Application of Control, Modelling and Optimisation to Biomaterials Manufacturing

A thesis submitted to the University of Manchester
for the degree of Doctor of Philosophy
in the Faculty of Engineering and Physical Sciences

Oliver Kaan Onel

2013

School of Electrical and Electronic Engineering
Contents

Contents....................................................................................................................................... 3
List of Figures ............................................................................................................................... 6
List of Tables .............................................................................................................................. 10
Abstract ..................................................................................................................................... 11
Lay Abstract ............................................................................................................................... 12
Declaration & Copyright Statement .............................................................................................. 13
  Declaration ............................................................................................................................. 13
  Copyright Statement .............................................................................................................. 13
Acknowledgements ..................................................................................................................... 14
Nomenclature ............................................................................................................................. 15
  Symbols .................................................................................................................................. 15
  Abbreviations ........................................................................................................................ 18
  Glossary ................................................................................................................................. 20
1  Introduction ........................................................................................................................ 21
  1.1  Biomaterials Manufacturing ........................................................................................... 23
  1.2  Solvent Casting ............................................................................................................ 24
  1.3  Electrospinning ............................................................................................................. 24
  1.4  Problem Definition and the Suggested Approach ............................................................ 25
  1.5  Summary and Main Aims .............................................................................................. 28
2  Background ......................................................................................................................... 30
  2.1  Nerve Regeneration and the Solvent Casting Method...................................................... 30
  2.2  Tendon Repair and Electrospinning ............................................................................... 37
    2.2.1  Solution Parameters ............................................................................................... 40
    2.2.2  Environmental Parameters ..................................................................................... 43
    2.2.3  Process Parameters ............................................................................................... 45
    2.2.4  Research Interests ................................................................................................ 50
    2.2.5  Summary .............................................................................................................. 51
  2.3  Control, Multivariable Systems and Modelling ................................................................. 51
    2.3.1  Systems and Control ............................................................................................. 51
    2.3.2  Model Predictive Control ....................................................................................... 61
    2.3.3  Electrospinning, Solvent Casting and MPC ............................................................... 71
    2.3.4  System Identification, Multivariable Systems and Modelling Methods ................. 73
    2.3.5  Multivariable Systems ............................................................................................ 78
    2.3.6  Statistical Methods ................................................................................................. 79
    2.3.7  Output Error and Prediction Error Methods ............................................................. 88
    2.3.8  Design of Experiments ........................................................................................... 91
  2.4  Summary ..................................................................................................................... 95
3 Solvent Casting

3.1 Properties of Solvent Cast Films and Manufacturing Issues

3.1.1 Solvent Cast Films

3.1.2 Manufacturing Issues

3.2 Development of the Experimental Setup

3.3 Experimental Methods & Results

3.3.1 Temperature and Humidity Test

3.3.2 Evaporation Rate Study

3.3.3 Cell Attachment Analysis

3.3.4 Discussion

3.3.5 Summary

4 Electrospinning

4.1 Development of the Experimental Setup

4.1.1 Hardware

4.1.2 Software

4.2 Experimental Methods and Results

4.2.1 Collection

4.2.2 Taylor Cone Formation and Jet Splitting

4.3 Summary

5 Numerical Methods

5.1 Theory of the Numerical Methods

5.1.1 Least Squares Estimation and the ARX Model

5.1.2 Partial Least Squares - PLS

5.1.3 Recursive Least Squares, Unbiased Recursive Least Squares and Unbiased PLS

5.1.4 OLS, PLS and RLS Modelling Examples

5.1.5 Output Error Method and Prediction Error Methods

5.1.6 Hybrid PLS Methods

5.1.7 Further Modelling Examples

5.2 Batch Process Methods

5.2.1 Multivariable Statistical Process Control

5.2.2 Pre-Treatment of Data and Unfolding

5.2.3 T² and Squared Prediction Error (SPE)

5.2.4 Product Quality Improvement with MSPC

5.3 Summary

6 Application of Numerical Methods to MIMO Systems

6.1.1 Application of PLS to PENSIM

6.1.2 Multi Input Multi Output System

6.1.3 Highly Correlated Multi Input Multi Output System
6.2 Guidance on the Application of MPC to Electrospinning and Solvent Casting
6.2.1 MPC for Solvent Casting
6.2.2 MPC for Electrospinning
6.3 Summary

7 Project Conclusions and Recommended Future Work
7.1 Conclusions
7.2 Future Work

8 Bibliography

9 Appendices

Appendix 1
Appendix 2
Appendix 3
Appendix 4
Appendix 5
Appendix 6
Appendix 7
Appendix 8

Word Count: 68,710
List of Figures

Figure 1.1; Solvent Cast Film and Electrospun Mat ................................................................. 23
Figure 1.2; Solvent Casting; Dispensing, Evaporation and Creation of Pits and Pores ............... 24
Figure 1.3; Electrospinning Overview ...................................................................................... 25
Figure 1.4; Microscope Images of Electrospun Fibres (1st row), Solvent Cast films (2nd row), Microscope images of Solvent Cast Films (3rd row) ......................................................... 26
Figure 1.5; Project Overview .................................................................................................... 28
Figure 2.1; Neuron .................................................................................................................... 31
Figure 2.2; Nerve Conduit and Nerve Repair (M. Sun et al. 2010) ........................................ 35
Figure 2.3; Water Content ....................................................................................................... 44
Figure 2.4; Taylor Cone ............................................................................................................. 46
Figure 2.5; Receding Taylor Cone and Multiple Jets ................................................................. 47
Figure 2.6; System Block Diagram ........................................................................................ 51
Figure 2.7; System with Disturbance ...................................................................................... 54
Figure 2.8; Control System Overview .................................................................................. 55
Figure 2.9; PID Loop ............................................................................................................... 55
Figure 2.10; Smith Predictor Overview ................................................................................... 59
Figure 2.11; IMC Overview ..................................................................................................... 60
Figure 2.12; Moving Horizon MPC ....................................................................................... 62
Figure 2.13; Conventional Control vs. MPC ....................................................................... 66
Figure 2.14; MPC Loop ......................................................................................................... 70
Figure 2.15; PRBS and GBN Signals ..................................................................................... 75
Figure 2.16; PRBS and GBN Power Spectrum ....................................................................... 76
Figure 2.17; Raw and Low Pass Filtered White Noise ............................................................... 77
Figure 2.18; Raw and Low Pass Filtered White Noise Power Spectrum .............................. 77
Figure 2.19; Multivariable Systems ....................................................................................... 79
Figure 2.20; Full Factorial Map for Three Variables .............................................................. 91
Figure 3.1; Time Lapse Picture of Solvent Casting ................................................................. 98
Figure 3.2; Inner Door and Glove Ports .............................................................................. 101
Figure 3.3; CLIMACELL Chamber ...................................................................................... 101
Figure 3.4; Environmental Chamber Controller Inverse Response 35°C 80% RH .................. 103
Figure 3.5; Environmental Tests .......................................................................................... 104
Figure 3.6; 15°C - 20% RH PCL-PLA (Mobasser, 2012) ....................................................... 105
Figure 3.7; 15°C - 50% RH PCL-PLA (Mobasser, 2012) ....................................................... 106
Figure 3.8; 15°C - 80% RH PCL-PLA (Mobasser, 2012) ....................................................... 106
Figure 3.9; 23°C - 50% RH, PCL (Mobasser, 2012) ............................................................... 107
Figure 3.10; Pit Area Distributions (Mobasser, 2012) ........................................................... 108
Figure 3.11; Evaporation Rate at Different Environmental Conditions ............................... 110
Figure 6.28; Variance of the Correlated System Identified with RGS 0.01 ........................................ 232
Figure 6.29; RMSE of the Correlated System Identified with GBN 64 ............................................. 233
Figure 6.30; Variance of the Correlated System Identified with GBN 64 ........................................ 234
Figure 6.31; RMSE, GBN Correlated - Step Dynamic 10% ................................................................. 235
Figure 6.32; Variance, GBN Correlated Step Dynamic 10% ............................................................. 236
Figure 6.33; RMSE, GBN Correlated INMA Disturbance ................................................................. 237
Figure 6.34; Variance, GBN Correlated INMA Disturbance ............................................................. 238
Figure 6.35; Solvent Casting MSPC Control ..................................................................................... 240
Figure 6.36; Batch MPC for Solvent Casting ..................................................................................... 242
Figure 6.37; Model Based Approach Overview for Electrospinning .................................................. 243
List of Tables

Table 2.1; Comparison of Fibre Manufacturing Techniques ............................................................... 39
Table 2.2; Three Variable Factorial Study Table ............................................................................ 92
Table 2.3; Fractional Design Table ............................................................................................... 93
Table 3.1; Water Content of Applied Tests .................................................................................. 104
Table 3.2; Pit Diameter Study Results ........................................................................................ 107
Table 3.3; Evaporation Test Results ........................................................................................... 110
Table 3.4; Cell Test Samples ...................................................................................................... 112
Table 4.1; Experiment Settings .................................................................................................. 132
Table 4.2; Taylor Cone Observations .......................................................................................... 140
Table 4.3; Experiment Settings .................................................................................................. 141
Table 4.4; Results of the Experiment ......................................................................................... 142
Table 4.5; Experiment Settings for the Modified Electric Field Tests ............................................. 148
Table 4.6; Protrusion Test Results .............................................................................................. 150
Table 4.7; Experiment Settings .................................................................................................. 153
Table 4.8; Conditions and Fibre Diameter Results ..................................................................... 154
Table 4.9; Example Input Data Matrix- Part 1 ............................................................................. 157
Table 4.10; Example Input Data Matrix- Part 2 ......................................................................... 158
Table 4.11; Example Output Data Matrix .................................................................................... 158
Table 5.1; Excitation Signal Properties ........................................................................................ 188
Table 5.2; Mean Squared Error of the Repeated Tests ................................................................. 189
Table 6.1; Electrospinning PLS Data ........................................................................................... 210
Table 6.2; Identification Signal Properties ................................................................................... 214
Table 6.3; Output 1 ................................................................................................................... 220
Table 6.4; Output 2 ................................................................................................................... 220
Table 6.5; Output 1 ................................................................................................................... 221
Table 6.6; Output 2 ................................................................................................................... 222
Table 6.7; Output 1 ................................................................................................................... 225
Table 6.8; Output 2 ................................................................................................................... 226
Table 6.9; Output 1 ................................................................................................................... 226
Table 6.10; Output 2 ................................................................................................................. 227
Table 6.11; Identification Signal Properties ................................................................................. 228
Abstract

Institution: The University of Manchester
Candidate Name: Oliver Kaan Onel
Degree Title: Doctor of Philosophy (Ph.D)
Thesis Title: Application of Control, Modelling and Optimisation to Biomaterials Manufacturing
Date: 20 October 2012

This thesis presents the work conducted during a three year research project in the field of Control Systems and Biomaterials Engineering. The findings are presented over seven chapters, starting with a thorough literature review of the existing methods and key technologies, and following through by highlighting the existing problems with the current methods and how they have been overcome. The data is presented in tables, figures and photographs to enhance understanding and clarification.

The research focuses on two relatively new manufacturing methods in the field of Tissue Engineering. Both of the methods are used for creating materials for regeneration of human and animal tissue, with the aim of replacing the current surgical methods. The methods are viewed from a control systems perspective and improvements have been made with the implementation of new technologies and methods. Additionally, further advancements are presented on the theoretical modelling field of control systems, where the shortfalls of existent modelling methods are highlighted and solutions proposed.
Lay Abstract

Tissue engineering is a relatively new field of research, concerned with creating biological structures that can be used to repair and further improve damaged or existing organs. Biomaterials manufacturing is a key element of tissue engineering, as the supporting structures need to be created in a reliable and consistent manner to support the biological tissue.

