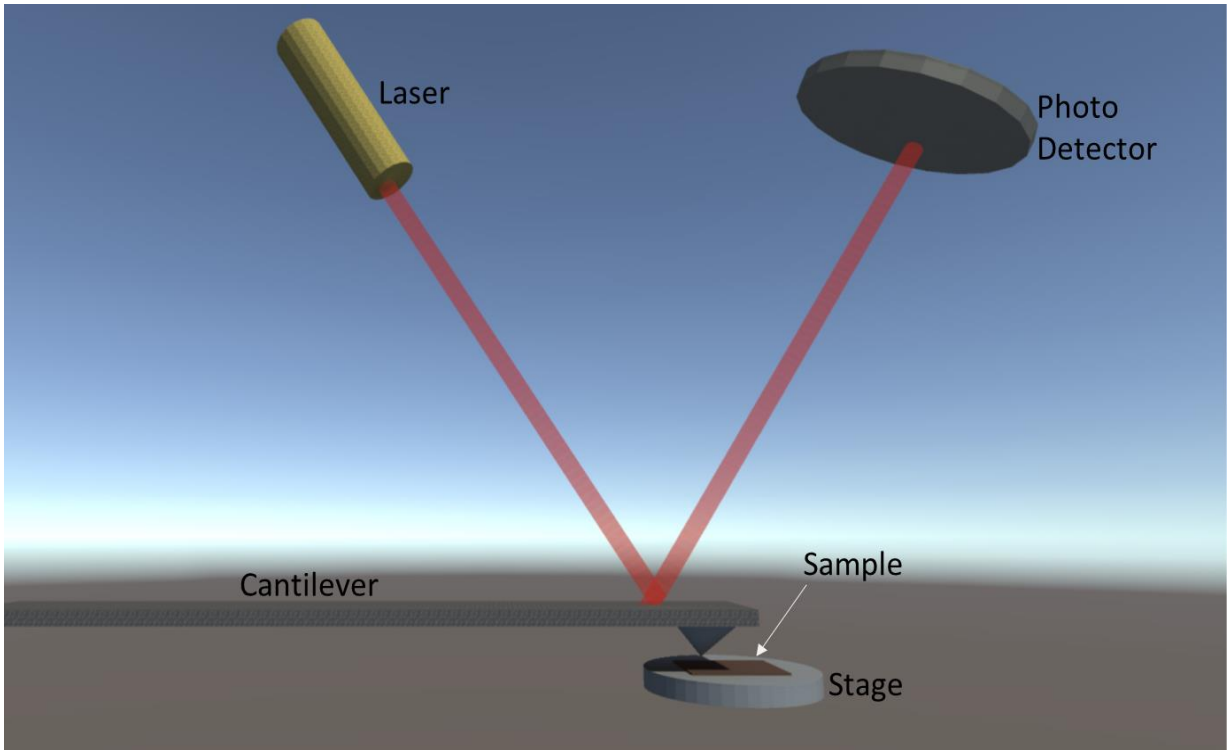


Figure 1-1. A schematic of an atomic force microscope; the laser generates a beam which bounces off the rear of the cantilever and onto a photodetector which measures how much deflection has occurred. The cantilever is connected to a z-piezo actuator which, when calibrated, enables a reading of the applied force to be obtained.

The Atomic Force Microscope (AFM) was first developed in 1986 by Binnig and Quate. The AFM was developed from the earlier scanning tunnelling microscope (STM) (Binnig and Rohrer, 1986). The AFM used in the experimentation performed as part of this research programme was an Asylum Research (Santa Barbara, USA), Molecular Force Probe-3D Microscope (MFP-3D). The AFM can operate in several different modes, the most common being contact mode, non-contact mode and an intermittent-contact, or tapping mode. The AFM can be used to capture information about just a single point, or for an area by taking readings multiple times, i.e. by scanning each point in a raster fashion.

1.1.1.1 Contact Mode

Contact mode operates by keeping the AFM tip in contact with the sample at a pre-defined set-point value, this set point is maintained by a z-piezo actuator via a feedback mechanism. Both height and deflection information is recorded in this mode and if scanning mode is engaged this builds up a picture of the 3D height of the sample, as measured by the z-piezo, and the cantilever deflection, as measured by the photodetector, in a grid-like raster scan pattern.

1.1.1.2 Tapping mode

Tapping mode works by the cantilever oscillating at its resonant frequency with a pre-set amplitude. The phase and height are measured during imaging of an area, the phase shift being the oscillating cantilever's difference in phase from the drive signal and the height measurement is again achieved via a feedback mechanism from the z-piezo that acts to maintain a constant set-point cantilever deflection.

1.1.1.3 Non-Contact Mode

Non-contact mode works by having the cantilever oscillating at its resonant frequency with a pre-set amplitude. The difference in the frequency from the resonant frequency is measured as the scan over an area progresses and used to build up a picture of the height of the sample being examined.

1.1.1.4 Force Measurements

In this thesis contact mode AFM was mainly used and it was used once only at a single point for any given cell. In this case, the force is measured using a cantilever (a OMCL-TR400PSA-1, Olympus, measuring 200μm in length) mechanically coupled to the z-piezo actuator, which has a small silicon nitride pyramidal tip attached to it, measuring only a 15 nm radius at the tip's point (the apex), with a tip height of approximately 2.9μm. This is contact mode AFM performed at a single point and therefore, as shown in Figure 1, the cantilever would be kept at a constant set-point by the z-piezo actuator, while the cantilever deflection would be measured by using the most common AFM cantilever position-measuring technology, an optical lever. Here a photodetector is used to assess how much the cantilever has deflected via a measurement of the deflected laser spot's position on the photodetector. This is combined with the force measured using a z-piezo actuator that was applied to the cantilever to produce a deflection verses force graph. A conversion using Hooke's Law, with prior knowledge of the spring constant of the cantilever, is then used to transform the deflection data to produce a force verses indentation graph (Radmacher , 2002).

One of the benefits of AFM is that it is possible to take measurements of cells living under physiological conditions with no chemical fixation process required (which would kill the cells). The substrate not only acts as a surface for the cells to anchor themselves upon, but also acts here as a physical restraint to stop the movement of the cells while they are being measured. Force verses indentation curves can also be acquired relatively quickly, typically at a rate of around 50 force curves per 30-40 minutes, depending upon the level of operator

skill and with the additional requirement to comply with conditions imposed by the subsequent automated Matlab processing step.

1.1.2 Cell mechanics.

The eukaryotic cytoskeleton is understood to be comprised of three main protein-based elements, namely microfilaments, intermediate filaments and microtubules. The primary roles of the cytoskeleton are structural support, movement and rapid signalling (Janmey, 1998).

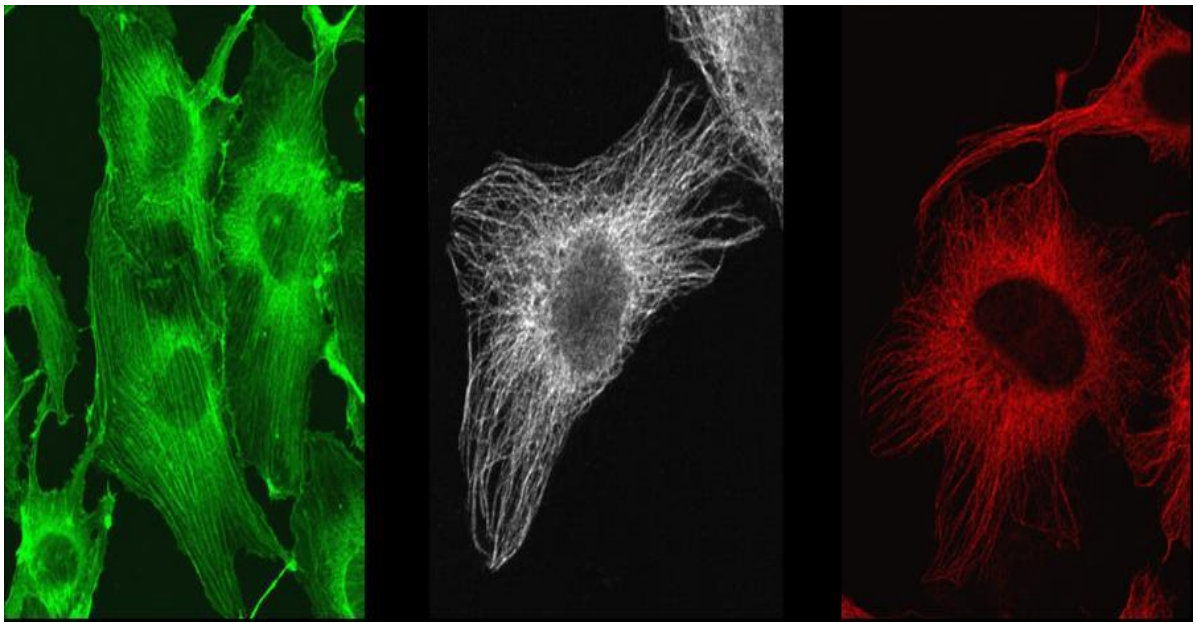


Figure 1-2. Fibroblast cells showing Microfilaments (left), Intermediate filaments (centre) and Microtubules (right), images from (Murphy M, 2010).

A typical cell undergoes continuous changes throughout its lifecycle. The state of a cell in terms of its position in the cell cycle influences both its short term response to external forces, but also the longer term internal and external responses. A population of cells typically exhibits a range of different phases in terms of their current stage in their cellular growth cycle, with different distributions of proteins, signal molecule concentrations, and other biochemical molecules (Janmey *et al*, 2007). Therefore, due to these differences that are present within any typical population of cells, the response of any given individual cell will differ within the population when all the cells are exposed to the same external insult, whether that is by AFM force application, or by ionising irradiation insult (Gabryś *et al.*,

2007; Mizuno *et al*., 2007; Vlashi *et al.*, 2009; Li, *et al.,* 2008). Temperature is another variable that must be considered, as differences in ambient temperature change the way that a cell responds to force (Sunyer, R. *et al.* 2009). Furthermore differences between any responses that are shown by pathological cells and non-pathological cells must therefore only be considered as being a property of a population of cells, not on an individual basis (Prasad, S. *et al*., 1998). It has been purported that the cytoskeletal state can influence the cell's behaviour and thereby alter force response to indentation (Heidemann *et al*.,1999 ; Olson. & Sahai, 2008); therefore examination of the cytoskeleton after a cell is irradiated to varying levels may give an insight into a cell's eventual pre-programmed cell death (apoptotic response).

Pathological cells and non-pathological cells exhibit differing responses to the same radiation dosage and to the same applied force (Suresh, 2007 ; Kumar *et al*., 2009; Woloschak *et al.*, 1991; Ahmed *et al*., 1997). A high level of apoptosis causes problems in the radiation treatment of cancer, particularly when normal cells are damaged. The inability to selectively cause apoptosis in cancer cells, while allowing most surrounding normal tissue to remain intact is dependent upon radiation dosage and the time between the applications of fractional radiation dosages (Jain, M. et al., 2007). The ability to be able to predict the response of a population of epithelial cells to repeated radiation dosage could allow for tailored treatments to be given to individual patients. One of the key differences between these cell states lies in the organisation of the cellular cytoskeleton (Hall, 2009).

Current biomechanical models of the cytoskeleton can be divided into two distinct approaches, namely continuum and discrete methods (Lim, C. *et al*., 2006). Which of the

two approaches is most applicable is dependent upon the level of detail that is required within the model.

1.1.2.1 Continuum Model.

A continuum model describes a model that can be either uniform, in which case the changes to the model are spread over an area, or compartmentalised, in which case the model has distinct areas of application. A continuum model is generally considered to be applicable when large macroscopic and passive responses, typically with an indentation greater than 1μm, are required (Mohammad *et al*., 2006). A continuum model therefore becomes a less viable solution when considering smaller indentations, such as those of a typical unmodified AFM tip, which usually range between 0.1 μm to 1.0 μm (Mohammad *et al.*, 2006). When a non-uniform continuum model is used, one in which compartmentalisation occurs, this allows for certain cell structures to be modelled at higher/lower densities according to their relative importance. However, as continuum models do not describe the individual cytoskeletal fibres, continuum models are still not suited to ultra-high resolution measurements (Mohammad *et al.*, 2006). The limited scope of this approach therefore makes this method impractical when investigating such high resolution nano-structural changes that are occurring within the cell.

1.1.2.2 Discrete model.

Discrete (non-continuum) models are often detailed and therefore are generally complex, and there are various types of such models that have been proposed for use in modelling cell mechanics. Each model typically describes one aspect of the cell quite well, for example, tensegrity models the actin cytoskeleton (Mohammad *et al.*, 2006[2]). Generally non-continuum models do not make the assumption that cell characteristics will apply to

the entire cell, for example in the case of the cytoskeletal force, which may be unevenly spread within the cell, the tensegrity model allows uneven force distribution to be modelled by multiple connected sub-tensegrity structures (Ingber, 2003b).

1.1.2.3 Tensegrity

Tensegrity, the name being derived from 'tensional integrity', is an approach to structural relationship that involves a balance of compression and tensional elements, as shown in Figure 3. Tensegrity does not evenly distribute force, and this model is one of the more robust approaches when applied in cell mechanics, principally as done by Ingber (Ingber, 1991; Ingber, 2003a; Ingber, 2003b; Ingber, 2008a) who pioneered this approach and with it has managed to help elucidate cell features before they were seen *in vivo* (Heidemann *et al*., 1999;Wang *et al*., 2001).

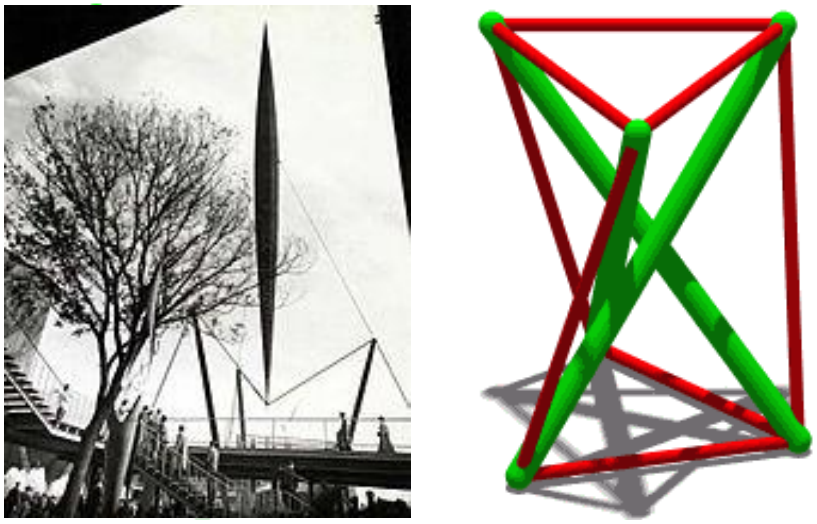


Figure 1-3. (Left) Skylon tower - a tensegrity structure, licensed under public domain via wikipedia commons https://en.wikipedia.org/wiki/File:In16695.jpg, (Right) Shows a simple tensegrity structure exhibiting tension and compression, with the tension shown in red and the compression in green, licensed under public domain via wikipedia commons https:// en.wikipedia.org /wiki/File: Tensegrity_simple_3_RL .png.

Mechanotransduction, the rapid transmission of a force to the cell nucleus, was shown to be observable *in vivo* (Wang *et al.*,2009; Maniotis, *et al*. 1997), but it was predicted first in the tensegrity model (Ingber, D., 1991). This pre-stressed model also incorporates a key feature that is observed experimentally in living human cells (Heidemann *et al.*, 1999), in that as more force is applied to the cell then more resistance to that force is encountered. While each of the discrete models may generally exceed the others in terms of explaining a certain single cell attribute (Mandadapu *et al.*, 2008 ; Hoffman *et al.*, 2009 ; Pullarkat *et al.*, 2007; Kroy *et al*., 2006 ; Brangwynne, *et al.* 2006; Schaus, *et al*. 2007), the tensegrity approach models many cell characteristics well, with the exception of the power law viscoelastic response of the cell, which has been attributed to much smaller elements within the cell, such as thermal fluctuations of individual actin filaments, but has not been linked to the structure as a whole (Mohammad *et al.*,2006[2]).

1.1.3 Ionizing Radiation Studies

Ionising radiation, such as x-rays, carries enough photonic energy to remove an electron from an atom and in practice can require shielding of lead, or another dense material, to absorb the energy.

One of the uses of ionizing radiation is for the treatment of cancer in the process of radiotherapy and this causes apoptosis (pre-programmed cell death) as a consequence of irreparable DNA damage for a targeted area. Typically there can be considered to be two main types of ionizing radiation, namely particulate types, such as alpha and beta particles and photon (electromagnetic) sources, such as gamma rays and X-rays. Alpha particles carry the least amount of energy and can be stopped by a piece of paper, beta particles can be stopped by relatively thin layers of aluminium/Perspex, whilst gamma/X-Rays can only be blocked by significant thicknesses of lead shielding. The level of radiation dosage is measured in Gray (Gy). One Gray is defined as being the absorption of 1 joule of ionizing radiation energy per 1 kilogram of matter. For the experimentation carried out in this research programme ionizing radiation was used in the form of a 30MBq, ^{90}Sr, beta (electron) radiation source (IBA Dosimetry GmbH, CDP Radioactive Check Device For Parallel Plate Ionization Chambers, Serial Number: P-1314/SM106). In the experimentation carried out as part of this research programme into the effects of therapeutic levels of ionizing radiation upon cell mechanics, four dosage levels were used 1 Gy, 2 Gy, 3 Gy and 10 Gy. The 1 Gy dose is equivalent to a single treatment dosage that is typically given in radiotherapy. The 10 Gy dose was used as a positive control and represents a dose that is well above the therapeutic level. The type of ionizing radiation used in the experiment was beta for safety and ease of access with care to avoid the

Bremsstrahlung effect which will create X-Ray radiation. The potential for adverse Bremsstrahlung effects was reduced by adding a layer of >1cm thick Perspex surrounding the source and to this end an experimental enclosure was created with a moveable sliding tray system for sample insertion, thereby avoiding directly the requirement for placing any part of the human hand near the radiation source aperture.

1.1.4 Cancer and cell mechanics.

The cell is a complex unit of life that is mechanically supported by a group of structures that are collectively termed the cytoskeleton. The name 'cytoskeleton' was first proposed by Koltsov in 1903 (Soyfer, 2001). The cytoskeleton is comprised of three main constituent parts in human and animal cells. These three components are actin microfilaments, microtubules and intermediate filaments (Holmes, 2009).

Many diseases are associated with cellular mechanical weaknesses, for example cancer. Structural weakness of the cytoskeleton is often seen in cancer development and this is particularly true in the case of high metastatic potential cells, that is cells which are highly invasive to other tissues and areas throughout the body.

Various researchers have conducted studies into the physical response of cancer cells. Early work was performed by Lekka et al, (1999), where they studied two normal, and three cancerous types of bladder cells. In their study they showed that the cancerous cells had an elastic modulus that was about one order of magnitude lower than that of normal cells, i.e. the normal cells were stiffer and therefore less malleable than the cancerous cells. They suggested that the reason for this may be due to the organisation of the actin cytoskeleton.

Later in a paper by Li et al, (2008), benign (normal) cells and cancerous human breast cells were studied. They found that the cancerous breast cells had a significantly lower Young's (elastic) modulus (i.e. they were less stiff) and that the level of organisation (quantified from confocal images) of actin filaments in cancerous cells was reduced. This increase in actin disorganisation could be inferred to correspond to a reduction in the mechanical strength of the cytoskeleton, leading to cancerous cells being more easily deformed than normal cells. This could in turn explain why cancerous cells are able to migrate more easily than normal cells by being able to pass through smaller spaces.

More recently in a paper by Kim et al, (2013), they showed that liver carcinoma (cancerous) cells and normal cells had a difference in terms of their response to probing by an Atomic Force Microscope (AFM). The normal cells had a higher elastic modulus (i.e. they were stiffer) when compared to the carcinoma cells. They also noted that there was a larger difference between the results obtained when using an AFM probe with a conical tip than there was when compared to results from a spherical tip. They also investigated the structure of the cytoskeleton and found that the carcinoma cells have more actin stress fibres than are found in normal cells. They concluded that changes in actin were the primary reason for the differences in force response, between carcinoma cells which were more malleable and normal cells that were mechanically stiffer. This was tested by chemical disruption of the microtubules and actin filaments which form part of the cytoskeleton. Intermediate Filaments (IFs) were not tested in that study, as the chemical disruption of actin indicated that IFs were not primarily responsible for the changes in force response that were seen within the cells.

In a paper by Gabryś et al, (2007), they explored the effects of radiation upon the cytoskeleton of endothelial cells and concluded that radiation insult enhanced cell migration and produced additional actin stress fibres, whereas the microtubules and intermediate filaments were not visibly altered.

Since the cancer cells become mechanically more malleable when exposed to radiation, and this is believed in part to be caused by the production of actin stress fibres, there is a clear link between cancer, cell mechanics and radiation insult.

1.2 Aims and Objectives

The aim of this research programme is to automate, where possible, the analysis of force response data from an AFM, by providing suitable software tools that may be readily and easily used by the life sciences community. This over-arching aim may be achieved by meeting various objectives which are listed below.

- Development of new automation strategies for automated data collection and automatic analysis tools for the examination of AFM data. The production of a suite of software tools to aid in the capturing and processing of AFM data such that the majority of the work is automated.
- Verification as to whether or not there is a statistically significant difference between PNT2 and PC3 cell lines in response to force probing with the AFM tip after radiation exposure and whether this difference is significant enough to proceed with further investigations.
- Developing tools to automatically analyse the orientation of F-actin and to extract structural information including angles between actin filaments. To achieve this purified actin polymerised in vitro was used as a model.
- A study accounting for the effects of the time taken to measure the force response. The measurement time taken during the experimentation to measure the AFM force responses could have an influence on the cells' response, i.e. a cell measured after an hour may not respond in the same way that a cell does that has been measured within 5 minutes of radiation exposure.
- Exploration of alternate formulae for modelling the AFM force-distance data. The standard Hertz formula may not be the best option when examining the data when

indenting the cells. Using the Akaike Information Criterion (AIC) to ascertain which model function from an empirical selection, scores highest and therefore which is the best model fit for those models that were under test.

- Transformation of the AFM data to give normal distributions. Attempting to convert the data to a normal distribution, using a mathematical method in order to enable the biologists to be able to apply standard population tests and standard deviations.

1.3 Thesis Structure

The structure of the thesis is broken into seven chapters.

Chapter 1: Introduction.

This introduces at a very top level the concepts behind the cytoskeleton, and gives a brief background to the thesis, describes the aims and objectives of the research programme and the structure of the thesis.

Chapter 2: Automated force curve analysis.

This chapter determines whether differences exist between the cell lines that were under test in terms of their respective response to force probing when using atomic force microscopy. It also describes the way in which the atomic force microscopy results were processed automatically for each of the various types of cells, i.e. both human and yeast cells in various states.

Chapter 3: Automatic contact point extraction methods.

This chapter describes the different methods used for the extraction of the contact point in AFM force-distance curves and the problems encountered when attempting to automate this analysis process without human supervision.

Chapter 4: A new model for investigating the biomechanical responses for human cells.

This chapter explores whether the commonly used standard Hertz model provides a reasonable fit for the data using the Akaike Information Criterion (AIC) approach. AIC is

a numerical method that tests the fits for the data and formulae given and produces a comparative numerical number.

Chapter 5: Normalisation transformation.

This chapter explains how to transform the non-normal data from the AFM to a normal distribution. This is an attempt to simplify the process of testing between populations of mechanical responses to atomic force microscope force probing.

Chapter 6: F-actin filament orientation.

This chapter summarises the work performed on F-Actin orientation and determines if any useful data from the height images obtained from the AFM can be ascertained.

Chapter 7: Conclusions.

This chapter summarises the work that has been carried out and presents conclusions based upon the work performed.

Chapter 8 : Further Work.

This chapter presents possible additional work that could be carried out to extend this research further.

1.4 References

Ahmed, M.M. et al. (1997). Ionizing Radiation-inducible Apoptosis in the Absence of p53 Linked to Transcription Factor EGR-1. *Journal of Biological Chemistry*. **272** (52): 33056-33061

Binnig, G. & Quate, C.F. (1986). Atomic force microscope. *Phys. Rev. Lett.* **56** (9): 930-933.

Binnig, G. & Rohrer, H. (1986). "Scanning tunneling microscopy". *IBM Journal of Research and Development* **30** (4): 355–69

Brangwynne, C.P. et al. (2006). Microtubules can bear enhanced compressive loads in living cells because of lateral reinforcement. *Journal of Cell Biology*. **173** (5): 733

Gabryś, D. et al. (2007). Radiation effects on the cytoskeleton of endothelial cells and endothelial monolayer permeability. *International Journal of Radiation Oncology, Biology, Physics*. **69**(5): 1553–1562

Hall, A. (2009). The cytoskeleton and cancer. *Cancer and Metastasis Reviews*. **28** (1): 5-14

Heidemann, S.R. et al. (1999). Direct Observations of the Mechanical Behaviors of the Cytoskeleton in Living Fibroblasts. *The Journal of Cell Biology*. **145** (1): 109-122.

Hoffman, B.D. & Crocker, J.C. (2009). Cell Mechanics: Dissecting the Physical Responses of Cells to Force. *Annu. Rev. Biomed. Eng.***11** (1): 259-288

Holmes, Kenneth C. (2009). Structural biology: Actin in a twist, *Nature*, **457**: 389-390

Ingber, D. (1991). Integrins as mechanochemical transducers. *Current Opinion in Cell Biology*. **3** (5): 841-848

Ingber, D.E. (2003a). Tensegrity I. Cell structure and hierarchical systems biology. *Journal of Cell Science*. **116** (7): 1157

Ingber, D.E. (2003b). Tensegrity II. How structural networks influence cellular information processing networks. *J. Cell. Sci*. **116** (Pt 8): 1397-1408

Ingber, D.E. (2008a). Tensegrity-based mechanosensing from macro to micro. *Progress in biophysics and molecular biology*. **97** (2-3): 163–179

Ingber, D.E. (2008b). Tensegrity and mechanotransduction. *Journal of Bodywork & Movement Therapies*. **12** (3): 198–200

Jain, M. et al. (2007). Optimization of radioimmunotherapy of solid tumors: biological impediments and their modulation. *Clinical Cancer Research*. **13** (5): 1374

Janmey, P.A., (1998). The cytoskeleton and cell signaling: component localization and mechanical coupling. *Physiological Reviews*, **78** (3), 763.

Janmey, P.A. & McCulloch, C.A. (2007). Cell mechanics: integrating cell responses to mechanical stimuli. *Annu Rev Biomed Eng*. **9:**1-34

Kim Y, Hong JW, Kim J, Shin JH. 2013 Comparative study on the differential mechanical properties of human liver cancer and normal cells. *Anim. Cells Syst*. **17**: 170–178

Kroy, K. (2006). Elasticity, dynamics and relaxation in biopolymer networks. *Current Opinion in Colloid & Interface Science*. **11** (1): 56–64

Kumar, S. & Weaver, V. (2009). Mechanics, malignancy, and metastasis: The force journey of a tumor cell. *Cancer and Metastasis Reviews*. **28** (1): 113-127

Lekka, M. et al, (1999) Elasticity of normal and cancerous human bladder cells studied by scanning force microscopy. *European Biophysics Journal*. **28** (4): 312-316

Li, Q.S. et al. (2008). AFM indentation study of breast cancer cells. *Biochemical and Biophysical Research Communications*. **374** (4): 609–613

Lim, C. et al. (2006). Mechanical models for living cells--a review. *Journal of Biomechanics*. **39** (2): 195-216

Pullarkat, P.A. et al. (2007). Rheological properties of the Eukaryotic cell cytoskeleton. *Physics Reports*. **449** (1-3): 29-53

Mandadapu, K.K. et al. (2008). On the cytoskeleton and soft glassy rheology. *Journal of biomechanics*. **41** (7): 1467–1478

Maniotis, A.J. et al. (1997). Demonstration of Mechanical Connections between Integrins, Cytoskeletal Filaments, and Nucleoplasm that Stabilize Nuclear Structure. *Proceedings of the National Academy of Sciences of the United States of America*. **94** (3): 849-854

Mizuno, D. et al. (2007). Nonequilibrium Mechanics of Active Cytoskeletal Networks. *Science*. **315** (5810): 370-373

Mohammad R. K. Mofrad, Roger D. Kamm.(2006a) Cytoskeletal mechanics: models and measurements. pp 71-83.

Mohammad R. K. Mofrad, Roger D. Kamm.(2006b) Cytoskeletal mechanics: models and measurements. pp103-128.

Murphy M, www.ljmu.ac.uk/BML/BML_Images/Picture17.jpg., Accessed 29/4/10.

Murphy M. (2007). Investigating the Mechanical & Structural Properties of Human Cells using Atomic Force & Confocal Microscopy, Phd, LJMU, Liverpool

Olson, M.F. & Sahai, E. (2008). The actin cytoskeleton in cancer cell motility. *Clin Exp Metastasis*. **26** (4): 273-287

Radmacher, M. (2002). Measuring the elastic properties of living cells by the atomic force microscope. *Meth. Cell Biol.* **68:** 67-90.

Ricci, R. et al. (2009). Biomodulation with low-level laser radiation induces changes in endothelial cell actin filaments and cytoskeletal organization. *Journal of Photochemistry & Photobiology, B: Biology*. **95** (1): 6–8

Schaus, T.E. et al. (2007). Self-organization of actin filament orientation in the dendritic-nucleation/array-treadmilling model. *Proceedings of the National Academy of Sciences*. **104** (17): 7086

Soyfer, Valery N. (2001) The consequences of political dictatorship for Russian science. *Nature Reviews Genetics* **2**: 723-729

Suresh, S. (2007). Biomechanics and biophysics of cancer cells. *Acta Biomater*. **3** (4): 438, 413

Sunyer, R. et al. (2009). The temperature dependence of cell mechanics measured by atomic force microscopy. *Physical Biology*. **6** (2): 025009

Wang, N. et al. (2001). Mechanical Behavior in Living Cells Consistent with the Tensegrity Model. *Proceedings of the National Academy of Sciences of the United States of America*. **98** (14): 7765-7770

Wang, N. et al. (2009). Mechanotransduction at a distance: mechanically coupling the extracellular matrix with the nucleus. *Nat Rev Mol Cell Biol*. **10** (1): 75-82

Woloschak, G. & Chang-Liu, C. (1991). Expression of Cytoskeletal Elements in Proliferating Cells Following Radiation Exposure. *Int J Radiat Biol*. **59** (5): 1173-1183

Chapter 2 : Force curve analysis.

2.1 Measurement Protocol.

In order to use the AFM the first calibration requirement was to use thermal noise method to calculate the spring constant for that particular cantilever which is saved in the IBW file. The cells are grown to confluence from seeding at a density of $1x10^5$ cells/cm^2 on glass coverslips and left until a uniform monolayer formed. The control and non-control cells undergo the exact same treatment being placed and handled exactly the same with the exception of not being exposed to radiation. Once the coverslip is placed onto the AFM the cells are ready to measured at room temperature which was 25 degrees.

In order to measure each cell the AFM must be positioned correctly using an Olympus IX50 inverted optical (IO) microscope to adjust the horizontal plane position of the AFM tip to the centre of the cell being probed and the vertical indentation level must be set correctly or the result may not be suitable for automation.

For a generic AFM force curve, to be considered to be valid for automatic processing it must pass three tests.

1) It must contain a starting approach region (the green section of curve in Figure 2.1) where the curve is flat, and this is necessary to ensure that the starting point for the curve is present within the measured range. Although this can be smaller than 0.5μm for most functions the energy transformation function requires at least 0.5μm in order to successfully function correctly.
2) The red portion of the curve that represents the force response that we are interested in must exceed a total indentation depth of 1.75μm. This is the minimum indentation range that we wish to measure. The average cell height was about 3μm so a minimum indentation depth of 1.75μm was chosen to make sure that we were significantly indenting the cells, while reducing the chance of the stiffness of the underlying hard substrate from affecting the results.

3) It should contain no large-scale noise levels, that are typically caused by external vibrations.

If any of the three tests described above fail, the IBW force curve must be ignored.

The AFM tip vertical position is adjusted by changing the initial z-piezo position. Practically, a force curve is measured by starting a force-curve approach-retract scan from a specific start position and with a pre-set scan length. This often results in a resulting force curve that doesn't pass the three tests described above. When this situation occurs the invalid curve is ignored by the user, although the data files are still gathered by the system, and the z-piezo position is adjusted by the user. This process is repeated until the curve passes all three tests, however all of the files are saved automatically including those that do not pass the tests and therefore the failed curves must be either deleted, or separated from the good data. The curves that pass the three tests above are placed into a directory for automated processing in order to characterise the cell's stiffness parameters from each force-indentation curve by application of the extended Hertzian model, as described below in Section 2.3.

2.2 Analysis of initial AFM force curve data to determine whether significant differences exist in force response between experimental cell lines.

Figure 2.1 shows a typical AFM force curve. It shows raw force in nano-Newtons on the vertical axis plotted against raw indentation in microns on the horizontal axis. The raw data is transformed by using the estimated contact point as the zero, zero point on the graph in

terms of both force and indentation. This would mean that the graph would be adjusted to reflect the new origin (0,0) point.

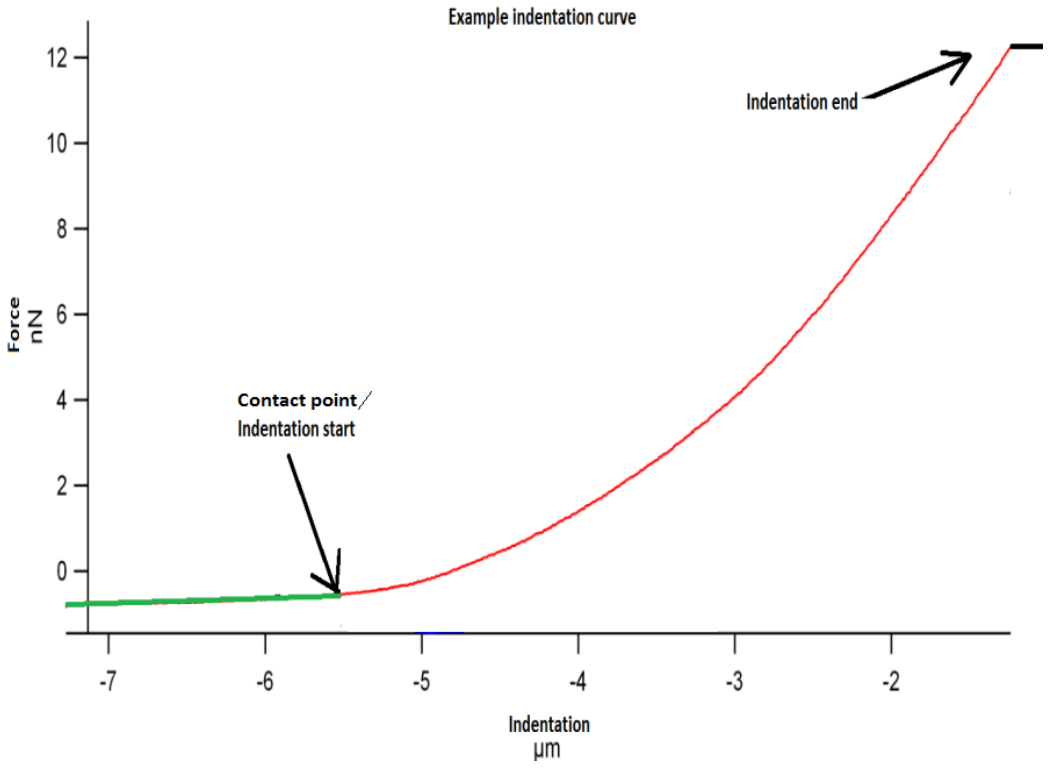


Figure 2-1 An example of a typical AFM force curve showing raw force verses raw indentation.

The result of the transformation to the origin (0,0) is that only positive indentation, along with positive force, would be shown on the new transformed graph, with the result that the superfluous green section of the graph would be clipped and removed. The negative numbers that shown on the indentation axis in Figure 2.1 therefore reflect the initial raw state of the graph before the transformation process to the contact point occurs.

The curve shown in Figure 2.1 has three distinct regions, the flat green starting region (representing the approach of the AFM probe towards the cell's surface which is clipped after the origin/contact point is found), a curved region (shown in red) which is the force

response that we wish to measure where the AFM probe first contacts and then indents the cell. Optionally at the end of the indentation another flat region can sometimes occur, which is caused by the attempted AFM scan depth being greater than the actual depth of the cell being measured and the AFM probe therefore encountering the hard underlying substrate. These flat regions can successfully be automatically processed (and indeed are!), but they are avoided in the current manual protocol by adjusting the positioning of the AFM's starting indentation point.

The rationale behind the initial data analysis of AFM force curves was to determine if statistically significant differences in mean force response existed between three distinct populations of living human prostate cells in different disease states, namely highly metastatic prostate cancer cells (PC3), prostate cancer cells having a low metastatic potential (DU145) and normal prostate cells having no metastatic potential (PNT2). It has been previously shown that breast cancer cells exhibit differences in the distribution of actin within the cell compared to normal cells, causing a significant difference in mean force response (Cross et al, 2008). It was hoped that analysis of the three prostate cell lines would be statistically different in terms of the force response measured from the AFM, otherwise the experimental work based on this hypothesis would have to stop here, as there would be no differences in mean AFM force response to radiation insult for the cells lines under test. In order to collect the data the following protocol was followed.

Stage 1 : Generation of the data.

Stage 2 : Copying the data from the AFM to the correctly named directory.

Stage 3 : Automatic removal of extra files captured during the generation stage.

Stage 4 : Extraction of the data from the IBW file.

Stage 5 : Finding the AFM contact point, median of force and determining statistical information.

2.2.1 Verification that the AFM measurement of force is not affected by measurement time.

The time taken to measure the AFM force response of a population of 50 cells was around 30-40 minutes. During this time the cells are removed from incubation at (37°C) and are measured under the AFM on pre-warmed medium (Dulbecco's modified Eagles Medium) (DMEM) at ambient temperature at approximately 35°C. They could therefore respond differently depending upon the duration of the time period over which they are being measured. In order to verify if this was a significant effect, a test was designed to measure the response of random cells grown on a monolayer over an hour long period, which was chosen to be longer than the duration of any practical AFM experimentation involving force curve generation. Shown below are the results of those measurements and it can clearly be determined from Figure 2.2 below that the measurement time plays only a minor role here as a factor having a 1-3% positive variation in force response over the hour period. The variation in force response for time was calculated by drawing a ‘best fit’ trend line for all the points for each type of cell, the PC3 had a 1% positive slope overall and the PNT2 cell line had a 3% overall positive slope. Therefore measurement time is not a primary factor in the measured cellular force response over an hour.