This project is concerned with determining the key variables, which will be used for improving and applying control to two manufacturing methods currently used by the Biomaterials Research Group at the School of Materials. The two processes are electrospinning, for creating fine fibre mats used in tendon repair, and solvent casting, for creating nerve support tubes used to heal injured nerves. Both processes use biodegradable materials, which after a period of time are harmlessly dissolved in the body of the patient.

In their original configurations the two methods were affected by inherent problems and limitations, which have been rectified in this work. The processes mainly suffered from a lack of repeatability and varying process parameters that were not previously measured, or controlled. Before investigating the underlying dynamics of these manufacturing methods, the inherent problems were solved by the introduction of regulated environments, data logging and computer control, where most of the systems were designed and built in house. With the implemented improvements and changes in their configuration, the most important variables of the systems were determined. In this work, the control of surface morphologies of solvent cast products and fibre diameters of the electrospinning process are demonstrated. These findings open further areas for investigation to allow these methods to be utilised to their full potential.

The main theme of this thesis is to view these processes from a control systems context, where modelling them with the aim of applying advanced control techniques is of interest. However, for this to take place, improvements implemented in the physical systems and the discovery of influencing variables was necessary. The introduction of model based control will enable optimising and calculation of process parameter values for creating biomaterials with desired specifications. To demonstrate this concept, structurally similar simulation studies were undertaken, where novel hybrid modelling methods are proposed. Both of these manufacturing methods are multivariable and there are challenges associated with modelling multivariable systems that have correlations and disturbances. Through the analysis of literature the Partial Least Squares method has been identified as a suitable method to model electrospinning and solvent casting. In the theoretical analysis and the simulation parts of this work, the Partial Least Squares for modelling the multivariable systems is analysed, and its shortcoming are highlighted. Other existing conventional time series methods are also shown to be improved with the newly proposed hybrid methods.
Declaration & Copyright Statement

Declaration

No portion of the work referred to in the thesis has been submitted in support of an application for another degree or qualification of this or any other university or other institute of learning.

Copyright Statement

Copyright in text of this thesis rests with the Author. Copies (by any process) either in full, or of extracts, may be made only in accordance with instructions given by the Author and lodged in the John Rylands University Library of Manchester. Details may be obtained from the Librarian. This page must form part of any such copies made. Further copies (by any process) of copies made in accordance with such instructions may not be made without the permission (in writing) of the Author.

The ownership of any intellectual property rights which may be described in this thesis is vested in the University of Manchester, subject to any prior agreement to the contrary, and may not be made available for use by third parties without the written permission of the University, which will prescribe the terms and conditions of any such agreement.

Further information on the conditions under which disclosures and exploitation may take place is available from the Head of School of Electrical and Electronic Engineering.
Acknowledgements

There are many people I would like to thank for their continuous support during the completion of my research and during the writing up of this thesis, but sadly it would be impossible to mention everyone here. Firstly, I would like to thank my supervisor Barry Lennox for his constant support and constructive criticism of my work, as well as the time and help he has given me for finalising the thesis. I would also like to extend my thanks to the Biomaterials Research Group at the University of Manchester’s School of Materials, particularly Brendan Robb and Atefeh Mobasseri. Others I would like to mention, for their input during the different stages of the project, are William Heath, Zhengtao Ding, John Medland, Nigel Schofield and Ian Cotton. I would also like to acknowledge the efforts of the mechanical workshops at the School of Materials and at the School of Electrical and Electronic Engineering. I would like to thank the Engineering and Physical Sciences Research Council for providing the funding for this project. Finally, I would like to thank my wife Charlotte Allen-Onel, my family (both in the UK and Turkey) and the Allen family for the love, encouragement and the patience they have shown me throughout this project.
Nomenclature

The symbols defined here are presented in appearance order for reference purposes. The notation used is explained throughout this thesis.

Symbols

t Time denoting one sample in continuous time representations

\( u^m \) \( m^{th} \) element of the input vector

\( y^n \) \( n^{th} \) element of the output vector

\( a \) Input parameter vector

\( b \) Output parameter vector

\( e \) Gaussian white noise vector with zero mean

\( d \) Disturbance vector

\( Y(s) \) Laplace representation of Output vector

\( U(s) \) Laplace representation of Input vector

\( G(s) \) Laplace representation of the transfer function

\( g(\tau) \) Impulse response at time \( \tau \)

\( d\tau \) Derivative with respect to \( \tau \)

\( q^{-1} \) Backwards shift operator

\( G(q) \) The discrete transfer operator

\( G(z) \) The discrete transfer function

\( h(k) \) The impulse response representation of the disturbance

\( H(q) \) The discrete transfer operator of the disturbance

\( P \) Output of the proportional part of PID

\( I \) Output of the integral part of PID

\( D \) Output of the derivative part of PID

\( K_p \) Proportional gain of PID

\( K_i \) Integral gain of PID

\( K_d \) Derivative gain of PID

\( T_i \) Integral time or reset time

\( T_d \) Derivative time
$G_c(s)$ Transfer function of the controller
$G_p(s)$ Transfer function of the process
$K$ Process gain
$e^{-\alpha s}$ Laplace representation of the time delay $t_s$
$\alpha$ 1/time constant of the system
$\hat{G}^+_p$ Invertible part of the process transfer function
$\hat{G}^-_p$ Non invertible part of the process transfer function
$G_f$ Transfer function of the low pass filter for IMC
$\tau_f$ Time constant of the low pass filter for IMC
$D(s)$ Laplace domain representation of the disturbance transfer function
$k$ Discrete time sample
$J$ Cost function
$A$ The state matrix
$B$ The input matrix
$C$ The output matrix
$P$ The prediction horizon
$M$ The control horizon
$y_{ref}$ The reference trajectory
$\hat{y}$ The predicted output
$\Delta u$ The control move added to the current output $u(k)$ at the next sampling instant
$Q, R$ Weighting matrices
$S$ Dynamic matrix of step responses, where $S_1$ is the first element
$y_{set point}$ Setpoint of the output
$m$ Number of rows in a data matrix
$n$ Number of columns in a data matrix
$X$ Input data matrix
$Y$ Output data matrix
$T$ Score vector, where $T_1$ is the first element
$P$ Loadings vector, where $P_1$ is the first element
$E$ Residual vector
\( \beta \) Parameter vector
\( \hat{\beta} \) Estimated parameter vector
\( \hat{R} \) Regression vector
\( h_k \) Finite impulse response representation of the system
\( \epsilon_{oe} \) Output error
\( \hat{y}(t) \) Predicted output
\( \epsilon_{pem} \) Prediction error
\( A(q) \) Discrete representation of the output parameter vector
\( B(q) \) Discrete representation of the input parameter vector
\( \lambda_k \) \( k^{th} \) Eigenvector of the covariance matrix
\( \theta \) System parameter vector
\( \hat{\theta} \) Estimated system parameter vector
\( J_{ARX} \) ARX cost function
\( J_{LS} \) Least Squares cost function
\( \Phi_u \) Power spectrum of the input signal
\( \Phi_r \) Power spectrum of the disturbance
\( F \) Frequency weighting parameters
\( L(q) \) Discrete transfer function representation of a stable filter
\( u_{filt} \) Filtered input \( u(t) \)
\( y_{filt} \) Filtered output \( y(t) \)
\( u_k \) \( k^{th} \) score vector for input variables for PLS
\( w_k \) \( k^{th} \) weighting vector for input variables for PLS
\( t_k \) \( k^{th} \) score vector for output variables for PLS
\( q_k \) \( k^{th} \) loading vector for input variables for PLS
\( e \) error tolerance
\( p_k \) \( k^{th} \) loading vector for output variables for PLS
\( b_k \) \( k^{th} \) model coefficient for the inner PLS scores
\( E \) Residual matrix for the input predictions
$F$ Residual matrix for the output predictions
$r_k$ $k^{th}$ regression vector for PLS
$S$ Initial cross product matrix for SIMPLS
$\lambda$ Forgetting factor for RLS
$J_{RLS}$ Cost function of RLS
$L$ Lower triangular matrix (in RLS)
$D$ Diagonal matrix (in RLS)
$v(t)$ Disturbance signal used for simulations
$R^2$ Coefficient of determination

### Abbreviations

This list shows the abbreviations used in this thesis in their alphabetical order.

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Full Form</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>AC</strong></td>
<td>Alternating Current</td>
</tr>
<tr>
<td><strong>ARIMAX</strong></td>
<td>Autoregressive Integrating Moving Average Exogenous model</td>
</tr>
<tr>
<td><strong>ARMAX</strong></td>
<td>Autoregressive Moving Average Exogenous model</td>
</tr>
<tr>
<td><strong>ARMAX-PLS</strong></td>
<td>ARMAX – Partial Least Squares hybrid model</td>
</tr>
<tr>
<td><strong>ARX</strong></td>
<td>Autoregressive Exogenous model</td>
</tr>
<tr>
<td><strong>ARX-PLS</strong></td>
<td>ARX – Partial Least Squares hybrid model</td>
</tr>
<tr>
<td><strong>BRG</strong></td>
<td>Biomaterials Research Group</td>
</tr>
<tr>
<td><strong>BJ</strong></td>
<td>Box Jenkins method</td>
</tr>
<tr>
<td><strong>bPLS</strong></td>
<td>Biased Partial Least Squares</td>
</tr>
<tr>
<td><strong>bRLS</strong></td>
<td>Biased Recursive Least Squares</td>
</tr>
<tr>
<td><strong>CA</strong></td>
<td>Cellulose Acetate</td>
</tr>
<tr>
<td><strong>CAD</strong></td>
<td>Computer Aided Design</td>
</tr>
<tr>
<td><strong>DAQ</strong></td>
<td>Data Acquisition</td>
</tr>
<tr>
<td><strong>DMC</strong></td>
<td>Dynamic Matrix Control</td>
</tr>
<tr>
<td><strong>DOE</strong></td>
<td>Design of Experiments</td>
</tr>
<tr>
<td><strong>DPCA</strong></td>
<td>Dynamic Partial Least Squares</td>
</tr>
<tr>
<td><strong>ECM</strong></td>
<td>Extracellular Matrix</td>
</tr>
<tr>
<td><strong>FDA</strong></td>
<td>Federal Drugs Agency (of USA)</td>
</tr>
<tr>
<td><strong>FEA</strong></td>
<td>Finite Element Analysis</td>
</tr>
<tr>
<td><strong>FIR</strong></td>
<td>Finite Impulse Response</td>
</tr>
<tr>
<td><strong>FPGA</strong></td>
<td>Field Programmable Gate Array</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Description</td>
</tr>
<tr>
<td>--------------</td>
<td>------------------------------------------</td>
</tr>
<tr>
<td><em>GBN</em></td>
<td>Generalised Binary Noise</td>
</tr>
<tr>
<td><em>GLS</em></td>
<td>Generalised Least Squares</td>
</tr>
<tr>
<td><em>HFIP</em></td>
<td>Hexa - Fluoroisopropanol</td>
</tr>
<tr>
<td><em>IMC</em></td>
<td>Internal Model Control</td>
</tr>
<tr>
<td><em>LQG</em></td>
<td>Linear Quadratic Gaussian control</td>
</tr>
<tr>
<td><em>MIMO</em></td>
<td>Multiple Input Multiple Output system</td>
</tr>
<tr>
<td><em>MISO</em></td>
<td>Multiple Input Single Output system</td>
</tr>
<tr>
<td><em>MPC</em></td>
<td>Model Predictive Control</td>
</tr>
<tr>
<td><em>MPCA</em></td>
<td>Multiway Principal Components Analysis</td>
</tr>
<tr>
<td><em>NI</em></td>
<td>National Instruments</td>
</tr>
<tr>
<td><em>NIPALS</em></td>
<td>Nonlinear Iterative Partial Least Squares</td>
</tr>
<tr>
<td><em>OE</em></td>
<td>Output Error method</td>
</tr>
<tr>
<td><em>OLS</em></td>
<td>Ordinary Least Squares</td>
</tr>
<tr>
<td><em>OE-PLS</em></td>
<td>OE-Partial Least Squares hybrid model</td>
</tr>
<tr>
<td><em>PCA</em></td>
<td>Principal Component Analysis</td>
</tr>
<tr>
<td><em>PCL</em></td>
<td>Polycaprolactone</td>
</tr>
<tr>
<td><em>PCR</em></td>
<td>Principal Component Regression</td>
</tr>
<tr>
<td><em>PDLA</em></td>
<td>Polyoxy(D,l-lactic acid)</td>
</tr>
<tr>
<td><em>PEM</em></td>
<td>Prediction Error Method (using the Box Jenkins structure)</td>
</tr>
<tr>
<td><em>PEM-PLS</em></td>
<td>PEM – Partial Least Squares hybrid model</td>
</tr>
<tr>
<td><em>PGA</em></td>
<td>Polyglycolic Acid</td>
</tr>
<tr>
<td><em>PID</em></td>
<td>Proportional Integral Derivative control</td>
</tr>
<tr>
<td><em>PLA</em></td>
<td>Polylactide</td>
</tr>
<tr>
<td><em>PLLA</em></td>
<td>Poly(l-lactide-co-glycolic Acid)</td>
</tr>
<tr>
<td><em>PLS</em></td>
<td>Partial Least Squares / Projection to Latent Structures</td>
</tr>
<tr>
<td><em>PRBS</em></td>
<td>Pseudo Random Binary Sequence</td>
</tr>
<tr>
<td><em>PVA</em></td>
<td>Polyvinyl Alcohol</td>
</tr>
<tr>
<td><em>PVP</em></td>
<td>Poly(vinylpyrrolidone)</td>
</tr>
<tr>
<td><em>RLS</em></td>
<td>Recursive Least Squares</td>
</tr>
<tr>
<td><em>SEM</em></td>
<td>Scanning Electron Microscopy</td>
</tr>
<tr>
<td><em>SISO</em></td>
<td>Single Input Single Output system</td>
</tr>
<tr>
<td><em>UoM</em></td>
<td>University of Manchester</td>
</tr>
<tr>
<td><em>uRLS</em></td>
<td>Unbiased Recursive Least Squares</td>
</tr>
<tr>
<td><em>uRLS-PLS</em></td>
<td>uRLS – Partial Least Squares hybrid model</td>
</tr>
</tbody>
</table>
Glossary