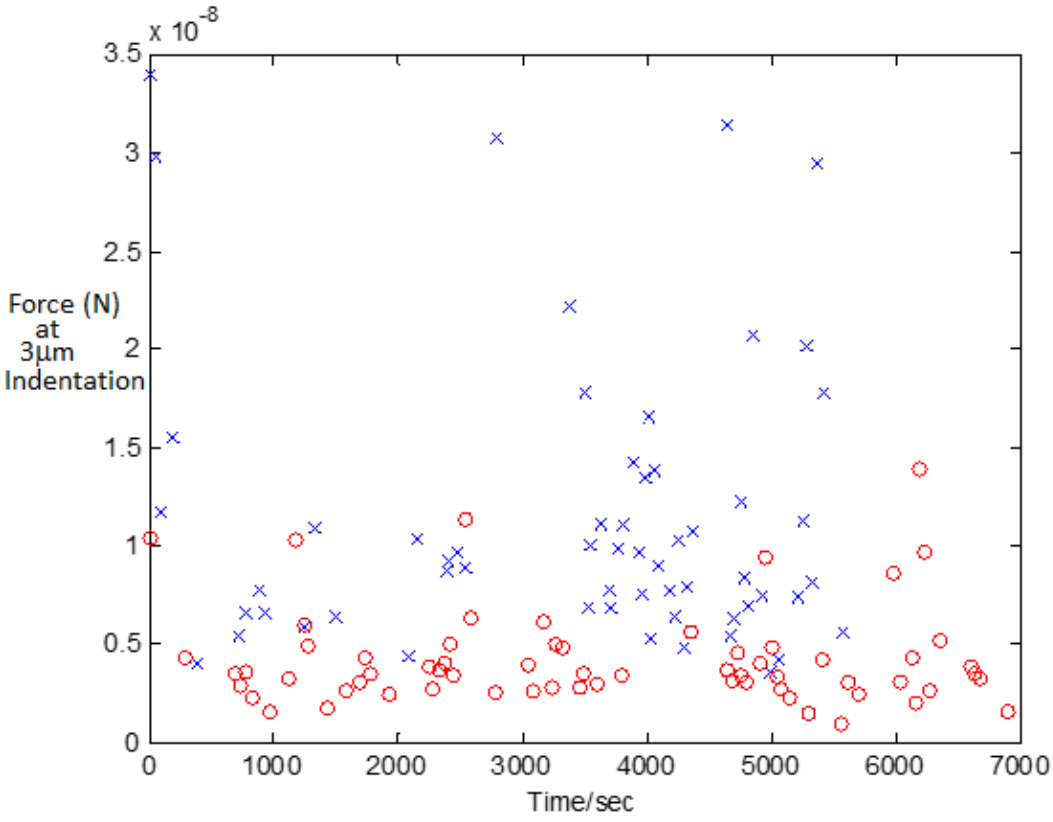
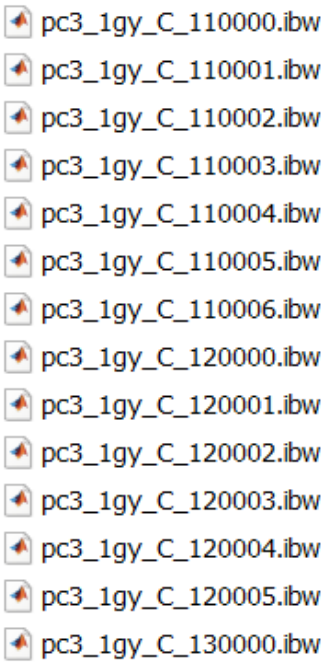


Figure 2-2 The AFM force response over an hour long period. This shows no difference for the response to the time over which the measurement occurs. The prostate cancer cells (PC3) are denoted by red circles and the prostate normal tissue cells (PNT2) are denoted by blue crosses. Both were grown on a poly-l-lysine substrate.

2.2.2 Generation of the AFM Force data.

The process to capture the AFM force data is manual and the process involves empirically determining how far the AFM tip is raised above the sample, which typically takes multiple attempts before the AFM user achieves a satisfactory result. This empirical process therefore generates multiple files, some of which are superfluous and for the example given below the following files are not needed (110000-110005, 120000-120004); therefore, the first task is the removal of these superfluous files. The files that are generated have a sequential format that is predictable and therefore the extra files can be purged before processing begins.



Sequential numbering.
Three different valid cell records are within this list -:
110006, 120005 and 130000. The other files are generated by the empirical nature of searching for the correct measurement start point/scan length, dependent upon operator skill and the actual cell height and as such do not represent valid data.
The length of the string data type storing the experimental filename varies per experiment.
The number of numerical digits varies at the end of the string.

Figure 2-3 A sub-sample from one of the AFM force data directories showing the sequential nature of the numbering format.

The Matlab code employed to process the IBW force curves parses the list that is shown in Figure 2-3 and extracts the highest numbered file for each block of 100 consecutive units, so in Figure 2-3 this would be 110006,120005 and 130000. The code relies on converting

the numeric portion at the end of the string into a number and testing this value against any number within a range of 100 units, then selecting the highest number and rejecting the rest. The 100 unit block was chosen because measuring 50 discrete cells (one force curve per cell) takes around 30-40 minutes, therefore scaling up to one hour would lead to a maximum number of 100 cells (force curves) being measured during an hour long period. The automatic Matlab processing code relies upon the fact that the end of the IBW filenames is numeric in form and that the files are created in a consecutive numerical sequence. The automatic numbering within the AFM machine defaults back to 1 when the AFM PC is rebooted. The maximum numerical difference of 100 in terms of the individual measurement file names was used as a divider between different valid readings as it was deemed to be very unlikely that 100 individual files (measurement attempts) would be required to attain a single correct force-distance curve reading. The IBW files that are rejected are moved to a sub-folder to remove them from being included in subsequent processing.

2.2.3 Copying the data from the AFM to the correctly named directory.

The number of individual AFM records acquired during this research programme numbers in the thousands in terms of different saved datasets and therefore a database was created by the author to hold all of the information generated by the AFM. In addition to these files is extra information held within the directory structure wherein each of the data files is located. The extra information contains experimental details such as experiment number, cell type, radiation exposure dosage and whether the data is control data or non-control data. Control data undergoes the same experimental processes as the non-control data with the exception of actually undergoing radiation exposure and is used to verify that the data

changes are solely caused by the effects of radiation exposure. This is covered in more depth in Section 2.2.2. Initially the extra information was input manually in the SQL (Set Query Language) database. However, as this was a time-consuming task, the manual method was subsequently replaced by an automated process designed by the author, involving ordering the directory structure and using the directory headings themselves as fields for the database.

2.2.4 Extraction of the data from the IBW file and placing it into the database.

The data from the MFP-3D, Molecular Force Probe-3D AFM (Asylum Research, Santa Barbara, CA) is output as a proprietary format IGOR Binary Wave (IBW) file. This is specified by the technical notes TN0003.zip from Asylum (Wavemetrics, 2004), and the initial code to read the raw files was available on 'Matlab Central' and was written by R.Naud in 2008 (See Appendix A1\IBW_Read.m). This code was used as a baseline in this research.

The baseline code by R.Naud was altered to read and parse the text contents within the IBW file in addition to the AFM force curve data. Primarily this was required to extract the spring constant value from the IBW file, as the IBW file does not in fact store the force data. In order to automate the process of force curve extraction the spring constant is required as the IBW force curve file stores the cantilever deflection data and this must be multiplied by the spring constant in order to convert the deflection data to the force data via Hooke's Law. Since the project only used IBW version 5.0 data, it should be noted that only that section of the code was altered to read the data, which means the text reading

code doesn't work for versions prior to version 5.0. The force data for a particular IBW file was then verified to be an identical match to that which was grabbed from the AFM machine interface and placed directly into an excel spread sheet and tested by visual inspection. The matching data verified that the process of extracting the force data was working successfully. At this stage the IBW file can be read along with the spring constant.

2.2.5 Defining the initial AFM contact point, mean force and determining statistical information.

The ability to read IBW files along with the spring constant meant that the only other parameter required to be determined was the initial AFM probe contact point before calculations could be made involving the force response for the cells. The contact point was initially calculated using a tangent to the smoothed force curve as described below. This was a quick method to extract a rough starting point that would be reliable enough to perform tests on the data and could be improved upon later in terms of its accuracy.

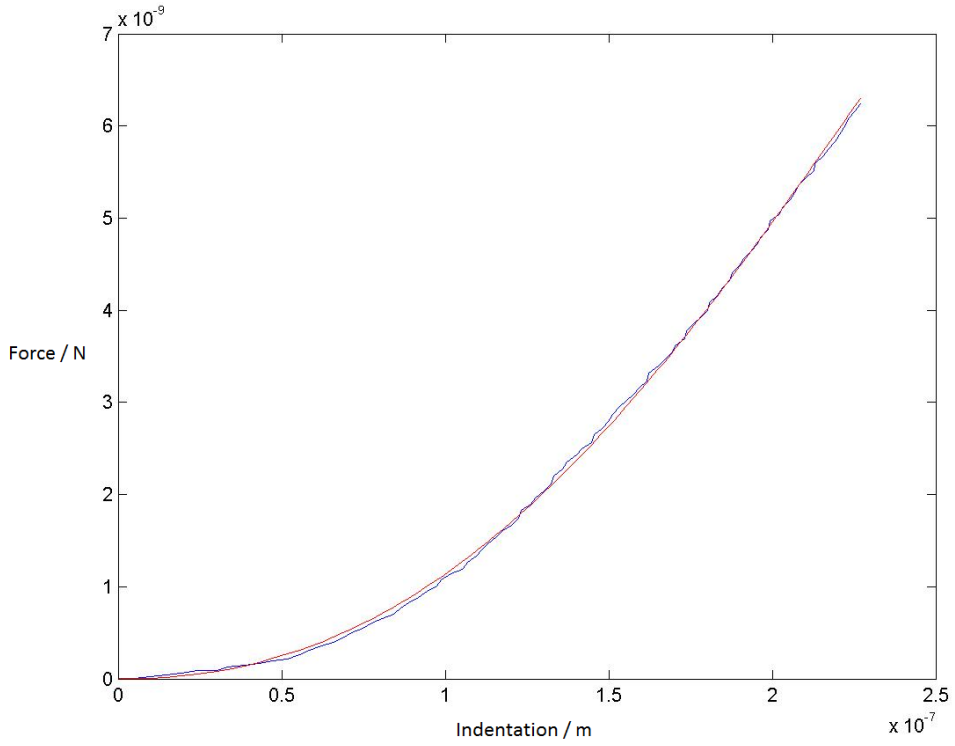


Figure 2-4. Showing the original AFM force curve data plotted in blue and the smoothed data curve plotted in red.

The initial contact point location method takes the original data and first performs a standard Matlab function *smooth* moving average filter, with a span of 51 from a dataset containing approximately 1000 entries. (Later replaced with a 9th degree polynomial fit function to bypass the need for a span number for differing span sizes ranging from 1000 to 50000 individual data points), as shown in Figure 2.4. In the case for the early work the moving average calculates a new point by averaging 51 points, so for the 1st point this would be 1-51st point, the next 2-52nd and so on, this could lead to slight shift to the right for the contact point. It then calculates the curve's gradient and sets the estimated initial contact point to be the position on the x (or 'indentation') axis where the tangent of the smoothed red curve has increased to an angle of 5 degrees (w.r.t. the x axis, when reading

from left to right along the x axis). This is shown in Figure 2.5. This angle was deemed to be sufficient such that the contact point would not be triggered by noise, or outside vibrations. The 5 degree slope angle contact point is determined by calculating the slope for all points on the smoothed or polyfit curve, then working from left to right advancing each point until 5 degrees is matched or exceeded.

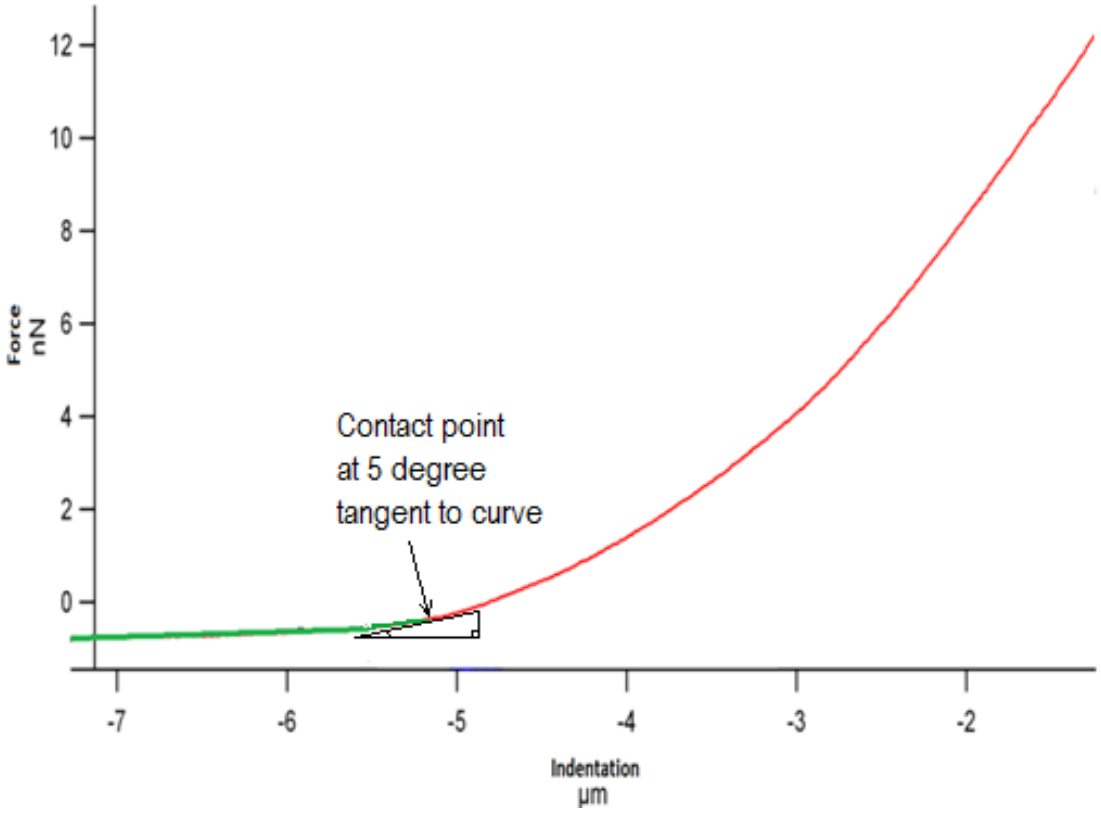


Figure 2-5, Shows an example tangent taken to the smoothed red curve, the contact point is estimated to be the position on the x axis where the tangent reaches a value of 5 degrees; this avoids problems with sharp changes to the gradient that occur in the raw data.

Once this estimated contact point is found, the total length of the curve is recorded from the estimated contact point to the end of the curve. Since the total length of the curve is recorded the end point is therefore known. This end point is used to ensure that any extrapolation of the force curve (trying to work from estimated data that is outside the range

of real data that has actually been recorded by the AFM) does not occur, as extrapolation is subject to large errors and is therefore undesirable.

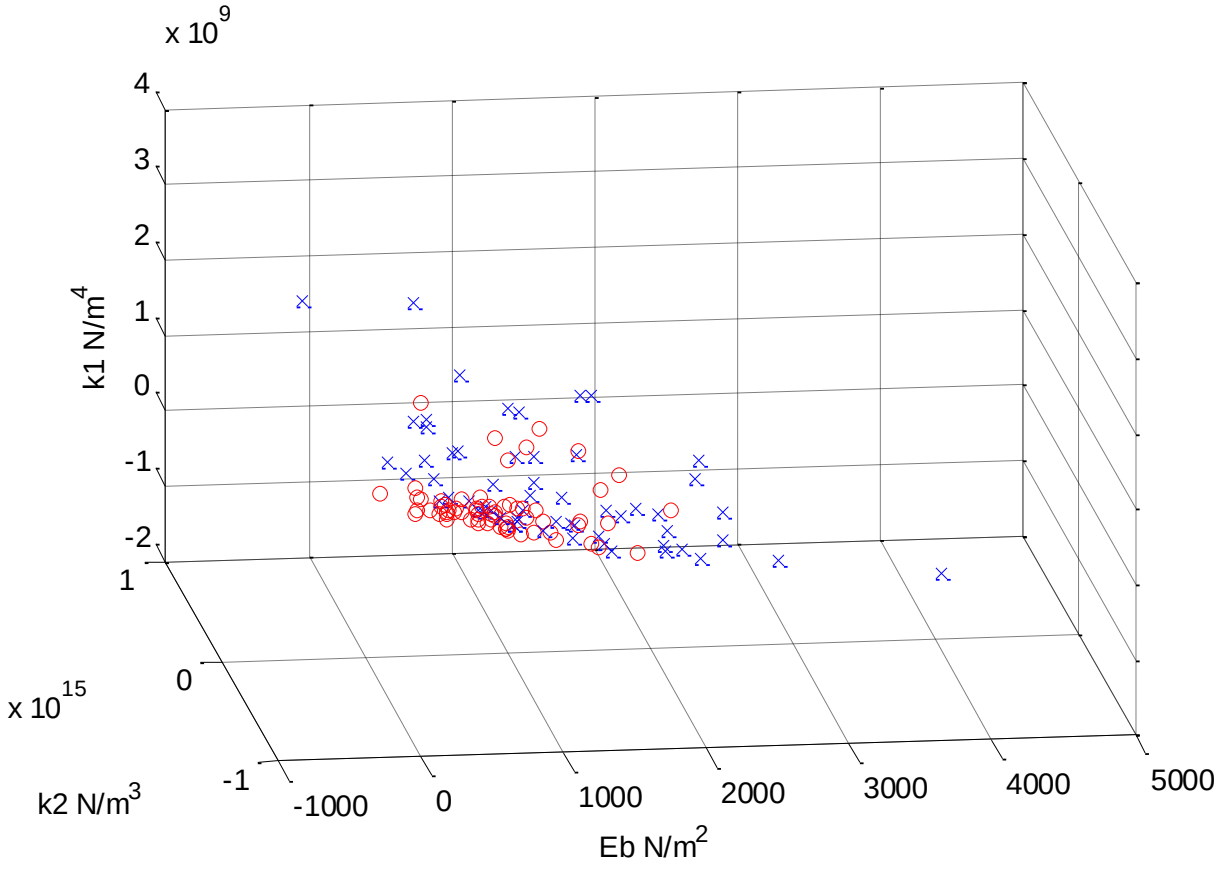


Figure 2-6 This shows the stiffness of 41 prostate cancer cells (PC3) denoted as red circles and 45 of prostate normal tissue cells (PNT2) denoted as blue crosses. Here the maximum value for cell indentation was 3 μm, the cellular substrate was poly-l-lysine. The cellular stiffness is expressed here as being a function of three parameters Eb, k1 and k2 in a new and better fitting non-linear Hertzian model that was developed at LJMU.

Once the contact point had been attained, albeit roughly at this time, the next stage in the automated process is to generate AFM force response curves and determine the stiffness parameters using the extended Hertz model. The Hertz formula for a conical indenter is given by,

$$F = \frac{2}{\pi}\frac{E}{(1-\nu^2)} tan\alpha\ \delta^2 \quad (2,1)$$

The extended Hertzian model extends the traditional Hertz model by allowing E to vary non-linearly by adding an x^2 and x component. Therefore letting E vary as a quadratic gives: $E(\delta) = k_1.\delta^2 + k_2.\delta + E_b$ (2,2)

$$F = \frac{2}{\pi}\frac{E_b^2+k_2 d^3+k_1 d^4}{(1-\nu^2)} tan\alpha \quad (2,3)$$

These stiffness parameters are shown in Figure 2.6. The mean value for the force at 3.00μm of indentation from the assumed contact point was taken to be the range at which the AFM probe was deemed to be sufficiently deep for a non-linear force response to be obtained. By using this value there were only a few rejections based on the total valid curve length being less than 3.00μm from the contact point. The distributions of the two different cell populations are not deemed to be normal using a Lilliefors statistical test at a probability of 0.05 (the *lillietest* test within Matlab 2010b). The two-sample Kolmogorov-Smirnov test (*kstest2* in Matlab) showed significant statistical differences between the prostate cancer cells and prostate normal tissue cells and this can also be seen visually from the wider blue cross clustering pattern evident in Figure 2.6.

2.2.6 Section conclusion.

The Kolmogorov-Smirnov statistical test confirmed that there was a statistically significant difference between the stiffness for populations of PNT2 and PC3 cell lines in response to force probing with the AFM tip. Experimental measurement time was eliminated as being a possible influencing factor on force response and a test was carried out that the data generated automatically was the same as that generated by the AFM machine manually,

this was verified against data output visually on screen and served as a check to ensure that the automatic data matched the manual output from the AFM. This PNT2 verses PC3 cell population result was expected as this matches similar work performed previously by other researchers (Gabryś et al, 2007; Li et al, 2008). However this knowledge is novel work as the cell lines here are from prostate, not breast, cells and the radiation dosages employed are different.

2.3 Biological AFM Data (processing and storage)

The biological data for all cell lines is generated from the AFM and produces data in the format of Igor Binary Wave (IBW) files. The IBW files contain a single curve that is typically of the form that is shown in Figure 2.5. Prior to the work that has been carried out by the author, IBW force curves had to be pre-processed manually in order to identify the extent of the curve that represented the force response of the cell to the indenting probe. This was achieved by eliminating data that was gathered prior to the AFM probe's initial contact with the cell surface, as well as any superfluous data that is gathered after the full indentation range of interest is achieved. This was a very slow and subjective process that typically needed to be carried out for many dozens, even hundreds of datasets. At this stage, any data that was obviously erroneous, caused by experimental error, environmental disturbances, or uncontrolled biological factors (such as impending cell death) was filtered out. The author aimed to automate this process, hence speeding up the procedure and removing subjectivity from the pre-processing stage.

2.4 The Hertz and extended Hertz model.

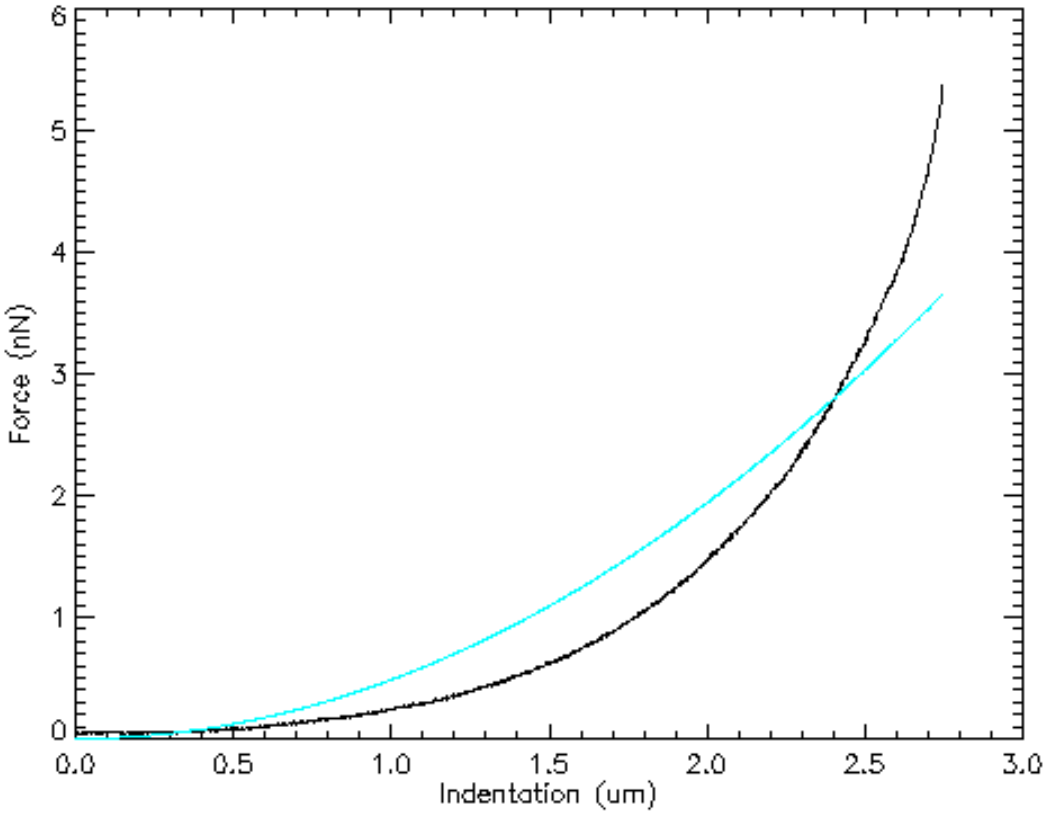


Figure 2-7. Force-indentation curve taken on the surface of a living LL24 cell growing on glass under physiological conditions and showing experimental data (plotted in black) and fitted data using the Hertz model (plotted in blue). (Murphy, 2007).

The Hertz model is a model that describes the normal contact between two spheres. The Hertz model requires the assumption of many different things (Dintwa et al, 2008).

- The material is linearly elastic. This is untrue as can be seen in Figure 2-7 where the curves are far from being straight and linear. The cells behave in a non-linear fashion and display strain hardening.

- The material is isotropic and homogenous. In the case of cells this is obviously not true. As they have different internal structures inside them and therefore fail the homogeneity test.
- The loads applied are static. If the cell dynamically reacts to the applied pressure (and there is some evidence to support this behaviour (Murphy,2007)) then this could be untrue.
- The object undergoing compression behaves in a linear elastic fashion when deformed.
- The contact radius is much smaller than the radius of curvature of the contacting bodies. For the cell this is true, however for the AFM tip however this is untrue.
- There is no friction on the surface, the contact surfaces are smooth. There is friction on the cell surface though it is not thought to produce problems when measuring the cells with the AFM.
- The deformations are small. This is violated in this thesis as the deformations are indeed larger than the 500nm normally used and are around 1750nm in indentation length.

The formula for the Hertz model for a conical indenter is given by:

$$F = \frac{2}{\pi}\frac{E}{(1-\nu^2)} tan\alpha\ \delta^2 \qquad (2,4)$$

Where F represents the force, E the elastic modulus, α the angle between the plane and side of the cone, δ represents the depth of indentation and υ the Poisson ratio.

The extended Hertzian model extends the traditional Hertz model by allowing E to vary non-linearly by adding an x^2 and x component. Letting E vary as a quadratic gives:

So letting $E(\delta) = k_1.\delta^2 + k_2.\delta + E_b$

Gives…

$$F = \frac{2}{\pi}\frac{E_b^2+k_2d^3+k_1d^4}{(1-\nu^2)}tan\alpha \qquad (2.5)$$

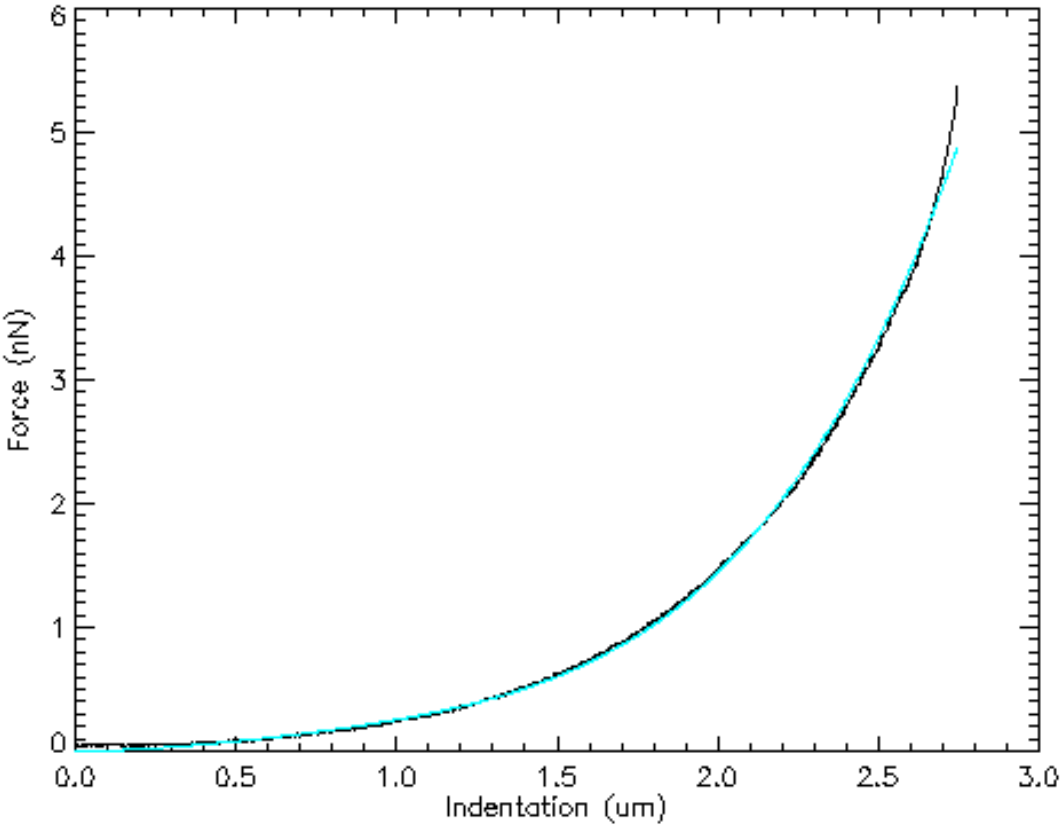


Figure 2-8. Force-indentation curve on LL24 cell showing the experimental data (black curve) and the fitted data (blue curve) using an extended Hertz model. (Murphy, 2007).

As shown in Figure 2.8 the fit is considerably improved compared to that of Figure 2.7. This enables larger indentations to be carried out. For the work carried out in this thesis the extended Hertz model is therefore used throughout, as the deformations are not small and the standard Hertz formula is insufficient to accurately model the curvature of the force indentation plot.

2.5 AFM Database

The AFM data pipeline is critical due to the large volume of data being generated, with 500 records typically being generated each week, and it is estimated that around 50,000 records have been produced by the end of the research programme, representing a hugely impractical task for manual processing.

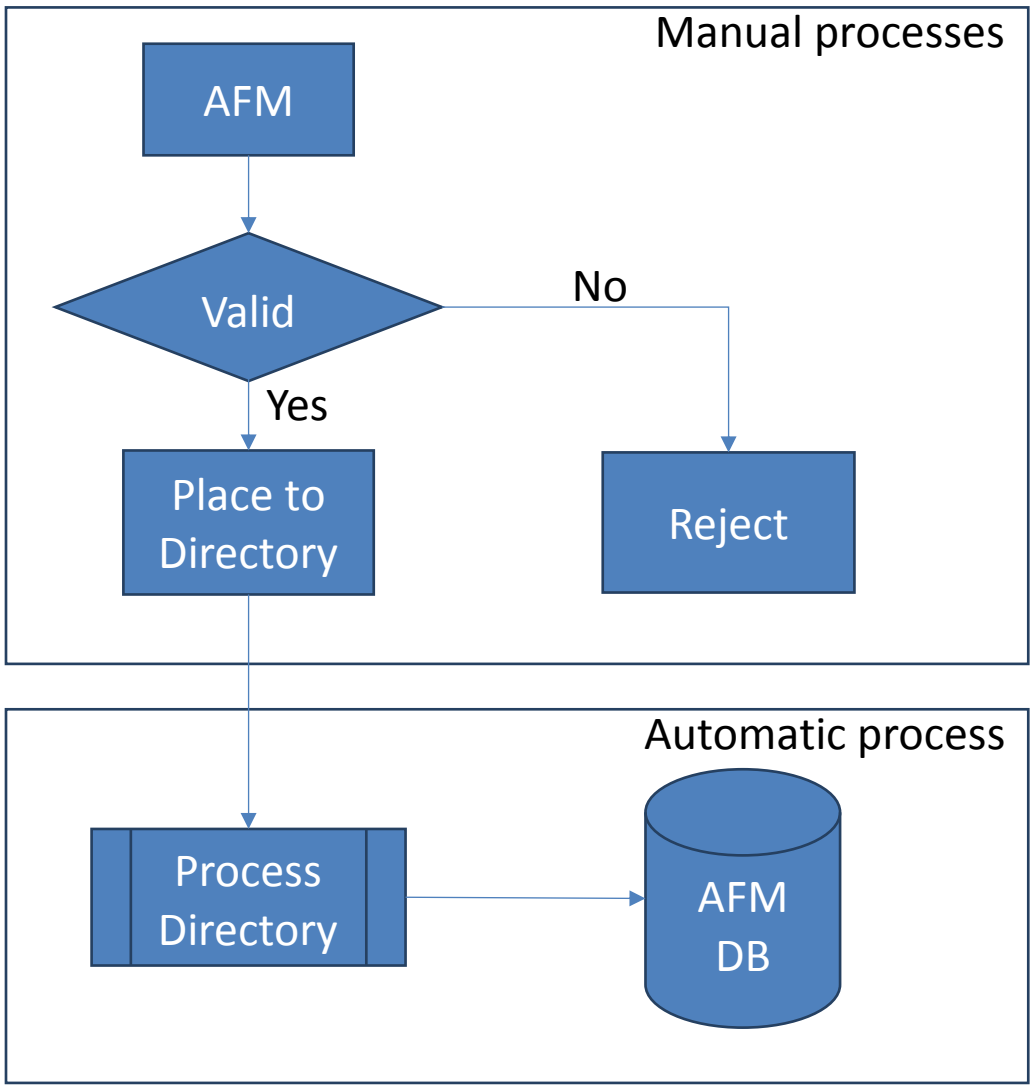


Figure 2-9. The AFM data pipeline to AFM database (DB)

An SQL database has been created by the author to efficiently store and organise the AFM data. The AFM database is regenerated whenever new data is processed. It is implemented as a Microsoft Access Database and is accessed by 32-bit ODBC drivers under Matlab's in-built database toolbox. This is not ideal, as there are no 64-bit drivers for ODBC, however the database size is projected to be around 1.5 GB and 32-bits is therefore sufficient for the task. Optimisation of strings can be incorporated if size becomes a

limiting issue later. This entire process seems to now be mature and static and no further changes to this database pipeline are anticipated at present.

In Figure 2.9 the AFM is first used to generate a force indentation curve. A visual test is then performed on the data on the AFM's dedicated host computer to determine if it passes all of the automation requirements mentioned in Section 2.2. If the curve is rejected a new AFM force curve measurement must be performed. This process is repeated until the AFM indentation curve passes the automation tests, then the AFM data is placed into a directory for automatic processing by the computerised IBW directory processes.

2.5.1 The processing of the IBW directory.

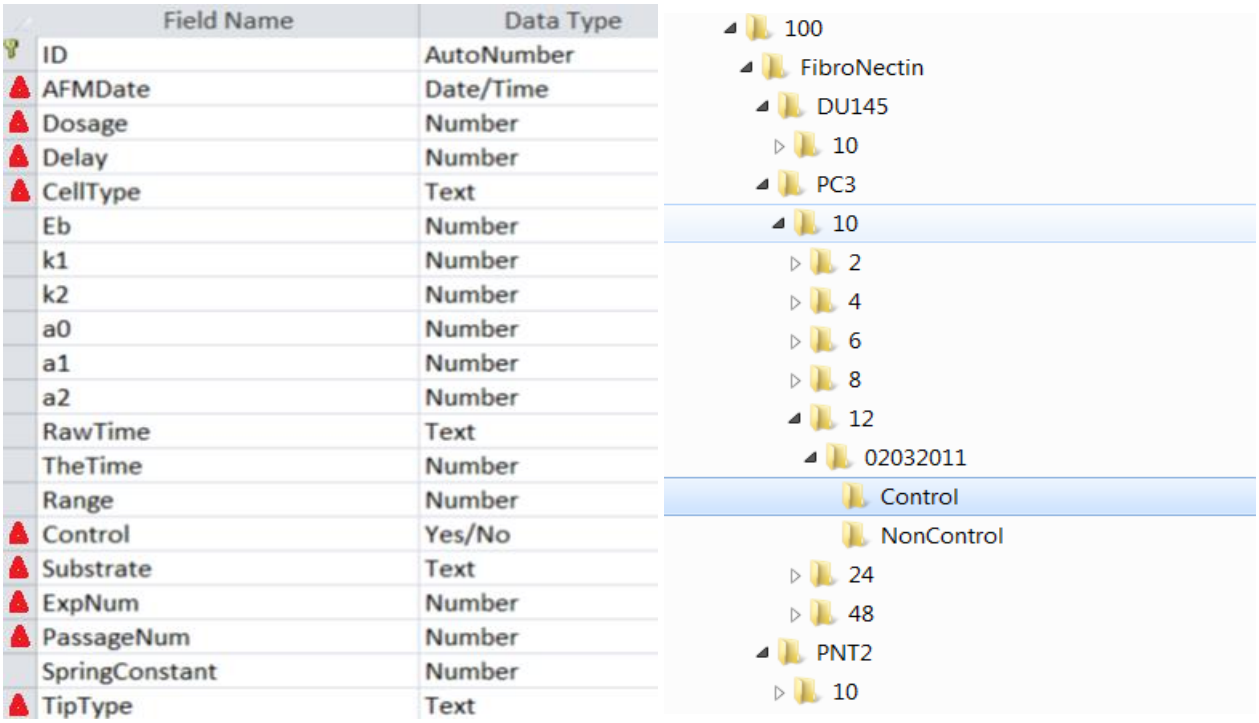
The AFM produces a curve that is fitted using an extended Hertzian model. The model extends the linear Young's modulus parameter into a quadratic form, so that stiffness varies non-linearly with indentation. This has been shown to improve the fit to biological cell data (Murphy, 2006; Randall, 2010). The fit from the normal Hertz model is still produced and saved in addition to that of the extended Hertzian model. However, this is only to act as a baseline, as all analysis at present is performed upon the extended Hertzian model so as to enhance accuracy.

The procedure for automatic processing of an IBW curve file can be described as follows:

1) Read in the file.
2) Extract and populate the raw force indentation curve.
3) Clip the curve, if the end of the Z-piezo range is reached. Tested by curve gradient becoming zero.
4) Obtain a good start point if one can be found, reject and flag curve as being rejected if not.

5) Fit both the extended Hertzian model curve and traditional Hertz model curve to the data.
6) Save output jpeg images of both fits for later comparison.
7) Save any required data to a much smaller file containing the extended fit parameters.
8) If this file is considered to represent an outlier (i.e. outside of 95% of the current data mean for the Hertz E_b) delete it and flag it as having been removed. The majority of outliers are thought to be due to cells being in different states and phases in the cell cycle thus causing the cell response to be radically different to that of normal cells, meaning that they do not form part of the population of cells that at present we wish to analyse.

2.5.2 The Database.



Field Name	Data Type
ID	AutoNumber
AFMDate	Date/Time
Dosage	Number
Delay	Number
CellType	Text
Eb	Number
k1	Number
k2	Number
a0	Number
a1	Number
a2	Number
RawTime	Text
TheTime	Number
Range	Number
Control	Yes/No
Substrate	Text
ExpNum	Number
PassageNum	Number
SpringConstant	Number
TipType	Text

Figure 2-10. Left -: These are the database fields, where the red triangles represent data that is inferred from the directory tree where the files are placed, the rest is obtained from the IBW file. Right -: Shows a subsection of the structured directory tree.

The IBW file does not contain all the data that is required to fully identify each cell type and the experimental conditions under which it was measured. Therefore the aforementioned directory structure is the current solution to this problem and this approach

also allows for extra fields to be added at a later date. The directory structure is also planned to be used in the future for the processing of confocal microscopy data, since confocal data can be placed into the same directory and processed in the same manner with minimal code changes. Consideration was given to re-naming the files, where this would be a manual task of re-naming each file with the red triangle fields shown in Figure 7, so that these were included within the filename. Although this solution would be initially much easier to code, with this approach adding additional information at a later date would prove to be more difficult, as every file would require adjustment rather than just a top level directory name and the approach was therefore rejected on those grounds, as a top level directory structure applies to all the files below it. Other simplistic solutions, such a configuration file within a subdirectory, suffer from same problem as file renaming, in that they need to be altered at the subdirectory level rather than at a higher directory level, thereby increasing the level of effort required to add new fields in the future.