This glossary defines a list of terms and names that are used throughout this thesis. Some of the definitions are from the Oxford English Dictionary (Oxford Dictionaries 2012).

**Aqueous** : A substance that is containing or is of water.

**Biodegradable** : A substance or object capable of being decomposed by bacteria or other living organisms.

**Biomaterial** : A biological or synthetic substance that is compatible with living tissue and is suitable for implantation.

**Electron Beam Lithography** : A method of creating patterned films using electron beams.

**Electrospinning** : A method of creating thin fibres in the nano and micro scale by electrostatically charging a conductive solution.

**Fibremetric** : A commercial image analysis software that is part of the Phenom Scanning Electron Microscopy system.

**Hydrophilic** : A substance that has a tendency to mix with, dissolve in, or be wetted by water.

**Hydrophobic** : A substance that tends to repel or fail to mix with water.

**ImageJ** : An open-source image processing and analysis software.

**Lead-Lag compensator** : A frequency response compensation method used in control system and operational amplifier circuit design.

**Liquid phase separation** : The process of separating various particles and liquids that are homogenously mixed into one solution.

**Luer lock** : A standardised fluid fitting for small scale applications such as medical hypodermic needles.

**Motor function** : Nervous system activity.

**Nerve grafting** : The method of transplanting a living nerve tissue.

**NG108 nerve cell** : A commonly used and commercially available nerve cell used in research.

**Orthogonal** : Two vectors or lines that are perpendicular to each other, which also means they are unrelated.

**Peptide** : A biological compound consisting of two or more amino acids linked in a chain, the carboxyl group of each acid being joined to the amino group of the next by a bond.

**Polysaccharide** : A type of carbohydrate whose molecules consist of a number of sugar molecules bonded together.

**Sublimation** : A type of phase change, where a solid directly becomes gaseous.

**Solvent casting** : A method of creating thin films by evaporation.

**Tissue engineering** : A multidisciplinary field consisting of materials science, engineering, and biochemistry where the main aim is to engineer living cell structures to heal damaged organs.
1 Introduction

This thesis presents the work undertaken during a three year research project on the topics of biomaterials manufacturing and control systems. The main aim of the project is to investigate how control systems and engineering methods can be applied to improve the operation of two biomaterials manufacturing techniques, electrospinning and solvent casting, which are currently used at the University of Manchester Materials Science Centre’s Biomaterials Research Group (BRG).

The main research field of the BRG is the repair of damaged tissue, such as nerves, tendons and bones. Their current focus is on the manufacture of artificial support structures for cell growth, that are biocompatible with the human body. The main reason for researching artificial structures for tissue repair is for these products to eventually surpass the currently used conventional techniques, and offer a better solution in the long term. The two methods used for manufacturing artificial support structures in this project are, electrospinning for tendon repair and solvent casting for nerve repair studies. Electrospinning is a method of creating fine fibrous mats using electrostatic forces, and solvent casting is an evaporation based process used for creating thin films. These methods initially suffered from a lack of repeatability, mainly caused by an incomplete understanding of the underlying processes.

The contributions of this thesis are summarised below:

- Identification of methods to manipulate the diameters of electrospun fibres by using a modified experimental set up.
- Identification of a novel method for monitoring the electrospinning process using measurements of the electric current.
- Identification of how environmental conditions affect the properties of thin films made from a specific polymer solvent mixture and subsequently used for nerve regeneration.
- Identification of an ideal surface morphology for cell attachment on thin films used for nerve regeneration.
- Development and analysis of new methods for modelling highly correlated data sets.

Publications relating to this thesis are:


• B. Robb, O. Onel, B. Lennox, S. Downes, “Optimising Electrospun Biodegradable Scaffolds to Encourage Specific Tissue Regeneration”. 2012, World Biomaterials Congress, China


The contributions of the work described in this thesis towards the first paper was the introduction of a consistent environment and data logging to the manufacturing process, the selection of the manufacturing conditions and manufacturing the films used in the experiments and their analysis. The second paper presents a summary of the work detailed in Chapter 5 of this thesis related to the application of PLS models within MPC controllers. The third and fourth publications are on the topic of electrospinning. The contribution from this thesis is the design and build of a new electrospinning set-up with computer control and data logging, the selection of tests to manufacture the fibres and the analysis of fibres using a Scanning Electron Microscope (SEM).

This thesis is presented over seven chapters. The introductory chapters, Chapter 1 and Chapter 2 are aimed at providing the reader with an understanding of the main aims, the justification of the work and the background research of the topics encompassing the scope of this project.

Chapters 3 and 4 explain in detail the manufacturing methods of solvent casting and electrospinning respectively. Both chapters identify the current problems associated with each method, and present the modifications that have been proposed and implemented in this research. This is followed by a description of the work conducted to better understand these processes, the experimental methods and the results obtained.

Chapter 5 presents the numerical methods proposed for modelling the manufacturing processes in detail. The theory of the numerical methods is explained together with the new hybrid modelling methods, and a short example is presented to demonstrate the theoretical concepts. Chapter 6 then presents the application of the modelling methods to more complex systems, which are similar in structure to the electrospinning and solvent casting processes. Chapter 7, the final chapter, critically analyses the work undertaken, assesses the contributions of the work and presents future work proposals.

Before describing the aims and the details of the technical work completed in this project, a brief background to the biomaterials manufacturing processes investigated is provided. This is presented in the next section, and is followed by a description of the identified problems and suggested solutions.
1.1 Biomaterials Manufacturing

Advances in technology since the late 1970s in the medical, biomaterials and engineering sectors have given rise to the multidisciplinary field of biotechnology. This is described by the Organisation for Economic Co-Operation and Development as “the application of scientific and engineering principles to the processing of materials by biological agents to provide goods and services” (Bud 1994).

Biomaterials research at the University of Manchester (UoM) is conducted collaboratively between research teams from the School of Materials, Medicine and Electrical and Electronic Engineering. The biomaterials research in the School of Materials is primarily concentrated on tendon, nerve and bone regeneration. Patents on certain materials and techniques have been filed and promising results have been achieved with the manufactured products tested under previous research (Downes and Bosworth, 2011), (Downes et al., 2011), (Downes et al., 2012).

The manufacturing techniques that are of interest to this study are electrospinning and solvent casting. Both of these tissue engineering techniques are used in the creation of biological platforms, such as tissue scaffolds and nerve regeneration structures (Yoshimoto et al., 2003), (Yang et al., 2005a), (Mo et al., 2004). Tissue engineering is described as the "application of the principles and methods of engineering and life sciences towards the fundamental understanding of structure-function relationships in normal and pathological mammalian tissues and the development of biological substitutes that restore, maintain or improve tissue function” (Shalak and Fox 1988). The biological platforms are created using support structures made from biodegradable polymers, and these engineered biodegradable polymer structures aid the growth of cells once inserted into the patient. The material naturally degrades inside the body without the need for removal, and the end product is a robust and organised structure of cells imitating organic tissue (Liao et al., 2002). Examples of the final materials are shown below for solvent casting and electrospinning, in Figure 1.1 respectively. The solvent cast film on the left has a porous surface used for planting cells, and similarly the electrospun scaffold on the right has aligned micro-fibres imitating a fibrous biological structure like a tendon. The manufacturing techniques are described in the following section.

![Figure 1.1; Solvent Cast Film and Electrospun Mat](image-url)
1.2 Solvent Casting

The solvent casting method of manufacturing for plastic films is well established. Since the introduction of this technique by the photographic plastic film industry, the application of solvent casting has become popular in specialised industries, which require high specification materials (Siemann, 2005). Solvent casting has now become an accepted biomaterials manufacturing process for creating ultra-thin biodegradable films. This method is used by the BRG for creating thin films to form nerve regeneration support structures.

The solvent casting process begins with making a chemical solution consisting of a biodegradable polymer and a liquid solvent to dissolve it, as the polymer is originally manufactured in a powder form. Following mixing on a magnetic stirrer for 24 hours, a predetermined volume of the solution is injected on to a glass plate with the aid of a pipette. The solvent then evaporates leaving behind a thin film of dried polymer, which forms a porous structure, that can be planted with cells for the purposes of nerve regeneration. An overview of this process, from dispensing through to evaporation is illustrated in the diagrams from left to right in Figure 1.2.

![Figure 1.2; Solvent Casting; Dispensing, Evaporation and Creation of Pits and Pores](image)

1.3 Electrospinning

The electrospinning method was first patented by Anton Formhals (1934) in the field of textiles manufacturing. This method works by charging a polymer solution that is pre-loaded into a syringe. The solution is charged by applying a high voltage to the needle, also termed the *spinneret*. This solution is then encouraged to flow through the needle by a syringe pump or gravity, and is drawn out by the electrostatic force, shooting out of the syringe tip in fine fibres on to a grounded metal collector plate (Ramakrishna, 2005). Following this, the produced fibre mats are planted with cells, where the aim is to form a tendon-like structure that can be used to heal muscle injuries in patients.

A representative set up of an electrospinning system is shown in Figure 1.3. The conductive needle is connected to the positive side of the power supply, where the solution that is pushed through with a syringe pump, becomes charged and shoots out to the grounded collector plate.

The School of Materials is working to perfect the method for manufacturing strong and aligned fibres of uniform diameter ranges, which can be used to form structures that will help heal damaged
tendons once planted with cells. Uniformity is important when growing cells and providing mechanical support. The scaffold will replace part of a tendon and as it is a load bearing structure, it should ideally not have any weak spots.

The methods explained in the previous section are currently used for biomaterial scaffold production. These will be presented in more detail in the following chapters.

![Electrospinning Overview](image)

**Figure 1.3; Electrospinning Overview**

### 1.4 Problem Definition and the Suggested Approach

At the start of this project, the quality of the final products created using electrospinning and solvent casting suffered from a lack of consistency, as there was limited understanding of the underlying cause and effect of the relationships within the processes. Prior to the work conducted in this thesis, both processes were undertaken in ambient conditions without any environmental control or data logging. The films and mats manufactured varied depending on environmental conditions in the laboratory and the raw material used. For example, on certain occasions the electrospinning process would not be consistent with the same solvent polymer mixture, and the same process settings used on a previous day. Some example images of the final products from the previous configurations are shown in the images of Figure 1.4.
In Figure 1.4, the top two images are Scanning Electron Microscope (SEM) close ups of two electrospun fibre mats, where the distributions and diameters of the fibres are varied and inconsistent. The second row of images show solvent cast films made on the bench top on two different days, using the same solution mixture. There are clear visible differences in appearance and
surface morphology. The third row shows SEM images of two solvent cast films made on the bench top on different days, and again there is a clear difference in the sizes of the pits and pores. These images demonstrate the variability of manufacturing films under uncontrolled conditions.

The experiments conducted by the BRG involved producing artificial scaffolds and then implanting these with cells. The growth of these cells was then analysed with the hope that a strong distribution of cells would encourage active tissue growth. Unfortunately, despite some scaffolds being produced that encouraged cell growth, due to the lack of repeatability and data it was not possible to reach any conclusions on the conditions necessary to produce high quality materials.

The initial experimental set up for solvent casting at the UoM either used a fume hood with a constant air extraction rate, or experiments were carried out on the laboratory bench top. These approaches were inconsistent, created problems of repeatability and did not provide a way of relating the environmental variables contributing to the evaporation of the solvent, which influenced the final product. Additionally, there was no quality measure in place for the assessment of solvent cast films, which made comparison and analysis difficult.

The electrospinning method is a continuous and dynamic process with multiple inputs. The output is a fibre mat of a desired range of fibre diameters, either aligned or randomly ordered. Like the solvent casting method, this process was also undertaken in a fume hood without environmental control, with electrical disturbances caused by insufficient electrical isolation of the process, and an absence of data logging.

Both methods suffered from a lack of understanding caused by the lack of data logging and control. Some of the key variables were not strictly controlled, which contributed further to the inconsistencies. This project aims to view the two manufacturing techniques from a control systems perspective and develop engineering solutions to address the repeatability problems. Both of the processes are multivariable systems and lend themselves conveniently to be viewed as a process in the control systems sense. The initial assessment of both methods highlighted some essential areas common to both manufacturing techniques, which needed to be addressed in the design and build of the new experimental systems for this project. These were:

- Implementation of environmental control.
- Recording of process measurements.
- Accurate control of certain variables for electrospinning.
- Electrical isolation for electrospinning.
- Solvent evacuation from the manufacturing site for health and safety reasons.
Following the completion of these tasks, the approach was the determination of important variables for both processes. This was to be achieved using small scale factorial studies and the application of Design of Experiments (DOE) where possible, to help reduce the number of tests and save time.

Both methods were viewed as potential Model Predictive Control (MPC) applications and the next chapter explains the reasons behind this. The further parts of this thesis highlight the limitations encountered with current modelling techniques in MPC applications for identifying process models that have highly correlated, time varying and limited amount of data, as identified with these two processes. The diagram in Figure 1.5 shows an overview of how different parts of this project tie together, where engineering solutions to ensure repeatable testing is applied in parallel with the control systems approach. To form a ground study aimed towards applying advanced control to the two manufacturing methods, similar multivariable systems were simulated for use in system identification studies and for calculating models.

1.5 Summary and Main Aims

The School of Materials uses two processes to produce biodegradable artificial tissue support structures. However, these suffer from inconsistency and a lack of in-depth understanding of the effects of the variables on the final products.

The main aims of this project can be summarised as conducting designed experiments and data analysis to understand the solvent casting and electrospinning processes, with a view to improve the quality of the final products. For solvent casting this entailed, identifying conditions for the
evaporation of the solvent cast films for successful cell growth, and the determination of a quality variable so different films could be compared. Additionally, the possibility of applying advanced control techniques to the electrospinning and solvent casting processes was assessed, as a method for identifying and regulating suitable conditions for high quality materials. In the electrospinning part, the measurement and control of the variables was implemented following the development of a new experimental configuration, as the equipment used in past studies was not suitable for reasons previously highlighted. Similar to solvent casting, this system was analysed for its suitability for control. In parallel to the engineering design and development part of the project, highly correlated systems that are similar in structure to electrospinning and solvent casting were modelled through simulations. This acted as a ground study where the models were built using different algorithms, conceptually proving that the electrospinning and solvent casting processes can be modelled in the presence of relevant data. The simulation work also addressed the issues traditional modelling methods face with highly correlated systems and ill conditioned data sets, and suggests how these can be best overcome. These simulated systems are an important foundation for any future work in this research field of modelling the two manufacturing methods.


2 Background

This chapter aims to provide background information on the different areas studied in this project and to give the reader an understanding of the methods; why they are used and how they fit in to the overall picture.

The first section presents the past work in tissue engineering and how it evolved into today's research. Then, both solvent casting and electrospinning methods are explained within their respective tissue regeneration fields and from a control systems perspective. This leads on to the final section, presenting the background of system identification, modelling and control for similar multivariable systems.

2.1 Nerve Regeneration and the Solvent Casting Method

In Europe alone, 300,000 cases of trauma relating to nerve injury are reported annually (Mohanna et al. 2003). A study conducted at a regional trauma centre in Canada showed that 2.8% of all patients, who are mostly affected with lifelong disability as a result of their accidents, had peripheral nerve injuries (Noble et al. 1998). These numerical indicators show the extent of the effect nerve injuries and the size of the impact a reliable nerve regeneration method could potentially have on peoples’ lives.

Today, autogenous nerve grafting is described as the “gold standard” of repairing damaged nerves (Wang et al., 2002). Nerve grafting is the process of using a separate piece of nerve tissue to patch the injured nerve. The new piece is taken from the patient themselves or from a donor and is planted into the wounded area where the injured gap is bridged with this tissue. The first nerve graft was performed more than a century ago by Philipeux and Vulpian in 1870, where an attempt was made to heal nerve defects in dogs. This study resulted in some degree of success, with two out of the seven dogs showing signs of healing (E. S. Dellon and Dellon 1993). Although today, with somewhat scientifically more advanced surgical methods and tools, grafting is still the primary method for healing neural injuries (Chiu et al., 1982), (Roganovic et al. 2007).

However, nerve grafting has its drawbacks. Using nerve grafting limits the amount of nerve tissue that can be taken from the same person, as this method causes morbidity on the donor site, which can result in a permanent loss of nerve function. This also means more than one operation needs to be carried out on the same patient, which is not ideal. If the nerve injury gap is more than 3 cm, the grafting method of treatment is not as successful as for shorter injuries. While it is possible to obtain longer nerve tissue from donors or even from other species, the availability of it is limited when compared to the amount of reported injuries (Fan et al., 2008). Although some successful grafting cases have been reported (Mackinnon et al. 2001), (Young et al. 1980), the recovery rate of motor
function of the healed nerve is estimated to be around 40% (Omer et al., 1998). Additionally, in some cases the patient’s body rejects the graft completely.

These disadvantages highlight the need for alternative methods of treatment such as artificial conduits. Nerve conduits guide the nerves in the correct growth path and promote healing by providing physical support and the correct growth environment. This method of neuroregeneration does not require donor nerves and can be used to heal nerve injury gaps of up to 8cm in length (Wang et al., 2005), (Matsumoto et al., 2000), (Luis et al., 2007).

Nerve regeneration using artificial biodegradable conduits is a field of active research that is mainly at laboratory level, with a small number of commercial products. Researchers are regularly reporting and patenting successful methods, yet the artificial conduits do not have widespread clinical use when compared with nerve grafting. The main reason for the artificial products not being used regularly at the clinical level is that they do not offer a significantly improved alternative over the nerve grafting technique. If the success rate of artificial conduits increases, they will become viable alternatives to nerve grafting in the future.

The main reasons for failure are the material mechanical properties, biodegradation times and the inability to mimic the nerve growth environment successfully (Wang and Lineaweaver 2002), (Jiang et al., 2010), (Pabari et al., 2011). The main view among the research community is that conduits can be improved with interdisciplinary work. By combining the knowledge of materials science, process control, general engineering and medicine, then better materials and methods can be developed. Further potential modifications could also include the customisation of the interior of the conduit for each injury type by using biological platforms such as proteins and other nerve growth enhancing substances. To aid understanding of the concepts explained in the following sections, an overview diagram of a neuron is given in Figure 2.1.

![Figure 2.1; Neuron](image)
Healing of a severed nerve depends on a number of factors. The most important three are:

- The survival of the neuron and for it to switch from "signalling mode" to "growth mode" (Fu and Gordon 1997).
- The axon growth environment around the severed axon stumps must provide enough support for new axons (Müller and Stoll 1998).
- The new axons need to reinnervate the target, meaning they need to form new connections to pass electrical and chemical signals, and the target needs to retain its ability to accept the reinnervation (Fu and Gordon 1997).

Artificial guides need to be able to provide the correct conditions and support in order to meet the points listed above. To achieve this, the nerve guide needs to provide an environment that promotes growth and offers mechanical support. When a nerve is injured, a process called Wallerian degeneration takes place. This is where the severed axon eventually degrades and disappears from the cell body it was attached to (Waller, 1850). The first aim of the nerve conduit needs to be to slow down Wallerian degeneration. This depends on the existence of Schwann cells and the correct extracellular matrix ingredients such as fibronectin and collagen. Schwann cells are the equivalent of caretakers for neurons. When Schwann cells detect they are not in contact with axons they quickly multiply and form a Schwann cell column, which helps provide an ideal environment for axon regeneration (Thoenen et al., 1988), (Heumann, 1987). Therefore being able to accommodate Schwann cells in the nerve conduit is one of the main criteria for successful regeneration.

The initial research on nerve conduits started in the 1980s when researchers tried using silicone tube guides to grow nerves in (Lundborg et al., 1982), (Williams et al., 1983). In due course the disadvantages of using silicone tubing began to emerge, mainly linked to silicone's impermeable structure, which restricts the flow of nutrients that may be needed for further healing. Since then it has been found that silicone can cause compression of the nerve axons during their growth, which can lead to late-stage nerve problems. Additionally, some patients have complained of localised discomfort where the silicone tube is located, leading to surgical removal (Ichihara et al., 2008).

Despite these problems, silicone remained a suitable platform for nerve growth research. More successful results were obtained using a silicone tube with holes in its walls that let oxygen and nutrients through (Jenq and Coggeshall, 1985). The maximum injury gap the silicone tube could heal was found to be limited at 15 mm. Therefore, silicone can be viewed as only a tool to study basic nerve regeneration principles and can be deemed unsuitable for clinical work. Ichihara (2008) lists three main characteristics an ideal nerve regeneration conduit should have, and these are summarised as follows:
• The conduit should be able to provide enough support for axon regeneration but not be so hard that it damages the nerves. It should have similar mechanical properties to real nerve tissue.

• The material should be permeable so the internal and external environments can mix in a controlled manner.

• The conduit should be able to degrade without causing any harm to the patient. The degradation needs to be at a suitable rate for correct healing, as degrading too early can cause the nerve to remain premature. This also eliminates the need for a secondary operation to remove the guide and creates a platform for controlled release of growth factors (Pfister et al., 2007).