2.5.3 The Database Interface (GUI)

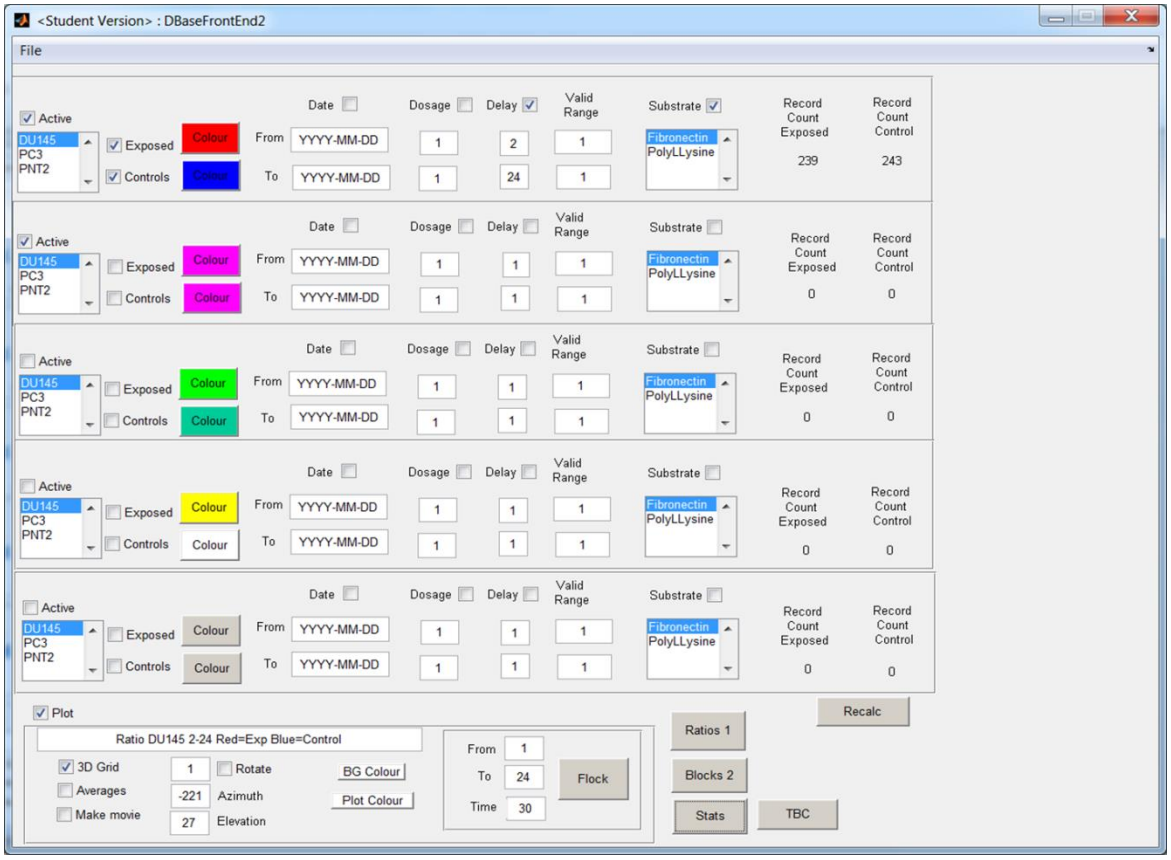


Figure 2-11. Screenshot from the database's graphical user interface (GUI).

The database front-end GUI that has been developed by the author during this research project is shown in Figure 2.11 and this allows the various combinations of different plots and statistics to be generated by a user of the database. As more methods and visualisations have been developed and completed throughout the course of this research programme, they were added as options to the database interface. Currently, up to 5 SQL queries can be performed at once and the output of these may be coloured independently within graphical outputs, giving 10 options when including controls/non-controls. Output plots that are currently available from the database GUI include: 3D Plots, Flock Plots, Ratio distance plots and Ratio percentage contribution plots, which form part of the visualisations to aid in the statistical analysis of the cellular data.

2.6 Initial statistical analysis and visualisations.

The initial statistical tests that were performed upon the datasets were designed to test if there were any statistical differences between the datasets. A single point on the force-indentation curve was used, namely the maximum indentation range, which for Figure 2.12 represented an indentation depth of 3.0 μm. The rationale behind choosing the point of maximum indentation depth to compare the force-indentation curves lay in the belief that if there were to be a difference in AFM force response, it would most likely be seen at the maximum indentation range, where typically AFM curves would exhibit the most force variance. This is a single point test of the force and does not describe the entire curve, so therefore this can only be used as an indicator, rather than as a full comparison of the cell curve indentation responses.

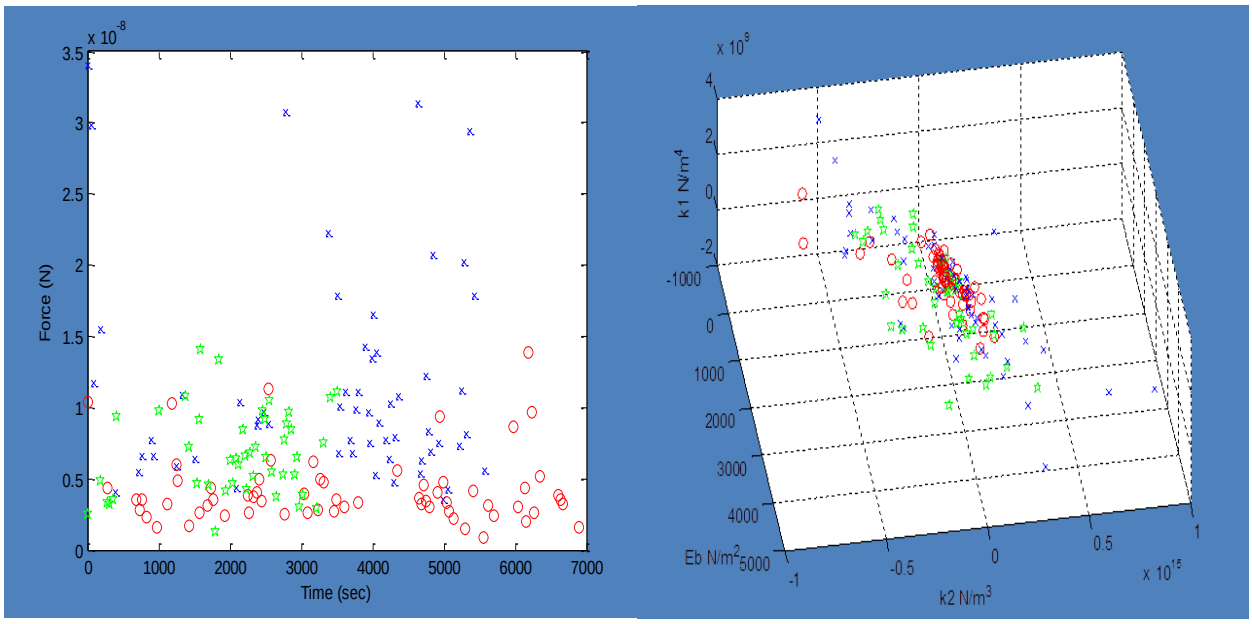


Figure 2-12. Left: Shows a test at a single indentation of 3.0um for PNT2 cells (denoted by blue crosses), DU145 cells (denoted by green stars) and PC3 cells (denoted by red circles). Right shows the corresponding Eb,K1,K2 quadratic stiffness parameters from the extended Hertzian model.

Figure 2-12 left represents the time at which the cells were measured and the corresponding force response. It can be seen that time has little bearing on the force response results, having only a 1-3% influence over an hour long period. In Figure 2-12 right the corresponding extended Hertz model fit is shown for the same indentation data.

An experiment was performed to test if the force responses of populations of cells were different at a maximum radiation exposure dose of 10 Gy. Primarily the experiment was aimed at determining if any statistically significant differences existed between the control cells and irradiated cells in terms of their force response at various durations after exposure. The results are summarised in Table 2-1 below.

Duration after Expsoure	2hr exp.	2hr ctrl	4hr exp	4hr ctrl	6 hr exp	6hr ctrl	8 hr exp	8 hr ctrl	12 hr exp	12 hr ctrl
Normal distribiution	n	n	y	n	n	n	n	n	n	n
Mean Force (N) x10^{-9}	3.61	4.71	7.49	4.98	5.03	5.55	4.46	4.84	5.41	6.06
Populations different	yes, p=0.0326		yes, p=2.1192 e-004		no, p= 0.6617		no, p= 0.1448		no, p= 0.5043	

Table 2-1. PNT2 normal cells, with a single point test at 1.75µm indentation, used to test the force responses. Kolmogorov-Smirnov tests whether the populations have differing distributions of response for force, at the static indentation point 1.75µm. At each time frame radiation exposed (exp) at 10 Gray and non-exposed control (ctrl) populations of cells are measured by AFM.

Table 2-1 above shows a population of PNT2 cells that is tested at a single indentation point (1.75µm) against the force required to attain that level of indentation by the AFM probe. Table 2-1, row 1, shows the statistical normality test (Lillefors at 5%) and it can be seen that only the 4hr control population (the non-irradiated cells) are normally distributed. The 'exp.' column identifies the cells which were exposed to 10Gy of radiation in a single dosage. The 'ctrl' columns represent the non-irradiated cells, i.e. the controls. These control

cells undergo exactly the same experimental protocol as the exposed cells but are not actually exposed to the radiation insult. The non-normality of the cell distribution is the reason that a non-parametric test (Kolmogorov-Smirnov) is used to test for a population distribution difference in the mean force response. Table 2-1, row 2, is the mean force required for the cell population to indent to 1.75μm. Table 2-1, row 3, is the population difference (Kolmogorov-Smirnov) test for force distribution at an indentation depth of 1.75 μm. This was the initial approach, comparing differences in the force response of various populations of cells. The control cells and the radiation exposed cells in Table 2-1 are statistically different at durations of 2 hours and 4 hours after radiation exposure finishes. At a duration of 2 hours post-irradiation the force required to indent the cells to 1.75 μm for the exposed cells is less than that required to indent the control cells to the same level, reduced by a statistically significant 1.1 Nano-Newtons (nN). However at a duration of 4 hours post-irradiation this situation is reversed and the radiation exposed cells require a further 2.5 nN of force to indent them to the same indentation of 1.75 μm compared to normal control cells. This could be related to the high radiation dosage used, which at 10 Gy exceeds the therapeutic dosage of 1Gy by a factor of 10.

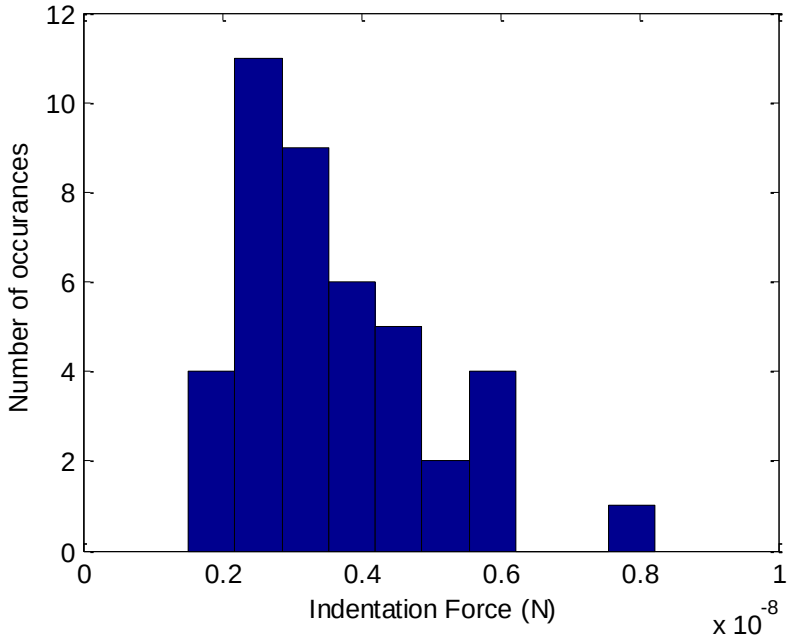


Figure 2-13. This is the histogram for the indentation force at 1.75μm for PNT2 cells exposed to radiation after 2 hours. The total population is 42 entries and is non-normal by testing with Lilliefors normal test. The mean is 3.61 x 10^{-9} N of force and has a median force of 3.36 x 10^{-9} N.

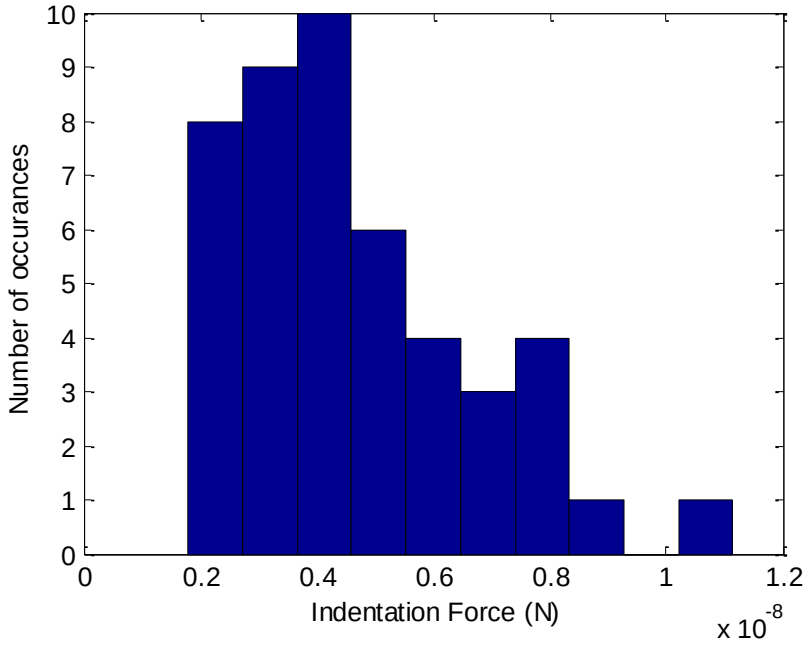


Figure 2-14. This is the histogram for the indentation force at 1.75μm for PNT2 cells control cells (Not exposed to radiation) after 2 hours. The total population is 46 entries and is non-normal by testing with Lilliefors normal test. The mean is 4.71 x 10^{-9} N of force and has a median force of 4.33 x 10^{-9} N.

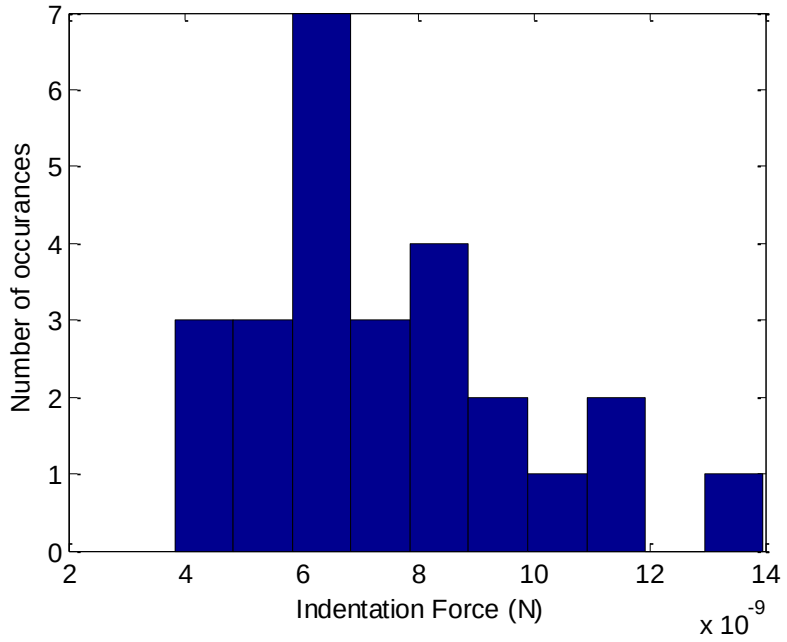


Figure 2-15. This is the histogram for the indentation force at 1.75μm for PNT2 cells exposed to radiation after 4 hours. The total population is 26 entries and is normal by testing with Lilliefors normal test. The mean is 7.49 x 10^{-9} N of force and has a median force of 6.78 x 10^{-9} N.

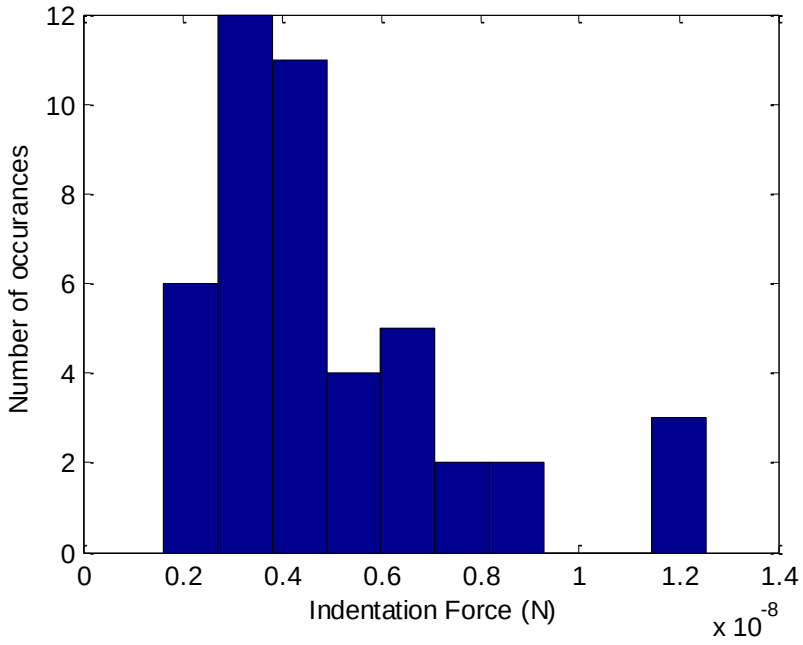


Figure 2-16. This is the histogram for the indentation force at 1.75μm for PNT2 control cells (Not exposed to radiation) after 4 hours. The total population is 45 entries and is not normal by testing with Lilliefors normal test. The mean is 4.98 x 10^{-9} N of force and has a median force of 4.53 x 10^{-9} N.

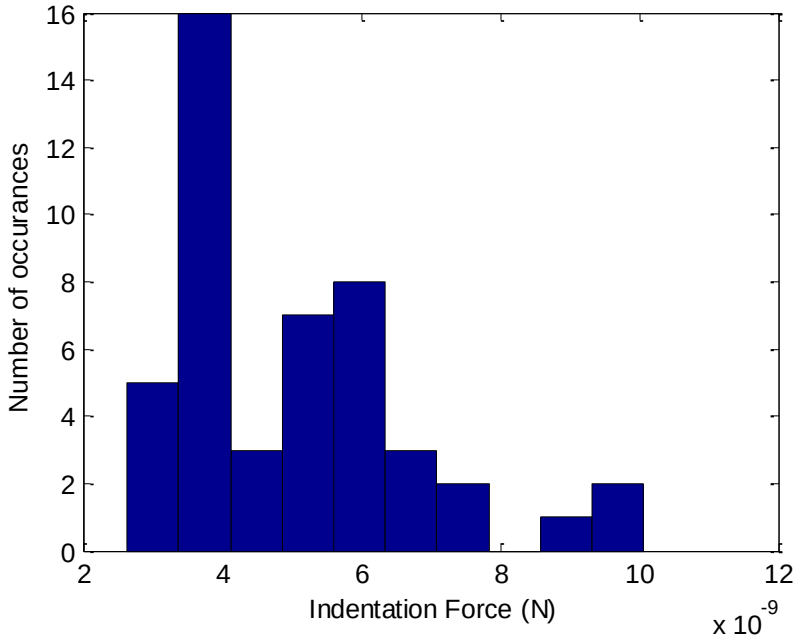


Figure 2-17. This is the histogram for the indentation force at 1.75μm for PNT2 cells exposed to radiation after 6 hours. The total population is 47 entries and is not normal by testing with Lilliefors normal test. The mean is 5.03 x 10^{-9} N of force and has a median force of 4.56 x 10^{-9} N.

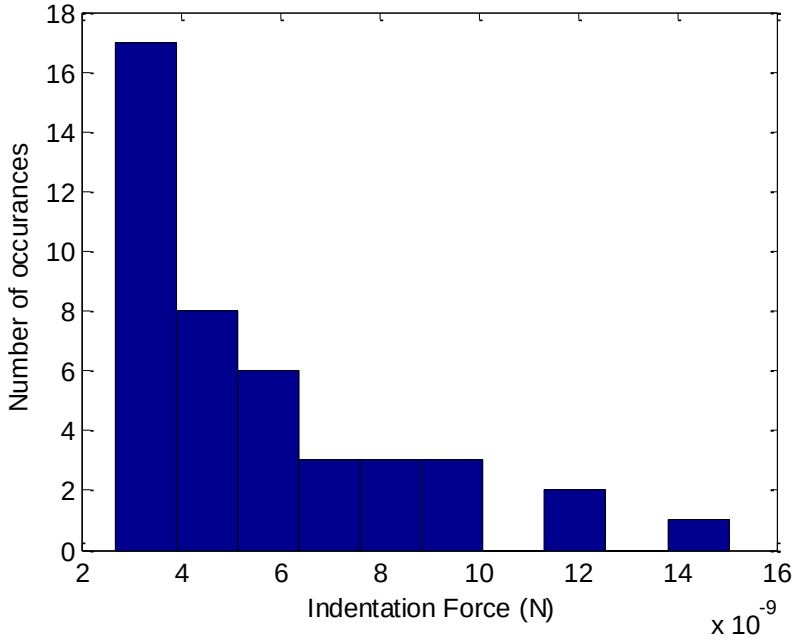


Figure 2-18. This is the histogram for the indentation force at 1.75μm for PNT2 control cells (Not exposed to radiation) after 6 hours. The total population is 43 entries and is not normal by testing with Lilliefors normal test. The mean is 5.55 x 10^{-9} N of force and has a median force of 4.80 x 10^{-9} N.

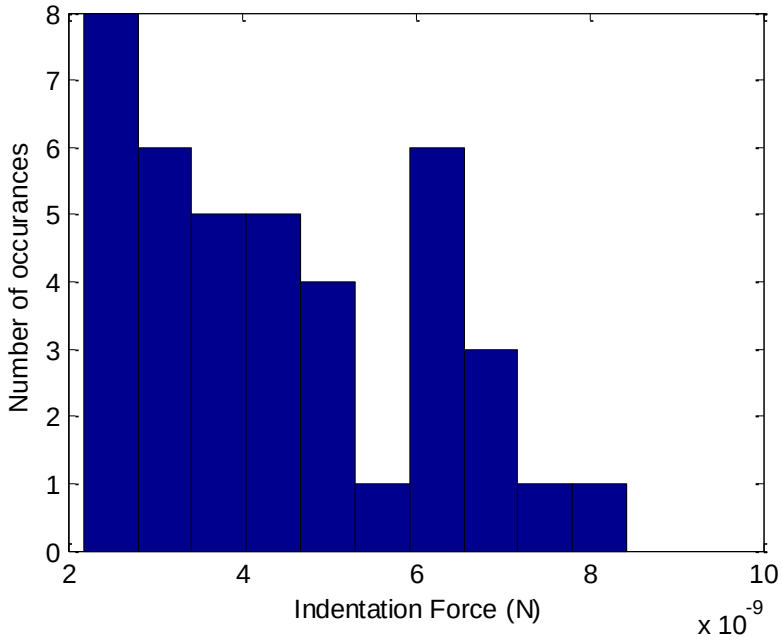


Figure 2-19. This is the histogram for the indentation force at 1.75μm for PNT2 cells exposed to radiation after 8 hours. The total population is 40 entries and is not normal by testing with Lilliefors normal test. The mean is 4.46 x 10^{-9} N of force and has a median force of 4.16 x 10^{-9} N.

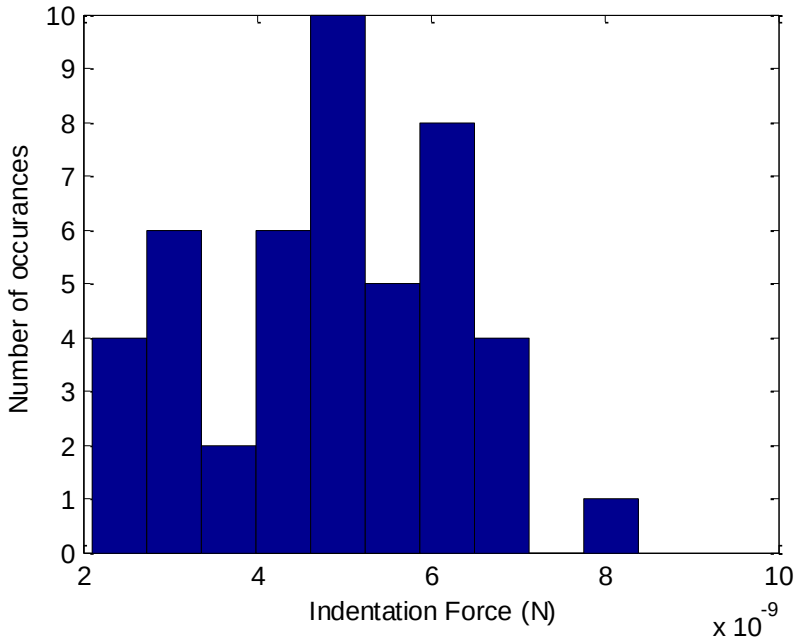


Figure 2-20. This is the histogram for the indentation force at 1.75μm for PNT2 control cells (Not exposed to radiation) after 8 hours. The total population is 46 entries and is not normal by testing with Lilliefors normal test. The mean is 4.84 x 10^{-9} N of force and has a median force of 4.96 x 10^{-9} N.

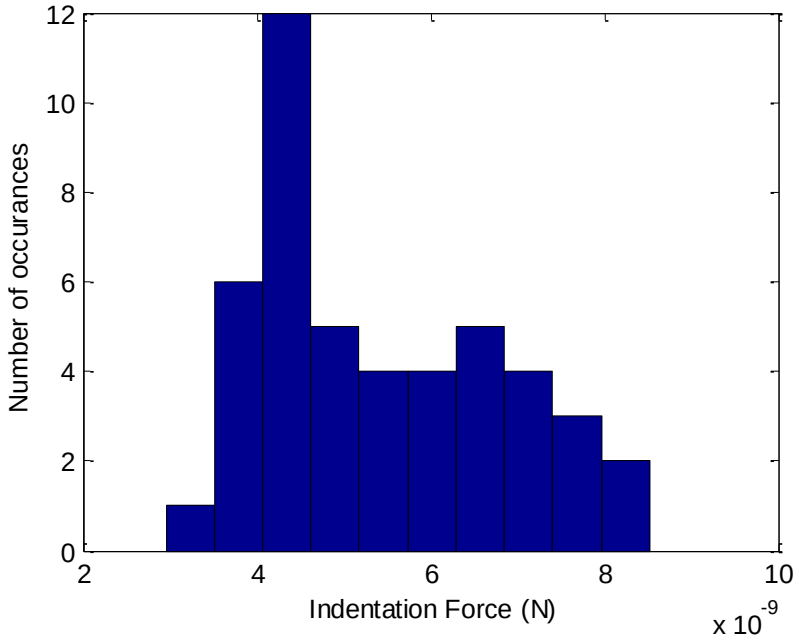


Figure 2-21. This is the histogram for the indentation force at 1.75μm for PNT2 cells exposed to radiation after 12 hours. The total population is 46 entries and is not normal by testing with Lilliefors normal test. The mean is 5.41 x 10^{-9} N of force and has a median force of 4.87 x 10^{-9} N.

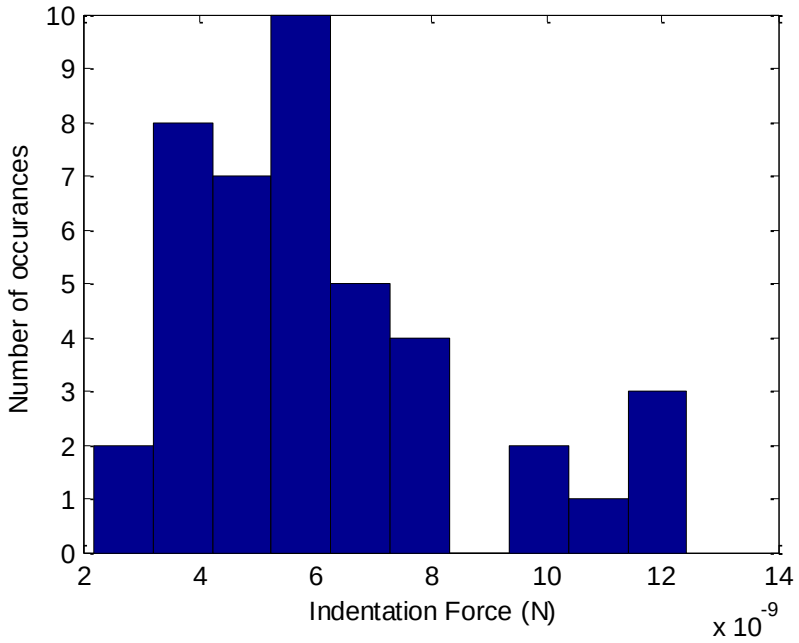


Figure 2-22. This is the histogram for the indentation force at 1.75μm for PNT2 control cells (Not exposed to radiation) after 12 hours. The total population is 42 entries and is not normal by testing with Lilliefors normal test. The mean is 6.06 x 10^{-9} N of force and has a median force of 5.70 x 10^{-9} N.

2.6.1 3D Plots and Flock plots.

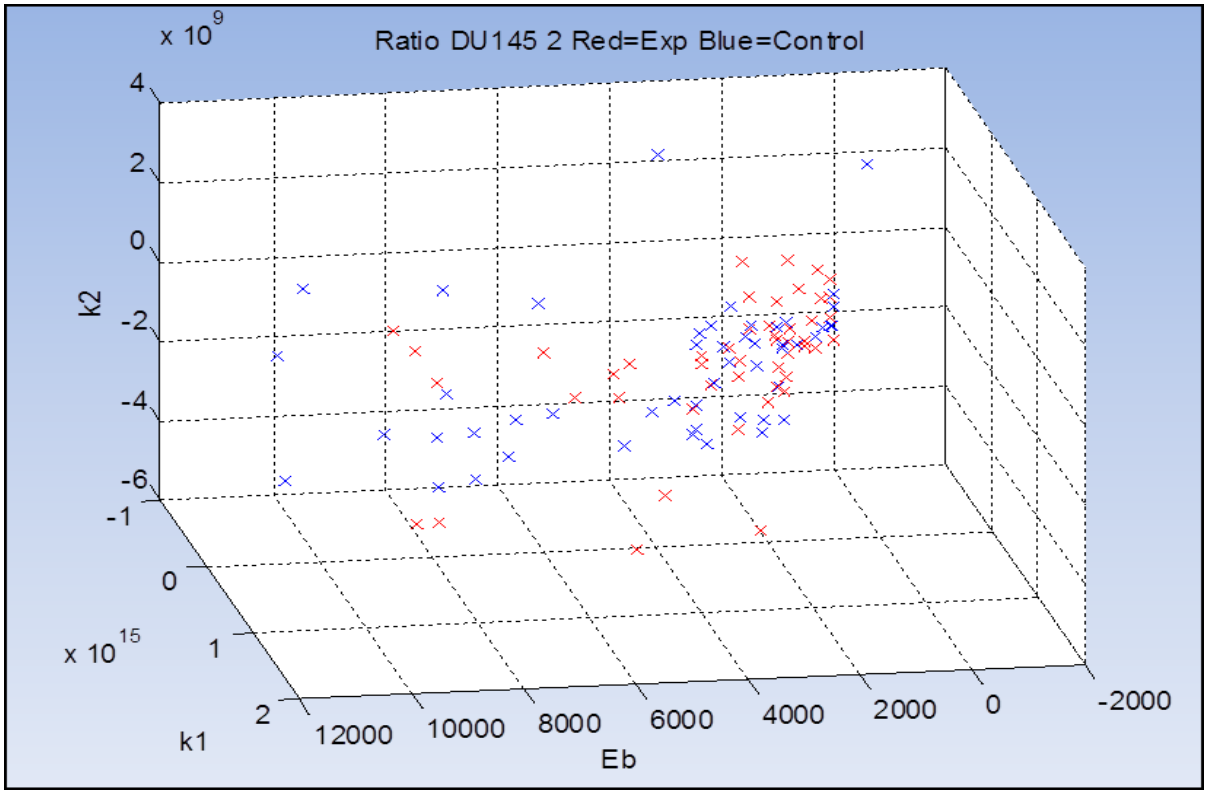


Figure 2-23. This is a typical plot generated in Matlab, for a cell line showing the extended Hertzian model stiffness parameters K1,K2 and Eb for both Exposed cells (red crosses) and Control cells (blue crosses). These plots can be generated from the GUI, either as a static image, or made into a rotating plot movie.

Flock plots are an extension of the 3D plot, to show the cells changing temporally, from a selected start time to an end time. By analysing the movement of the flock it is hoped to elucidate overall features that may not be seen within static plots. Given that cells actively react to force by reorganisation of their internal architecture to counteract the applied force, no cell can be measured twice. This means that the flock movement is actually representing one cell population morphing into a different cell population, measured at different times. The morphing is performed by interpolation and the interpolation algorithm attempts to minimise the total distance travelled for all the cells. A set of images are saved out at 15

frames per second and these images are turned into an uncompressed AVI movie file using standard Matlab libraries.

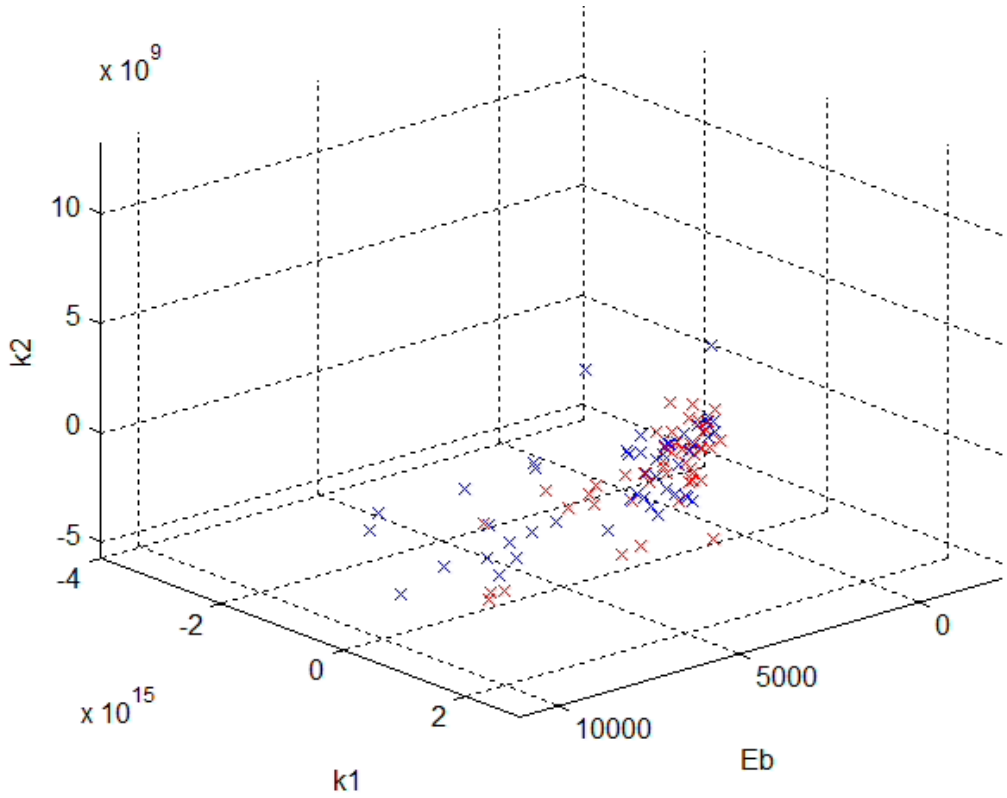


Figure 2-24. A plot showing the Eb,K1,K2, values for PNT2 cells after 2.1 hours. The blue 'x' are PNT2 control cells, the red 'x' are the PNT2 cells exposed to 10 Gy radiation. Of note is the blue 'x' greater clustering along the Eb axis than the red 'x' indicating a lower level of stiffness for the exposed cells. This is confirmed in Table 2.1 which showed that at 2 hours the exposed cells (red 'x') were less stiff than the non-exposed cells.

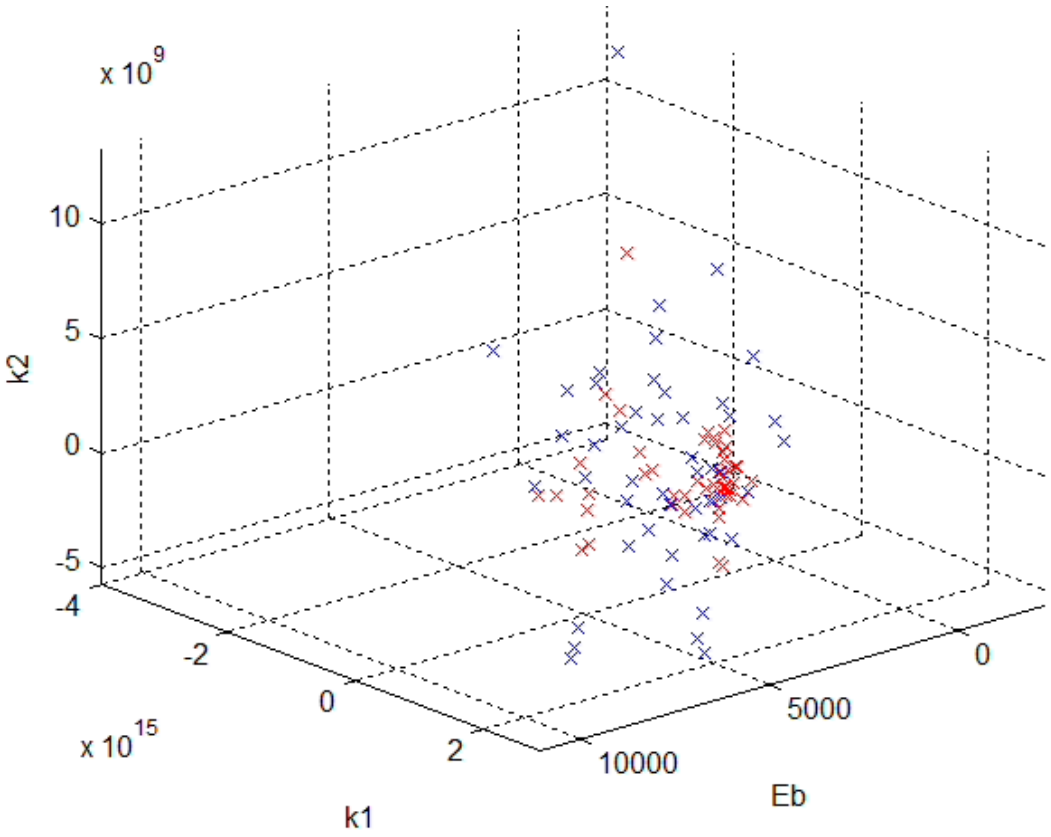


Figure 2-25. A plot showing the Eb,K1,K2, values for PNT2 cells after 4.2 hours. The blue 'x' are PNT2 control cells, the red 'x' are the PNT2 cells exposed to 10 Gy radiation. The spread of the cells has altered from figure 2-14, the cells are less stiff in terms of Eb but the K1,K2 have more of a spread indicating more stiffness. The table 2-1 showed that the exposed cells (red 'x') are stiffer at this point, this is difficult to see from this figure.

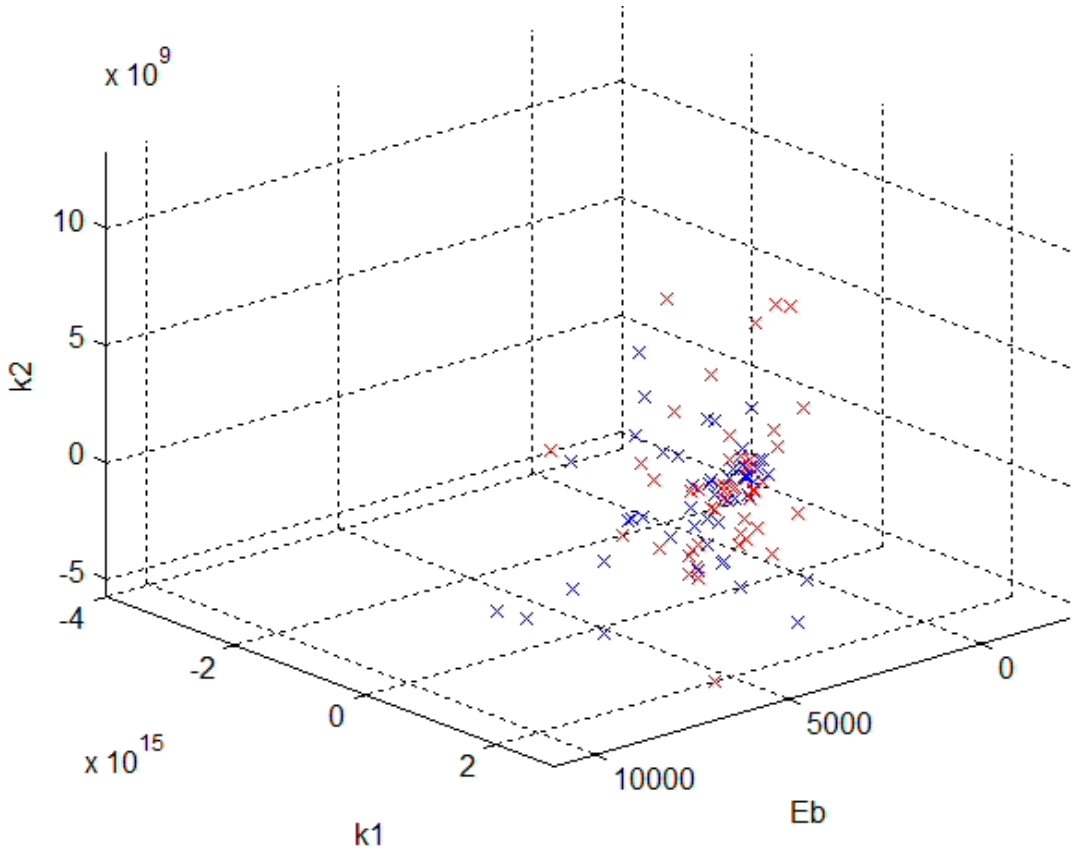


Figure 2-26. A plot showing the Eb,K1,K2, values for PNT2 cells after 6.4 hours. The blue 'x' are PNT2 control cells, the red 'x' are the PNT2 cells exposed to 10 Gy radiation. The spread of the cells has altered from figure 2-15, the blue and red cells are roughly identical in spread indicating that there is no difference in the populations, this is confirmed in the analysis of the mean force in table 2-1.

It can be seen from Figures 2-24, 2-25 and 2-26 that the cells respond differently to AFM force probing after exposure based on the time after that exposure. Table 2-1 therefore represents a snapshot look at the exposure at 2 hour intervals to see what happens to the cells response for exposed and non-exposed cell lines. In the appendix B1 (Appendix\B1\topone all 2_24.avi) a flock plot that these static images were taken from can be played to see the dynamics of how the snapshots of the cells change over time.

2.6.2 Ratio Distance Plot.

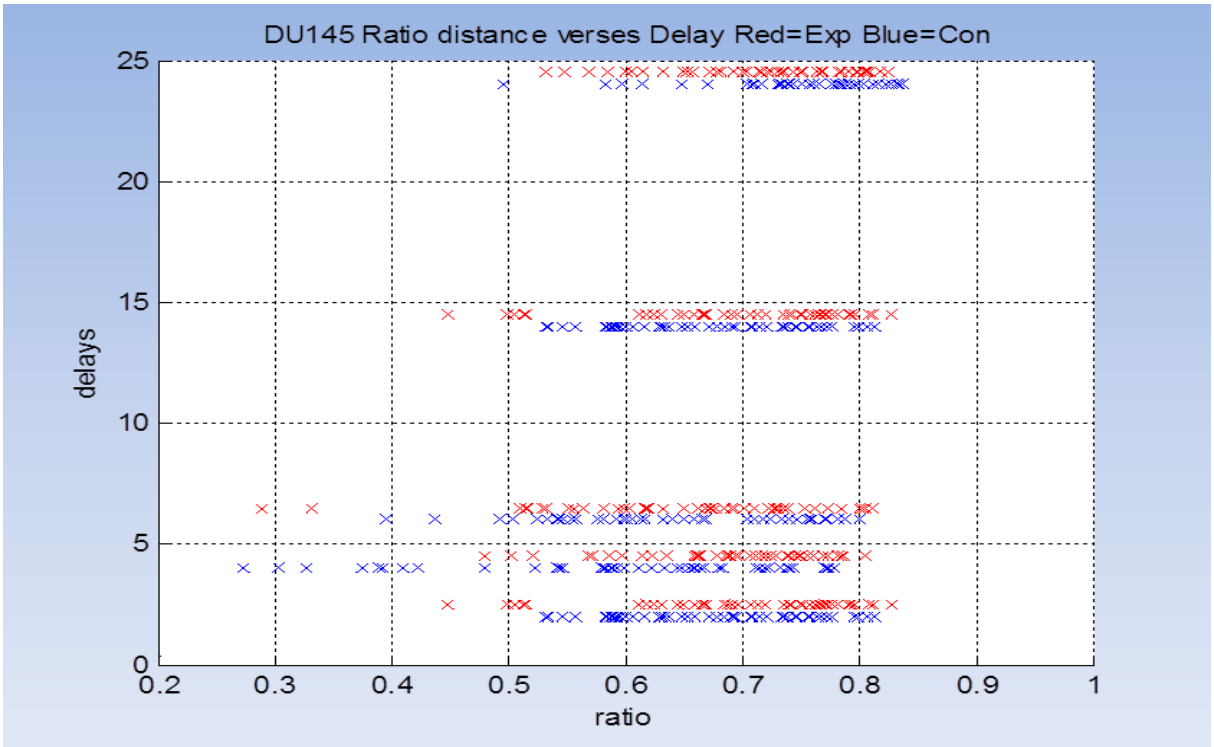


Figure 2-27. A typical ratio distance plot. Measurement delay is measured in hours. Ratio is the ratio of the normalised data point's position compared to the distance from the centre of a cube for Eb, k1, k2, where Eb, k1 and k2 are normalised to lie in the range 0.0 to 1.0. This is intended to compare the variance in distance between the stiffness parameters Eb, k1, k2 in a systematic manner that is not biased by the relative magnitudes of the individual parameters.

In figure 2-27 the exposed red 'x' cells have time added to the delay in order to make them visible in the figure, the actual delay is the same as for the non-exposed cell line. The figure above indicates that after 2 hours the difference in force response is small. At 4 hours it indicates that the position is reversed and the non-exposed cell line is less stiff having a smaller average ratio. At 6 hours the cell lines are both about the same, the same is true for 12 and 24 hours. There is a general shift to the right from 4 hours onward perhaps indicating a stiffening of the cells.

The original data points cannot be used, due to the widely varying scales of the three stiffness parameters, as the value of stiffness parameter Eb is typically 9 orders of magnitude greater than both the k1 and k2 parameters and therefore a direct comparison would result in the value of Eb dominating the others. Therefore the original triplet points (Eb,k1,k2) are first normalised independently into a scale ranging from 0.0 to 1.0. The ratio in Figure 2-12, is determined as being the three dimensional Pythagorean distance from the point at the centre of the normalised cube (0.5,0.5,0.5) to the re-scaled triplet points (Eb,k1,k2). It is hoped that this approach will be able to provide a method to investigate alterations in ratio variations between exposed and control cell populations.

2.6.3 Ratio Percentage Contribution Plots.

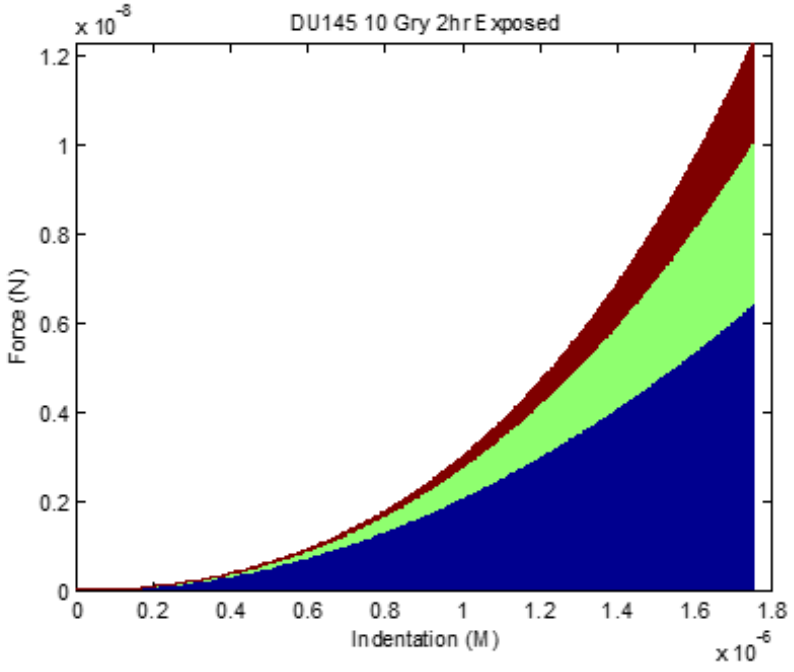


Figure 2-28. Typical ratio percentage contribution plot. This represents the percentage force contributions of each of the three stiffness parameters Eb (Blue), k2 (Green) and k1 (Red) that make up the overall force response versus indentation distance.

The purpose behind the ratio percentage method is to explore what happens to the ratio of Eb, k1, k2 over the entire indentation range and also what changes occur when the cells are exposed to radiation. This has shown some promise in that differences do appear between the percentage contribution of responses for the three stiffness parameters between PNT2 cells and PC3 cells when they are exposed to radiation. However the data is still preliminary at this stage and requires confirmation by repeatability testing if it is to be considered to be conclusive. Figure 2-28 also shows that the original linear Hertz model is not sufficiently accurate to describe the force-indentation curve adequately, because at an indentation of 1.75um the force contribution from the Eb component of the model is approximately 50% of the total force.

The initial population analysis was performed using the single point approach. The main analysis problems lie with the number of variables to be analysed and how to both display them and interpret them meaningfully. Since Eb, k1, k2 are not independent variables, it is difficult to see the effect that the values are having on the force response on a static 3D plot.

2.7 Automated Curve fitting.

2.7.1 Automatic curve fitting of the human cell data.

Once the IBW files had been verified as correctly accessing the AFM data, the task of automatically fitting the data could then be begun. The typical force-distance curve shape

shown in Figure 2.1 has a flat start section progressing into a shallow curve becoming steeper the further the AFM tip progresses in its indentation of the cell. The curve needed to be fitted for both the standard Hertz model and the extended Hertz model (Randall, 2009). The fitting process that was chosen initially was a least squares approach and this was fitted using Matlab.

The least squares fitting method used the right slash matrix method within Matlab. The right slash method within Matlab is actually QR decomposition which will compute the eigenvalues for the matrix. Initially the process of fitting the Hertz and extended Hertz model were performed without auto-adjusting the scale of the numbers. The matrix solution however often became rank insufficient due to the k_1 term and the nanometer scale 10^{-9} involved. In order to solve this problem, a sample was taken near the start of the curve, typically the 25^{th} position of the curve. The sample was then multiplied by 10 until the sample become larger than 1, then the matrix solver was adjusted in terms of the power. For example, if the scale was 15 to scale the value to 1 unit E_b within the matrix would be multiplied by a factor of 10^{30} and an k_2 value would become 10^{45}. This would be fed into the matrix and solved using the least squares method. The values output from the solver would then have to be scaled down to match the original data set. This solved the rank insufficiency problem while maintaining the original scale. This method was replaced later when the *fit* command was available within Matlab's curve fitting toolbox and the auto-generated code. The two different methods are functionally equivalent given that both still solve using the same least squares method.

Initially there were two different functions fitted, the baseline Hertz, which is an $Eb*x^2$ function, and the extended Hertz function which involves an $Eb*x^2 + k1*x^3 + k2*x^4$ form.

The reason for fitting both models was the rationale that the Hertz function, while being suitable for lower indentation levels, fails to accurately match the curve data when the indentation depth is increased beyond 500nm. This is shown in Figure 2-7.

2.7.2 Automatic curve fitting of yeast data.

Yeast data was used as a 'simple' model to test out the automation when applied to a completely new project that didn't adhere to the rigid capture rule-set for human cell lines. The goal was to enhance the automation tool to enable it to be tailored to automate other cell lines with different characteristics, including smaller indentation levels.

The yeast AFM force data presented a different set of problems with regard to automation, the primary one being that the indentation level used was actually indenting against the mica substrate and therefore the curves were different in terms of their basic structure. They were exhibiting a folded shape, which is an artefact and not a desired input to the automation processing code. This is shown in Figure 2-29. In order to process the yeast data curves, the curve was therefore made to match the shape of the original human cell curves, as described in the following section.

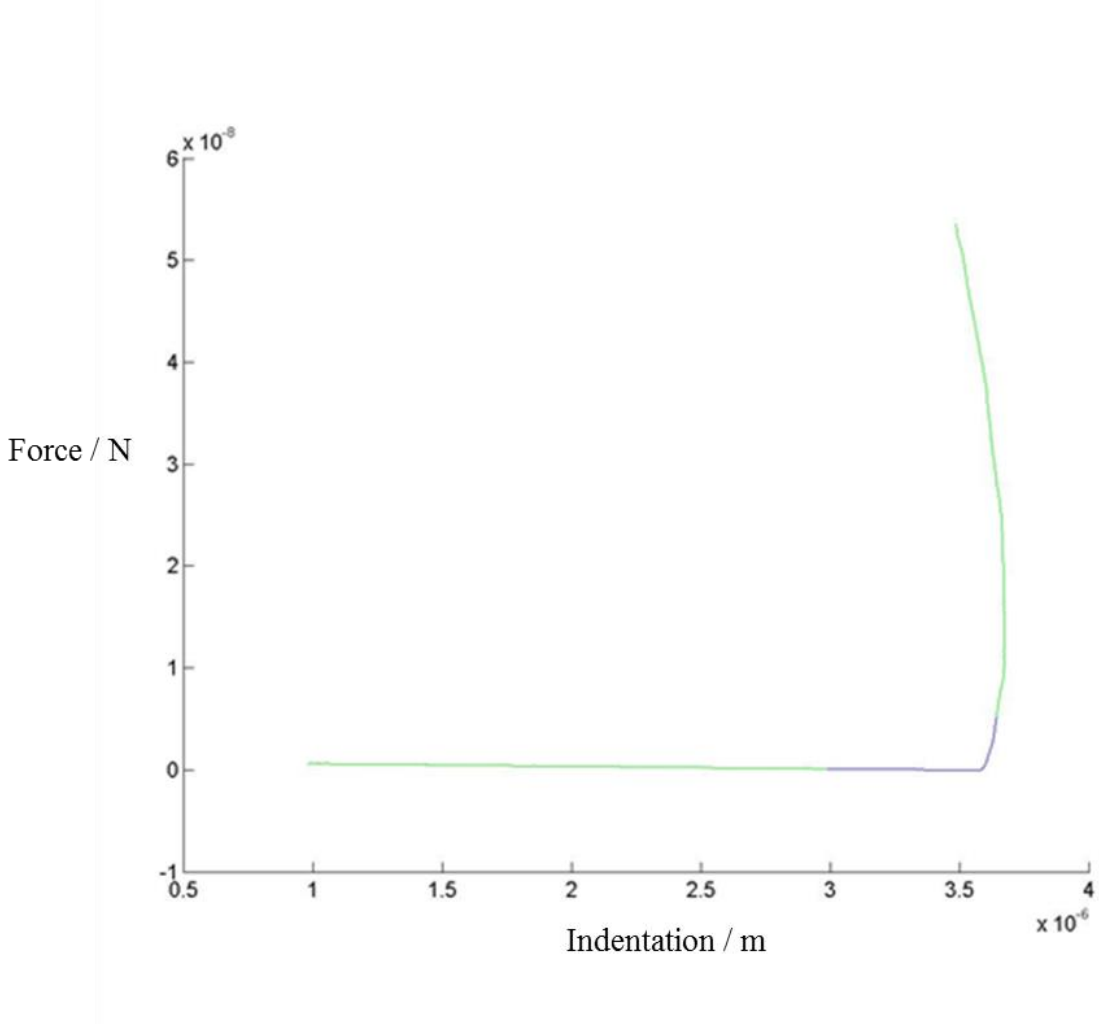


Figure 2-29 An example yeast data curve, showing the region selected to be processed in blue and the green regions which are discarded. The curve slopes back on itself possibly indicating that the AFM tip angle passed the middle 0 degree region and therefore the indentation looks like it backtracks.

The first attempt at solving this problem was to look for the maximum rate of change of the curve, unfortunately while working for the specific curve shown in Figure 2-29, this failed when the yeast data hit the end of the piezo and became flat, thereby becoming the largest rate of change over the curve.

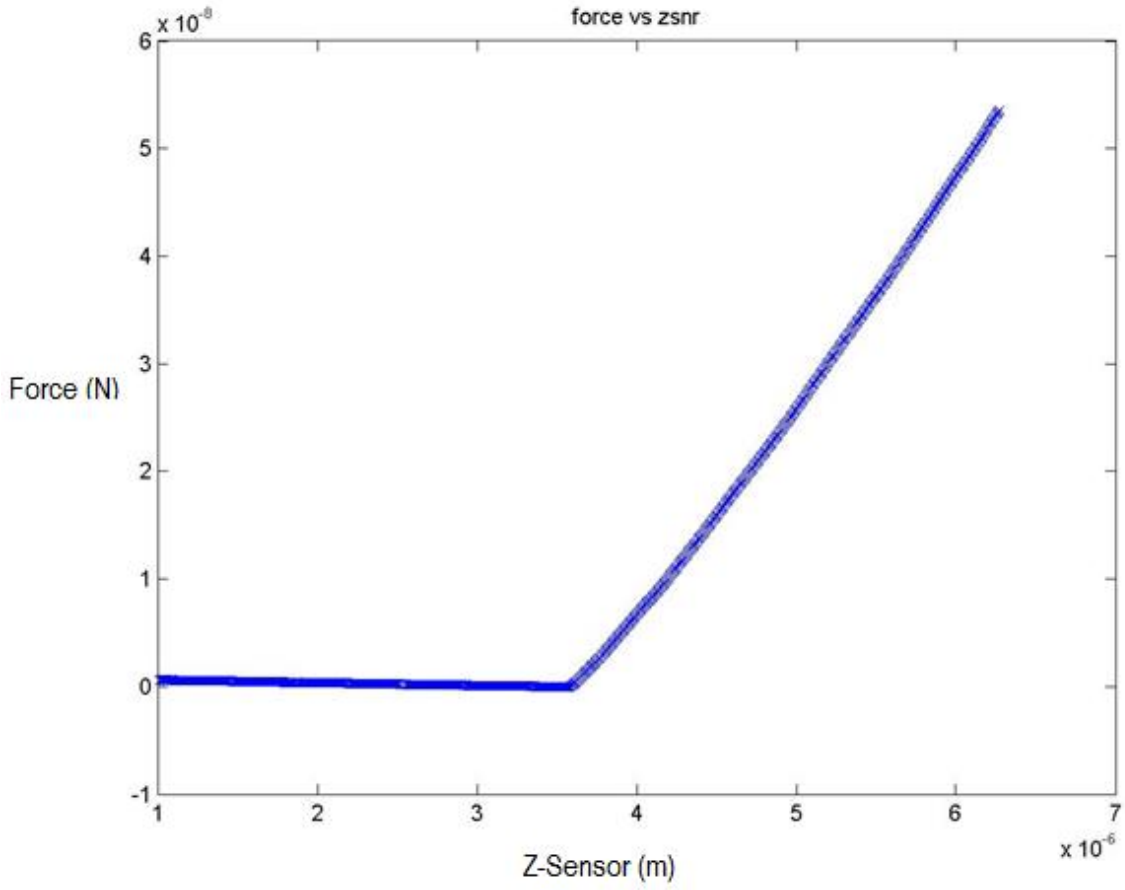


Figure 2-30 The force verses z-sensor that shows that the indentation does not loop back on itself. However, the slope is constant which indicates that the response has become linear and representative of the spring constant for the IBW indentation file.

The solution to the problem was a two-tiered approach, the first aspect was to exclude data above a certain force threshold in order to restrict the maximum amount of force to which the yeast cell was being subjected. The actual amount of force restriction was only the first step and was primarily aimed at excluding those curves where the piezo was at the end of its range.

The author also believed that the substrate upon which the cells are grown was possibly influencing the results; as slope of the curve became flat and the indentation range was too high for yeast cells causing the force curve to loopback upon itself as shown in Figure 2-

28 Therefore in order to help alleviate this problem, at the position where rate of change of the slope of the curve becomes flat towards the right topmost end of the curve this segment of the curve was removed as this area is not required by the experiment. The difference in fit can be seen visually. The original data fit for a yeast cell using the entire data range after clipping is shown in Figure 2-32.

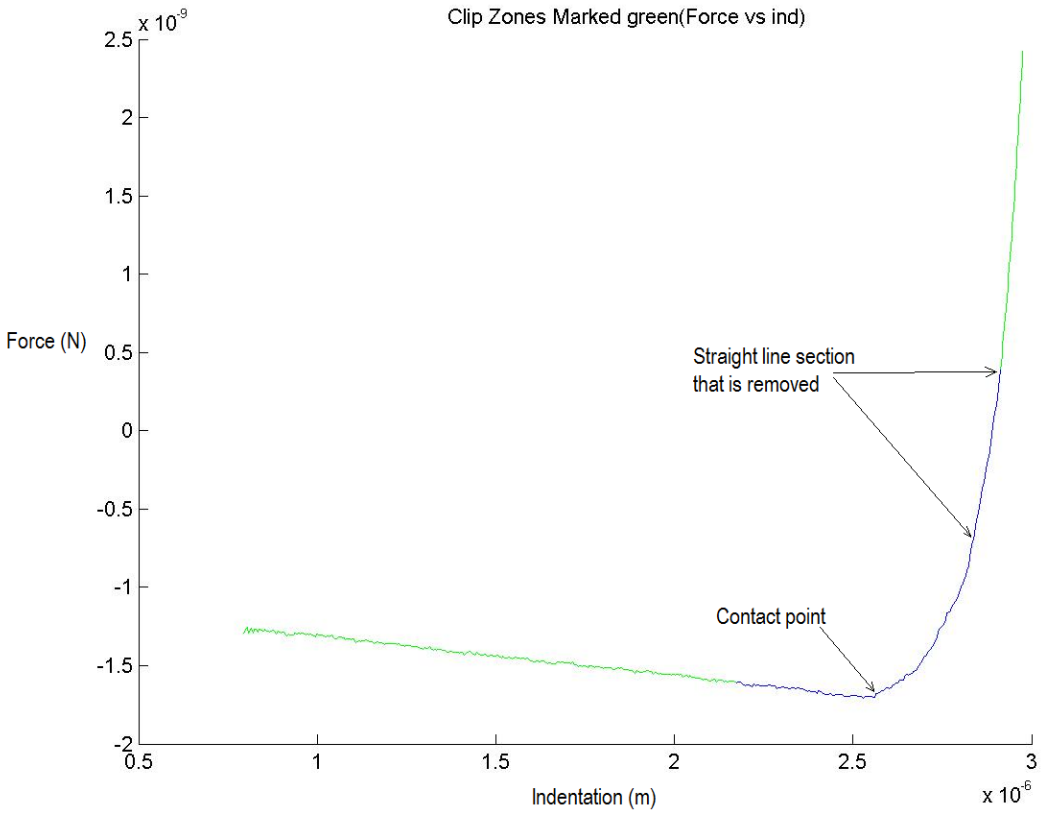


Figure 2-31 The original data for the yeast cell line is processed in several stages. The first stage is the removal of the regions marked in green. The contact point is then found and the region to the left of the contact point is removed. Then the straight line portion is found and also removed to improve the fit.

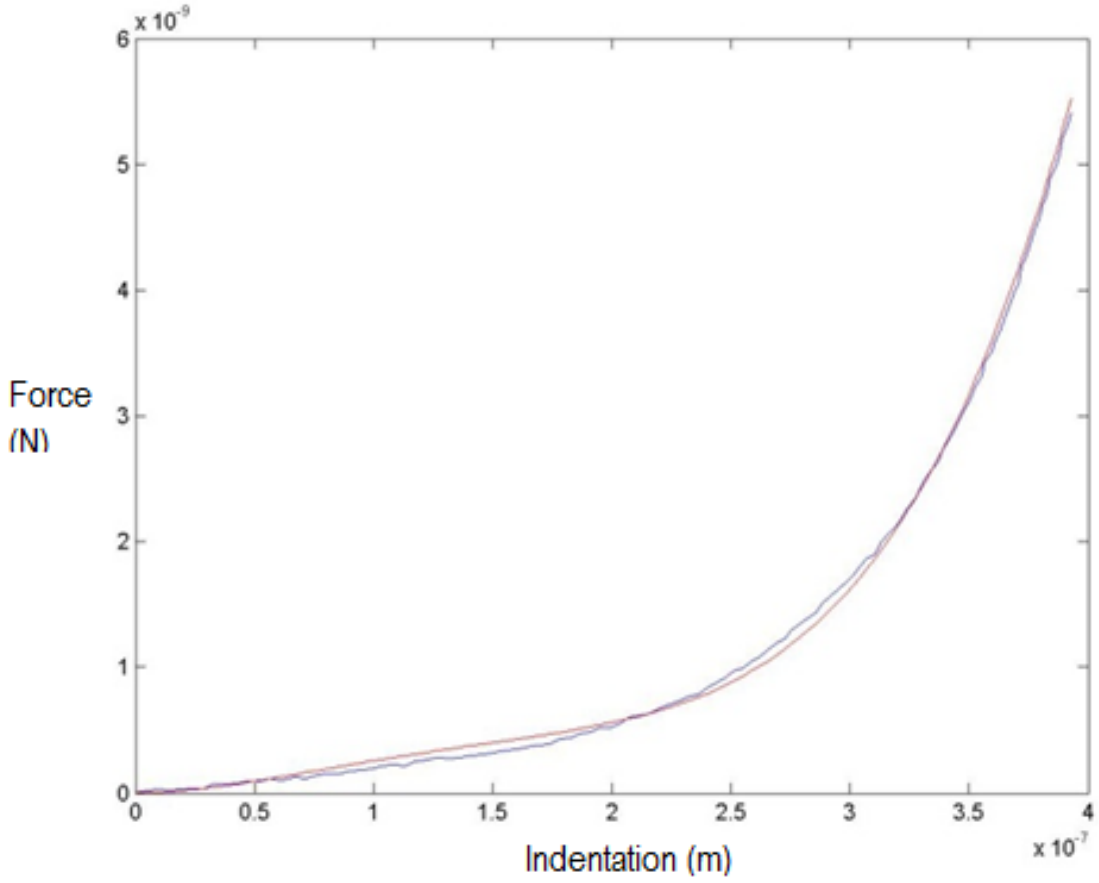


Figure 2-32 The original fit that includes the entire range of the fit for the same data as that shown in Figure 2-31. The blue curve represents the original AFM force curve data and the red curve the fit which can be compared to Figure 2.33 and is clearly not as good a fit.

When the top rightmost curve segment region is clipped and removed (shown in Figure 2-33) the extended Hertzian model's fit to the data is far superior to the data fit that is shown in Figure 2-32, which still includes the straight portion at the end of the curve.

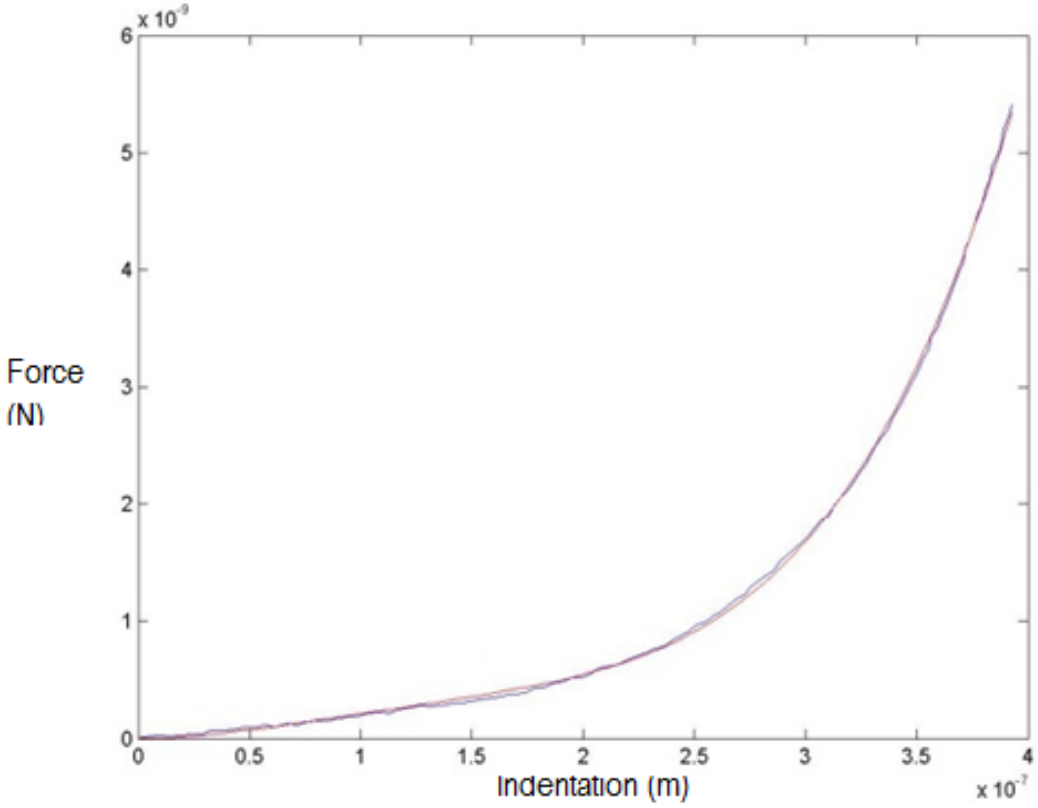


Figure 2-33 The improved extended Hertzian fit for the yeast data that removes the tail end of the data that is not of concern to the experiment. The blue curve is the raw AFM force data and the red curve is the extended Hertzian fit.

2.8 Discussion

The AFM measurement of force is not affected by measurement time at least in the short term (under an hour). The positive skew is only 1-3% meaning that for the experiments performed the time was not a factor, for experiments that could run longer than this would need to be extended to ensure that the force difference doesn't become significant.

The manual process to adjust the height above the cells for the AFM force data is cumbersome, involving trial and error for setting the setpoint and therefore results in multiple 'bad' results for one 'correct' final reading. While an automated process was developed that removes these extra readings it would be preferential that they not be there in the first place.

The initial contact point method was only a stop gap in order to enable testing to be completed. The tangent method is not ideal as it is right shifted along the x-axis from the contact point although this method was sufficient to produce results that were significantly statistically different.

The AFM force curves that were tested do not fit the Hertz model, in that they break many of the assumptions and exhibit strain hardening. The 3 parameter extended Hertz model fits improves the curve fitting but introduces 3 linked parameters which are difficult to analyse. The different plots (Ratio, 3d and flock) attempted to resolve the issue but it is difficult to determine one curve from another by simply looking at the Eb,K1 and K2 parameters in isolation. Instead of comparing the 3 parameters directly, the mean force at a certain indentation point was used to determine if significant differences existed between the populations.

2.9 References.

D.H. Ballard, "Generalizing the Hough Transform to Detect Arbitrary Shapes", *Pattern Recognition*, Vol.13, No.2, p.111-122, 1981.

Cary (ed), SAS User's Guide, Version 5, NC: SAS Institute Inc, 1985.

Conover, W. J. Practical Nonparametric Statistics. John Wiley & Sons, Hoboken, NJ, 1980.

Dintwa E, Tljskens E, Ramon H, On the accuracy of the Hertz model to describe the normal contact of soft elastic spheres. 2007. *Granular Matter.* 10:209-221. 2008.

R.O. Duda and P.E. Hart, "Use of the Hough transform to detect lines and curves in pictures", *Communs Ass. comput. Mach,* Vol. 15, pp. 11–15, 1975.

Gabryś, D. et al. (2007). Radiation effects on the cytoskeleton of endothelial cells and endothelial monolayer permeability. *International Journal of Radiation Oncology, Biology, Physics*. **69** (5): 1553–1562

Hall, A. "The cytoskeleton and cancer". *Cancer and Metastasis Reviews*. Vol 28, No. 1, pp 5-14. 2009.

Lam, L., Seong-Whan Lee, and Ching Y. Suen, "Thinning Methodologies-A Comprehensive Survey," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol 14, No. 9, pp 879, 1992.

Lin, D. C., E. K. Dimitriadis, and F. Horkay (2007a). Robust strategies for automated AFM force curve analysis—I. Non-adhesive indentation of soft inhomogeneous materials. J. Biomed.Eng., 129, 430–440.

Murphy, M.F. et al., (2006). Comparative study of the conditions required to image live human epithelial and fibroblast cells using atomic force microscopy. *Microscopy Research and Technique*, **69** (9), 757-765.

Pollard, T.D. & Berro, J. “Mathematical Models and Simulations of Cellular Processes Based on Actin Filaments”. *Journal of Biological Chemistry*. Vol 284, No. 9, pp 5433-5437 , 2008.

Randall, C.M., (2010). Investigating the Mechanical and Structural Properties of Human Cells by Atomic Force Microscopy. Ph.D. Thesis. *Liverpool John Moores University*.

Rudoy, Daniel, Shelten G. Yuen, Robert D. Howe, and Patrick J. Wolfe. "Bayesian change-point analysis for atomic force microscopy and soft material indentation." *Journal of the Royal Statistical Society: Series C (Applied Statistics)* 59, no. 4 (2010): 573-593.

Spudich & Watt. “Preparation of actin from rabbit skelletal muscle”. *JBC*, Vol. 246, pp 4866, 1971.

Wavemetrics. (2004) Technical Notes [online]. Available FTP: ftp.wavemetrics.net

Directory: IgorPro/Technical_Notes File: TN003.zip [2 March 2010]

Chapter 3 : Automatic contact point extraction methods

The contact point for an AFM curve is difficult to determine with high accuracy (Lin et al,2007; Rudoy et al,2010). The importance of the contact point is highlighted in the literature (Lin et al, 2007) and certainly in the case of automation this is only now beginning to be explored (Rudoy et al,2010). The methods which this project used progressed from an initial placeholder (angular) approach to the more robust (line projection) technique.

3.1 Angular contact point.

The initial method of determining the AFM tip contact point within the AFM force curve was via the tangent angular method. This was envisaged as being a relatively fast and simple preliminary method in order to be able to capture early data, due to the large accuracy error of the calculated contact point and would be refined later. The slope of the tangent at all points on the force-distance curve would be determined from left to right and when the slope's tangent at the point exceeded a predetermined 5 degree value the contact point would be assumed to be at this position.

The tangent method was attempted first without smoothing, though it was fully expected that due to noise the method would find the contact point too early. As expected the AFM noise triggered the contact point early, therefore smoothing was added to the pre-processing stage to stop noise from triggering the slope tangent's contact point position.

Smoothing the force-distance curve however introduces an accuracy error in that, once smoothed, the original contact point will appear to start earlier than it would in the noisy AFM force curve. The change in the contact point was approximately 3 nm when compared to an evaluation by human eye. The smoothing accuracy error however was minimal when

compared to the larger error involved in using the tangent slope, as this resulted in a larger change in the apparent position of the automatic contact point estimation compared to that which was obtained manually by human eye. Using the tangent slope made the automatically estimated contact point appear to start later and therefore be offset positively than the manually estimated contact point by 10nm. The precision of the method is acceptable being repeatable exactly, however the accuracy of 10nm was not for the purposes of this study.

3.2 Bayesian Change-point analysis

In the paper by Rudoy et al,2010 they implemented a Bayesian change point method for soft materials. This work was used here by the author, along with that of Lin et al, 2007, as a baseline to test against by comparing estimated contact points. The basic principle of the Rudoy method is the use of the post-contact region fitting in the model that is assumed to be in place for the tip, a power of x^2 or $x^{3/2}$ for the pyramidal and spherical indenters respectively. This leads to a switching regression model by comparison of the pre-contact and post-contact regions. In Rudoy's implementation there is assumed to be only one change point and though this is technically true, bumps in the curve could prove problematic.

The difference in the Rudoy model for the switching regression is based upon the polynomial degree. Assuming a pyramidal tip type, then this would be a degree of 2 post-contact and for the pre-contact data, a degree of 1. This method, as implemented in Matlab 2011b, takes a considerably longer time to process than the simpler methods like tangent slope, i.e. for the data we collected about 5 days to complete. This was on a Windows 7, 1.6 GHz i7 Core Laptop with 4Gb of memory. While the method still might work in

combination with another method like line projection, problems occur when it is used on its own, the contact point will often be found long after the actual contact point as shown in Figure 3-1, leading to its rejection as a method for the data that we collected. This is summarised in Figure 3-1.

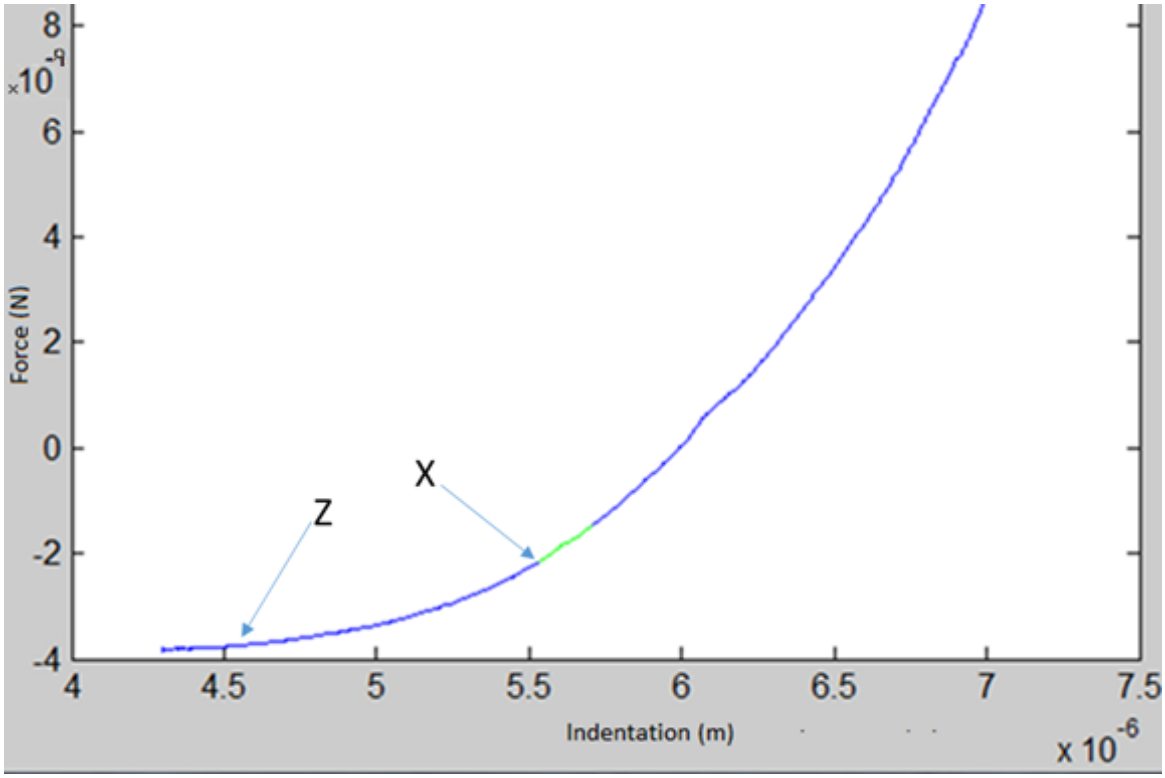


Figure 3-1. The contact point region X found from using the method by Bayesian switching regressions. While this matches with the x^2 portion of the curve to the right of the green section, in this case it's clear from the force applied that there is already contact before the green section with the area labelled Z being the most likely position of first contact. The force being almost 2 nano Newtons with respect to the Z point where the green section starts. It should be noted that the accuracy error in the apparent contact point is due to model unsuitability and not the method itself.

As shown in Figure 3-1, the start of the green section represents the contact point as found by the Bayesian change-point analysis approach. The force applied is approximately 1.5 nN at this point from the minimum force point and the travel distance is almost 1 μm from when the force curve starts to trend upwards. The author's conclusion is that the contact point is before the green section about 1μm, it should be noted that this is a model problem and not the algorithm itself, since the model assumes an x^2 portion to detect the change in slope, therefore the detection only triggers when the x^2 potion of the slope is encountered and therefore what is detected is the starting point for the x^2 portion of the graph.

3.3 Constrained, first derivative and power series.

The paper by Lin et al, 2007 gives an overview of alternate strategies for the automatic methods of determining the contact point and discusses their limitations. For constrained searches, the contact point must lie on the curve and within the ramp(The section of the graph which transitions from non-contact to engaging contact for the AFM), semi-constrained the contact point must lie within the upper and lower bounds of the measured data but is not necessarily an exact data point. The same is true for unconstrained sequential searches as semi-constrained searches. Lin at al. chose both a semi-constrained for positive slopes and constrained sequential search for non-positive slopes for their work due to the methods being amenable for automation. Other methods considered were extrapolation of the first derivative, which was rejected by the paper as the extrapolation is not always precise if there are errors in the contact region. Other factors affecting the derivative methods are that the uncertainty is higher when the sample is soft, which is certainly the case for our cellular studies. The power series with a Taylor series expansion. The power series has problems of bias if non-contact region data is included.