Advancements in the field of materials science opened a new area of using conduits made from biodegradable polymers, whose material properties aim to conform to the points made above. The first non-silicone trials were conducted using collagen tubes and similar results to grafting were obtained (Archibald et al., 1991), (Archibald et al., 1995). This meant that for small gap nerve injuries, that are less than 3 cm, it could potentially eliminate the need for grafting. Further experiments conducted with biodegradable synthetic polymer conduits, such as polyglycolic acid (PGA), showed improvements in the regenerated nerve quality and the healed nerve gap length (Dellon and Mackinnon, 1988), (Matsumoto et al., 2000), (Lee et al., 2006). Materials such as PGA, polyvinyl alcohol (PVA), poly(l-lactide-co-glycolic acid) (PLLA), polylactide (PLA) and Polycaprolactone (PCL) were eventually approved by the US Federal Drug Administration (FDA) for clinical trials. There are five commercial nerve guide products made from some of the approved materials. NeuraGen and NEUROMetrix Neuroflex, are made from type I collagen, which is the most common natural protein in the human body. Neurotube is made using PGA, Neurolac is made using PCL and SaluBridge is made from a PVA hydro gel (Schlosshauer et al., 2006a).

Clinics that use Neurolac have reported problems of biocompatibility, swelling and discomfort due to rigidity (Marcel F. Meek and Den Dunnen 2009), (Bertleff et al. 2005). Neurotube has given positive results for shorter nerve injury gaps that are less than 3 cm (Weber et al. 2000), (Donoghoe et al. 2007). NeuraGen has also been outperformed by processed allografts when tested for injuries smaller than 3 cm (Whitlock et al. 2009). A processed allograft is nerve tissue taken from a donor or a cadaver, which is conditioned to be non immunogenic and then treated with growth factors before being planted into the injury site. Out of these five products the most successful is Neurotube, showing a 75% success rate between 1990 and 2005 (Schlosshauer et al. 2006b). Nectow et al. (2012) point out that second generation of nerve conduits like these, following from the first
generation silicone tubes, only provide structural support for nerve growth, and not the full nutritional support environment the nerve needs for full recovery.

Today, in the development phase of the third generation nerve conduits, alongside the healing qualities and biocompatibility, research has focused on the actual conduit properties such as diameter, porosity and other physical characteristics and their modifications. As pointed out in the ideal nerve conduit specifications, being able to incorporate growth factors into the conduit is an important factor in providing a suitable environment for nerve regeneration. It is also possible to make the films so that as it slowly degrades, it releases certain chemical nutrients at a controlled rate, maximising recovery chances.

The most popular materials for manufacturing nerve conduits amongst researchers which offer good potential have been found to be PCL, PLA and PGA. All of the materials are FDA approved and biodegradable. The clinical results of Neurotube using PGA show that it can cause irritation and discomfort for the overlying skin of the injury site (Waitayawinyu et al. 2007) (Weber et al. 2000). This is not ideal and ensuring sufficient tissue coverage following every operation can be unachievable. An alternative study using PCL nerve conduits has shown promising results for nerve regeneration (Meek et al., 2003) followed by a long term analysis (Bertleff et al., 2005) where hand surgery patients were analysed in terms of touch and feel and the reported outcomes were good functional recovery and no irritation. One of the disadvantages of PCL is that it has a long degradation time and it can take years to completely degrade.

Another material used in research is PLA, as it is reported to have better mechanical properties when compared with PGA and is highly permeable, which is desirable for nerve recovery where the flow of nutrients is important. However, Sun (2009) reports that ultrathin PLA films are difficult to work on when implementing surface modifications. It was also found that better mechanical properties were desirable for nerve conduits. Broz et al. (2003) presented a PCL – PLA blend approach where the mechanical properties of the conduit could be fine tuned using the polymer mixture levels. Other researchers also reported the use of blended polymers to get the best out of the materials, rather than strictly choosing one material (Rodríguez et al. 1999), (Koshimune et al. 2003), (Millett et al. 2011).

Following this trend the Biomaterials Research Group also started using the PCL-PLA blended films, which were created using solvent casting. Images of the in vitro study of a previously developed film using only PCL are shown in Figure 2.2. The image on the top shows the nerve conduit implanted in to a 1 cm gap in a laboratory test rat’s sciatic nerve. The second image was taken after the conduit was removed from the test animal after 14 days, and shows the nerve had started recovering. However in this work the films were grown on the bench top and under the fume hood. The authors
reported that the size of the pits and pores may have had an influence on the cell attachment success. The purpose of the solvent casting section of this thesis is the next stage in developing the PCL/PLA blend films and investigating the effect of surface properties for cell growth.

Figure 2.2; Nerve Conduit and Nerve Repair (M. Sun et al. 2010)
Various methods for nerve conduit manufacturing are being used by researchers, such as injection moulding (Sundback et al., 2003), dip coating (Liao et al., 2002), melt extrusion (Chiono et al. 2008), inkjet printing (Radulescu et al., 2007) and solvent casting (Sun and Downes, 2007), (Valeria et al., 2009). Today, there is not one standard method for conduit manufacturing and the preferred methods vary amongst research groups, as the materials that are used for manufacturing the conduits tend to govern the manufacturing method. Not all the manufacturing techniques are suitable for all materials, for example melt extrusion cannot be used with PLA due to its thermal instability. Dip coating and injection moulding do not allow for internal surface modifications for successful cell attachment and creation of the ideal environment for nerve growth.

Inkjet printing and solvent casting allow for growth factor planting and internal surface modifications. Using the inkjet printing method layers of materials can be printed on top of each other, incorporating nerve growth factors such as proteins. However, the final internal structure for chemical modifications is limited. Solvent casting is the most versatile and rapid technique for such research. This method produces ultra-thin films made from biodegradable synthetic materials that can be planted with cells that enable axon growth. Additionally the surface of the film can be conditioned further for cell growth and modified with chemo-tactic cues which can guide the growth direction of the nerve cells. Following the modifications the film is planted with nerve cells, which can then be implanted into the patient and rolled up to form a conduit. However, once the final desired conduit morphology is identified and finalised then an investment can be made to upgrade the process to a high volume production method similar to the inkjet approach or an automated version of solvent casting.

Research towards modifying the internal properties of the conduits to increase nerve growth success also took place in parallel to the material properties studies. Some examples are, the use of Schwann cells (Hadlock et al., 2000), extracellular matrix (ECM) (Rosen et al., 1990), (Maililiche et al., 2002), muscle tissue (Meek et al., 2004) and growth factors (Pfister et al., 2007).

Solvent casting is the process of dissolving a biodegradable polymer such as PLA in a solvent such as dichloromethane, and letting the solvent evaporate once the mixture is spread on to the substrate. When the solvent evaporates, it leaves behind a thin film of PLA. The surface morphology of this film is of high importance for cell growth. Researchers have used various methods to influence cell growth. Miller et al. (2001) showed that grooved films made from poly(D,L-lactic acid) (PDLA) polymer, on grooved silicone substrates helped Schwann cells align in the direction of the micro grooves. Nishikawa et al. (2002) presented how using honeycomb structured porous PCL films with different pore diameters affect cell growth.

Solvent cast films produced using substrates manufactured with electron beam lithography with cylindrical nano-pillars have been trialled with different diameters and patterns on stem cells.
(Gadegaard et al., 2006). The pits on the films ranged from fully ordered symmetrical patterns to offset patterns. The stem cells preferred the slightly offset pattern and 120 µm diameter pits (Dalby et al., 2007).

These examples highlight the need for artificial conduits and show the potential to manipulate the surface morphology of the films using the solvent casting method, which is directly linked to cell growth success. The solvent casting method is the current method of choice for manufacturing nerve conduits in this project and will be analysed in further detail, where the nerve cell growth success of various surface types will be investigated. The following section presents the electrospinning method.

2.2 Tendon Repair and Electrospinning

Tendons are a type of tissue connecting muscle to bone, and they absorb and transmit the forces exerted during the movement of body parts. Due to the rising elderly population and more people taking part in recreational sporting activities, tendon injuries are on the increase in the developed world. More athletes are suffering from tendon problems caused by an increase in training schedules which leads to the overuse of body parts (Kader et al., 2002).

Inflamed tendons cause pain and discomfort for the patient. These are generally treated with physiotherapy or in more severe cases with anti-inflammatory drugs, orthotic devices (Carter et al., 1992) and platelet rich plasma injections (de Vos et al., 2010), (Sampson et al., 2008). However, in the case of ruptured tendons surgical procedures are necessary. With the advancement of surgical techniques success rates have increased, but the same disadvantages exist with tendon grafting as with nerve grafting. The main issues are injury to the donor site, localised pain and rejection of the tissue by the body of the patient. Similar alternative biomaterial structures have previously been trialled with limited success (Millett et al., 2011), including platforms made from carbon fibre, PLA and small intestine sub-mucosa (Badylak et al., 1995).

The tissue engineering approach has the potential to eliminate the drawbacks of grafting and traditional surgical procedures, and tendon repair techniques are part of ongoing research. Recently, electrospun fibres have been considered as an alternative platform for tendon repair (Zong et al., 2003), (Bhattarai et al., 2004), (Xu et al., 2004). This method has the potential to create tendon like strong fibrous structures from biosynthetic materials, which can be implanted into the body.

Other methods with the potential to create tendon-like fibres also exist. Drawing from a solution melt is one of these methods. The drawing process involves melting the polymer using heat and dipping a thin rod into this melt pool and then taking the rod out, drawing out a molten polymer string that dries rapidly on contact with air. This method has created fibres as small as 60 nm (Xing et al., 2008), (Jayakumar and Nair, 2012). Another alternative method is templating. Templating uses custom
aluminium oxide moulds that are made to the required dimensions, where the polymer melt is injected or gravity fed into the mould. Once the polymer is dry, the mould is either mechanically detached or chemically dissolved to release the final product (Sachlos and Czernuszka, 2003). Phase separation is another commonly used method for making porous and fibrous materials. This method works by dissolving the polymer in molten organic acid, such as phenol, followed by lowering the temperature and quenching to encourage liquid phase separation. The solvent is then removed by sublimation and the product left behind is a porous structure that can be used for planting cells (Nam and Park 1999). Another method is molecular self assembly, which is a complex but a more direct approach to tissue engineering. Rather than building biodegradable support scaffolds, then modifying them with biologically active molecules to encourage cell growth, these scaffolds are entirely made of proteins or peptides. Although the nano-manufacturing methods may differ, an example can be given as the self assembly of ionic peptides. When released in a liquid solution they form two distinct surfaces, one hydrophobic and the other hydrophilic. The hydrophobic parts self assemble and help form ionic bonds with regular intervals and peptide fibres can be manufactured (Zhang, 2003). Three dimensional printing has also been proposed as a biomaterial scaffold manufacturing method (Landers and Mühlaupt, 2000). Where the solvent polymer mixture is forced out using compressed air to create fine fibres and moved through the x – y – z axis to build a three dimensional structure, which is pre designed using a computer aided design package. A comparative table showing the advantages and disadvantages of each method is given in Table 2.1.
<table>
<thead>
<tr>
<th>Method</th>
<th>Advantages</th>
<th>Disadvantages</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Templating</strong></td>
<td>• Relatively simple process.</td>
<td>• Not continuous, limited by template size.</td>
</tr>
<tr>
<td></td>
<td>• Fibre dimensions and scaffolds can be varied by creating different templates.</td>
<td>• Requires destruction of the template.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Manufacturing the template is an extra step, time and resource expense.</td>
</tr>
<tr>
<td><strong>Drawing</strong></td>
<td>• Relatively simple process.</td>
<td>• Not continuous, governed by rod dimensions and evaporation rate of the solvent.</td>
</tr>
<tr>
<td></td>
<td>• Low cost.</td>
<td>• Fibre dimensions not controllable.</td>
</tr>
<tr>
<td><strong>Phase Separation</strong></td>
<td>• Continuous simple process.</td>
<td>• No control of fibre diameters and lengths.</td>
</tr>
<tr>
<td></td>
<td>• Scaffold structure can be varied by changing the polymer properties and solvent polymer mixture levels.</td>
<td>• Only limited to specific polymers and some may not satisfy mechanical properties required of the scaffolds.</td>
</tr>
<tr>
<td><strong>Molecular self assembly</strong></td>
<td>• Very fine nano-scale fibres can be produced.</td>
<td>• Complex process which requires specialist knowledge.</td>
</tr>
<tr>
<td></td>
<td>• Ideal for cell growth.</td>
<td>• Length of fibres cannot be controlled and are generally short, several µm in length.</td>
</tr>
<tr>
<td><strong>Three dimensional printing</strong></td>
<td>• Similar to templating, without the added complexity of making a template.</td>
<td>• Expensive process.</td>
</tr>
<tr>
<td></td>
<td>• Custom scaffold shapes and sizes.</td>
<td>• Complex integration of hardware and software required for fine controlled actuation.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Potentially expensive to maintain, and time consuming to clear polymer blockages.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Minimum fibre thickness is restricted to 200 µm.</td>
</tr>
<tr>
<td><strong>Electrospinning</strong></td>
<td>• Simple and continuous process.</td>
<td>• Presence of toxic solvents during manufacture.</td>
</tr>
<tr>
<td></td>
<td>• Scalable for mass production.</td>
<td>• Jet instability and unknown effects of variables.</td>
</tr>
<tr>
<td></td>
<td>• Fibre diameter ranges can be from nm to µm</td>
<td></td>
</tr>
</tbody>
</table>