Among the tested contact models chosen were the powers of x^2 and $x^{3/2}$. The $x^{3/2}$ power version for spherical indenters and the x^2 version for other tip types.

The contact model is not accurately described by the constrained fitting, first derivative and power series approaches, at least for the soft samples that were under test. There are portions of the force-distance curve which the model would fit, but the 'ramp' section of the curve would be excluding data that belongs to the cell. This can be seen in Figure 3-1, by a straight section pre-contact, followed by a gentle slope gradually becoming steeper as the curve progresses. The gradual sloping section does not fit an x-squared contact model (the case would be for a pyramidal indenter) or a spherical $^{3/2}$ contact model (an Argument could be made that due to the manufacturing, the tip can not be precisely crafted the tip is often spherical for the initial indentation). Either contact model leads to problems when automation is applied in that the gentle slope, which to the eye is obviously part of the indentation in Figure 3-1, is ignored as 'ramp' the section of the graph where the AFM tip has not contacted the surface. The problem at least from the perspective of a specific contact model is not to fit one at all but to use something that relies on the shape of the curve. It should be noted however that if stiffer samples are being measured, a more generic approach may still give good results.

3.4 Line projection.

The contact method named ‘line projection’ was an effort to increase the accuracy of the automatic estimation of the contact point position by utilising the rate of change of the difference in slope. Although it’s a first derivative method, this algorithm smoothes the data locally. The smoothing that occurs locally over 100nm is due to the 50nm span from each section which will lead to some uncertainty in the contact point but allowing for faster processing 2 hours for all the data collected.

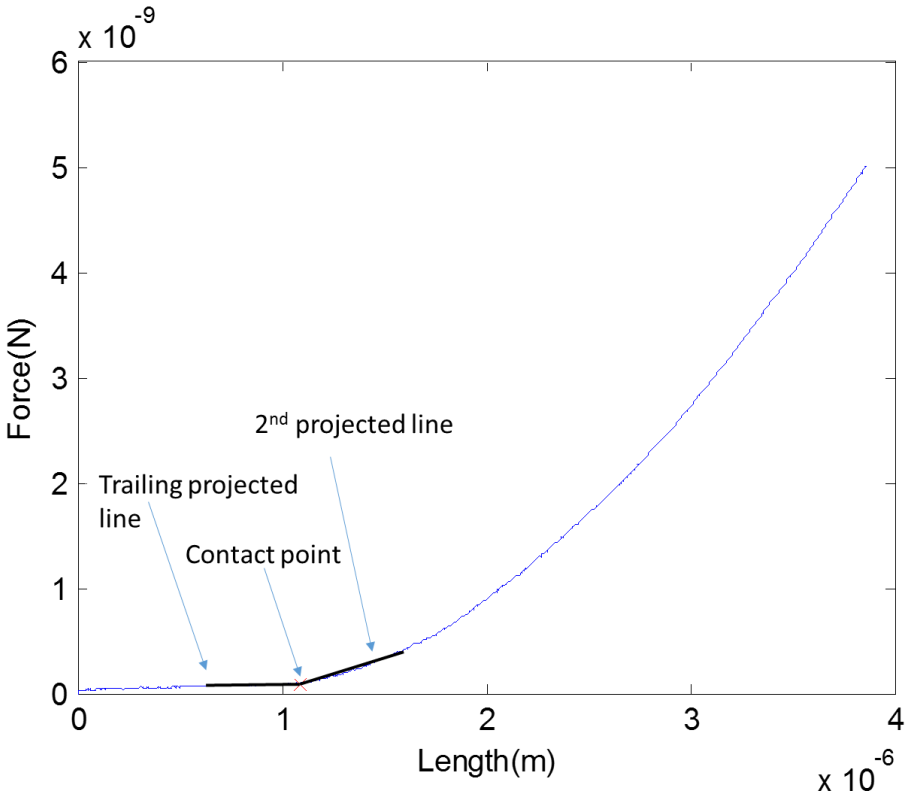


Figure 3-2. The projection of the line and the contact point found by the line projection method. The first (trailing) projected line acts as a predictor for the 2nd projected line, with the difference in height between the 2nd projected line and the trailing line being summed as an integral. The sum errors are back projected which creates the requirement that the lead-in be at least 50nm in length, or otherwise the method can fail to achieve a good prediction of where the contact point is located.

The method works by having three points that are equidistant from each other in the x axis (indentation). The points are spaced 50nm apart for a total range of 100nm. As shown in Figure 3-2, the points are placed on the noisy data, but it was hoped that they were located sufficiently far apart that the noise would even out, hence reducing the error. The method works by taking the first two points and calculating where the next point would be if they represented a straight line over the span from the mid-point to the end point of the projected line. The y difference between the calculated points and the points from the AFM force curve are stored in an array.

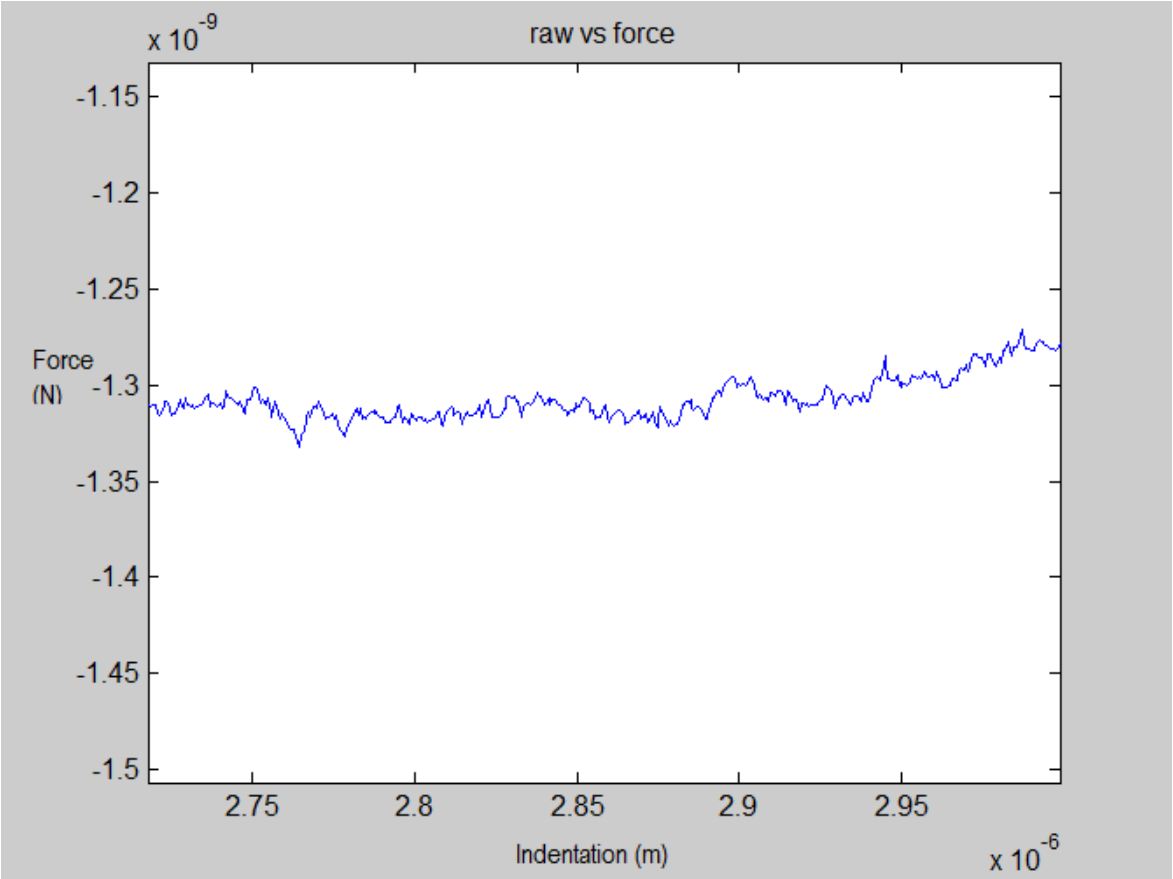


Figure 3-3. This is the indentation verses raw force 'zoomed in' to show the noise that occurs near the contact point for the AFM force curve.

Once the differences were found between the first trailing projected line and the 2nd projected line, as shown in Figure 3-2, the contact point was found, but this time not working from left to right, but instead working from right to left. Attempting to find the contact point by working from left to right was rejected as an approach, as this can include an occasional false positive due to AFM noise from curves that have non-random errors (bumps) in the initial pre-contact region. Therefore, that approach is not suitable without first avoiding the bump regions. Since working from right to left doesn't have this problem it was deemed to offer the best solution that avoided this issue.

The method works conceptually by assuming that when the change from the projected line stabilises, then the contact point has been reached. That was the original intention, i.e. to pick up the position at which the sum over the projected line area starts to deviate from the points before the line. This would in theory work as a possible approach, as the curve would gently trend upwards, thereby increasing the error as the line progressed. Then it would be a matter of calculating when the first point over the area started to deviate from the projection and this would become the estimated contact point. The theory here is that due to the noisiness of the pre-contact region the method instead goes down the curve from right to left until the zero point is determined for the large error sum array. The contact point would often be found late if the zero point was used due to the noise. The contact point instead of being calculated as the zero point is instead projected by measuring the difference between 2% and 1% then subtracting that distance from the location of 1% leading to the prediction of where the zero point would be this resulted in better accuracy to measurement by eye. This approach improved the accuracy of the contact point for the

data tested, when compared to finding the zero point directly from the summed array of errors.

3.5 Performance of the three methods used to calculate contact points.

A typical curve is shown below in Figure 3-4 with the various instances of the contact points that were found using Angular contact point, Bayesian change point and line projection.

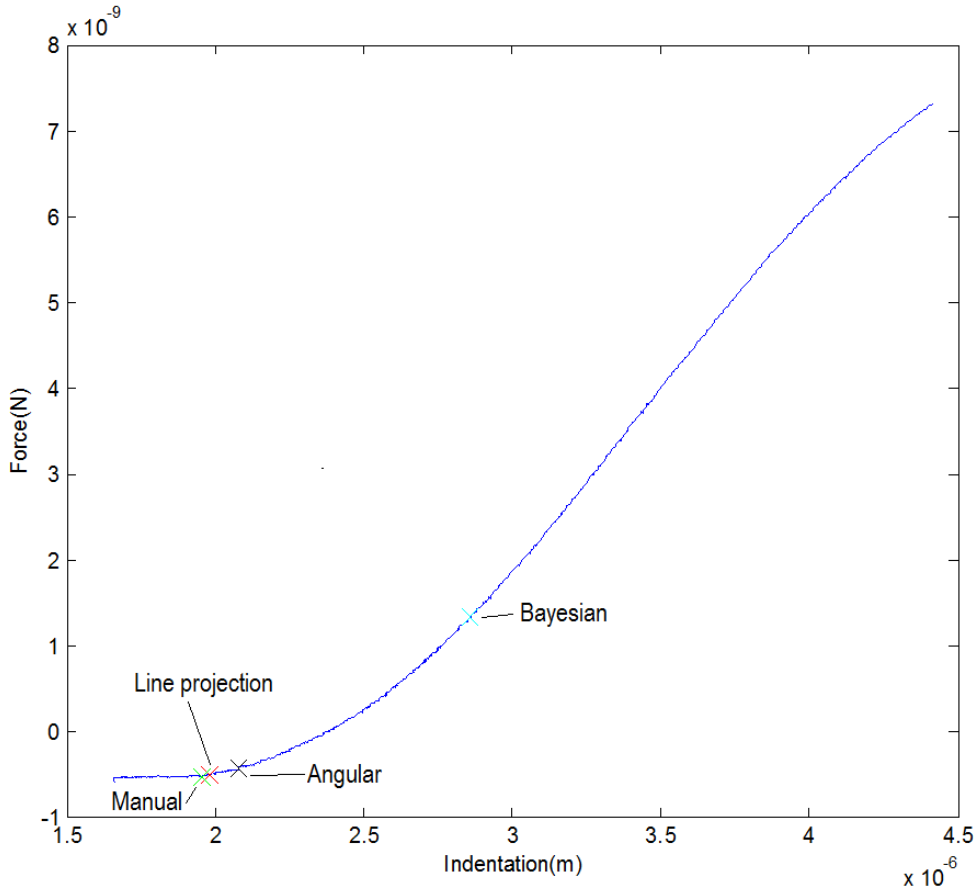


Figure 3-4. This is an example of the contact points found with each of the contact point methods. Manual is green and is the leftmost. Line projection is closest in red, the angular contact point is black and finally the Bayesian is in cyan.

	Line	5 degree	Bayesian
Difference	3.02E-08	1.26E-07	9.03E-07
From	4.86E-09	1.46E-07	9.62E-08
Contact	-1.98E-07	-1.43E-06	1.47E-08
Point	-9.83E-08	2.94E-07	1.28E-07
	2.67E-07	4.41E-07	1.81E-07
	7.90E-08	7.41E-07	2.51E-07
	1.60E-07	1.04E-07	4.38E-07
	-3.67E-08	1.01E-09	6.55E-08
	1.29E-07	5.68E-07	3.26E-07
	1.51E-07	-9.52E-07	5.75E-07
	7.58E-08	2.21E-07	1.27E-06
	1.70E-07	7.28E-07	2.34E-07
	2.52E-07	8.28E-07	6.00E-07
	3.46E-11	3.97E-08	6.25E-07
	9.77E-08	7.61E-07	9.51E-08
	1.99E-07	4.07E-07	2.57E-07
	8.93E-08	5.91E-07	3.34E-08
	8.62E-08	4.06E-07	1.40E-07
	8.27E-08	1.54E-07	1.05E-06
	1.85E-07	1.88E-07	4.23E-07
	-7.94E-08	2.14E-07	1.63E-07
	6.49E-08	4.97E-07	3.31E-07
	2.70E-08	2.13E-07	6.61E-08
	1.18E-07	8.48E-07	7.60E-07
	5.59E-08	9.89E-07	4.55E-07
	-1.81E-08	1.86E-07	3.00E-08
	1.76E-08	1.97E-07	1.06E-07
	1.35E-07	6.13E-07	4.12E-07
	-1.13E-07	3.23E-07	1.17E-07
	2.95E-08	8.00E-07	1.88E-07

Table 3-1 The raw contact point deviations from a sample of 30 curves. Column 1 is the line projection method deviations which are generally smaller than the other two methods. Column two is the 5 degree angular method, column three is the Bayesian change point analysis function.

The mean for each of the absolute errors from table 3-1 are as follows -:

1.017×10^{-7} for the line projection method.

4.67×10^{-7} for the angular method.

3.443×10^{-7} for the Bayesian change point method.

The minimum is the line projection method by a factor of 3 to the next closest the Bayesian change point method, therefore this is the reason that the line projection method was used as the primary way to automatically determine the contact point as given a random sample it typically provides the closest contact point with the least variation.

3.6 Section Conclusion

The assumptions of the line projection method as implemented are that the contact point lies on the force-distance curve, and that the contact point is within the lower and upper bounds of the measured data. Large positively sloped gradients judged manually to have already started are rejected and therefore projection would be the only way to ascertain the contact point which was not desirable for this project. The author assumed that the back projection would be difficult to fit and that, due to the exponential nature of the curves, as the contact point is reached there would be expected to be accuracy issues, dependant on the contact model chosen.

The line projection method robustness was increased by elimination of the large positive slopes. The method therefore doesn't have to handle the problems created by inaccurate AFM indenter tip positioning. The line projection method is also relatively fast compared to other techniques, such as the Bayesian approach 2 hours for line projection verses 5 days

for the Bayesian approach. Also the results can be achieved without any human intervention, the contact points generated are more accurate when judged empirically and the algorithm will generate the same figure for the same data so therefore it's more repeatable as well.

3.7 References.

Lin D, Dimitriadis E, Horkay F. Robust Strategies for Automated AFM Force Curve Analysis—I. Non-adhesive Indentation of Soft, Inhomogeneous Materials. *Transactions of ASME*. 2007 ; 129 : 430-440.

Rudoy, Daniel, Shelten G. Yuen, Robert D. Howe, and Patrick J. Wolfe. "Bayesian change‐point analysis for atomic force microscopy and soft material indentation." *Journal of the Royal Statistical Society: Series C (Applied Statistics)* 59, no. 4 (2010): 573-593.

Chapter 4 : A new model for investigating the biomechanical responses of epithelial human cells from AFM force-displacement curves.

4.1 Introduction

The typical model used for measuring Atomic Force Microscopy (AFM) force-displacement curves is the Hertz model. The Hertz model in the context for AFM is a primarily a method to determine the bulk Young's modulus for the cell undergoing investigation. The Hertz model makes a number of assumptions, for example that the deformation is small, the deformation is within the elastic limit of the material and linear, the surface is frictionless, the volume being deformed is small compared to the overall size and that surface is continuous (Dintwa *et al*,2008).

The Hertz assumptions are mostly violated within this application, as cells are not continuous homogenous entities, instead having various discrete organelles within them. The cellular surface is not frictionless and the level of deformation caused by a probing AFM tip can be large compared to the overall size of a cell. Also, the elasticity of cells behaves non-linearly.

As the biological cells behave in a non-linear fashion, a one parameter model may not fully describe all of the underlying features. A balance must be achieved however between increasing the number of parameters to capture underlying features and over-fitting random elements present in AFM measurements of the force displacement curves, causing a decrease in the ability of the fitting function to generalise.

The error in the fit can be calculated and this is used in various fitting models to give an error term that can be used to assess how model fits the data under test. In the case of the

AFM force curves the error is the square of the force difference summed over the span from the contact point to the end of the AFM force curve. This gives an error value which can be used to judge the quality of the models fit is compared to another model. The problem is that the more parameters are used the better the fit will become, this is where AIC and BIC are useful in that they penalise extra parameters to avoid overfitting the model. AIC and BIC are slightly different in the way they handle extra parameters, while AIC follows information theory, BIC does not and penalises the extra parameters more strongly, due to the break from information theory AIC was chosen as the model selector rather than BIC.

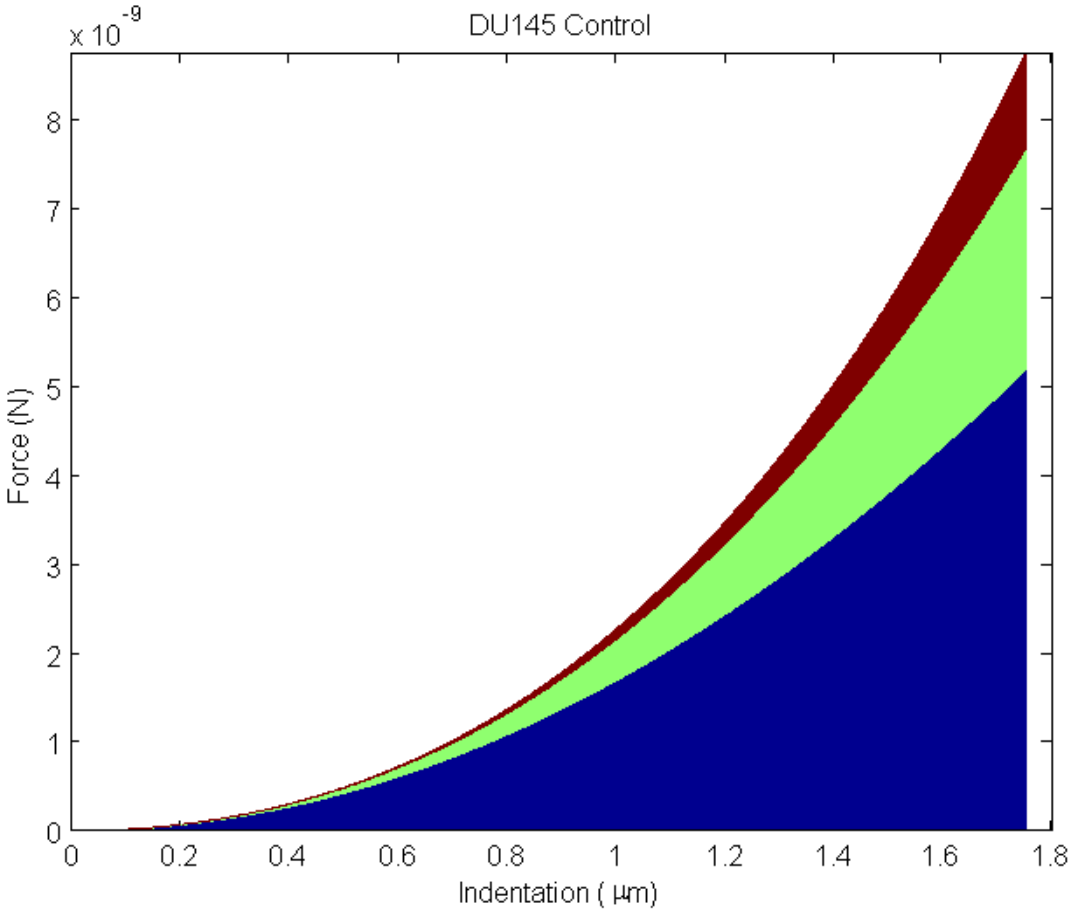


Figure 4-1. The relative force contributions to total force for DU145 human prostate cancer cells for the Young's modulus extended to a quadratic, the blue region represents the contribution from the x^2 term and the light green from the cubic term and finally the red region represents the quartic term.

The model first implemented was the extended Hertzian model (Randall, 2009). This model has 3 parameters that allows the Young's modulus to vary as the force increases, as shown in Figures 2-28 and 4-1. Analysis revealed that the 3 parameter model, while obtaining an excellent Goodness of Fit (GOF), had interpretation problems in a physical sense. When the model was least squares fitted, negative components for the force were

generated. This fact, while increasing the GOF, also led to the parameter values oscillating. They would sometimes oscillate between negative and positive values, depending upon the specific tip contact point being used. The negative component for the cubic term does not make logical sense, as typically any cell process will attempt to utilise the minimum amount of energy required to perform a specific function or task. Would a cell have two forces pushing out and one acting against the other two? This seems unlikely and therefore it was deemed necessary to limit the force contribution of the cell model to positive forces only. Also it was decided to allow other types of variation of the Young's modulus to be considered. Hence, it was decided to allow the Young's modulus to vary in powers and incorporate exponentials, as detailed later in Section 4.2.2.

4.2 Materials and Methods

4.2.1 Cell Culture

Prostate cancer cells (DU145), were obtained from the European Collection of Animal Cell Cultures (ECACC), Porton Down, Salisbury UK. The DU145 cells were grown in RPMI medium supplemented with 10% foetal calf serum (Biowest Ltd. UK), penicillin (100 IU/ml) and streptomycin (100 μg/ml)at 37 °C in a humidified 5% CO2/95% air atmosphere.

The AFM measurements were performed using an Asylum Research MFP-3D microscope. The cantilever was calibrated using the thermal noise method and an Olympus IX50 inverted optical (IO) microscope was used to position the AFM tip at the centre of the cell to be probed. The AFM cantilevers used were V-shaped silicon nitride Olympus cantilevers OMCL-TR400PSA-1. The DU145 cells were seeded onto glass coverslips at a density of $1x10^5$ cells/cm^2 and left until a uniform monolayer formed.

The force curves measured from the AFM were then processed automatically. Briefly, force curves for the DU145 cell measurements were scaled to nN for the force and μm for the indentation to help ensure that rank deficiency errors did not occur when solving. The point of contact was determined by a gradient method similar to (Lin et al. 2007).

4.2.2 Fit functions.

The following are the parameters that were passed into the Matlab 'fit' function, with the minimum value set to 0.0 and using the damped least squares Levenberg-Marquardt fitting function from the curve fitting toolbox in Matlab 2011a.

The formulae for the one parameter fit models used in Matlab are as follows:

ad (4.1)

ad^2 (4.2)

ad^3 (4.3)

ad^4 (4.4)

$ae^d - a$ (4.5)

d^a (4.6)

a^d (4.7)

$aLog(d)$ (4.8)

For the two parameter fit models:

$ad + bd^2$ (4.9)

$ad^2 + bd^3$ (4.10)

$ad^3 + bd^4$ (4.11)

$ad + be^d$ (4.12)

$ad^2 + be^d$ (4.13)

$ad^3 + be^d$ (4.14)

$ad^4 + be^d$ (4.15)

$ad^3 + b^d$ (4.16)

$ad^3 + d^b$ (4.17)

$ad^2 + b^d$ (4.18)

$ad^2 + d^b$ (4.19)

$ad + b^d$ (4.20)

$ad + d^b$ (4.21)

$ad^4 + b^d$ (4.22)

$ad^4 + d^b$ (4.23)

And finally, for the three parameter fit models:

$ax + bd^2 + cd^3$ (4.24)

$ad^2 + bd^3 + cd^4$ (4.25)

$ad^3 + bd^4 + cd^5$ (4.26)

$ad^2 + bd^3 + cd^5$ (4.27)

$ad + bd^3 + cd^4$ (4.28)

$ad + bd^4 + cd^5$ (4.29)

$ae^d - a + bd + cd^2$ (4.30)

$ae^d - a + bd^2 + cd^3$ (4.31)

$ae^d - a + bd^3 + cd^4$ (4.32)

$ae^d - a + bd^4 + cd^5$ (4.33)

$ae^d - a + bd^2 + cd^4$ (4.34)

$ae^d - a + bd^2 + cd^5$ (4.35)

$$ae^{d} - a + bd^{3} + cd^{5} \quad (4.36)$$

$$ae^{d} - a + bd + cd^{3} \quad (4.37)$$

$$ae^{d} - a + bd + cd^{4} \quad (4.38)$$

$$ae^{d} - a + bd + cd^{5} \quad (4.39)$$

$$ae^{d} - a + be^{d^{2}} - b + cd \quad (4.40)$$

$$ae^{d} - a + be^{d^{2}} - b + cd^{2} \quad (4.41)$$

$$ae^{d} - a + be^{d^{2}} - b + cd^{3} \quad (4.42)$$

$$ae^{d} - a + be^{d^{2}} - b + cd^{4} \quad (4.43)$$

$$ae^{d} - a + be^{d^{2}} - b + cd^{5} \quad (4.44)$$

$$ae^{d} - a + bd + cd^{5} \quad (4.45)$$

Each of these 45 individual formulae had the AIC score calculated and then was tested to work out which function had the highest relative likelihood with respect to the others for a particular cell population. These functions are used to extend the classical Hertz model equation into a non-linear form.

4.2.3 Extending the Hertz equation.

The Hertz formula for a conical indenter is given by,

$$F = \frac{2}{\pi}\frac{E}{(1-\nu^2)} tan\alpha\ \delta^2 \qquad (4.46)$$

The Poisson ratio (ν) for the cells was taken to be 0.5 meaning that the material is perfectly incompressible at small strains, and an opening angle (α) for the indenter was taken from the manufacturers specification to be 35 degrees. The force (F) is the force for the conical indenter as a result of the indentation deformation (δ). The Young's modulus (E) is a measure of the tensile stress divided by the strain. The AFM force curves were fitted for the 45 different formula variations in Section 4.2.2. For each formula, the AFM force curve was fitted using the Levenberg-Marquardt option within the curve fitting toolbox in Matlab with the minimum value set to zero for the force contribution parameters. Setting this to zero ensured that no negative force components could be attained for each of the customised formulae. The Levenberg-Marquardt algorithm was used for the 1 parameter model, as this was empirically the best option that fitted the cell population data consistently without adjusting the starting parameter. For the 2 and 3 parameter models the default non-linear least squares option (finite differencing) worked successfully. The empirical starting parameter for the 1 parameter model was a= 0.185. For the 2 parameter model, the empirical starting parameters were a=0.234054 and b=0.355939, these parameters were calculated from actually fitting the curve for formula 4.9 and using the fit itself for the parameters, this is why the numbers for a and b are not rounded up. For the 3

parameter model, the empirical starting parameters were a=0.1340541, b=0.155939 and c=0.109191, again these were calculated from a curve fit this time using formula 4.25 the fit itself becoming the starting parameters used for the other formulae.

Once the fits were obtained, the formulae were compared against each other. From previous work (Randall, 2009) the 3 parameter model had been shown to have an acceptable fit level, so therefore the Akaike information criterion (AIC) was determined to be an appropriate method for comparing the respective models. The AIC compares the accuracy against the complexity of the formula model, where the number of parameters is used as a negative weight against the accuracy of the model (Akaike, 1973). The AIC value for each of the formulae was then calculated and compared to all the other formulae for that specific AFM force-distance curve, under the assumption of normally distributed errors. The number of times each formula was the 'best' fit was then measured against the cell population and a histogram produced against each of the formula under test. A measure of the statistical relative likelihood was also calculated for the classical Hertz model against the attained model to ensure that the relative likelihood was increased significantly enough to support the validity of extending the model to 3 parameters.

4.3 Results.

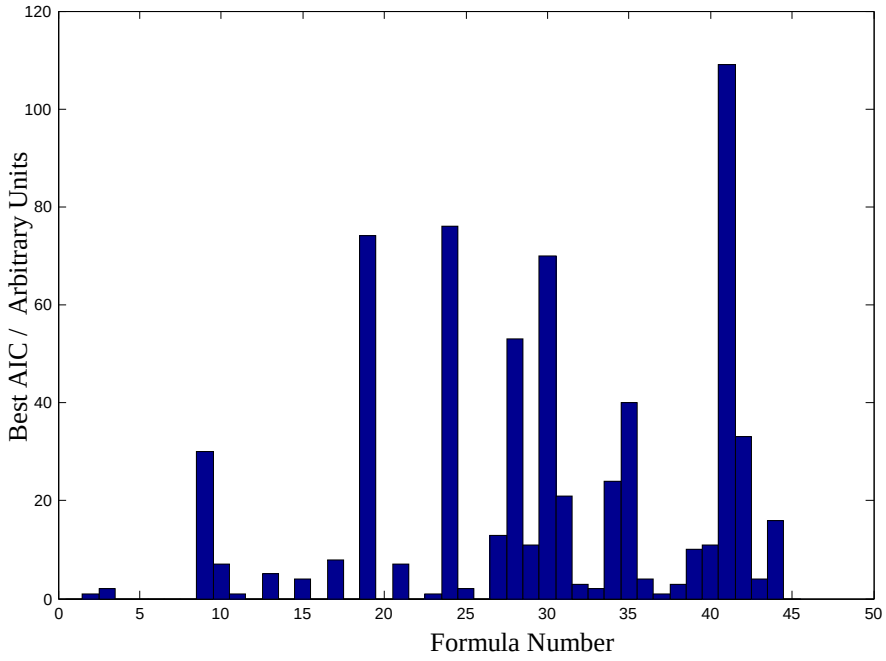


Figure 4-2. Akaike Information Criterion versus the different non-linear model formulae being tested. This shows the formula being tested verses the number of times the formula was chosen as the maximum likelihood for that cell population. The formula with the highest peak is formula 41. This model is represented by Force = ae^x – a + be^{x*x} –b + cx^2. This approach should choose the model that matches the general trend in the underlying data, that also has the fewest number of parameters.

There were 646 cells tested for the DU145 prostate cancer cell population each having 45 different AIC values calculated after the contact point had been determined automatically from the raw force curve Igor Binary Wave files that were produced by the AFM. At a 44% likelihood level the formula ae^x – a + be^{x*x} –b + cx^2 has a greater likelihood at a level that is greater than 0.1% better than that of the classic Hertz model, based on the AIC criterion. The ability of AIC to determine the best fit is limited in that it does not indicate how good the actual fit is only that the fit is best among those being tested. This is not really a concern for the data in that the 3 parameter fit for formula 4.25 (The extended Hertz formula) was

already seen to be a good fit to the data being tested. The penalisation of extra parameters in AIC is less than that of the Bayesian Information Criterion(BIC) this means that it could prefer more parameters and therefore may lead to overfitting. BIC however was rejected in that it doesn't follow information theory. AICc which also adds an extra penalty to additional parameter was also considered but wasn't required as the number of data points was sufficiently large 300+ that the change to the formula for AICc would be insignificant.

	Eb	Extended	Formula 41
Rmse (nN)	0.062	0.027	0.031
	0.041	0.034	0.009
	0.032	0.026	0.014
	0.058	0.058	0.058
	0.032	0.032	0.028
	0.066	0.066	0.043
	0.065	0.061	0.019
	0.063	0.063	0.034
	0.082	0.078	0.018
	0.098	0.073	0.023
	0.219	0.134	0.029
	0.119	0.083	0.044
	0.078	0.078	0.051
	0.176	0.176	0.170
	0.088	0.032	0.038
	0.377	0.377	0.377
	0.279	0.279	0.250
	0.152	0.152	0.085
	0.157	0.157	0.131
	0.165	0.130	0.115
	0.591	0.161	0.210
	0.092	0.089	0.050
	0.457	0.179	0.105
	0.333	0.124	0.219
	0.185	0.077	0.107
	0.227	0.227	0.076
	0.048	0.048	0.016
	0.611	0.514	0.403
	0.233	0.233	0.201
	0.203	0.203	0.158
	0.246	0.246	0.200
Mean	0.182	0.136	0.107

Table 4-1 A table showing the root mean square errors(rmse) found from a sample of 30 curves. The hertz equation (4.2) typically has the highest rmse and therefore has the worst fit in general. The Extended Hertzian fit (4.25) is typically the second best fit and finally formula (4.41) is typically the best fit having the lowest rmse. The mean rmse also reflects this trend with (4.41) having the lowest mean of 0.107 nN, 0.136 nN for the Extended Hertz(4.25) and finally 0.182 nN for the Hertz equation(4.2).

4.4 Discussion.

Criteria for measuring how to determine a model's suitability include: Does it make sense physically? Can it be explained within the context of cell mechanics? Do the parts make sense? Does the model describe the observed data and can it predict future data? How complex is the model ? (Pitt et al., 2004).

The low likelihood of 0.1% for 44% of DU145 cells for the classical Hertz model results is surprising when compared to the results for a three parameter model, given that the model chosen includes an x^2 term. Previous unpublished work by the author had showed that when allowing the stiffness parameters to become negative resulted in a roughly 50% contribution from the x^2 term when allowing the Youngs modulus to vary quadratically.

In terms of the model complexity, the classic Hertz 1 parameter model is obviously simpler, in terms of having just a single changing parameter. However, the three parameter AIC relative likelihood is much smaller than the single parameter model when the AIC criterion is applied. The additional parameters are providing a more accurate fit to the extent that the likelihood ratio for the AIC criterion is beyond the 0.1% likelihood for 44% of the DU145 cells measured. The goal of the AIC comparison was to choose the most accurate model, versus the model having the least parameters. AIC most chosen from repeated measurements was selected as the model that explained data consistently well, that is the AIC with the largest score. The formula Force = $ae^x - a + be^{x*x} - b + cx^2$ is not

based on physical reality, but fits the data more consistently than the other alternatives. This means that among those formulae considered here it has the best AIC score. From a random sample of 30 curves the formula (4.41) also has the lowest mean root mean square error as shown in table 4-1. Interestingly it is not always the best formula for a particular curve as occasionally formula (4.25) has a lower root mean square but this only occurs three times from thirty or about 10% of the time. Thus 90% of the time formula (4.41) provides the best fit for the data that was sampled. The root mean square errors in the fit align with the results from the AIC scoring with formula (4.41) scoring the best, (4.25) lower and finally (4.2) being poorly scored.

Among the formulae that attained the highest scores, the alternatives all had an x^2 component present and an x term. While an x^2 term can be explained in terms of the pyramidal tip type, that has a geometrical ratio of 1:1 and relating that to the volume of the cone, the x linear term presents a challenge to explain in physical cell terms, as it is actually a volume that is being displaced. It is possible that an exponential feature is present when indenting with a conical tip, but to gain further clarification on these issues would require more testing.

4.5 Section Conclusion

The standard model for measuring the force response of biological cells over relatively small distances, i.e. less than 500nm, is the classical Hertz model. The Hertz model is not sufficiently accurate to model indentation with respect to applied force for distances that are greater than 500nm due to the strain hardening effect that is readily observable in

experimental practice for living biological cells. The biological cells' non-linear response to applied force by the AFM tip becomes marked after approximately a 500nm indentation level. This non-linear force response leads to curve fitting errors that increase with the degree of indentation and therefore this necessitates a different method of modelling the cellular AFM force response. The initial method developed by GERI Cell Mechanics Group was an 'extended' Hertzian model(Randall,2010), which extends the Young's modulus to include a squared polynomial term varying with indentation. The extended model significantly reduced the fit error, but introduced a problem of physically explaining the model parameters. This was more apparent when considering that the best force curve fit would often lead to one of the parameters being negative. The search for a model that can be explained physically as well as mathematically was undertaken by the author and in this chapter the preliminary results have been shown for the human prostate cancer epithelial cell line DU145.

4.6 References.

Akaike, H. Information theory and an extension of the maximum likelihood principle. *Second international symposium of information theory*. 1973 :267-281

Cross, S.E. et al. AFM-based analysis of human metastatic cancer cells. *Nanotechnology*. (2008). **19** 384003 (8pp)

Dintwa E, Tijskens E, Ramon H. On the accuracy of the Hertz model to describe the normal contact of soft elastic spheres. *Granular Matter*. 2008;10(3):209–221.

Lin D, Dimitriadis E, Horkay F. Robust Strategies for Automated AFM Force Curve Analysis—I. Non-adhesive Indentation of Soft, Inhomogeneous Materials. *Transactions of ASME*. 2007 ; 129 : 430-440.

Pitt M,Myung I, Zhang S. Toward a method of selecting among computational models of cognition. *Psychological review*. 2002 ; 109 (3): 472-491.

Randall, C.M.,. Investigating the Mechanical and Structural Properties of Human Cells by Atomic Force Microscopy. 2010 Ph.D. Thesis, Liverpool John Moores University.

Chapter 5 : Energy Normalisation Transformation.

This chapter introduces a new approach to examine AFM force distance curves that is based upon a consideration of the curves as being representations of the energy exerted by the indenting AFM probe tip upon the soft living cell samples. In all cases here the AFM probe tip is positioned at a single spatial X/Y point over a living cell's surface and the cell is indented using a standard approach-retract AFM force spectroscopy cycle.

The 'energy normalisation transform' described here uses the raw AFM force-distance curve to calculate the area under the curve between two discrete indentation points considered along the continuous 'approach' section of the indentation cycle. The force-distance curve, when integrated between these two indentation points, becomes work done, since force x distance = work done. Therefore, the area under the slope of the force curve represents the energy, or work done, by the indenting AFM probe, with units of Newton metres (Nm). The conversion uses the indentation energy, i.e. the AFM probe tip energy exerted when indenting living cells, at two points, namely a starting energy level of 0.5 x 10^{-15} Nm and a maximum level of 2.25 x 10^{-15} Nm. The difference between the starting and maximum energy levels has been termed the "Energy Length" by the author, by virtue of the fact that it encompasses a specific energy level, which for the current experiment was 1.75 x 10^{-15} Nm. The number 1.75 x 10^{-15} Nm was empirically derived by ensuring that the staring energy level of 0.5 x 10^{-15} Nm was far enough from the contact point that the energy level was not subject to large changes. The maximum energy level was again empirically chosen to ensure the achievement of the most normally distributed results.

Another primary benefit to this approach is the effect of reducing the contact points influence on the resulting energy length. When the area under the contact point is considered or work done, the slope is generally very flat and ideally the work done would be zero but errors in measuring the contact point would still result in a low energy value since the shape of the force curve at this point is for the most part horizontal and therefore the amount of total energy for an error in measuring the contact point is low when considering the area under the curve with respect to the contact point.

The energy normalisation transformation examines the non-normal distribution of the force curve data and how to transform that data into a normal distribution. The primary motivation behind this conversion from a non-normal to a normal distribution is the importance of normalisation in most parametric comparative statistical tests, like the t-test. Non-parametric tests have less power to discern whether the distributions are significantly different and therefore it is preferable, if possible, to attain a normal distribution before such testing, or the experiments will risk determining that there is no statistically significant difference, when a difference actually exists. In biological investigations such as the radiation studies into cell biomechanics that have motivated this work, the ability to discern statistically significant differences between test samples is of great importance.