*Table 2.1; Comparison of Fibre Manufacturing Techniques*
Electrospinning is a method of creating fine fibres with small diameters (Ramakrishna, 2005) with the use of high voltage. The method was first patented by Anton Formhals (1934), although the effect of electrostatic forces on conductive liquids had been observed many years before by Rayleigh (1882). Taylor (1964) proved with theoretical calculations that with sufficient electric field strength a jet can be formed from a conductive liquid droplet, pulling the liquid and creating a cone shape which a jet shoots out of, and the cone shape is today known as the Taylor cone (Taylor and McEwan, 1965), (Taylor 1969). The work following Taylor’s, mainly investigated the properties of the jets emanating out from the Taylor cone (Gañán-Calvo, 1997), (Cherney, 1999). For the jet to form a Taylor cone, the electric force exerted on the liquid should be greater than the surface tension of the liquid. For the production of nanofibres however the stage after the Taylor cone jet is of interest. The liquid jet is accelerated towards the lowest potential point termed the collector, where the dry fibre jet is collected, and the stretching caused by the electric forces influence the fibre diameters. The resulting electrospun fibres are used in different applications such as filtering, tissue engineering and textiles manufacturing. The University of Manchester Materials Science Centre uses electrospinning for tendon repair research and holds a patent on electrospun tendon scaffolds (Downes and Bosworth, 2011).

The simplicity of the experimental set up and the ability to manufacture a range of fibres makes electrospinning ideal for tendon repair research and is the chosen method of the Biomaterials Research Group. The system is continuous and theoretically the parameters can be changed for manufacturing different size fibres. Large scaffolds with a number of different surfaces can be created, for example as an adapter tissue between bone and tendon. Another advantage of electrospinning is that it is ideal for mass production, with some commercial products already in the market such as the Nanospider from Nanorforce Technology, although it is not aimed at tissue engineering purposes.

Although convenient for fibre manufacturing the electrospinning process is still not fully understood. Researchers typically find the right conditions for manufacturing fibres of a certain diameter or alignment that suits their application, by trying to keep the variables they can influence constant, but still report inconsistencies and repeatability problems. In literature the electrospinning variables are grouped in to three categories. These are solution parameters, process parameters and environmental parameters, which are discussed in the following three sections.

### 2.2.1 Solution Parameters

The solution parameters are primarily governed by two factors, the solvent and the polymer that is mixed with it. The solvent is used to dissolve the biodegradable polymer so it is available in liquid form ready for electrospinning. Solvents that are more volatile will evaporate faster, which generally means the spinning distance between the needle and collector can be reduced, as the fibres will dry
quicker. Additionally for the same process conditions, changing the solvent can eliminate the *beading* phenomenon. Beads are droplets of polymer attached to the fibre strand with a regular period and are unwanted for tendon scaffold applications. Changing the polymer from acetone to hexafluoroisopropanol (HFIP) is shown to provide smooth fibres (Bosworth and Downes, 2011). The choice of polymer is the other main factor in influencing the solution properties. Generally the polymer is determined depending on the specific application of the electrospun product, for example the difference between fibres for biodegradable scaffolds and advanced filters that need to last a long time will be significant.

The biodegradable polymers used in electrospinning are the same as the ones used for solvent casting for nerve conduits, examples are, PLA, PCL and polyethylene oxide (PEO). For tendon repair scaffolds, PCL is the chosen polymer. Experiments have shown that PCL fibres provide the closest match for creating tendon like, flexible and strong scaffolds. In addition PCL has a slow degradation time in comparison with other polymers, which means the planted cells have support for a longer time on a high load bearing tissue, such as the tendon (Bosworth et al., 2008).

Once the solvent and polymer is mixed and the solution is created, electrospinning can take place. This solution will have four main properties that will influence the fibres generated through electrospinning. These are:

- Conductivity
- Surface tension
- Viscosity
- Elasticity
- Thermodynamic properties
- Mass transfer properties and coefficients
- Phase equilibrium coefficients

Numerous attempts have been made in the past to analyse the effect of solution properties on electrospinning. The polymer – solvent solution needs to be conductive, so fibres can be created through electrospinning. Conductivity of the solution depends on the polymer solvent mixture levels. Depending on which one the charge carrier is, increasing the amount of it in the mixture will affect the conductivity. Salts can also be added to the solution to increase the ion content, therefore increasing the number of charge carriers. However, if a large amount of salt is used, it would interfere with the intrinsic properties of the solution, such as surface tension (Choi et al., 2004). Conductivity also depends on viscosity which governs how mobile the ions are in the solution (Du and Zhang, 2008). Trying to alter the conductivity of the solution is more difficult during a manufacturing run, but by using temperature the viscosity of the solution can be changed, which in turn has an
effect on conductivity. Being able to carry more charge, allows more current to travel through the system, and is suggested to help with obtaining thinner fibres (Fong et al., 1999).

Viscosity of the solution affects all the other solution properties and fibre diameters (Gupta et al., 2005). Viscosity is the solution's ability to flow, which is dependent on the speed and the length of the polymer chains within the mixture. Viscosity is dependent on the polymer solvent mixture, the molecular weight of the polymer used, the impurities in the solution and the temperature of the solution (Fong et al., 1999), (Sperling, 2005). The viscosity of the solution could be tailored before electrospinning using the Mark – Houwink – Sakurada equation (Bosworth and Downes, 2011), given as:

\[
|\eta| = K \cdot M^a
\]

Where, \(|\eta|\) is intrinsic viscosity and \(M\) is molecular weight. \(K\) and \(a\) are the Mark – Houwink parameters for the polymer and solvent at a set temperature. These will change if the temperature varies, but an initial indication of the viscosity that will be achieved when the solvent and polymer are mixed can be calculated. A constant viscosity throughout the electrospinning run is desirable to create consistent fibres and for this the temperature of the solution must remain constant. In scenarios where a specific viscosity is needed, but is not achievable due to regulations of the Federal Drugs Agency (FDA) or availability, alternative additives such as salts can be used. However, mixing salts into the polymer mixture does change the fibre morphology and the chemical composition of the film, which can have an unwanted effect, such as biodegradability or mechanical properties (Du and Zhang, 2008).

Surface tension of a liquid is caused by the forces exerted on the liquid molecules by each other. This creates a resistive force on the surface of the liquid (Erbil, 2006). For electrospinning to take place, the electrostatic force needs to overcome the surface tension of the liquid droplet at the tip of the needle it is being dispensed from (Ramakrishna, 2005). If the surface tension of the solution is too high, instead of forming fibres, the liquid will shoot out in droplets and not form fibres. Surface tension is governed by the solution’s material properties. Viscosity also has an effect on surface tension, which in turn is affected by the solution temperature (Deitzel et al., 2001).

Fluid elasticity is another factor in electrospinning, and depends on the polymer solvent mixture levels. Rutledge et al. (2006) have shown how solutions with the same conductivity, surface tension and viscosity have different elasticity. The elasticity of the liquid determines whether a fibre is formed, or droplets are formed. Sometimes beading can be observed in the fibres if the liquid is not elastic enough for fibres but not so stiff to form droplets. Beads can also be created by the effects of surface tension of the liquid (Reneker et al., 1999).
The explained solution properties can also be affected by environmental conditions, which as a result have an effect on the fibre diameters. The environmental parameters are explained in the next section.

2.2.2 Environmental Parameters

Electrospinning can be significantly affected by the environmental conditions. As mentioned previously, one of the current problems in the previous electrospinning configuration was the variation in daily ambient conditions, which resulted in a change in fibre morphologies. Controlling these parameters would be the first step towards introducing consistency to electrospinning at the BRG. The environmental parameters that affect electrospinning are:

- Temperature
- Humidity
- Ambient air composition and air movement

Temperature has three main effects on the electrospinning process. An increase in temperature increases the evaporation rate of the solvent, decreases the viscosity and increases the moisture holding capacity of the ambient air, which directly affects humidity. With increasing temperature, the solvent evaporates quicker as the molecules of the solvent have more energy due to the heat, and they escape from the liquid form to gas form faster.

The viscosity of the solution decreases with the increase in temperature, as the polymers in the solution soften. This means that with increased heat, the polymer chains relax more, increasing the fluidity of the solution, therefore lowering the viscosity, which means lowering the resistance of the liquid against external forces. A faster flowing solution will result in greater elongation of the fibres with the pull from the electrostatic force before they dry, thus creating thinner fibres (De Vrieze et al., 2009). Evaporation and viscosity could be analysed as opposing effects (Bosworth and Downes, 2011), where increasing the temperature will evaporate the solvent quicker, decreasing the time for elongation. In return the decreased viscosity will make the fibres elongate at a quicker rate.

The humidity of the air is the measure of how much moisture is present in it. The calculations for exact humidity are complex, but the chart in Figure 2.3 (Shallcross, 1997) can be used if we assume the evaporating solvent will have a negligible effect on the volume of the air that is present in the electrospinning environment. The chart gives the maximum amount of moisture 1 kg of air can carry at different temperatures. Humidity is expressed in terms of Relative Humidity (RH) in percentages. 10°C at 50% RH means that volume of air at 10°C is carrying 50% of its maximum moisture capacity.
Humidity of the air will affect various solutions differently. Solutions that use aqueous solvents which are water based, will dry slower when the humidity is high, thus creating thicker fibres (Bosworth and Downes, 2011). If the air is at 100% RH it will mean that there is no more capacity for more moisture, and this will slow down the evaporation process. This will result in fibres that have not solidified properly, producing thicker or even congealed fibres. Some non-water based solutions can also absorb water during electrospinning, resulting in pores on the fibres (Jeun et al., 2007). De Vrieze et al. (2009) have shown the effects of different ambient conditions on two types of polymer, where Polyvinylpyrrolidone (PVP) and Cellulose Acetate (CA) were dissolved in ethanol and electrospun in different environmental conditions. Due to the different chemical compositions of the polymers, they reacted differently to the conditions. The study concludes that fibre diameters can be influenced using humidity as it affects solvent evaporation.

The ambient air composition is another factor that can affect fibre properties. If electrospinning takes place in closed conditions, for example in a cabinet over a long period of time, the evaporated solvent can build up in the atmosphere, thus changing the composition of the air. This will in return affect the evaporation rate of the solvent influencing fibre diameters. For high volume electrospinning over a long period of time, it is desirable to be able to measure the solvent in the air and control the atmosphere containing solvent with new air (Bosworth and Downes, 2011). However, care must be taken when designing such a system, so that the direction of the air around the electrospinning area does not affect the jet’s flight path and disturbs where the fibres land. In addition electrospinning studies under vacuum have been published (Reneker and Chun, 1996), where the vacuum provides a stronger electric field, due to the absence of air, which is a dielectric. The vacuum also reduces the
boiling point of the solvent, meaning it will evaporate quicker and the fibres will dry faster. Additionally there is no air resistance for the fibres to contend with, meaning they will fly quicker as well as drying faster.