The Young's modulus (E) distributions from the AFM force curves have the property that they are not normally distributed and therefore this means that normally distributed parametric tests aren't applicable. The energy normalisation transformation is a bespoke transformation function written in Matlab 2011a, which is a conversion function that takes a non-normal statistical distribution of force curve data and transforms it into a normal distribution.

The non-normal data in this case is the set of Young's modulus values from all the force curves for a given specific individual cell within a particular cell population (for example 'PC3 Control'). The value of the Young's modulus (E) is calculated for each AFM force-distance graph by using the AFM tip's contact point as the origin of the curve and subsequently curve fitting an x^2 fit to the force curve data. Each of the E values calculated in this manner are placed into the histogram and are deemed to represent the distribution of values for E and a typical example is shown in Figure 5-2. The normalised data is created from all the Energy Length values, where the Energy Length is calculated between two energy levels within the force curve namely 0.5×10^{-15} Nm and 2.25×10^{-15} Nm. The same cell population is used for both the raw E values and the Energy Length values to enable a comparison to be made between them.

5.1 Generation of the E histograms.

E is the value of the Young's modulus for a given cell. An E histogram is the distribution of these values for a collection of cells arranged automatically by Matlab 2011a into discrete groups. An example of an E histogram can be seen in Figure 5-2 for PC3-Control data. There are roughly 160 force curve graph entries for Figure 5-2 and it can be seen from the figure that the histogram has a positive skew. The E histograms are generated by using Matlab 2011a, by cycling through all the cells types within the author's custom SQL database for a given radiation exposure level (for example PC3-Control data), then using the contact point that is already stored in the database, and subsequently fitting using a bespoke curve fit routine to an x^2 fit for the force curve. The purpose of the initial E histogram is to show the distribution of the graph before the normalisation method is

applied and therefore this data acts as a baseline for subsequent comparison of the normalised data.

5.2 The Energy Normalisation Transformation Method.

5.2.1 Calculating the indentation length required to travel to a particular set energy level.

The energy normalisation transform works on an AFM force-distance curve - see Figure 5-1 for a typical AFM force-distance curve. When the force-distance curve is integrated, the integration represents work done, since force x distance = work done. Therefore the area under the slope of the force curve represents the energy, or work done, by the indenting AFM probe, which has units of Newton metres (Nm).

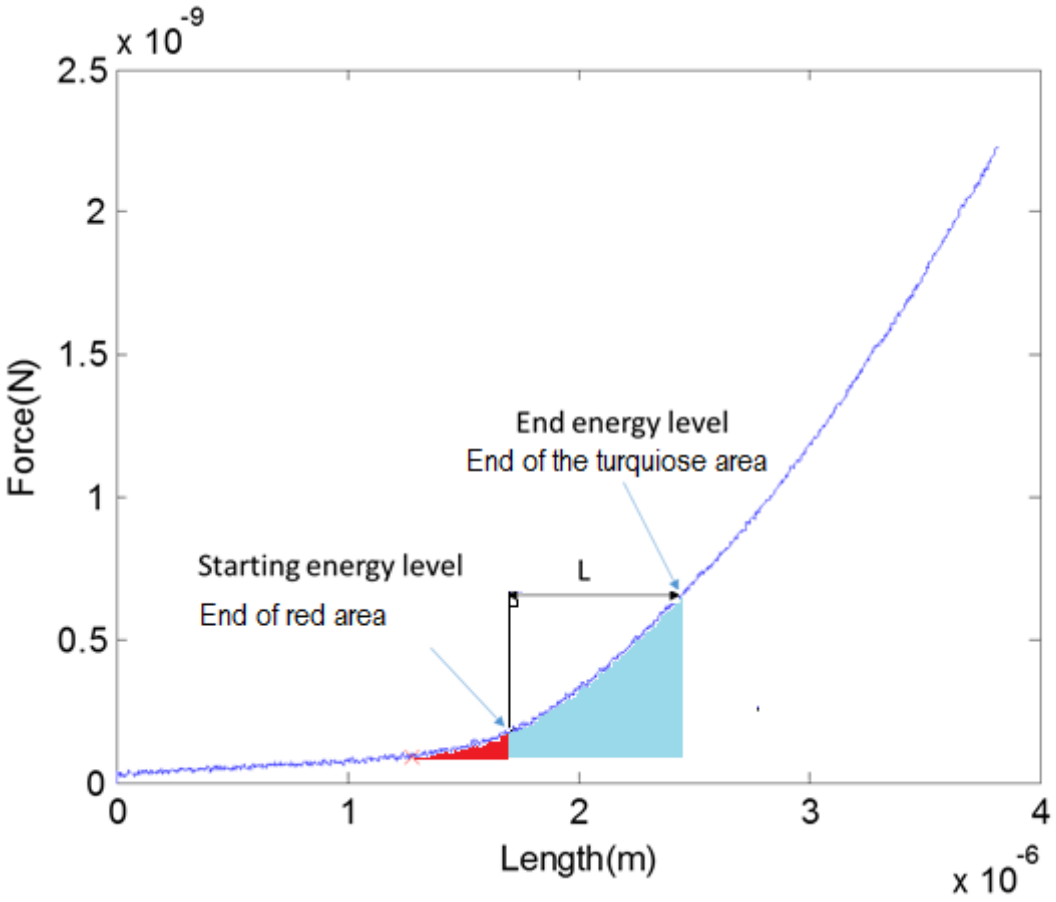


Figure 5-1 Example of the transformation method measured between two points on the AFM curve. The starting energy level position is the red area under the curve, the turquoise level represents the end energy level and includes the red area. L – The value that is obtained from the starting energy level to the end energy level positions, the force is held constant and the length is the variable.

To calculate the 'energy length' requires the calculation of two discrete energy levels, namely a starting energy level, and an ending energy level. The 'energy length' is calculated by measuring between those two energy levels. In the case of the energy normalisation transformation the starting energy level was 0.5 x 10^{-15} Nm also known as the starting energy point, and the ending energy point 2.25 x 10^{-15} Nm. Where these points fall on the curve is calculated by finding the closest array entry for the force curve from the force curve data. The calculation gives two entries in the array, which can then be used to ascertain the starting energy length and finishing energy length and therefore the total

energy length, L (shown in Figure 5-1 above), can be recorded and stored for an individual reading. Each of the individual entries for the AFM data like the one shown in Figure 5-1 are cycled through in this manner generating an energy length for each force-distance curve. The energy lengths are grouped by cell type and radiation exposure level, thereby creating a subset of the overall data within the database.

5.3 Levels of kurtosis, cell stages and looking at the unbounded data.

The kurtosis value for a normal distribution is approximately 3. As can be seen in Table 5-1 below, this is obviously not the case for almost all of the Young's modulus E data presented here. A clipping bound removing the top 5% of the total amount of data was introduced to minimise the effects of extreme outliers within the cell population data. The extreme outliers are caused by the very different force responses of a small number of cells that are at different stages within their mitotic cell cycle.

		Kurtosis	
Cell Type	Gy 1	Gy 2	Gy 3
PC 3 Control	15.03	13.82	13.82
PC 3	10.64	3.1474	5.074
PNT2 Control	6.861	6.659	6.6
PNT2	29.63	15.4	27.65

Table 5-1 The baseline levels of kurtosis for the raw E values at the corresponding irradiation Gray levels. PC3 are prostate cancer cells, Prostate normal tissue is PNT2. The distributions here are leptokurtic, a mesokurtic distribution would have a kurtosis of 3 for the Matlab implementation.

Table 5-1 shows the levels of kurtosis within the raw Young's modulus, E, data. Although the experimental control samples are not irradiated, the controls are exactly procedurally matched to the exposed irradiated cells. Normally there would be only one control, but as the controls were procedurally matched each time the experiment was performed, there are

therefore 3 controls, each matching the different levels of radiation exposure, even though the controls have actually received no radiation exposure. The reason behind the existence of triple controls here lies behind the fact that the waiting period for the control cells is different in each case. This protocol ensures that the controls are as closely matched as possible to the irradiated cells, having undergone every procedure and time delay that the irradiated cells went through except for any actual exposure to radiation.

	Kurtosis		
Cell Type	Gy 1	Gy 2	Gy 3
PC 3 Control	2.263	2.092	2.092
PC 3	2.724	2.748	3.626
PNT2 Control	3.481	3.571	3.571
PNT2	3.626	4.279	4.483

Table 5-2. The levels of kurtosis for the transformed energy length values at the corresponding Gray levels. PC3 are prostate cancer cells, Prostate normal tissue is PNT2. The resulting distribution is more spread out when compared to Table 6-1 but more importantly is close to the mesokurtic value of 3. The removal of outliers is not enough to lead to the data becoming normally distributed and requires the introduction of the energy length to make nearly all the distributions become normal for the 12 cells under test.

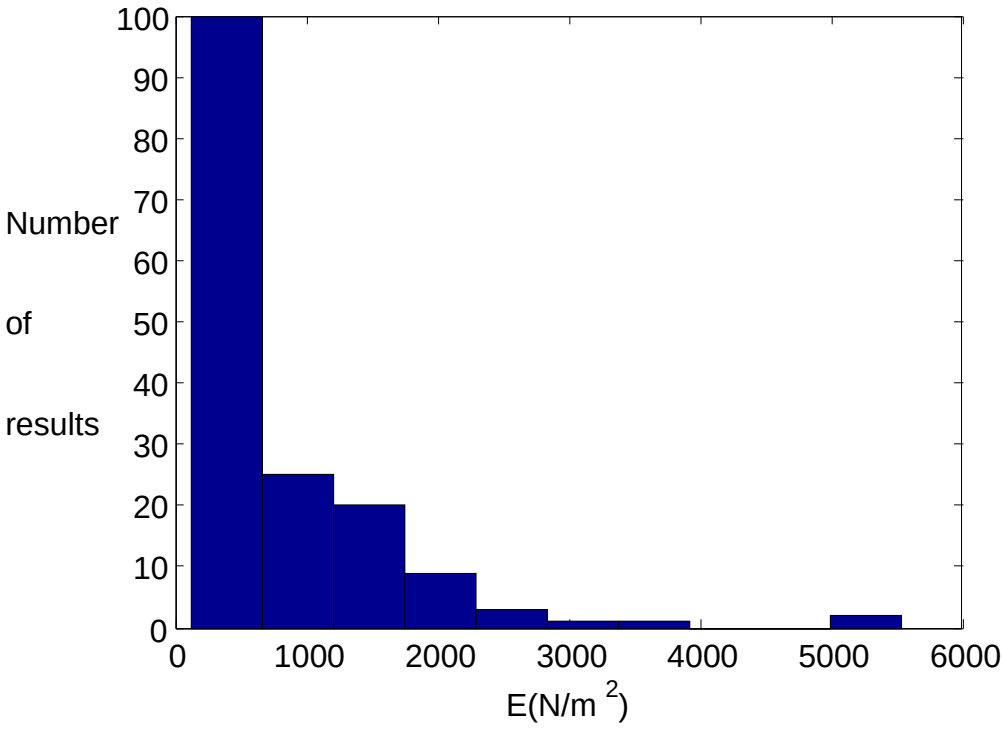
Table 5-2 shows the levels of kurtosis after both a 5% clip at the upper bound for the data and after undergoing the normalisation energy length transformation. The resulting kurtosis is much closer to the value 3, with one exception where the data was already close to 3.

5.4 Production of histograms for the energy values.

When the statistical outliers are removed 11 of the 12 energy length distributions become normal under a Jarque-Bera test (Jarque and Bera.,1987). This test is less sensitive to the tails of the data than the Anderson-Darling test (Matlab, 2015). This is a desirable feature given that outliers will still be present in the data (even in the case of 5% clipped data) and

would thus lead to leptokurtic data, as can be seen in the upper histogram shown in Figure 5-2.

The next several figures are for PC3 cells, which are a prostate cancer cell line.



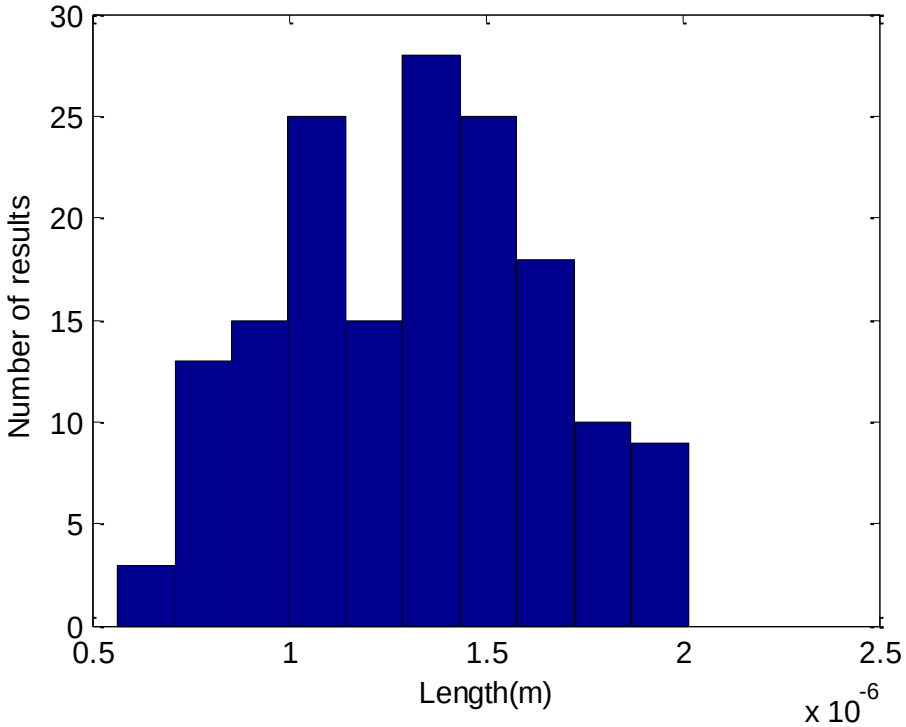


Figure 5-2 Histograms showing PC3 Control, before conversion (upper) and after transform to energy length conversion (Lower)

The PC3 control data presented in Figure 5-2 shows the raw Young's modulus, E, values depicted in the upper histogram. These values display a very high level of skew towards the left, whereas after the transformation to energy length is applied (as shown in the lower figure) a more centralised distribution is obtained. The same is true for both the controls and non-control data. The boxplots shown later in Figure 5-4 show that there are no statistically significant differences to be found in the energy length data for ionising radiation exposure at a dose of 1 Gy. It should be noted that the untransformed data showed the same results as the transformed data in that there were no statistical differences found in Youngs' modulus for the untransformed data or found in the energy lengths for the transformed data between the irradiated cells and the un-irradiated cells.

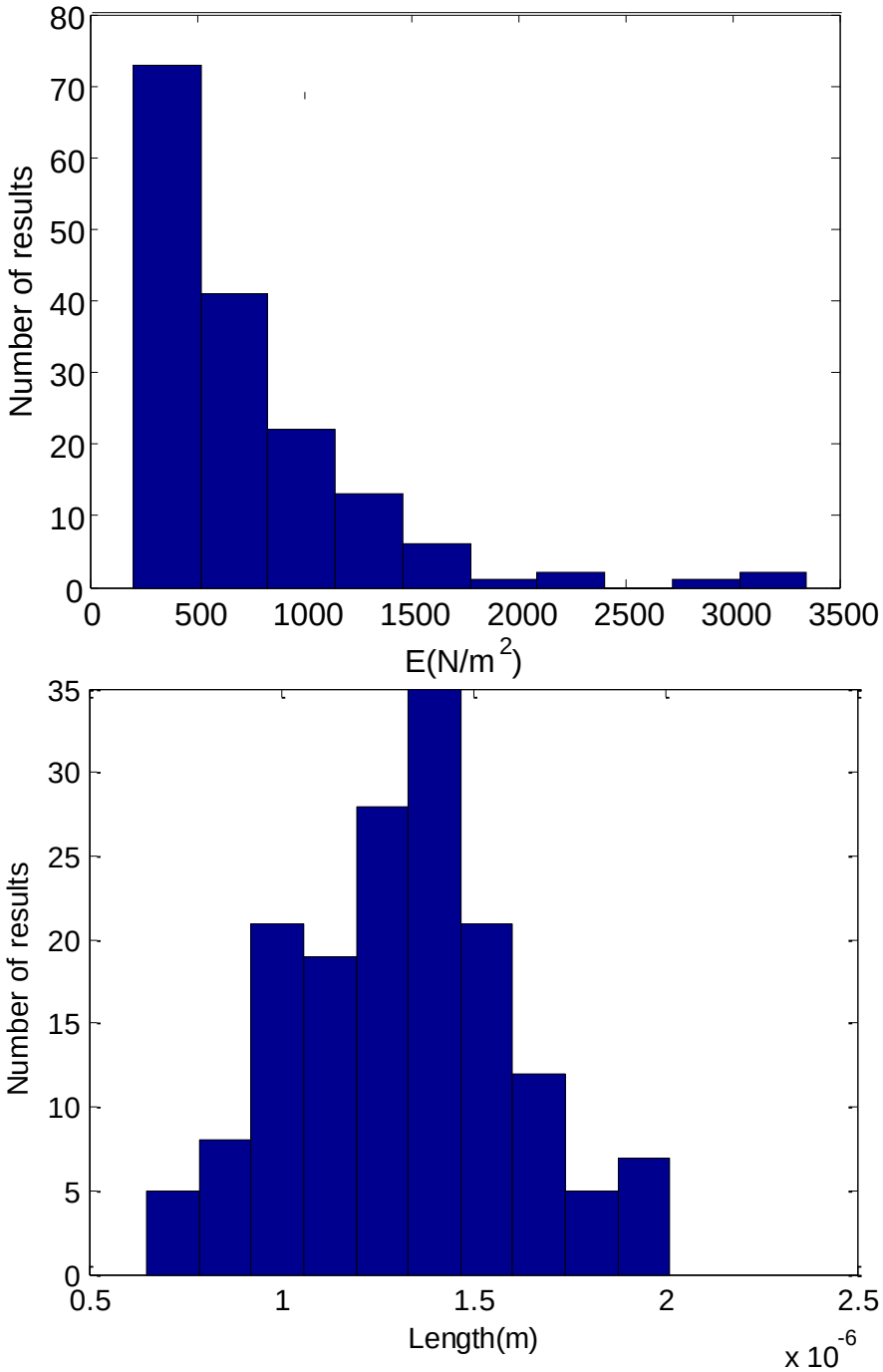


Figure 5-3 Histograms showing PC3 Non-Control with a radiation level of 1 Gy, before conversion (upper) and after transformation to energy length form (lower).

In Figure 5-3 the raw E values histogram (upper diagram) has a kurtosis value of 10.64 and this is also evident by the skew to the left as seen in the figure. Of particular note here are the outliers which are present around the 3000 N/m^2 mark on the x axis. When converted

using the energy normalisation transformation method (see lower figure) the statistical outliers are first removed. The resulting histogram has a kurtosis value of 2.724 and is normal when tested using the Jarque-Bera test (Jarque and Bera., 1987) within Matlab 2011a.

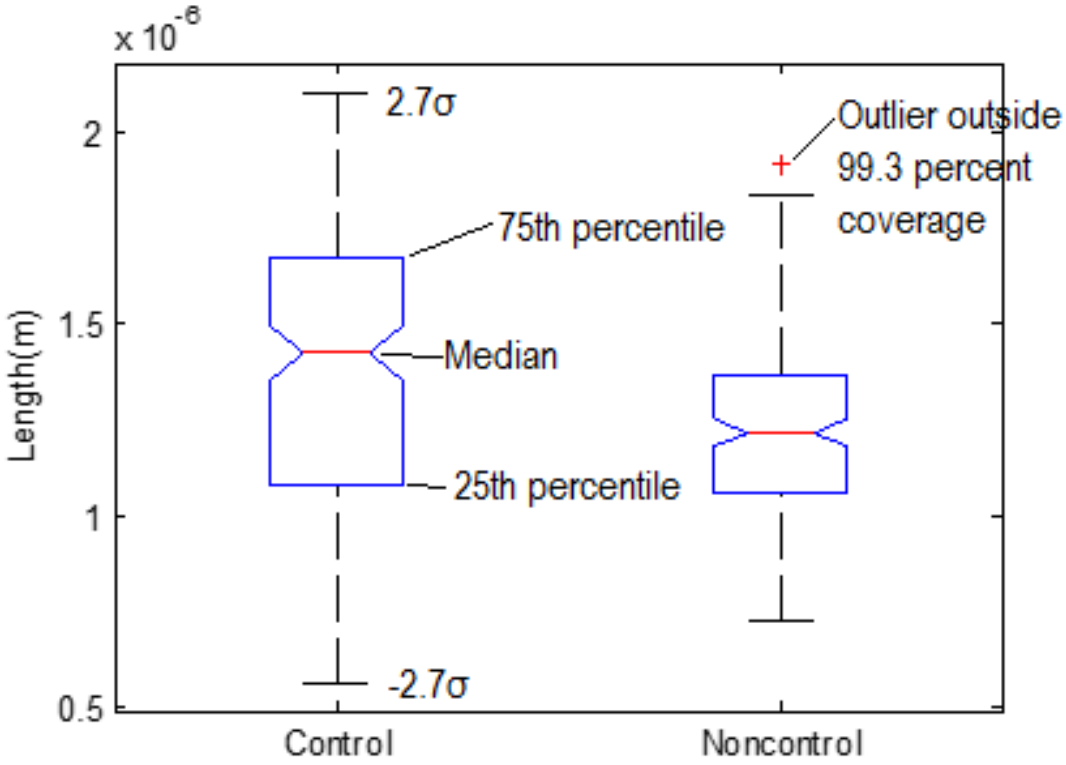
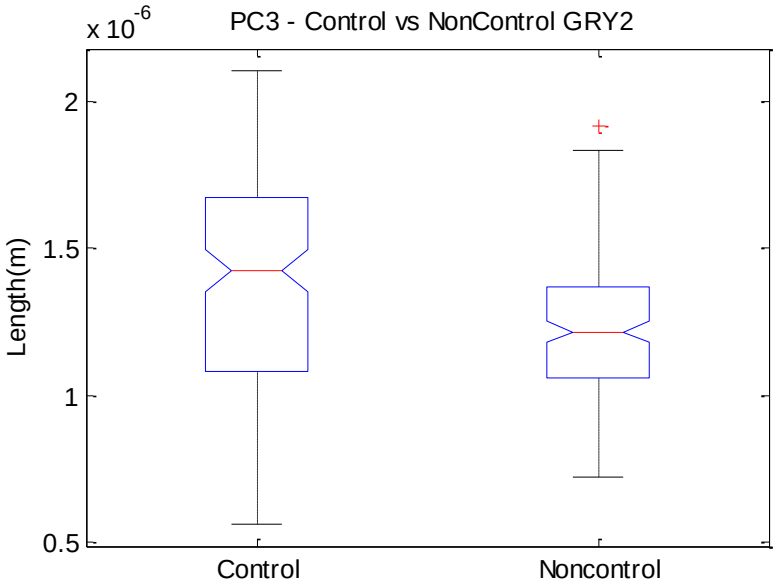


Figure 5-4 The standard box plot used throughout this chapter. The median is represented the red line in the middle of the box. The edges of the boxes are the 25th and 75th percentiles. The whiskers represent if the data is normally distributed the 99.3 percent coverage (+/- 2.7σ). The outliers are marked as red crosses which is the data that lies outside the whiskers.



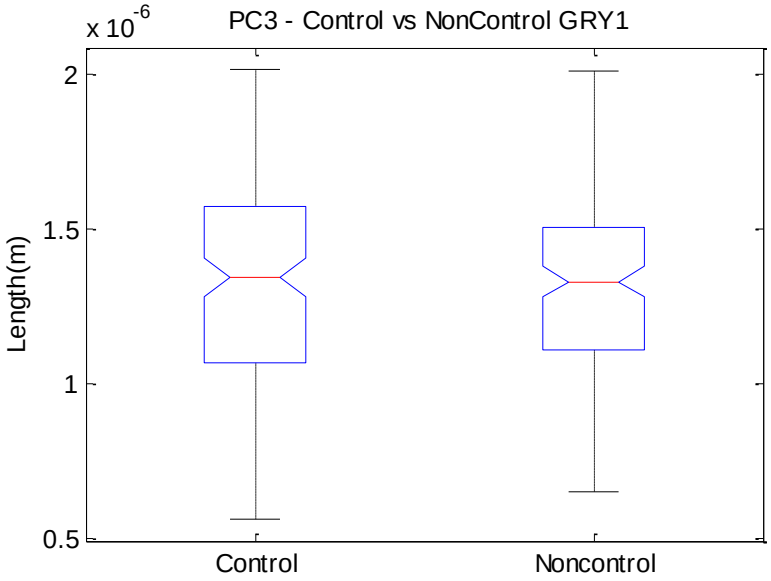


Figure 5-5 (Lower) Boxplot showing PC3 Control and PC3 Non-control with a radiation exposure level of 1 Gy. (Upper) PC3 Control and PC3 Non-control data with a radiation level of 2 Gy.

The PC3 control and non-control data presented in Figure 5-5 shows the boxplots for both PC3 1Gy and PC3 2Gy readings, versus appropriate controls. There is no statistically significant difference between the energy lengths at 1 Gy, as can be seen by the overlapping notches within the boxplot diagram. At 2 Gy, however there is a statistically significant difference in energy lengths indicated by the non-overlapping sections of the boxplot. This shows that in the case of a 2 Gy dose of ionising radiation exposure the radiation dosage has altered the mechanical response of the cell to make it less malleable. A lower energy length value indicates that the energy is spread over a shorter distance and therefore the energy per m is higher than that of the non-exposed cell data, hence indicating a stiffening of the cell.

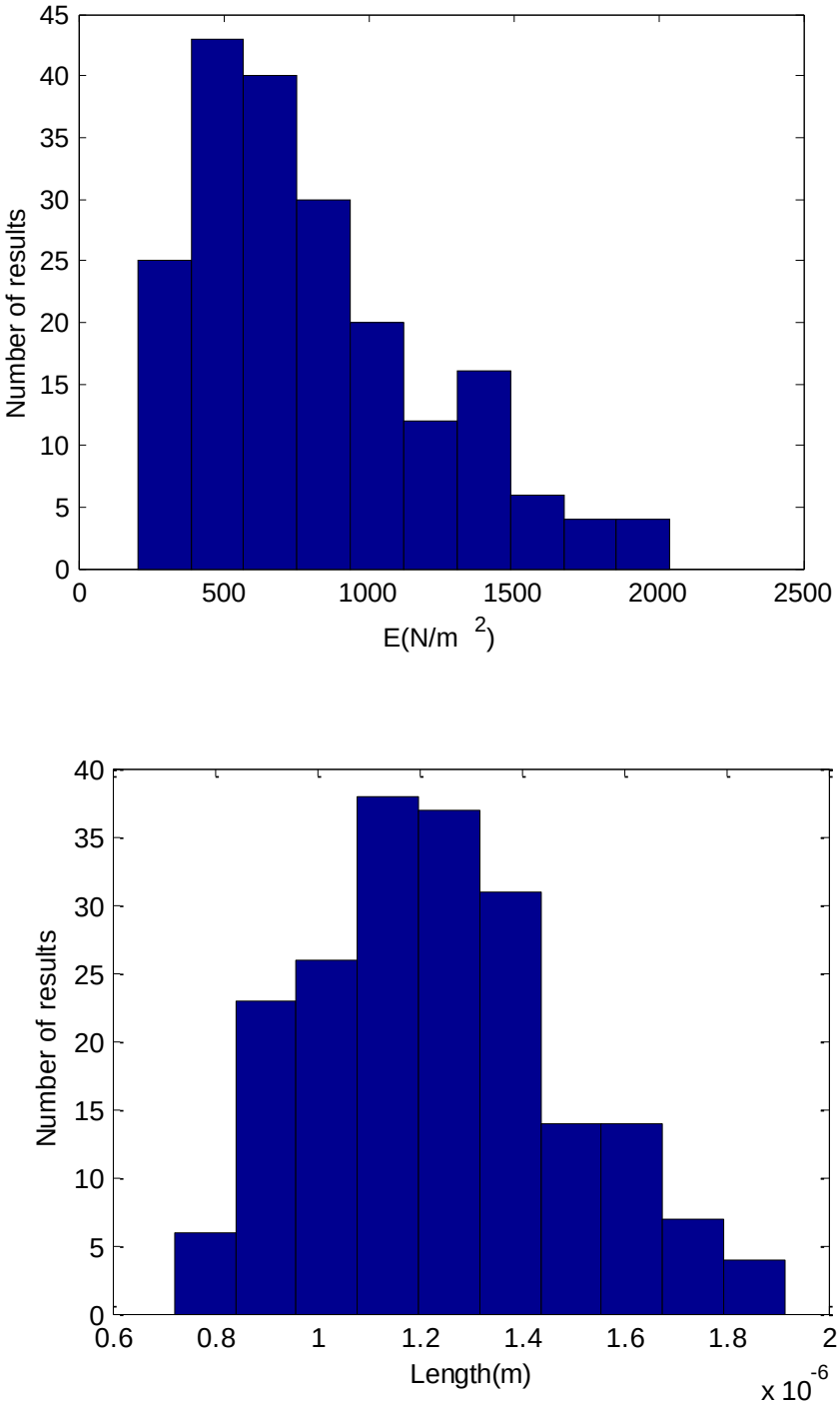


Figure 5-6 Histograms showing PC3 Non-Control with a radiation level of 2 Gy, before conversion (upper) and after transform to energy length conversion (Lower).

The distribution of raw E values for PC3 Non-control data irradiated at a dose of 2 Gy is presented in Figure 5-6 (upper) and shows a slight skew to the left. The conversion to

energy length however produces a more centralised grouping. The kurtosis factor for the upper figure was 3.15 and for the lower figure was 2.75. Therefore, it could be argued that in this case the energy normalisation transformation was not necessary. The energy normalisation transformed data was a normal distribution, whereas the raw Young's modulus data is not a normal distribution.

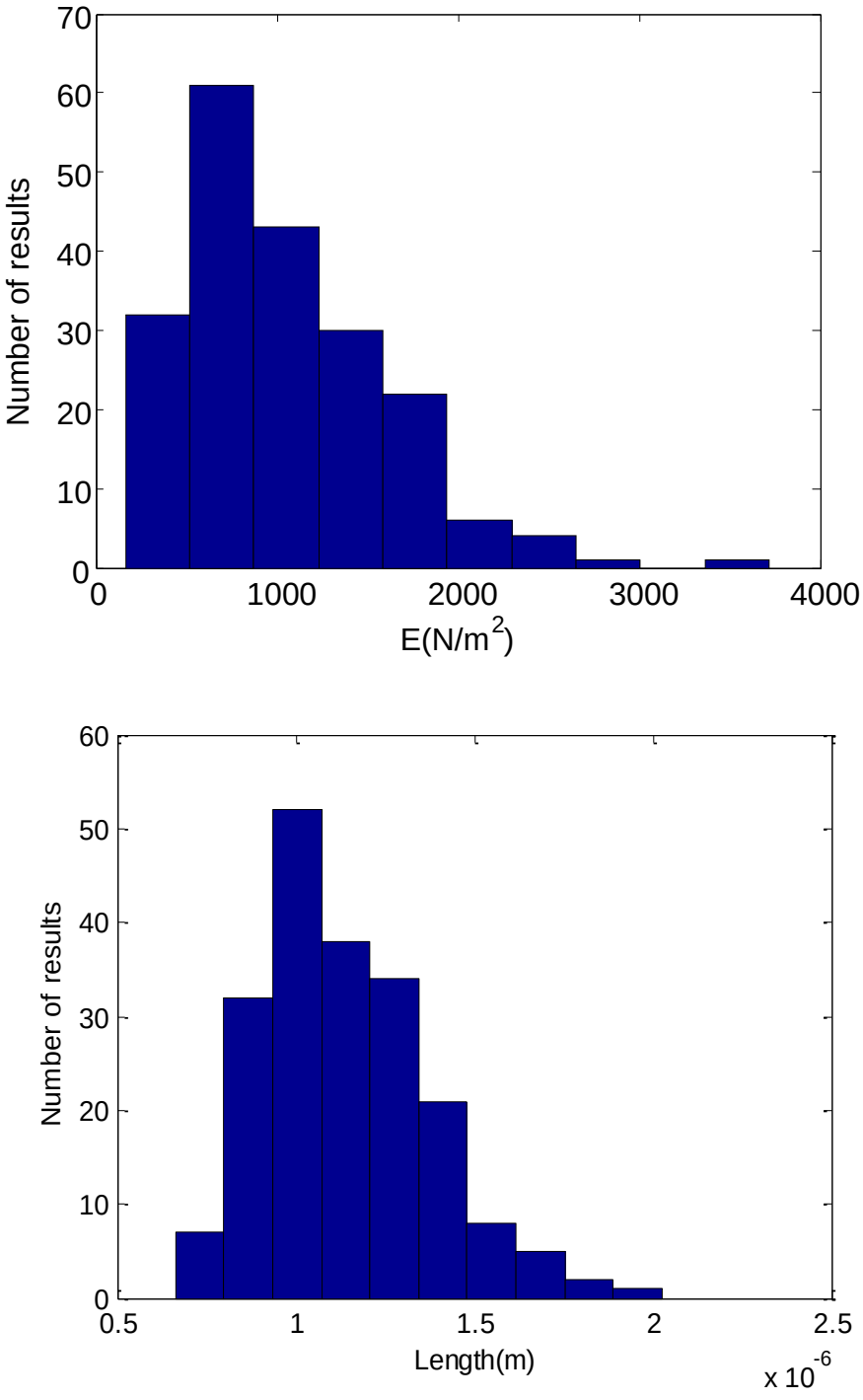


Figure 5-7 Histograms showing PC3 Non-Control with a radiation level of 3 Gy, before conversion (upper) and after transform to length conversion (Lower).

The PC3 non-control data in Figure 5-7 upper shows the non-normal data that is skewed to the left with a long right-hand tail. The conversion to energy length format produces a more centralised grouping, although there is still a tail to the right of the data distribution. The kurtosis factor for the raw E data in the upper figure was 5.07 and the kurtosis value for the lower figure was 3.63, representing the transformed data. Therefore, again the conversion to energy length format has caused the distribution to tend towards normality and this was confirmed with a Jarque-Bera test (Jarque and Bera.,1987) or JBtest in Matlab 2011a.

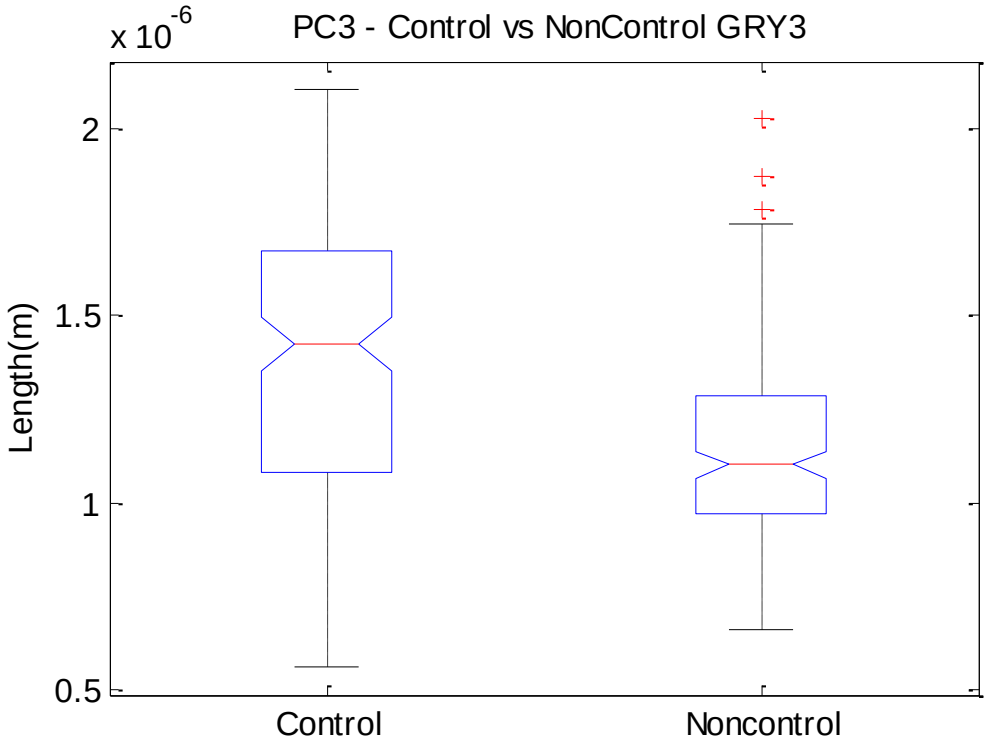


Figure 5-8 Boxplot showing PC3 Control and Non-control with a radiation level of 3 Gy.

The PC3 Control and Non-control data at 3 Gy that is presented in Figure 5-8 shows the boxplots for energy lengths of PC3 cells irradiated at a radiation dose of 3 Gy versus non-irradiated controls. There is a significant statistical difference between the responses of the irradiated cells and the non-irradiated controls, indicated by the control and non-control boxplot indentations not overlapping. These suggest that in the case of the 3 Gy irradiated samples the radiation dosage has altered the mechanical response of the cell to make it less malleable. A lower energy length value indicates that the energy is spread over a shorter distance and therefore the energy per m is higher than that of the non-exposed cell data, indicating a stiffening of the cell. It should be noted that this appears to indicate a progressive stiffening of the cell from 2 Gy to 3 Gy, with a 3 Gy radiation dose having a shorter energy length than is the case for 2 Gy.

The next figures show data for PNT2 cells, from a prostate normal tissue cell line.

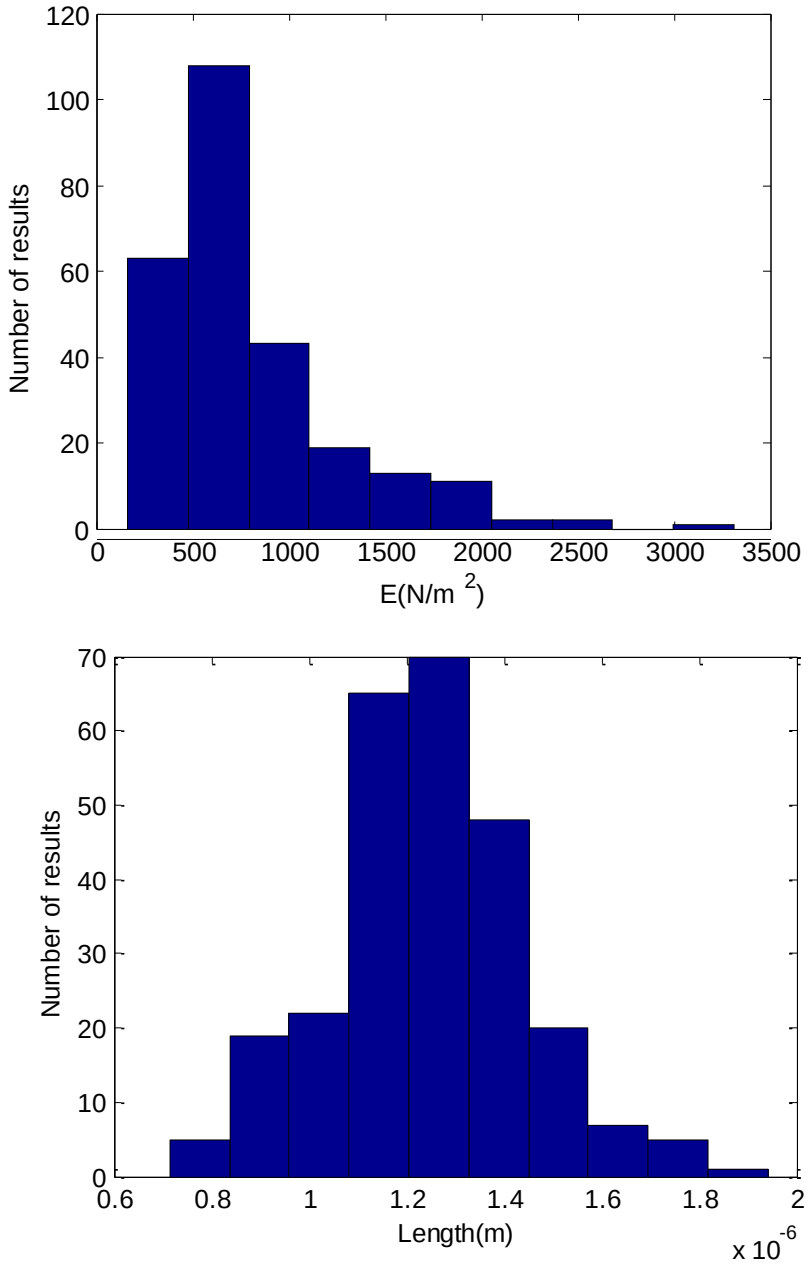


Figure 5-9 Histograms showing PNT2 Control, before conversion (upper) and after transform to length conversion (lower)

Figure 5-9 (upper) shows the non-normal raw E data that is skewed to the left with a long right-hand tail. The conversion to energy length format produces a more centralised grouping. The kurtosis factor for the raw E data in the upper figure was 6.86 and for the transformed data shown in the lower figure was 3.48. Therefore, once again the conversion

to energy length has caused the distribution to tend towards normality. The distribution in this case appears to be more mesokurtic.

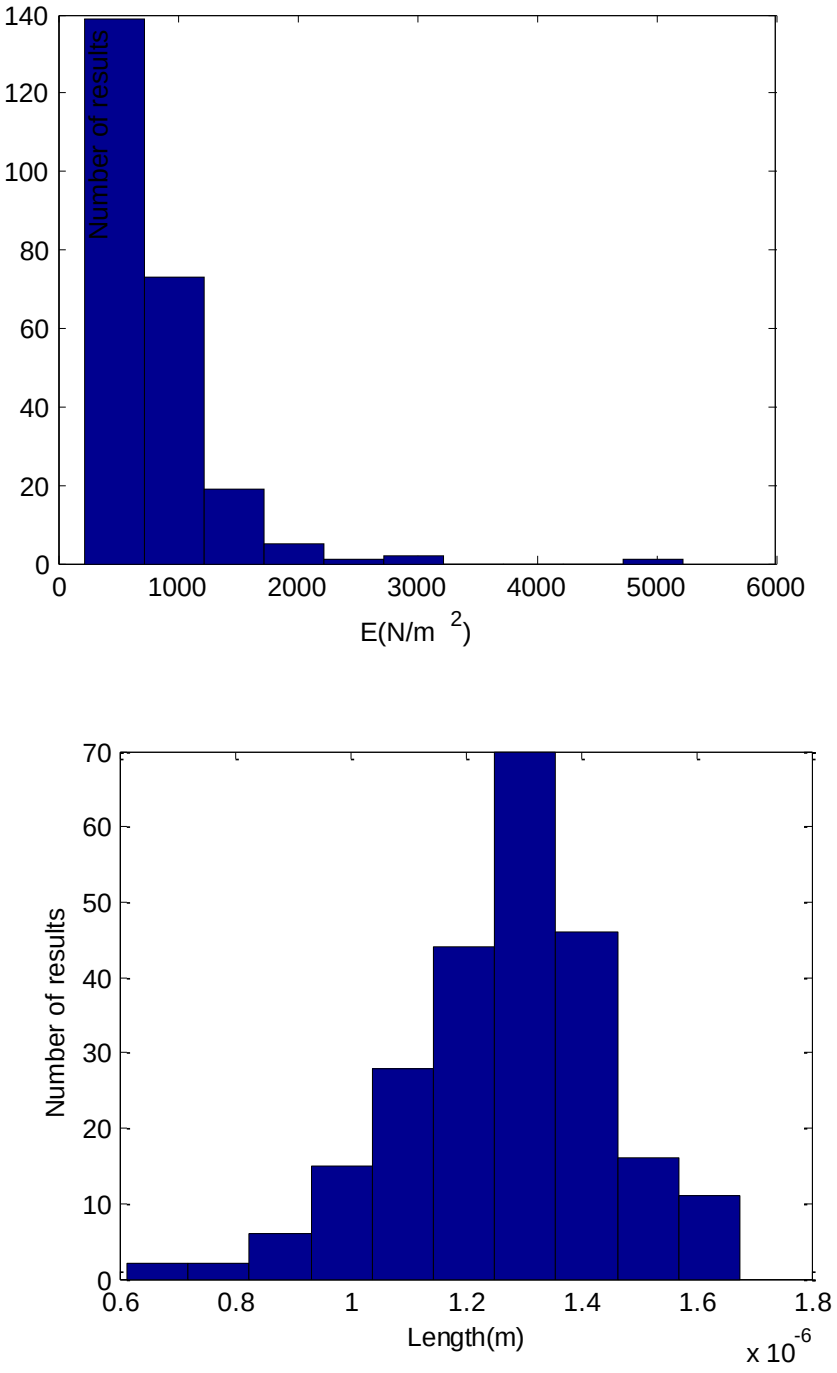
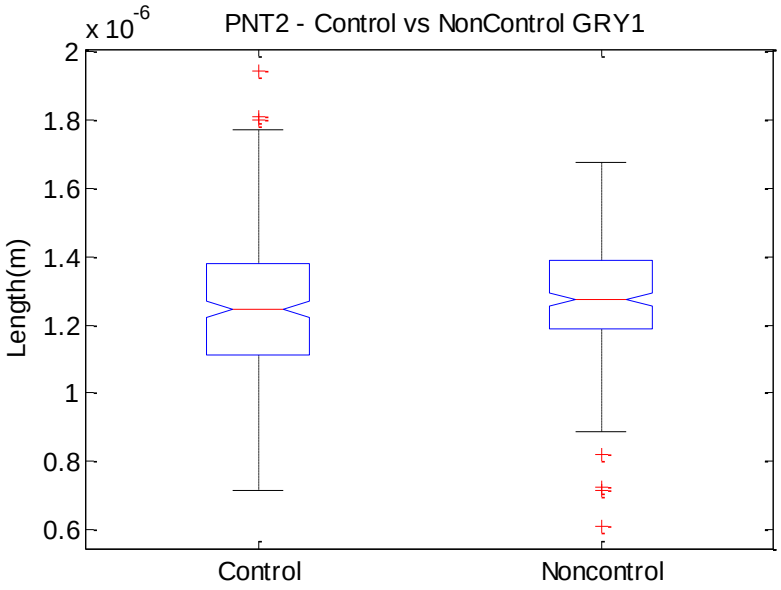


Figure 5-10 Histograms showing PNT2 Non-Control with a radiation level of 1 Gy, before conversion (upper) and after transform to length conversion (lower)

Figure 5-10 (upper) shows PNT2 non-control raw E data that is skewed to the left with a long right-hand tail. The conversion to energy length format produces a more centralised grouping. The kurtosis factor for the upper figure was 29.6 and for the transformed data in the lower figure was 3.63. Therefore, once again the conversion to energy length format via the energy normalisation transformation method has caused the distribution to tend towards normality.



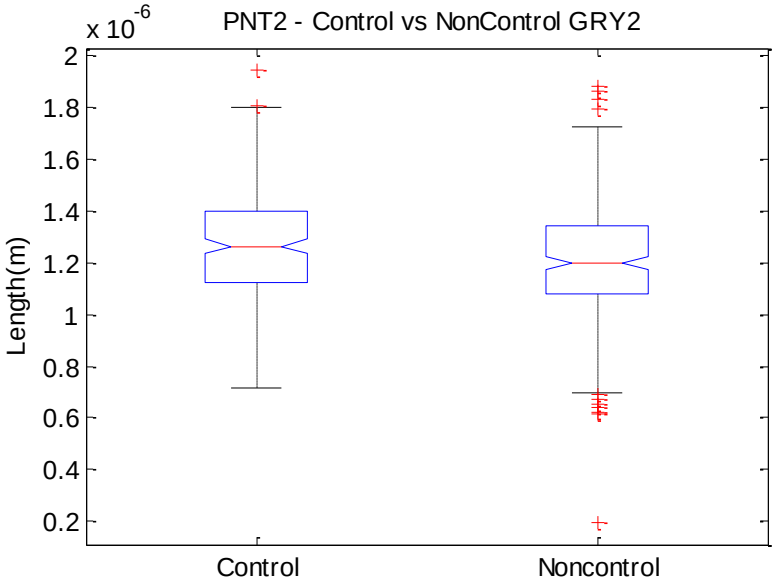


Figure 5-11 (Upper) Boxplot showing PNT2 Control and Non-control with a radiation level of 1 Gy. (Lower) Boxplot showing PNT2 Control and Non-control with a radiation level of 2 Gy.

The PNT2 control and non-control data presented in Figure 5-11 shows the boxplot for PNT2 cells at radiation exposure dose levels of 1 Gy (upper) and 2 Gy (lower), versus controls. This data is similar to that of the PC3 cells shown previously in that the 1 Gy dose showed no statistically significant difference in energy levels between the irradiated cells and the non-irradiated controls, while the 2 Gy dose showed a statistically significant smaller energy length indicating that cell stiffness has increased.

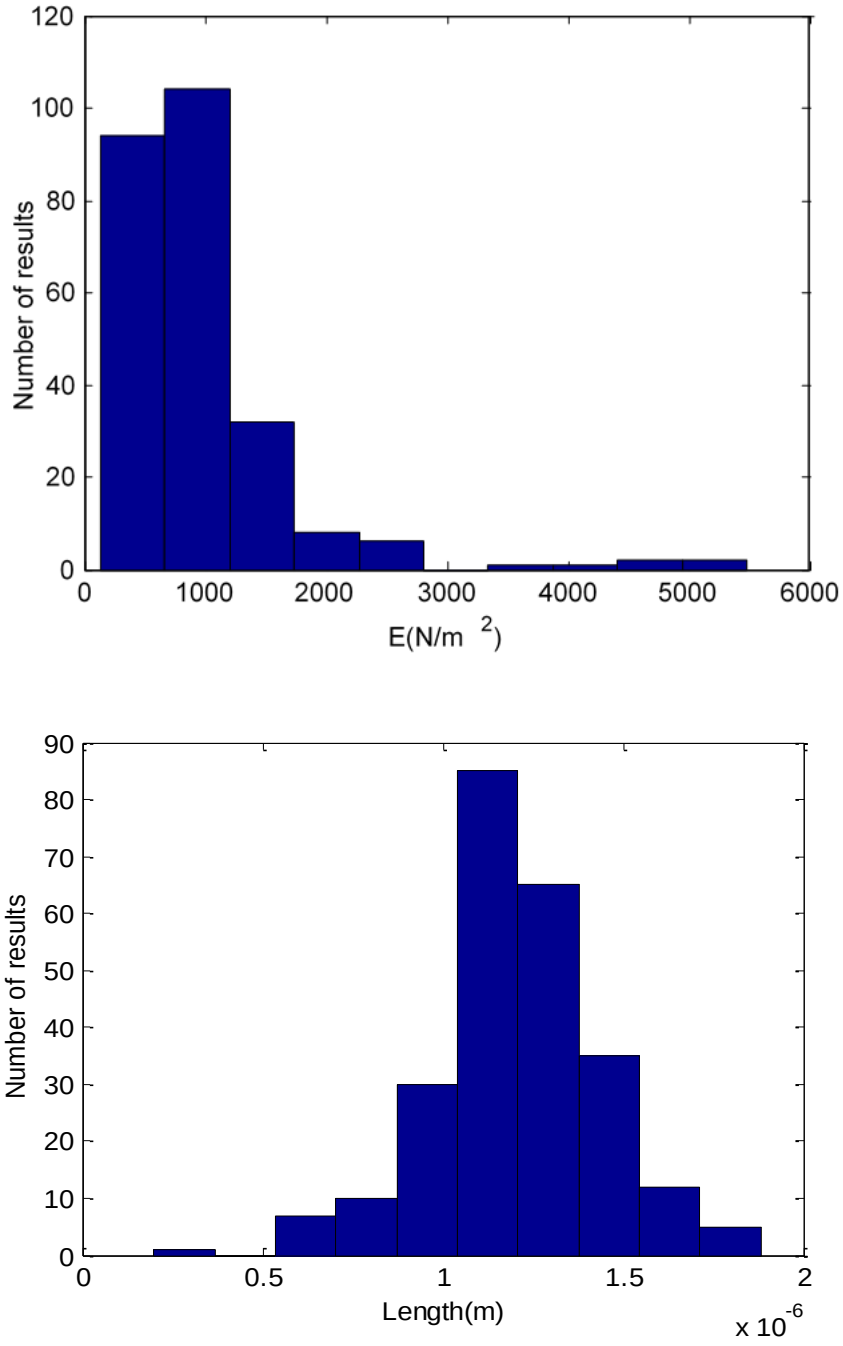


Figure 5-12 Histograms showing PNT2 Non-Control at 2 Gy, before conversion (upper) and after transform to length conversion (lower)

The PNT2 non-control data at 2 Gy presented in Figure 5-12 (upper) shows the non-normal raw Young's modulus, E, data that is skewed to the left with a long right-hand tail. The conversion to energy length format produces a more centralised grouping. The kurtosis factor for the upper figure was 15.4 and for the lower figure, showing transformed data, was 4.28. Of note is the outlier on the left of the lower figure showing energy length converted data - perhaps an indication that it may be necessary to perform a clip to remove 5% of the total data on the lower force data bound, as well as that already occurring to the upper bound.

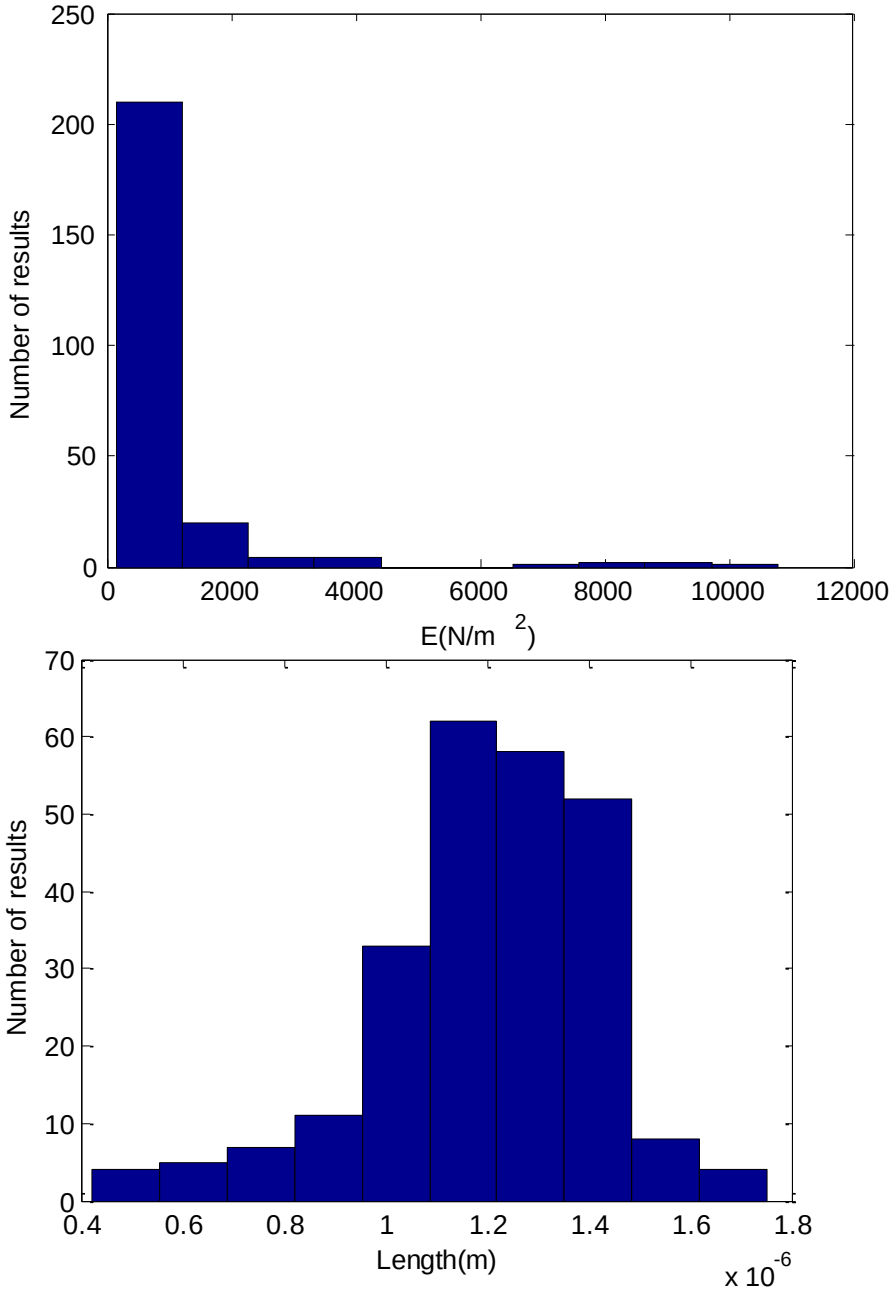


Figure 5-13 Histograms showing PNT2 Non-Control at 3 Gy, before conversion (upper) and after transform to length conversion (lower)

The PNT2 non-control data presented in Figure 5-13 shows the non-normal raw E data that is skewed to the left with a long right-hand tail. However once it has been transformed the data is still not normal unless the left-hand tail is also clipped at 5% of the total data size.

It is the view of the author that whilst the cutting away of 5% of the total amount of data at the lower end of the distribution makes logical sense, since we already remove 5% of data at the higher end the loss of another 5% of data is probably not advisable, even though the results would be improved and in this manner we would convert the last non-normal data to normal form. Further AFM force-distance curve studies are required to determine whether the results for the PNT2 3 Gy data are an anomaly, or whether they also require the lower end of the data distribution to be clipped at a level of 5% of the total data.

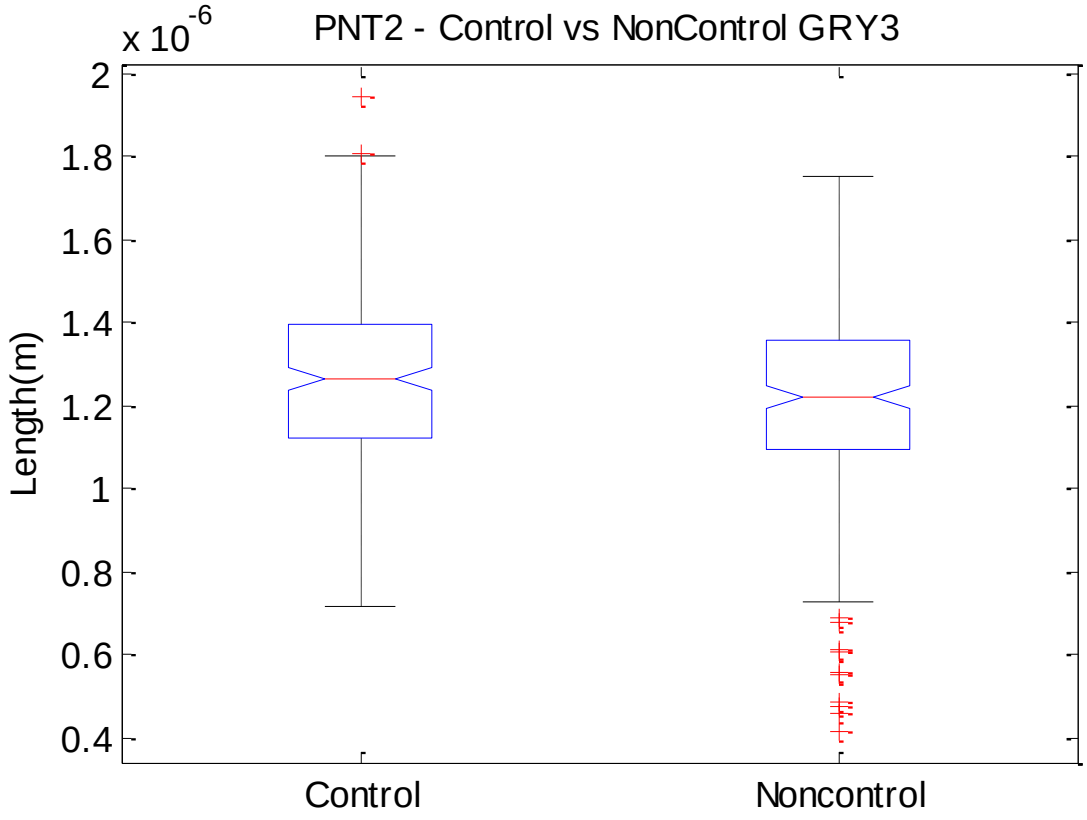


Figure 5-14 Boxplot showing PNT2 Control and Non-control with a radiation level of 3 Gy.

5.5 Testing for normality.

The Jarque-Bera test (Jarque and Bera.,1987), or JBtest in Matlab 2011a, is used to test for normality for the energy length data as well as the E histogram data. However, within

Matlab 2011a the skewness and kurtosis of the distribution due to the smaller sample size (< 2000 entries) uses a Monte Carlo simulation to estimate the parameters, rather than estimates of skewedness and kurtosis. This is built into Matlab 2011a as a way of not rejecting normal distributions due to the sensitivity of testing for normal distributions within small populations of less than 2000 entries. (Matlab, 2015)

	Control			Non-Control		
Cell Type	Gy 1	Gy 2	Gy 3	Gy 1	Gy 2	Gy 3
PC3 – NT	No	No	No	No	No	No
PNT2 – NT	No	No	No	No	No	No
PC3 – T	Normal	Normal	Normal	Normal	Normal	Normal
PNT2 – T	Normal	Normal	Normal	Normal	Normal	No

Table 5-3. The results of testing for normality before and after the transformation. The non-transformed data labelled NT and the transformed data labelled T. The last PNT2-T can be made to be normal by clipping away 5% of the total data towards the lower end of the data distribution. No = not normally distributed. Normal = normally distributed.

5.6 Discussion.

The transformation discussed here introduces normality to the majority of the AFM force-distance curve data. The sample size is roughly 130 for the smallest data collection, which is higher than is typical for studies of this type. Typically, a population size of 30 samples is capable of determining whether a significant statistical difference exists between two populations, provided that the differences between them are sufficiently large.

At lower sample resolutions however, if the values are relatively close together, i.e. the variance is small, a Type I statistical mathematical error would be introduced. The rationale for clipping away the top 5% of the total data is elucidated by looking at the original histograms, where the level of positive skew is high, but outliers exist on the right hand tail

of the distribution. This is due to the fact that there is no upper limit to the data values and therefore the higher levels of force for the cells are unbound. This is due to the cells being at different stages of their mitotic cycles and therefore they will respond differently to force dependent upon which particular part of their cellular life cycle they are undergoing. The cells that are stiffer than typical will unduly influence the normal distribution testing, resulting in higher than normal levels of kurtosis. The cell population we are considering is likely to be normally distributed for the cell stages that we are interested in, namely interphase stages G1, S and G2. This is reflected in the results in that by using the upper 5% clipping bound which removes the upper 5% of the total data set along with the energy length method almost all the data under test (with the same parameters used for all the data) returned a normal distribution.

The results could be perfect if we removed another 5% of the data from the lower end of the distribution. An argument could be made that if the cell outlier results can be too large then they can also be too small. The energy length histogram presented in Figure 5-12 shows that this is indeed what is happening. However, for the sake of a single result, losing another 5% of the total available data may not be advisable, but it is worth considering if this re-occurs with a different set of data.

The choice of the JBtest could be argued against as this test puts less importance to the tails of the distribution. However, the skewedness and kurtosis values are primarily what are of interest here. It is not a purely theoretical mathematical result that is of interest, but rather whether we can use the mean and standard deviation of the data and have it be a reasonable estimate of the data distribution once transformed. This could allow parametric tests such as the T-test to be performed.

	Significant Difference In Energy Length Between Control vs Non-control Cells At Various Radiation Exposure Dose Levels		
Cell Type	Gy 1	Gy 2	Gy 3
PC3	No	Yes	Yes
PNT2	No	Yes	No

Table 5-4. The results of testing for statistically significant differences in the energy lengths between irradiated cells and non-irradiated controls for the two cell lines under test after being transformed to energy length format. Yes = Statistically Significant Difference; No = No Statistically Significant Difference In Energy Length.

Table 5.4 above summarises the findings regarding the presence of any statistically significant changes in energy length between irradiated non-control cells and controls for the different cell lines under test. The Gy 1 cells, irradiated at a dose of 1 Gy, exhibited no statistically significant difference in the energy lengths at the 5% significance level for both the PC3 and PNT2 cell lines. For the Gy 2 and Gy 3 radiation exposure levels the non-control energy lengths all became smaller. For the PC3 cell line the Gy 2 and Gy 3 results were significantly different at 5% significance level in energy lengths. However, for the PNT2 cell line only the Gy 2 exposure level was significantly different in terms of the energy lengths at the 5% significance level. This matches a previous study (Gabryś, 2007) in which the exposure to radiation results in an increase in cell stiffness. However for the PNT2 cells under test at the Gy 3 radiation exposure dose the energy level wasn't found to be statistically significantly different from the controls.

5.7 Section Conclusion.

Presented here is a method that normalises the data for AFM force curves. The tolerance to changes of the estimated contact point are increased compared to linear methods, in that the contact point errors, both positive or negative, have a lower effect upon the starting position of the energy length. The reduction in the energy length starting position error is due to the nature of the method, in that the lowest energy is at the contact point. The method normalised almost all of the AFM force curve data and will normalise all of the data if the lower energy length data values are clipped by removing 5% of the total amount of data at the lower end for the data under test. The PC3 cancer cells under test behaved in a similar manner to the normal PNT2 cells when exposed to radiation, in that at a dose of 1 Gy both cell lines showed no statistical significance when compared to their respective controls for energy lengths at 5% significance. At a radiation exposure dose of 2 Gy both cell lines showed a statistically significant difference when compared to their respective controls at the 5% significance level in energy lengths, with the energy lengths becoming smaller for both of the irradiated non-controls. At a 3 Gy dose of ionizing radiation exposure the cell lines behaved differently in that the cancerous PC3 cells exhibited a statistically significant reduction in energy length, however the normal PNT2 cells, whilst displaying a reduction in energy length, were not statistically significantly different from the controls at the 5% significance level.

5.8 References.

Gabryś, D. et al. (2007). Radiation effects on the cytoskeleton of endothelial cells and endothelial monolayer permeability. *International Journal of Radiation Oncology, Biology, Physics*. **69** (5): 1553–1562

Janmey, P.A., (1998). The cytoskeleton and cell signaling: component localization and mechanical coupling. *Physiological Reviews*, **78** (3), 763.

MATLAB. (2015). *Anderson Darling Test*. [Online] Available from:http://uk.mathworks.com/help/stats/adtest.html [Accessed: 12th May 2015].

Chapter 6 : Investigating the orientation of actin ex-situ.

Cancer and many other diseases have been linked to a compromised cytoskeleton at the cellular level, i.e. a cytoskeleton that is mechanically weaker than that of a normal cell. There are 3 primary components of the cytoskeleton, namely actin filaments (F-actin), microtubules and intermediate filaments. F-actin is a key component of the cytoskeleton and is known to play a vital role within cell mechanics (Hall, 2009). In general actin filament orientation is usually determined non-quantifiably from inspection of images usually taken using fluorescent microscopy. However, this is often subjective and small changes in actin organisation/orientation are difficult to determine. Automating this would allow a more quantifiable approach to provide a more accurate interpretation of any changes to actin (or cytoskeleton) organisation in disease or following irradiation insult. Therefore, the orientational preferences of F-actin fibres are of interest within this research programme. In order to study F-actin fibre orientation an *ex-situ* experimental protocol was used as this would offer the most practicable approach for the study, so F-actin networks were produced *in-vitro*, on mica substrates, and subsequently imaged using AFM.

6.1 Actin preparation.

Following the protocol of Spudich and Watt [2], actin was purified from rabbit muscle by adding Mg^{2+}, K^{+}, Na^{+} & ATP for 2hrs at 30°C. The polymerised *ex-situ* actin was dried at room temperature on cleaved mica. Images of the polymerised actin were then acquired in AC mode using an MFP-3D Asylum Research AFM. The parameters for the AFM imaging process are shown below in Table 5-1. This work was carried out in the GERI laboratory.

Parameter	Value
Scan size	2.0 x 10^{-6}m
Integral gain	10
Scan rate	2.0032 nm/s
Image size	512 x 512 pixels

Table 6-1 The AFM parameters used to image the polymerised actin filaments.

6.2 Feature extraction methods

The feature extraction methods are only applied after pre-processing those being conversion to grayscale, estimation and subtraction of the the background from the image, histogram equalisation and finally thresholding to a binary image.

It can be clearly seen that the AFM 3D height image of polymerised F-actin filaments obtained by AFM, shown in Figure 6-1, has a non-uniform underlying pattern in the background intensity profile. In the work carried out here, the image shown in Figure 6-1 can be considered to be typical of a biological height image produced by the AFM, in that there is a non-uniform background intensity pattern under the ex-situ cellular F-actin fibres grown on mica. This non-uniform background intensity profile is shown in Figure 6.2 and requires removal before further processing can be considered. Removal of the non-uniform background intensity profile was automated into a series of steps, using Matlab 2010a. The first step that was necessary to enable further processing of the AFM image was to convert the raw AFM height image into grayscale format from 16-bit RGB format.

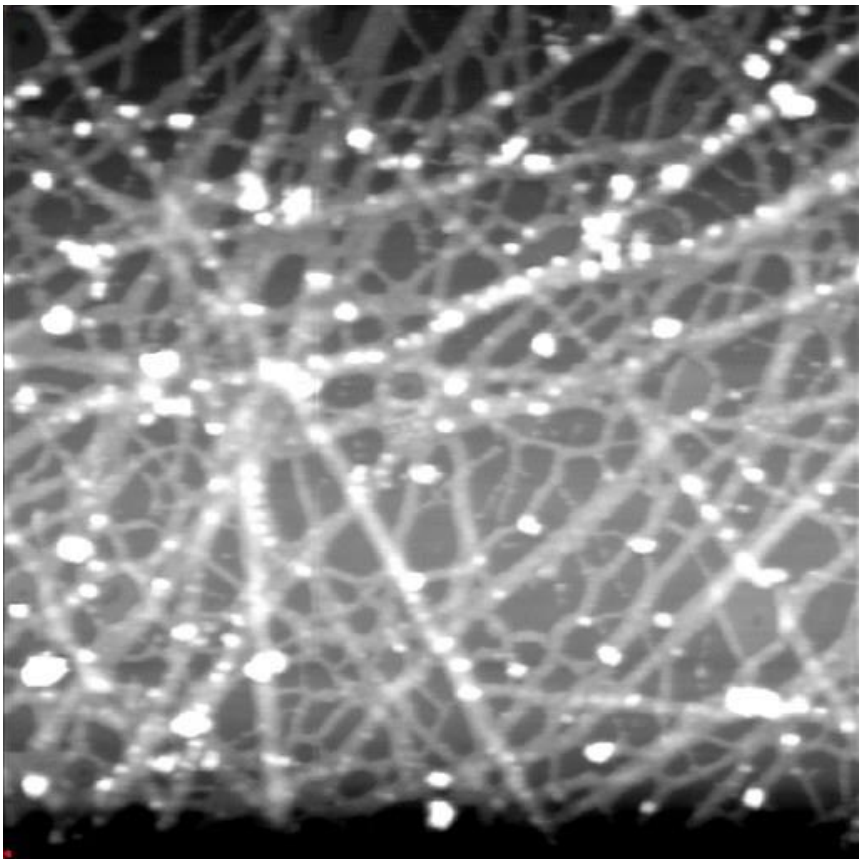


Figure 6-1 AFM height image of *ex situ* polymerised rabbit F-actin filament.

In the second processing step, the height map shown above in Figure 6-1 undergoes a morphological opening image processing operation to estimate the local background intensity level. Morphological opening averages the background over an area and therefore was deemed suitable as a method of background intensity level estimation. The parameters for this operation were arrived at by empirical investigation. A processing kernel element shape of a 20 pixel diameter disk was found to produce the best results. The morphologically opened background intensity map is shown in Figure 6-2. This was used as a basis to normalise the image by subtracting the opened background image from the original AFM image. This results in a darkened image compared to the original and the result is shown in Figure 6-3a.

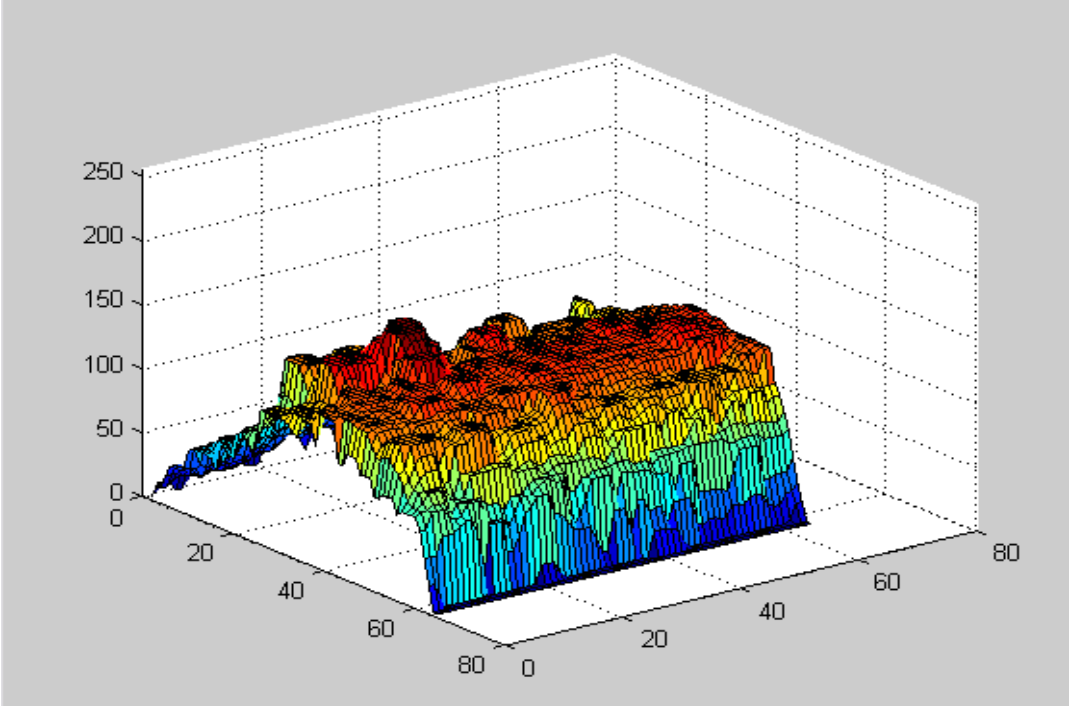


Figure 6-2 The estimated background intensity profile from the AFM height image, produced by morphological opening using a disk region of 20 pixels diameter. The surface displayed here is condensed and only shows every 8 pixels hence the size of 64 rather than 512 of the original data.

The darkened image was then histogram equalised in order to standardise the image in terms of its dynamic range and to enhance the image feature contrast. The image shown in Figure 6-3b was the result of 1% saturation of image intensity at both top and bottom intensity levels. In order to extract the features a number of alternate methods were considered as potential candidate techniques, including the Hough transform approach, morphological thinning, and a nearest neighbour area addition method.

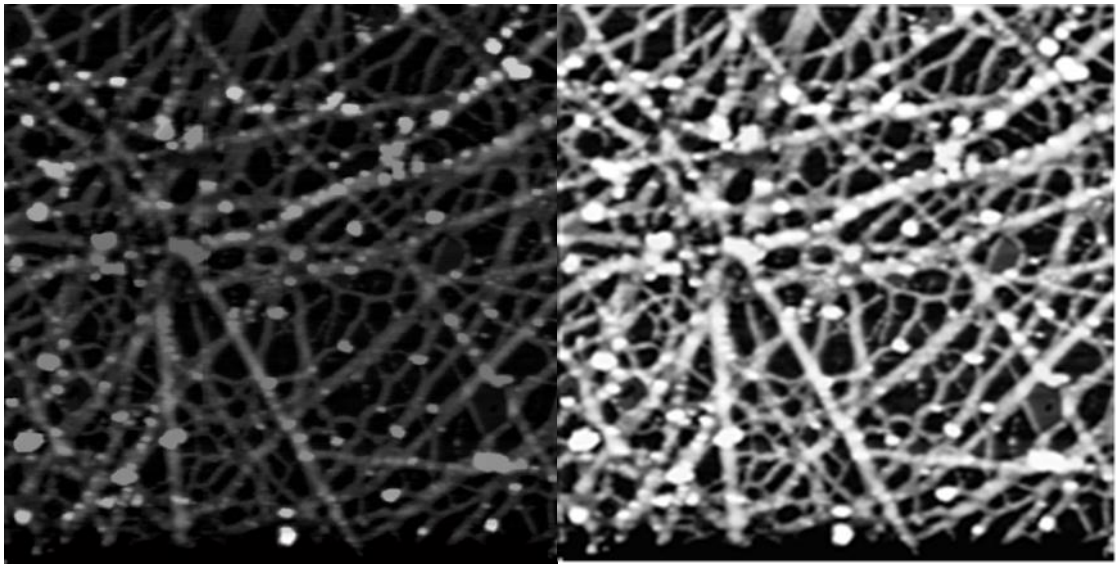


Figure 6-3 Left (a): The AFM height image of polymerised rabbit F-actin after removal of the estimated background intensity profile. Right(b): The histogram equalised image saturated to 1% image intensity of both top and bottom intensity levels.

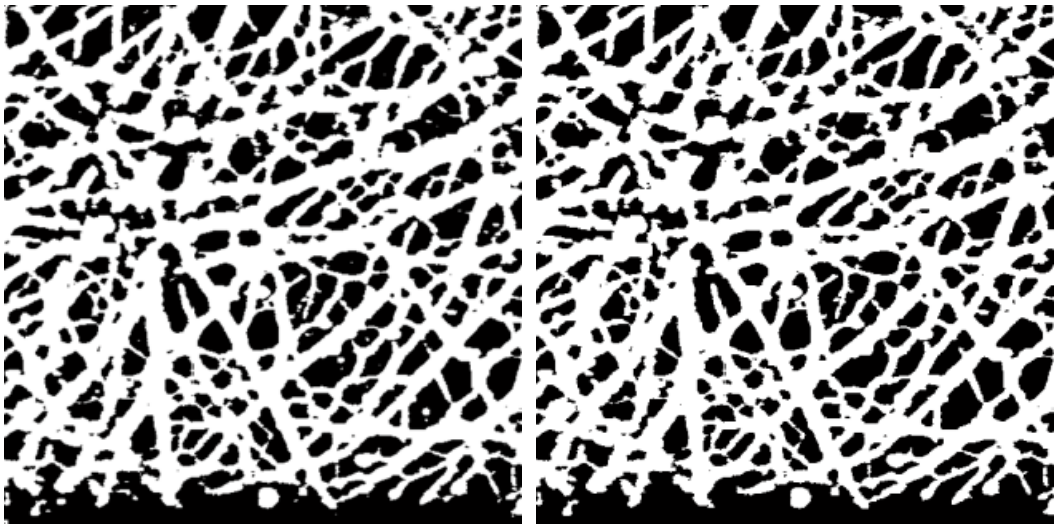


Figure 6-4 Left (a): The logical binary image of Figure 5-3(b) with an intensity threshold of 75% of maximum image intensity. Right (b): The result of removing smaller areas consisting of areas less than 50 pixels.

The binary image shown in Figure 6-4(a) was the next step in the image processing chain.

This was achieved by converting the image shown in Figure 6-3(b) to binary form at a

threshold of 75% of the maximum intensity level in the original image. Next the smaller globular areas of actin within the image shown in Figure 6-4(a) were eliminated by performing a morphological opening process with a circular kernel diameter value of 50 pixels, giving the final logical image shown in Figure 6-4(b).

6.3 Simple Methods for Feature Extraction

A number of relatively simple feature extraction techniques for automatically identifying the actin fibres within the AFM height images were investigated in the early stages of the research programme. However, all of these methods suffered from various technical shortcomings and were discounted in favour of more complex methodologies. The various simplistic feature extraction methods that were investigated are listed below.

A Neighbourhood method was attempted that summed the logical image shown in Figure 6-4(b) within a 5 by 5 matrix region. This produced peaks where the filament was present and zeros where there was no filament present. The time taken to perform this was not ideal and feature detection performance was not optimal. Other methods were subsequently tried resulting in this method becoming obsolete.

A Sobel edge detection method was also utilised. This method uses two 3x3 matrix kernels to iterate over the image to detect when an edge occurs. The resolution however was not acceptable and this method was therefore rejected.