There are a significant number of publications where the tests were carried out under ambient conditions or uncontrolled humidity levels, some of which are; You et al. (2006), Jeun et al. (2007), Shields et al. (2004), Hong (2007), Zhang et al. (2012). You et al. (2006) have analysed the effect of conductivity on the final fibre diameters. This study uses salts to increase the ion count in the solution therefore adding more charge carriers and changing the conductivity, which has shown to affect the fibre diameters. In this study there is no mention of temperature or humidity control of the electrospinning environment. The environmental conditions can vary significantly day to day, therefore affecting the fibre morphology, which is not ideal. Temperature is a major variable in the evaporation rate of the solvent and additionally it influences the viscosity of the solvent as presented in the previous section. These combined effects have not been considered in many research studies and require a more thorough assessment. Some selected studies are presented in more detail in the main electrospinning chapter, Chapter 4, where their shortcomings have been highlighted. The following section describes the process parameters of the electrospinning method.

2.2.3 Process Parameters

The parameters described under this section are the main components of electrospinning. They can be listed as follows:

- Electrospinning distance is the distance between the needle tip, where the solution is contained, and the collector where the fibres are accumulated. This also affects the electric field strength.
- The electric field strength is quantified by kV per cm over the spinning distance. When using a needle only configuration the field is directly proportional to the voltage. The shape of the electric field also affects the fibres and the way they land on to the collector.
- The shape of the spinneret, which in most cases is a needle, and the collector influence the electric field, which can influence the fibre diameters and orientation. To increase the volume of the material produced, spinnerets with multiple needles are sometimes used.
- Hydrostatic pressure of the solution governs the rate the solution arrives at the needle tip, and is usually controlled by a syringe pump, or in some cases is gravity dependent. There needs to be enough flow of solution for electrospinning to be continuous.

The following sub sections describe each of the points listed above in more detail, starting with the electric field.
2.2.3.1 Electric Field Strength

The electric field strength is controlled by the voltage. The electric field force acting on the solution must be able to overcome the surface tension of the solution for electrospinning to take place. When the force of the field and the surface energy of the solution are at an equilibrium, a Taylor cone is formed (Ramakrishna, 2005). This is observed as an elongated droplet, which changes shape depending on the strength of the field, and this is where the jet shoots out from towards the collector. An image of a Taylor cone and the jet is shown in Figure 2.4.

![Taylor Cone](Image)

The droplet at the tip of the needle is the Taylor cone, and to the right of this, a fine jet can be seen. The increase in the electric field strength further accelerates the jet, thus more stretching of the polymer chains takes place, creating thinner fibres (Bosworth and Downes, 2011). The higher electric field also removes the set volume of solution from the tip of the needle faster, requiring a higher flow rate. If using a gravity-fed system, the electrospinning process will be limited to a voltage range, which can be overcome by using a syringe pump and a variable power supply. Zong et al. (2003) report the effects of voltage on fibre diameters, but they use a gravity feed set up, which means the solution flow rate was not constant during the process.
The reason for this can be explained with the hydrostatic equation, which is:

\[ P = d \cdot g \cdot h \], where \( P \) is the pressure of the liquid in Pascals (Pa), \( d \) is the density of the solution (kg/m³), \( h \) is the height of the liquid (m) and \( g \) is the gravitational acceleration (m/s²). As the solution is used up the height of the solution will decrease and the hydrostatic pressure will decrease, affecting the flow rate. This is not ideal, and a syringe pump should be used to keep the pressure constant.

Different polymer-solvent mixtures will have different threshold voltages where the electrospinning process can start. This is due to their surface tension properties, which can be influenced with temperature. As pointed out in the previous section, it is important that the polymer-solvent mixture is analysed in terms of evaporation rate, surface tension, viscosity and conductivity before electrospinning the mixture. Additionally, if the field strength is very high, the Taylor cone starts receding into the needle and the jet splits into several subsidiary ones as shown in Figure 2.5. This is caused by the electric field and results in the jets repulsing each other on their way to the grounded collector, which creates unwanted disturbances in the collection process. Depending on the desired fibre diameters the repulsion is also a crucial part of the electrospinning process, as this further stretches the fibres and elongates their flight path. The extra stretching makes the fibres thinner, following the initial stretch after the Taylor cone (Wang et al., 2011), (Reneker et al., 2000). The following section discusses another main aspect; the electrospinning distance.

Figure 2.5; Receding Taylor Cone and Multiple Jets
2.2.3.2 Electrospinning Distance

The spinning distance is an important factor of electrospinning, which can be defined as the distance between the spinneret tip and the collector. When the jet leaves the Taylor cone, it travels in liquid form through the air. During its flight the jet elongates and gets thinner, which is due to the acceleration of the electric field and the charge repulsions within the solution itself (Bosworth and Downes, 2011). Although the jet starts its flight in liquid form, the solvent starts to evaporate as soon as it contacts the air. Throughout its flight, the jet elongates and starts drying. If the spinning distance is not sufficient, the jet will land on the collector in a semi-dry form. Eventually the fibres landing subsequently will congeal and form more of a porous web without individual fibres.

Distance contributes to electric field strength as much as the voltage. There is an inverse square relationship between the field strength and the distance; as the distance increases, the field strength weakens (Doshi and Reneker, 1995). Depending on the application the electrospinning distance can be fine tuned for a range of settings. The best method for this would be to simulate the electric field and compare this to the desired fibres, whilst taking the drying process in to account. However, in practice the spinning distance is not changed often as the users will establish their working ranges and alter the other variables instead, as this is more convenient. Alongside the electric field strength and the electrospinning distance, the flow rate of the solution is the next main component of the process.

2.2.3.3 Flow Rate

The flow rate can be described as the rate at which the solution flows through the needle. The effect of hydrostatic pressure was briefly mentioned in Section 2.2.3.1, where if the flow rate is not constant, variations may occur in the final fibre morphologies. The internal diameter of the needle tip and the solution viscosity both affect the flow rate. When the flow rate is insufficient, the solution can dry up in the needle and cause blockages, and in extreme cases can lead to the loss of the Taylor cone. If the flow rate is too high, and the electric field is not strong enough, the solution can start dripping. If the electrospinning is a top down configuration, the dripping solution can spoil the fibres that are already on the collector. Zong et al. (2002) suggest that flow rate affects the fibre morphologies when tested with the same voltage and two different flow rates. However this requires confirmation, as the force of the electric field influences how much solution is taken from the Taylor cone. As long as flow rate is sufficient to continually supply enough fluid, the fibre dimensions should not be affected significantly.

When commercialisation or up scaling of the electrospinning process is considered, the disturbances in the flow rate will be much more significant in comparison to the laboratory level set ups. This highlights an area where the process would benefit from a control system. With up scaling in consideration, Druesedow et al. (2010) compared infrared and ultrasonic pressure measurement
techniques, where the pressure fluctuations in the flow line were detected without any physical interference with the flow. They reported that flow pressure does not influence the fibre morphologies significantly, as opposed to Zong et al. (2002). This could be due to the different solutions used in the tests and highlights an area that would require further testing. Deitzel et al. (2001) report an increase in the electrospinning current as the flow increases, which is expected as there will be more charge carriers present, but there is no mention of an effect of this on fibre morphologies. This will be investigated in this project, as the current publications contradict each other. The following section explains the influences of the collector and the spinneret on the electrospinning process.

2.2.3.4 The Collector and the Spinneret

The collector and the spinneret are the two main parts of the electrospinning system. The spinneret is usually the needle, which is the last place the solution sits before being drawn out by the electric force. In most research studies a blunt medical needle is used, which is available in various internal diameters. Macossay et al. (2007) present an extensive study on the effect of needle diameters, where no significant changes were observed on mean fibre diameters. Yarin et al. (2001) present the effects of the shape of the droplet at the needle tip and how this influences the initial voltage to start electrospinning. As the needle diameter influences the shape of the Taylor cone, it can be argued that the spinneret’s internal diameter influences the electric field shape and therefore the final fibre diameters (Bosworth and Downes, 2011). Spinneret diameter is a parameter that cannot be changed during the process itself, but can be used to fine tune a process. Commercially, various types of spinnerets are made, and some examples include multi needle configurations and heated tips where the aim is to spin molten polymers using high temperatures. Termed **melt spinning**, this method has the additional advantage of not using any harmful solvents such as dichloromethane, but is out of the scope of this research study due to time and resource restrictions. Additionally co-axial spinneret configurations are also used to produce three dimensional scaffolds or hollow shapes, such as nerve conduits (Bosworth and Downes, 2011).

The complementing part of the spinneret is the collector, where the fibres coming from the spinneret are collected. In most configurations, the collector is the low potential part of the electrospinning system, which the electric field vector lines flow towards. The collector can be used to influence the orientation of the fibres. Depending on application, commonly used collectors are static plate or rotating drum collectors. The static plate is used to collect randomly oriented unwoven fibres, which are employed in tissue engineering research. More advanced plate collectors utilise robotic moving stages (Kim et al., 2007) where improved mechanical material properties can be observed. For tendon research and other applications where the fibres are expected to experience forces mainly in one direction, alignment is important. These aligned mats can be processed once again using textile engineering techniques, such as yarning or plating, to further improve their mechanical properties (Bosworth et al., 2008).
For the collection of aligned fibres, various configurations of rotating collectors are used. Katta et al. (2004) used a rotating wire drum to collect aligned fibres, where in a similar study Pan et al. (2006) rely on a fast rotating cylindrical collector. Teo and Ramakrishna (2005) use two fixed points separated by a fixed distance, where the fibres zigzag across the two points to create aligned fibres. From a tissue engineering perspective, aligned fibres have been proven to influence the orientation of the cells that are grown on the fibre mats, which is an important property for tissue repair (Yang et al., 2005). As the field of electrospinning is varied, the main research interests that are relevant to this thesis are briefly discussed in the following section to define the scope of the project.

2.2.4 Research Interests

The main tissue engineering research interests in this study are the ability to create mechanically strong and flexible fibre mats, where cells can successfully populate whilst promoting healing and regeneration in the wounded area. For tissue engineering applications, such as tendon repair, straight fibres are desirable. These properties are aimed to be achieved through a thorough investigation of the process parameters and investigating the methods to model these systems, which will be vital when applying optimisation and control.

If the electric field is not manipulated either by the collector or using extra magnets or electrodes, the whipping during the flight of the jet results in randomly collected fibres on the collector plate. Various methods have been trialled to align the electrospun nanofibres, some of which were discussed in the previous section. The main method to align fibres is manipulating the electric field by changing the collector type or rotating it, which ultimately affects the field. Alternative alignment attempts were made using a sharpened edge spinning disc collector plate, which concentrates the negative charges on the edge of the collector plate and concentrates the electric field vector lines to that point (Theron et al., 2001). Other manipulation methods include having switching electrodes along the flight path (Bellan and Craighead, 2011), using conductive rings to keep the jet focused, and using a three dimensional axis moving stage (Mitchell and Sanders, 2006). All the above mentioned manipulation methods concluded that influencing the electric field plays the most crucial role in fibre collection, alignment and diameters. Less than a handful of the presented existing research utilises feedback in the process. Saeidpourazar and Jalili (2008) present their work on a nano-manipulator where one of the potential applications is electrospinning. The system uses visual feedback to control the position of the jet. However, there is no mention of fibre diameters or the effect of other variables. Mitchell and Sanders (2006) present their work on controlling the pressure in the solution line to minimise the variance of the fibre diameters. The ability to continuously adapt with the changing variables, such as the humidity and the temperature of the environment, or pressure fluctuations within the system, or even disturbances to the electric field has not been addressed in any previous studies as a whole system. For this, a model based approach would be necessary and the groundwork towards achieving this is presented in the further sections of this thesis.
2.2.5 Summary

The studies discussed in this section summarise some of the main examples of research conducted on the parameters of electrospinning. When analysed as a whole, they demonstrate the existence of correlation between the variables but also a clear need to gain better understanding of the interactions within the process, which can be gained through conducting designed experiments, under controlled conditions using real-time data logging. Many of the previous studies contain at least one area that has not been considered, such as changes in the flow rate or the ambient conditions. The full effects of the variables as a whole are still not fully understood by the research community and there is no published work on controlling the electrospinning parameters using feedback, in real time, in order to change the properties of the manufactured scaffold. Mitchell and Sanders (2006) use a custom built system, where the flow pressure is in fact controlled using feedback. However, the other parameters such as the temperature, or the syringe pump flow rate are controlled independently, whilst the possible correlations between these variables have been ignored. In the following section a general overview of process modelling and control is provided, and the further chapters analyse a simulation which is similar in structure to the electrospinning process, and has highly correlated variables. The next section provides a brief introduction into control systems theory and some of the common methods that are used when controlling physical systems.