A Canny edge detection method was attempted. This method uses two different thresholds to detect strong and weaker edges by looking at the derivative of a Gaussian filter. Unfortunately, it had poor feature resolution.

The next method tried was an image Skeletonisation approach via the function ‘bwmorph’ within Matlab 2011a. This was quickly rejected as this removes pixels from the image that are at the end of lines and therefore gave poor feature resolution.

6.4 Hough Method

The Hough method was a tailored solution to detect the maximum number of lines in the logical image. The rationale here was that by recording all the lines (actin fibres) within the image and obtaining their angles and associated lengths via a Hough transform approach, any angular information contained within the image would be extracted. Hence the orientation of the actin fibres could be automatically ascertained. This was achieved by first performing a Hough transform upon a logical image shown in Figure 6-4(b), then isolating the peaks in the Hough (accumulator) space using houghpeaks function within Matlab 2010b. The houghpeaks function defaults were overridden for both the number of peaks to find (changed to 50 - note-: there was no effect on the peaks found by increasing this value) and also the threshold, which was changed to 0.3*max(H(:)). Then finally using the ‘houghlines’ function within Matlab 2010b to calculate the line segments that have been identified and display them back in ordinary image space. The parameters for the houghlines routine in Matlab 2010b that was used for this task were 32768 for ‘peaks’ and 0.3 for ‘threshold detection’. Any lines below a length of 40 pixels were discarded and the default pixel gap that was allowed for the lines was 5% of the minimum length (a minimum of 2 pixels). The result is shown in Figure 5-5 (a). However, it should be noted that not all of the lines present in the underlying actin structure are picked up by this automatic detection method and several lines that are visible by eye remain undetected.

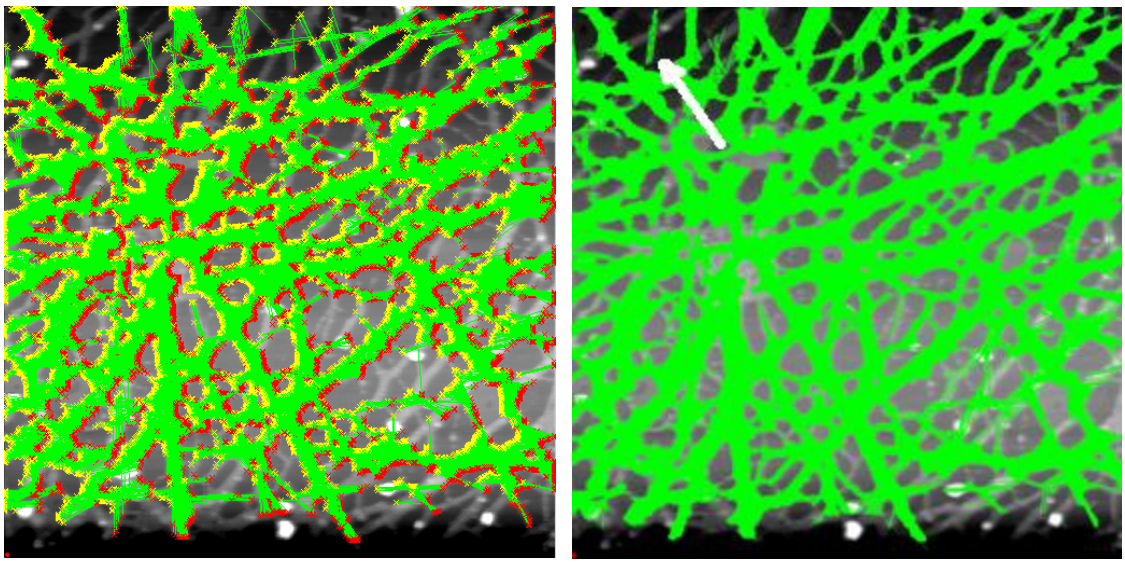


Figure 6-5 (a)(Left): are the lines that are generated by the Hough transform method, (b) (Right): Are the lines that are detected after performing the Hough transform at two different image rotations, 0 (original) and 90 degrees.

The amount of lines detected by the Hough transform can be made to increase if the process is repeated at a different image rotation. This procedure obviously increases the overall processing time and does increase the number of lines automatically detected, but still does not detect all the lines within the image. This situation with image rotation and subsequent re-processing is visible above, where Figure 6-5 (b) shows an image that has been processed as in Figure 6-5 (a) and then rotated by 90 degrees and reprocessed using the same Hough methodology. The line highlighted by an arrow at the top left of the double-pass processing (0 and then 90 degree rotated) Hough image in Figure 6-5 (b) remain undetected in the single pass Hough approach shown in Figure 6-5(a). Various different parameters for the peak detection and bins were tried, but certain lines would still not resolve using this approach.

6.5 Thinning method

The mathematics behind feature extraction of the thinning method in matlab are based on (Lam et al., 1992). The following section is from Mathworks, 2017 bwmorph section of the feature extraction algorithm that is used to calculate the thinned image.

“

1. In the first subiteration, delete pixel p if and only if the conditions G_1, G_2, and G_3 are all satisfied.

2. In the second subiteration, delete pixel p if and only if the conditions G_1, G_2, and $G3'$ are all satisfied.

Condition G1:

$$X_H(p) = 1$$

where

$$X_H(p) = \sum_{i=1}^{4} b_i$$

$$b_i = \begin{cases} 1, \text{ if } x_{2i-1} = 0 \text{ and } (x_{2i} = 1 \text{ or } x_{2i+1} = 1) \\ 0, \text{otherwise} \end{cases}$$

$x_1, x_2, \ldots, x_8$ are the values of the eight neighbors of p, starting with the east neighbor and numbered in counter-clockwise order.

Condition G2:

$$2 \le \min\{n_1(p), n_2(p)\} \le 3$$

where

$$n_1(p) = \sum_{k=1}^{4} x_{2k-1} \vee x_{2k}$$

$$n_2(p) = \sum_{k=1}^{4} x_{2k} \vee x_{2k+1}$$

Condition G3:

$$(x_2 \vee x_3 \vee \bar{x}_8) \wedge x_1 = 0$$

Condition G3':

$$(x_6 \vee x_7 \vee \bar{x}_4) \wedge x_5 = 0$$

The two subiterations together make up one iteration of the thinning algorithm. When the user specifies an infinite number of iterations (n=Inf), the iterations are repeated until the image stops changing. The conditions are all tested using applylut with precomputed lookup tables." (Mathworks, 2017)

In order to try to improve the number of lines being detected, the next method used was a thinning method, which sums a 3x3 neighbourhood region of the logical image shown in Figure 6-4 (b). The results were placed into a cumulative array of the same size as the original image. This cumulative array has the characteristics that a value of 2 correlates with the end point of a line, a value of 3 represents a continuation of a line, and a greater value than 3 indicates that there is a branching point upon the line. The thinned image that is shown in Figure 6-6 is used as a basis to process the lines, by growing each of the branch points until they either reach a branch point, or an end point. Following this process results in the image shown in Figure 6-7.

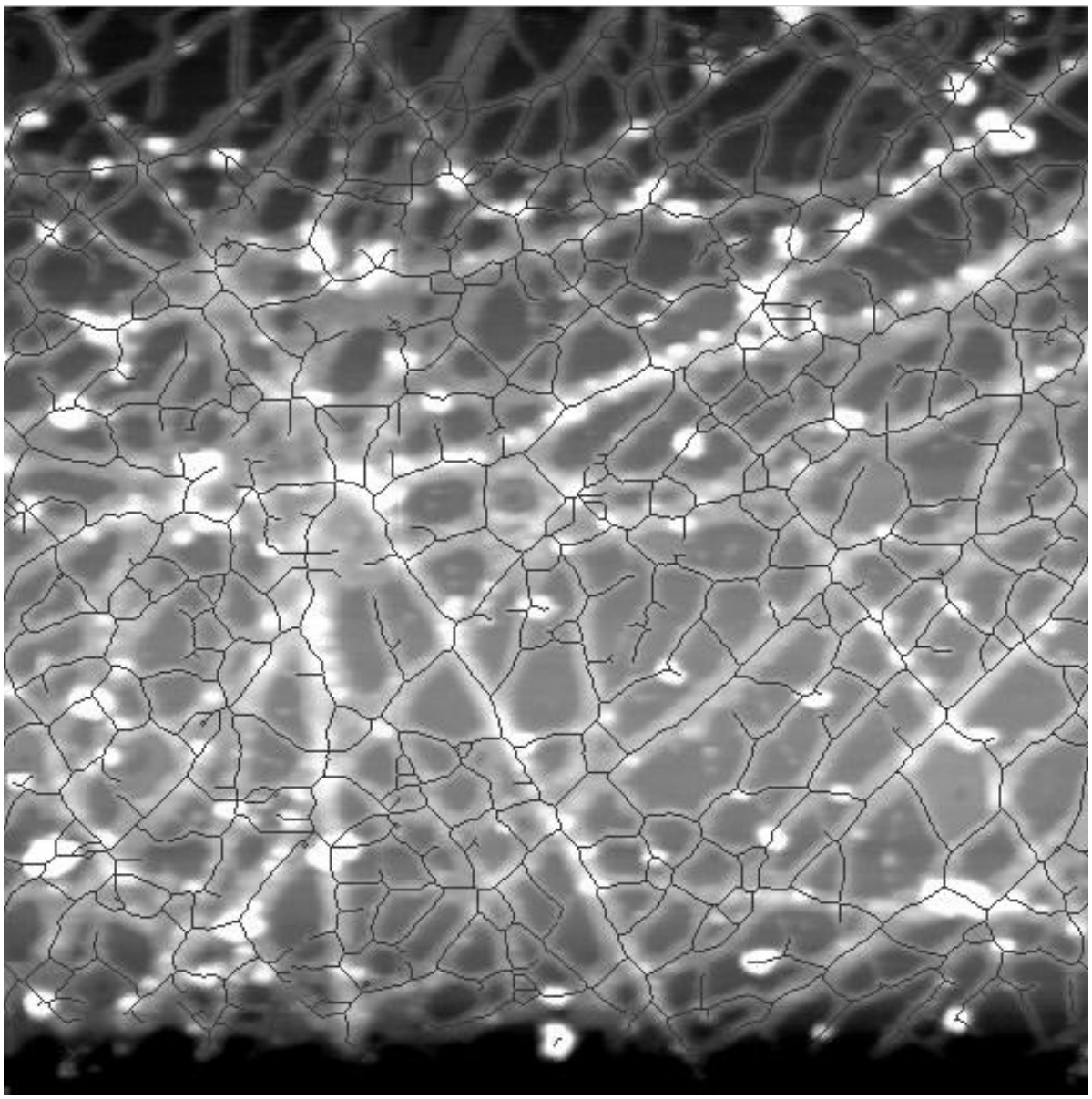


Figure 6-6 The thinned image overlaid on the original AFM background image.

The method for tracing the actin at this point is based on branching points and storing the number of pixels that are present in the image. The method works by counting the number of pixels in a 3x3 area and summing them. The end points are labelled first by counting the number of pixels in the 3x3 area, these have less than 3 pixels and are stored into an array. The branch points are labelled next and stored into an array, they have the property that they have more than 4 pixels in a 3x3 area. The actual lines are stored into an array by using variation filters for diagonal and straight combinations than enable an image to be built that has just lines within it again using a 3x3 filter. Each branch point is then traced from each point within it to go along the line image that was created until it reaches either an end point or another branch point. This builds up the image seen in Figure 6-7 which consists of a matched set of start and end points for each line in the image.

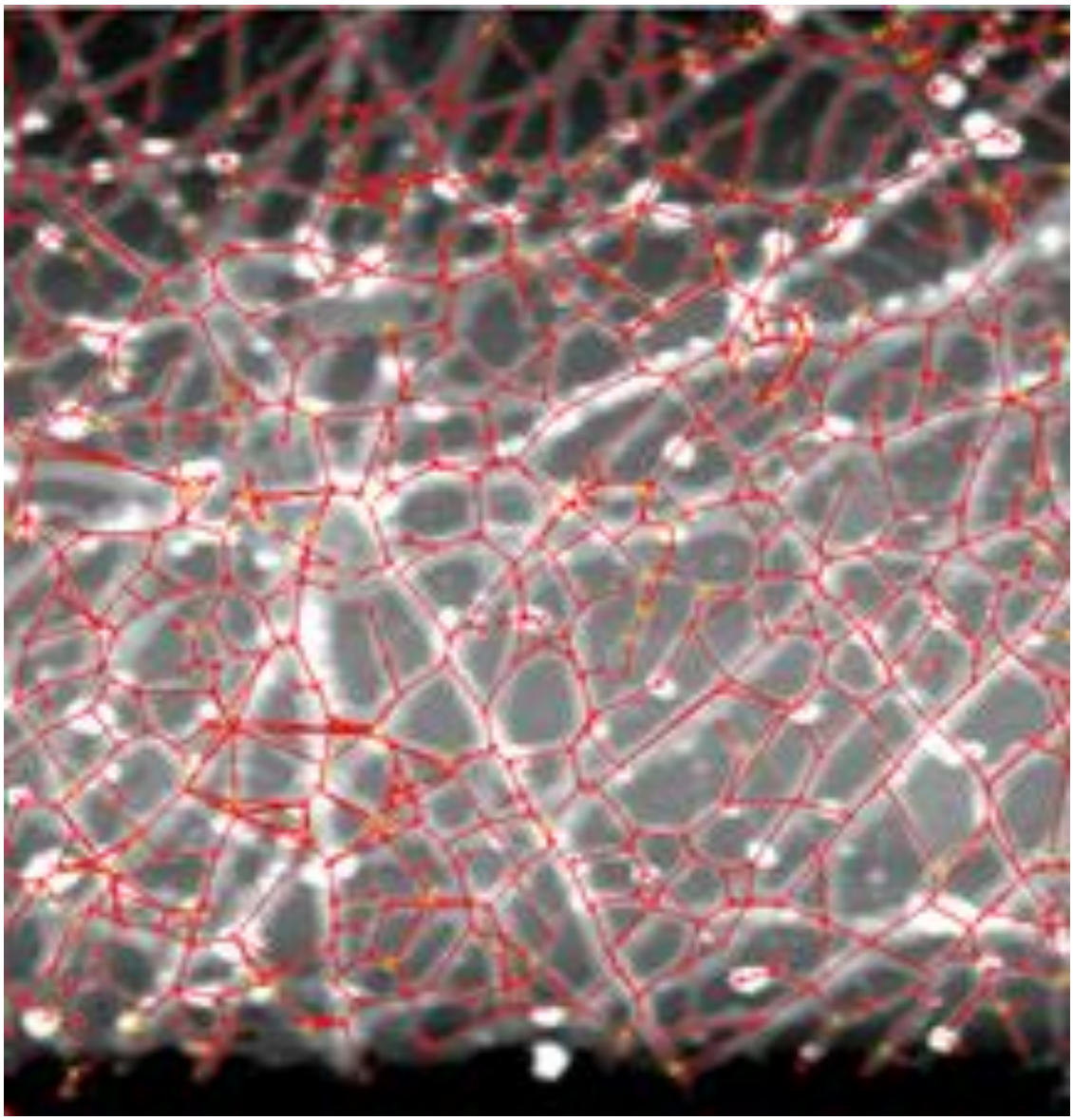


Figure 6-7 The lines generated from the thinned image.

6.6 Discussion

AFM biological images of filamentous actin networks present a problem to life scientists, in terms of being able to perform automated processing upon them, due to the presence of underlying structured noise in the background intensity distribution beneath the actin

filaments. Hence the background intensity distribution was estimated and removed from the source data by subtraction. The initial feature extraction methods that were tried included edge detection methods, like Sobel, and skeletonisation approaches, however these were rejected quickly due to poor feature resolution. They did detect the features of interest, i.e. actin fibres, within the AFM height images, but produced more noise than was the case for the thinning method and therefore these simplistic approaches were not investigated past the preliminary phases of the work.

The Hough transform method was at first thought to be the best choice for this task, as the length of the lines representing the individual strands of F-actin and the angle of those lines could be automatically extracted using Matlab 2010b. The number of lines appearing over an area would hence provide a measure of the thickness of the F-actin filaments and the Hough algorithm approach seemed to offer a good fit to the data. However not all the lines were detected within the image and more importantly the missing lines were linked to particular areas within the image, causing the lines within that area to become under-represented. The first attempt to resolve this issue was to rotate the image by 90 degrees clockwise and then to perform a second proceeding stage on the rotated image, subsequently combining lines from both images. This increased the number of automatically located lines from approximately 20,000 lines to around 40,000 lines after the second pass and the extended process took roughly one hour to complete on an Alienware Mx15 Quad Core i7 at 1.7Ghz with 4 Gb of RAM running on Windows 7 64-bit version 6.1 build 7601. No suitable parameter selection was found that could resolve all the lines in the image. Thought was given at this stage to perhaps creating a bespoke Hough transform algorithm, however different alternatives were tried first.

The first of these alternative techniques was the neighbourhood method. This method took around the same length of processing time, at approximately an hour, as the 2 pass Hough technique, on the same computer hardware. However, it did detect 8 more actin features than the 2 pass Hough method.

The next method attempted was morphological thinning. This reduced the overall processing time to around 5 minutes on the same computing platform. This significant decrease in the computational overhead was very attractive, as the automated processing speed would be an issue for large scale detection of filaments in images, which was a goal of the research. The results for the nearest neighbour method and thinning were equivalent in terms of quality, with both adequately returning the central spine of the F-actin network, i.e. the middle of the thick portion of the F-actin fibres. However, the thinning approach was the obvious best candidate, as it had a faster processing time on the same computing platform. The resulting image after the line extraction process is shown in Figure 6-7, capturing almost all of the F-actin filaments in the original AFM height image. The algorithm at present does not detect curves, instead representing them at straight lines, this was not required for the present data, but could be a future algorithm improvement.

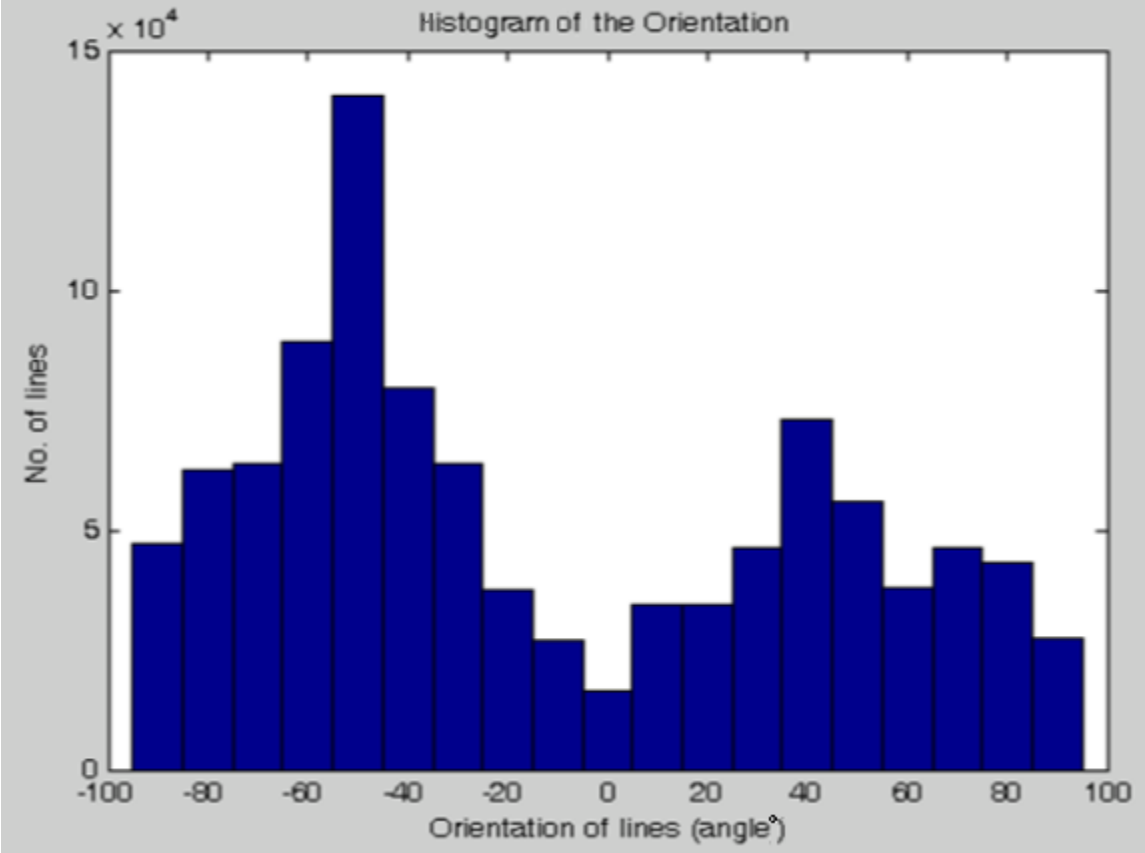


Figure 6-8 Histogram of the lines generated from the thinned logical image, weighted for line length and area.

Figure 6-8 shows a histogram grouped into bins of 10 angular units of the underlying angular orientation of *ex-situ* polymerised rabbit F-actin AFM height image data. The extraction of the values for the angles is produced using the atand (arc tangent) function within Matlab with care to avoid the singularity at zero. The function that extracted the angles was tested using the figure 6-9 to ensure that it was correctly assigning angles to the correct values.

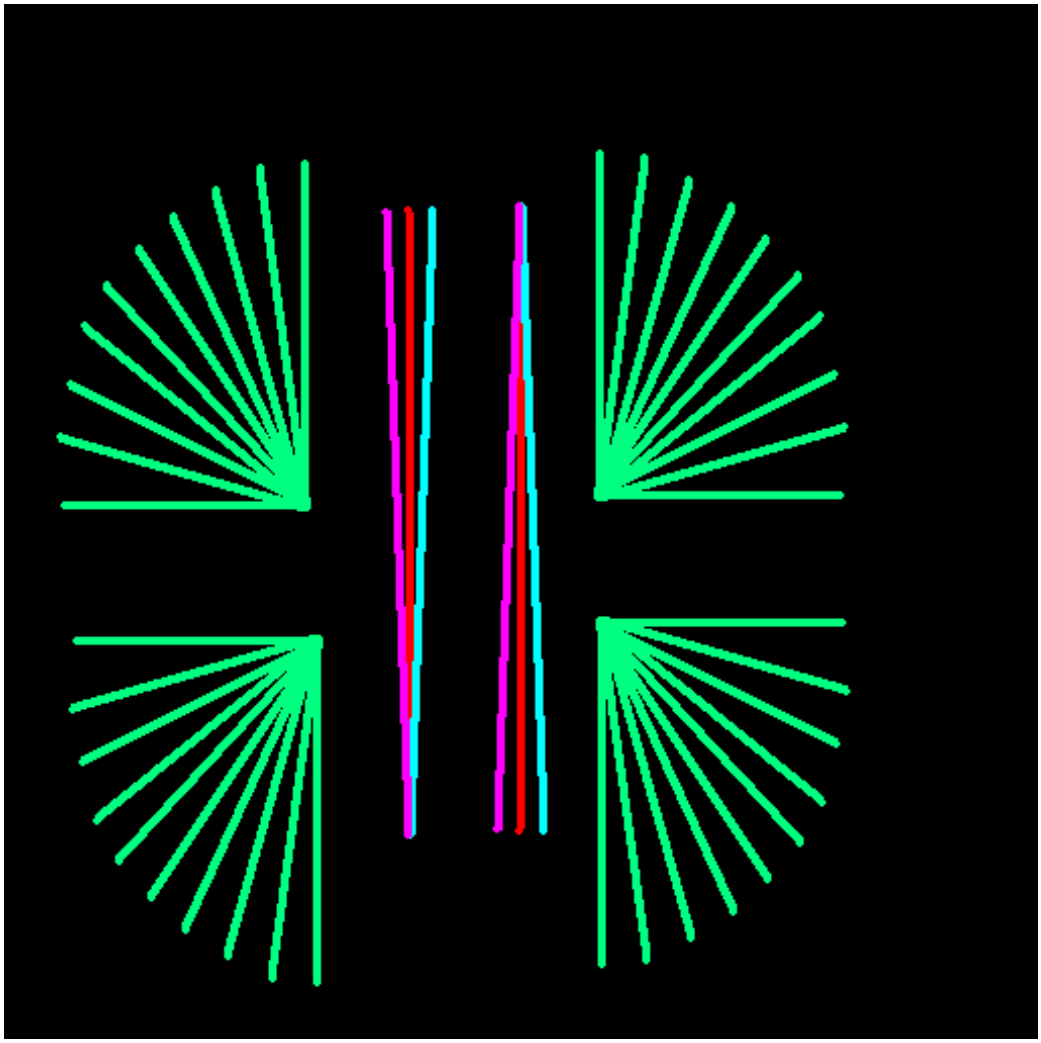


Figure 6-9 The image of the lines that were used to verify that the angles and values were being correctly assigned by the atand function within Matlab.

Analysis shows that the F-actin orientation data does not follow a normal distribution. This is shown in the histogram, as there are clearly two peaks, as opposed to a single peak, present in Figure 6-8. Dividing the histogram data into two segments, split at 0 degrees, and performing a test for normality does not result in a normal distribution for either split region (Lilliefors at 5% certainty). This could be explained by the large dominant orientations of the F-actin in the image. The coefficient of bimodality is a measure of the likelihood of there being more than one statistical distribution in the data. For the non-

grouped data a significant value of 0.65 indicates that there are at least two distinct populations present within the data (SAS, 2008). By visual inspection there do appear to be two peaks present within the data. Similar to the theoretical data of (Pollard, 2008) the experimental data from Figure 6-8 shows that there is a peak at an orientation angle of around 40 degrees and another at approximately 75 degrees. The distance between these histogram peaks therefore approximately matches Pollard's theoretical data 'gap' of 37.5 degrees between the major orientational axes. The *ex-situ* distribution of actin of course may not be indicative of actual *in-vivo* actin distribution, but the preference for its orientation and distribution is still of theoretical interest here in that it shows orientation peaks that match those that Pollard et al. predicted in 2008.

6.7 Conclusion

The standard Hough transform approach was not able to detect a complete map of all the lines in the ex-situ polymerised rabbit F-actin AFM height images. The automatic detection of an *ex-situ* F-actin network was possible if the thinning approach was used. The detection can be accomplished by either summing the immediate area using a 5x5 grid, or thinning approaches, but the latter process completes over ten times faster than the former in Matlab 2010b. The coefficient of bimodality was statistically significant and indicated that there are two regions that are separated by an angular rotation of 90 degrees. There were also some indications of correlation between the experimental work performed here and the theoretical work carried out by Pollard, 2008 in that there is a possible peak at 37.5 degrees. This method may in the future allow us to understand more about actin organisation and could be expanded to in situ from images of cells.

6.8 References.

Ballard D.H., "Generalizing the Hough Transform to Detect Arbitrary Shapes", *Pattern Recognition*, Vol.13, No.2, pp 111–122, 1981.

SAS Institute Inc. (1990). SAS/STAT User's Guide, Version 6, 4th Edn Cary, NC: Author. Pg 491.

Conover, W. J., Practical Nonparametric Statistics. John Wiley & Sons, Hoboken, NJ, 1980.

Duda R.O. and Hart P.E., "Use of the Hough transform to detect lines and curves in pictures", *Communs. Ass. Comput. Mach.* Vol. 15, pp. 11–15, 1972

Hall, A. "The cytoskeleton and cancer". *Cancer and Metastasis Reviews*. Vol 28, No. 1, pp 5-14. 2009.

Lam, L., Seong-Whan Lee, and Ching Y. Suen, "Thinning Methodologies-A Comprehensive Survey," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol 14, No. 9, pp 879, 1992.

Mathworks, Bwmorph, [online] Available at : https://www.mathworks.com/ help/ images/ ref/ bwmorph.html#bui7chk-1 , Accessed 17/04/2017

Pollard, T.D. & Berro, J. "Mathematical Models and Simulations of Cellular Processes Based on Actin Filaments". *Journal of Biological Chemistry*. Vol 284, No. 9, pp 5433-5437, 2008.

Spudich & Watt. "Preparation of actin from rabbit skeletal muscle". *Journal of Biological Chemistry*, Vol. 246, pp 4866, 1971.

Chapter 7 : Conclusions

Chapter 2 confirmed that there was a statistically significant difference between the prostate normal tissue PNT2 cell line and the cancerous high metastatic PC3 cell line in response to force probing with the AFM tip and time was eliminated as being a possible influencing factor during these experiments.

Using the Kolmogorov-Smirnov statistical test showed that the PNT2 and PC3 cells lines showed a significant difference for the stiffness in response to force probing with the AFM tip. The result for the PNT2 versus the PC3 cell lines is as expected for cells of this metastatic type as this matches previous work carried out by (Gabryś et al, 2007; Li et al, 2008), in that the cells of the high metastatic potential were less stiff and therefore more malleable when undergoing force probing using the AFM. Since these cell lines are from prostate, not breast, tissue, the result is novel, the technique however is not.

Experimental measurement time was tested to ascertain what effect if any it had upon the cells undergoing testing. The time taken to complete an experiment data recording was 30-40 minutes. The time used to test if there was a significant difference was 1 hour, thus ensuring that under experimental conditions the hour length would not be exceeded. Measurement time was shown to only influence the force response in a positive manner, playing only a minor role as a factor and having a 1-3% positive variation in force response over the hour long period.

Chapter 3 presents a method for AFM contact point estimation that has been termed ‘line projection’ by the author and which was found to be suitable for automation. This method

is faster than the alternative Bayesian method which was also presented and is suitable for the gently sloping force curves as found in this work with PNT2, DU145 and PC3 cell lines.

The assumptions used for the line projection method are that the contact point is within the upper and lower bounds of the measured data and that it is actually on a point of the measured force distance curve. Back projection was eliminated from the set of possible fittings as for the large positive sloping curves the starting point would not be within the bounds of the curve and therefore the starting point would become difficult to predict due to the exponential nature of the curves. In such cases this would lead to accuracy issues related to the contact point, which would depend on the particular contact model that was chosen. The elimination of the large positive slopes from the possible data set increases the robustness of the line projection method by not having to deal with inaccurate AFM indenter tip positioning. Compared to some of the other techniques the line projection method, due to its simplicity, is relatively fast, taking around 2 hours compared to 5 days for the corresponding Bayesian approach. Once the curves are recorded and match the required format no further human interaction is required and when judged empirically the results are more accurate than the Bayesian approach. In addition, the results are the same each time, as there is no sampling taking place and therefore the results are found to be highly repeatable.

Chapter 4 showed that the Hertz model is not sufficient for indentation distances greater than 500nm due to the strain hardening effect for biological cells and that the biological cells' non-linear force response becomes marked after the 500nm indentation depth region.

The classical Hertz model is often used for measuring a cell's force response over relatively small indentation distances, i.e. less than 500nm. When used to measure the living biological cells that were under test, namely PNT2, PC3 and DU145 cell lines, the strain hardening effect was readily observable for indentations above 500nm. Therefore the model is not sufficiently accurate to model indentation with respect to the applied force for the comparatively larger indentation distances that were under test, namely a distance of 1750nm. The non-linear force response leads to errors when fitting the curve using the classical Hertz model. The model developed by the GERI Cell Mechanics group (Randall, 2010) extends the classical model to include a squared polynomial term varying with indentation. The extended model introduced the problem of physically explaining the model parameters, but when implemented by the author it was found to significantly reduce the fit error. The best fit for a curve would often have parameters with a negative value, leading to questions about how that would physically function. The AIC was used to determine if the model undergoing testing could be improved upon in terms of fitting the data. It was found that the model $ae^{x} - a + be^{x*x} - b + cx^{2}$ best fit the data under test and that the classical Hertzian model had less than a 0.1% likelihood to be the best fit for the data under test.

Chapter 5 investigated actin orientation when an in-vitro polymerised actin network was measured by AFM ex-situ and subsequently analysed via automated image processing techniques. This work showed that the coefficient of bimodality was statistically significant

and thus indicated that there are two distinct regions for the major fibre orientations, which are separated by an angular rotation of 90 degrees. This experimental study also indicated some correlation with theoretical work carried out by Pollard, 2008.

A selection of common methods was used to detect the lines corresponding to the ex-situ polymerised rabbit F-actin AFM height images. Neighbourhood summing, Sobel, Canny, Hough and thinning transforms were used. Using the Hough transform from within Matlab led to lines not being detected, even when the transform rotated the image and was applied again. Neighbourhood, Sobel and Canny approaches were quickly rejected as being inferior once the thinning approach was used. Although the neighbourhood method worked in the same manner as the thinning method, it was slower in terms of computational time taking an hour to complete versus 6 minutes for the thinning approach.

The coefficient of bimodality was 0.65 for the data, which indicates that at least two distinct populations exist within the data set (Cary, 2008). The main region separation was 90 degrees, as indicated by the peaks within Figure 5.8. There was also some indication of correlation of peak at 80 degrees from the peak at 40 degrees, which corresponds to theoretical work carried out by Pollard, 2008.

Chapter 6 presents a method that normalises the data for AFM force-distance curves, thereby providing an approach to test arithmetic means and standard deviations for these curves for AFM data that has a contact point that is approximately in the correct position.

Also presented in Chapter 6 is a method based upon indentation energy that that alters the way that the choice of automatic contact point estimation influences the error in distance

when analysing AFM force-distance curves. The AFM probe tip contact point is deemed to be at zero in terms of indentation energy and therefore the curve region around this contact point has lower energy than the starting energy length (Which is a point to the right on the curve 0.5 x 10^{-15} Nm from the contact point) by at least a factor of 10. The contact point error therefore has less influence by at least a factor of 10 on the position of the starting energy length, thereby making the starting energy length more tolerant to slight changes in the starting contact point.

The different cell lines when undergoing ionising radiation exposure testing showed no behavioural differences except at the 3 Gy dosage, where at 3 Gy radiation exposure dose the irradiated cell lines showed a statistically significant difference in the PC3 cell lines by having a reduction in their energy length. In comparison the PNT2 cell lines, while showing a reduction in energy length, were not deemed to be statistically significant as this reduction was at the 5% significance level. At 2 Gy at the 5% significance level, the energy lengths for both cells lines, namely PNT2 and PC3, showed a statistically significant reduction in energy length. At 1 Gy both cell lines showed no statistical difference in energy lengths at the 5% significance level.

The Energy normalisation transformation method normalised almost all of the AFM force curve data and will normalise all of the data if the lower energy length data values are clipped by removing 5% of the total amount of data at the lower end for the data under test.

7.1 References

Gabryś, D. et al. (2007). Radiation effects on the cytoskeleton of endothelial cells and endothelial monolayer permeability. *International Journal of Radiation Oncology, Biology, Physics*. **69** (5): 1553–1562

Li, Q.S. et al. (2008). AFM indentation study of breast cancer cells. *Biochemical and Biophysical Research Communications*. **374** (4): 609–613

Cary (ed), SAS User's Guide, Version 5, NC: SAS Institute Inc, 1985.

Chapter 8 : Further Work

The work carried out in this research programme answered a few questions about the responses of cells to radiation and AFM probing. AFM probing studies however have a lot of influencing factors and therefore can give different results depending upon what particular set of conditions are applied to the cell line. Below is a set of possible additional work that could be carried out to extend this research further.

8.1 Cell lines and radiation

The cell substrate that is used could be a potential influencing factor and in this study both poly-l-lysine and fibronectin were utilised. The poly-l-lysine proved to be unsuitable as a substrate when the PC3 cell lines were irradiated due to poor attachment and therefore another substrate was used that the PC3 cell line could adhere to more strongly. Cells could be grown in a 3D scaffold and then tested to determine if they behave mechanically in a different manner than those that are grown on a substrate such as poly-l-lysine. Therefore, it may be possible to test the influence of the substrate upon the force response of the cell lines. In those experiments where the cell lines did not adhere to the substrate, other methods could be used to determine the mechanical response of the cells, particularly for those that do not adhere well to surfaces. Approaches such as optical tweezers could be used to measure the force of these cells and could be statistically tested in a similar manner to those that were performed in Chapter 2. Other factors, such as the temperature at which the cells are grown and the growth medium, whether the cells are grown to confluence, etc, all play a part in the cellular response, and therefore the effects of these factors could be isolated by subsequent experimental testing.

While the work in this thesis is primarily concerned with cancer and the PC3, DU145 and PNT2 cell lines, other cell lines could be tested to determine if they behave different mechanically. These cell lines could include muscle cells, nerve cells and stem cells. Other diseases could also be investigated that have a link to an altered cytoskeleton, e.g. Alzheimer's, Parkinson's, etc. If the AFM can measure the specific cell type, then the techniques described with this thesis can be used to determine if significant mechanical differences exist. The measurements of different cell lines, such as breast cancer, kidney cancer, etc, and subsequent determination of the mean energy length of each of these to determine if there is a constant ratio, or pattern, within the cells could be useful to see if there are any connections between the different cancer types.

The effects of ionising radiation upon the cell lines during the experiments described in Chapter 2, Table 2.1, where the mechanical recovery of the cell lines can be seen to occur from hour 6 post irradiation, prompted an interesting question about the recovery time. Thus an exploration of the relationship between recovery and radiation exposure could be useful to determine the optimum time between treatment dosages and whether there is a difference between recovery time for cancer cells and non-cancerous cells.

8.2 Line projection method.

The line projection method discussed in Chapter 3 assumes that the contact point lies on the force distance curve and as such does not interpolate to find the optimum point. This was due to the high rate at which points where generated. If however, a lower scan rate was used it could lead to problems with the contact point being inaccurate in the future and therefore the algorithm could be improved by determining if the scan rate is low enough to warrant interpolation. The line projection method also assumes that the contact point lies

within the upper and lower bounds of the measured data. Methods do exist to project both backwards and forwards in terms of the force-distance curve. Therefore a reliable method could be developed to project the data and therefore less measurements would be required, thus shortening the time taken to record the curve. In terms of the line detection methods described in Chapter 5, these could be improved to detect curves and treat them as curves, rather than as straight lines.

8.3 AIC and 37.5 angular difference.

The force-distance curve model formula which was determined to be the best fit to the data was $ae^{x} – a + be^{x*x} –b + cx^{2}$. While this formula was deemed to be the best fit, the formula does not have a physical interpretation and is purely empirical. Effort could be made to find a physical formula which would not only fit the data with the best AIC value, but would also be explainable in terms of physical cell mechanics.

The indications that a 37.5 angular difference exists for orientation of ex-situ actin fibres was not investigated and this could be looked at from both an ex-situ and in-vivo perspective to verify if a 37.5 degree difference exists and if the 90 degree dominance seen in ex-situ remains true in-vivo. This could be related to the cell cycle stage and therefore it may be worth looking at the fibres in different stages of the cell cycle to see if there is any difference to the angular preferences seen in the ex-situ F-actin.

8.4 Modelling

The model used for most of the testing was $ax^{3} + bx^{2} + cx$. Alternatives to this could be to instead simulate the model using a software package such as ANSYS, using this as a starting point to match the cell data to determine the role structures such as actin play in

the force curves generated. This could give clues to the roles organelles play in the force response of the cell.

Rather than focus on just one point upon a cell and a single AFM measurement, it is possible to perform multiple readings over an area using an AFM via force spectroscopy. This type of approach could allow the response of the cell to repeated probing to be tested, enabling a picture to be built up of the response over an area and perhaps allowing for an averaging out of any anomalous readings. This could then potentially be used as a model for a computer simulation to determine the role that actin actually plays within the cell in terms of force response and structural support.

Chapter 9 : Appendices

ON DISK

Directory 'A1' – Contains IBW_Read.m. This is the Matlab source code for the reading of IBW AFM data file based on the work of R.Naud.

Directory 'Actin' – Contains the Matlab source code for the image analysis of actin image based on thinning.

Directory 'Common_m_files' – Contains files that were reused in different parts of the project. This is required in the search path for almost all the project code.

Directory 'ContactPointEditor' – This contains a standalone contact point editor that reads a local AFM database and Matlab application that contains various contact fitting strategies used in the project.

Directory 'DatabaseProcessing_WithEnergyLevels' – This directory contains the code for processing the local AFM database and extracting the Energy lengths as described in chapter 6.

Directory 'Yeastfulldataset29092011' – This contains the raw data from the yeast cell lines that was used to test the contact point methods used in the project.