2.3 Control, Multivariable Systems and Modelling

This section initially presents what a system is in the control systems context, and how it could be controlled. The methods of control and modelling relevant to this work are presented in their natural evolutionary manner, with the aim of justifying the need for model based control strategies.

2.3.1 Systems and Control

Control systems in industry play a crucial role in ensuring safety, improving consistency and maximising throughput. A system, from a control systems theory perspective, can be described as an entity where variables create observable outputs. These variables can be independent or may interact with each other within the system. A representative block diagram of a dynamic system is shown in Figure 2.6, where \( t \) is time.

\[ \text{Input, } u(t) \quad \text{Output, } y(t) \]

*Figure 2.6; System Block Diagram*
All systems are by nature continuous and they can be described by their differential equations. The difference equation for the system shown in Figure 2.6 is assumed to be:

\[ y_n(t) + a_2 y_{n-1}(t) + \ldots + a_n y(t) = b_2 u_m(t) + b_3 u_{m-1}(t) + \ldots + b_n u(t) \]

where \( a \) and \( b \) are system parameters. There are two types of general models that are described commonly in the literature used in control systems (Söderström and Stoica, 1989), which are:

- **Graphical models**: These are suitable ways to represent linear systems using time domain testing methods such as impulse responses and step responses. Frequency domain methods can also be used, such as Bode plots.

- **Mathematical models**: These models are used to describe more complex systems such as nonlinear, linear - time variant and multivariable. The systems are modelled using time series or differential equations.

An example of a system can be given as an air conditioning unit in a room. The air conditioner will use a refrigerator coil and a fan to blow cold air into the room. The user would be interested in how the coolant temperature of the refrigerator and the fan speed will affect the room temperature. If the user wanted to control this system, the aim would be to make the output behave in a way the user desires. This is achieved by correctly selecting the input value sequence (Söderström and Stoica 1989).

The transfer function model, which is a mathematical representation of a system’s input – output relationship, can be obtained by taking the Laplace transform of the differential equation. Here \( U(s) \) and \( Y(s) \) are respectively the Laplace transforms of the input and the output. \( G(s) \) is the continuous time transfer function of the system, and \( a \) and \( b \) are system parameters.

\[
Y(s) \cdot \left( s^n + a_2 s^{n-1} + \ldots + a_n \right) = U(s) \cdot \left( b_0 s^m + b_1 s^{m-1} + \ldots + b_m \right)
\]

\[
G(s) = \frac{Y(s)}{U(s)} = \frac{b_0 s^m + b_1 s^{m-1} + \ldots + b_m}{s^n + a_2 s^{n-1} + \ldots + a_n}
\]

Today, control is applied using computers such as microprocessors, desktop computers and Field Programmable Logic Arrays (FPGA). For a computer to interpret a continuous analogue signal, for example the temperature value of the refrigerant, it will need to be digitally sampled, which is also termed as **discretising** the signal. Discretising is achieved by using an analogue to digital converter, where the measured signal is kept constant between the sample instants using a zero-order-hold. The sampled system representation can be shown using a causal, time invariant and linear system’s
The impulse response. The impulse is a short one sample signal inputted to the system, and the observed response is the system's impulse response. The impulse response \( g(\tau) \) at time \( \tau \) provides a linear map between the input \( u(t) \) and the output \( y(t) \) as a function of time. The system with the impulse response can be represented as the convolution integral:

\[
y(t) = \int_{\tau=0}^{\infty} g(\tau) \cdot u(t-\tau) d\tau
\]

2-4

If the sample instants are denoted as \( t_s = kT, \) where \( k = (1, 2, 3... ) \) and \( T \) is the sampling interval of one time unit, the sampled system can be described as:

\[
y(kT) = \int_{\tau=0}^{\infty} g(\tau) \cdot u(kT-\tau) d\tau
\]

2-5

This can also be represented in discrete time as the convolution sum, where the system is represented by the following equation:

\[
y(t) = \sum_{k=1}^{\infty} g(k) \cdot u(t-k), \text{ where } t = (0, 1, 2, 3...)
\]

To make the representation of this simpler, a backward shift operator derived from a forward shift operator can be utilised, where \( qu(t) = u(t+1) \) is the forward shift operator and \( q^{-1}u(t) = u(t-1) \) is the backward shift operator.

\[
y(t) = \sum_{k=1}^{\infty} g(k) \cdot u(t-k) = \sum_{k=1}^{\infty} g(k) \cdot (q^{-k}u(t))
\]

2-6

Since \( q^{-k}u(t) = u(t-k) \), \( y(t) \) can also be written as:

\[
y(t) = \left[ \sum_{k=1}^{\infty} g(k) \cdot q^{-k} \right] \cdot u(t), \text{ where } G(q) = \left[ \sum_{k=1}^{\infty} g(k) \cdot q^{-k} \right] \text{ and } y(t) = G(q) \cdot u(t)
\]

Here \( G(q) \) is the discrete time transfer operator of the linear system.

\[
G(q) = \frac{b_1q^{-1} + \ldots + b_nq^{-n}}{1 + a_1q^{-1} + \ldots + a_nq^{-n}}
\]

2-7

In 2-7, \( G(q) \) is termed the transfer operator as it is not strictly the discrete transfer function, which can be obtained using the \( z \)-transform. This would be \( G(z) = \left[ \sum_{k=1}^{\infty} g(k) \cdot z^{-k} \right] \). However, replacing \( z \) with \( q \) for discrete transfer function representations is not incorrect and it is often presented as \( G(q) \) in literature (Söderström and Stoica, 1989) and it will be used throughout this thesis. This
transfer function model of a system shows how the input (cause) variables relate to the output (effect) variables. The time invariant and linear system shown here is represented by its impulse response.

Systems are generally not as straightforward as shown in Figure 2.6 and by the corresponding transfer function presented in expression 2-3, as they can contain inputs that are not controlled by the user, called *disturbances* (Ljung, 1987). For linear systems the disturbances can be assumed to be an additive term to the output, represented in the diagram in Figure 2.7.

The disturbance $d(t)$ can be described as:

$$d(t) = \sum_{k=1}^{\infty} h(k) \cdot e(t-k),$$

where $e(t)$ is a random variable sequence with zero mean, and $h(k)$ is the impulse response representation of the disturbance. The disturbance can be described as:

$$d(t) = H(q) \cdot e(t) \quad \text{(2-8)}$$

Then the output of the system represented in Figure 2.7 can be written as:

$$y(t) = G(q) \cdot u(t) + H(q) \cdot e(t) \quad \text{(2-9)}$$

Systems that control is applied to are mostly physical, such as controlling the temperature of reactions in chemical plants, delivering the optimum amount of fuel to the engine of a car and controlling electric motors in robotic manipulators. Control is often applied by computers, and the application of computer control can be represented as a block diagram as shown in Figure 2.8. The input from the controller goes into the plant and creates an output. This output is sensed using a physical sensor, such as a tachometer or a thermocouple and is fed back into the controller. A closed loop system can be represented as shown in Figure 2.8, where in the presence of feedback the control loop takes corrective actions through the controller. These corrections are calculated by using many of the alternative methods that are available today, such as PID or model based approaches.
The controller can consist of different types of algorithms that have been developed over time. The simplest type of control is on-off control. This type of controller is utilised because of its simplicity and in some cases when the system is limited by its actuator, meaning it can either be only on or off, much like a solenoid valve that consists of an electromagnet and a spring. The electromagnet can be excited to open the valve and when power is removed the spring will return the valve to its closed position. If the air conditioned room is considered, the controller will run the air conditioner when the temperature goes above a setpoint and turn it off when it has reached the required temperature. This will result in an oscillatory behaviour of the temperature within the room rather than a steady temperature. It can be argued that on-off control is the optimum case in the room temperature example; assuming the room is large and there are no disturbances, the rate of change in temperature will be slow and the room temperature will remain close to the desired setpoint. The air conditioner will only operate when required, thus minimising power consumption. However, there are applications where on-off control is not suitable, such as position control, where accurate positioning of an automatic milling machine head is required. In addition, on-off controller is not adjustable to cater for time constants of the system, where large delays need to be taken into account.

An alternative control algorithm is the Proportional Integral Derivative (PID) controller. The PID controller is the most common controller used in industry today. More than 90% of industrial control utilises PID (Åström and Hägglund, 2001), where they are used for controlling the speed of electric motors, levels of liquid tanks, temperatures and many other industrial systems. PID controllers regulate the process based on the present, past and future values of the process error. An overview of the PID control loop is shown in Figure 2.9.
The three parts of the PID controller are proportional gain $K_p$, integral gain $K_i$, and derivative gain $K_d$. They are represented in the following equations.

$$P = K_p \cdot e(t)$$ \hspace{1cm} \text{where } e(t) = r(t) - y(t)$$

$P$ is the output of the proportional part, and $e(t)$ is the error between setpoint and the plant output. The output is directly proportional to the size of the gain and the size of the error.

$$I = K_i \cdot \int_0^T e(\tau)d\tau$$ \hspace{1cm} \text{where } I \text{ is the output of the integrator part. Here the output depends on the sum of past errors and the integral gain } K_i. \text{ The effect of this is that the plant will move towards the setpoint until the error is zero. Depending on the gain } K_i, \text{ the integrator will accelerate the response of the controller and may overshoot the setpoint. If the integrator overshoots excessively it can continue to accumulate the error and keep increasing its output to compensate. This is termed integrator wind-up and can be prevented with the use of filters. Additionally the integrator can wind-up in a similar manner if an actuator saturates as it can physically not move anymore and the error starts accumulating, this is termed actuator windup.}

$$D = K_d \cdot \frac{de}{dt}$$ \hspace{1cm} \text{where } D \text{ is the output of the derivative part. The derivative term calculates the rate of change of the process. The derivative gain will control the rate of response of the control system, where it can be used to reduce the overshoot caused by the integral gain. The derivative gain is sensitive to noise and can make the output unstable. This can be overcome by frequency domain design techniques, where the bandwidth of the differential part is designed to the required specification. The controller output is the sum of the three terms. A complete classical representation of the PID controller is:}

$$u(t) = K_p \cdot e(t) + K_i \cdot \int_0^T e(\tau)d\tau + K_d \cdot \frac{de}{dt}$$ \hspace{1cm} \text{2-10}$$

Often the controller is represented in terms of integral time $T_i$ and derivative time $T_d$. In many industrial controllers the integral part is represented as $\frac{1}{T_i}$, which is referred to as reset time, meaning repeats per unit time.

$$u(t) = K_p \cdot \left( e(t) + \frac{1}{T_i} \cdot \int_0^T e(\tau)d\tau + T_d \cdot \frac{de}{dt} \right)$$ \hspace{1cm} \text{2-11}$$
In the frequency domain the transfer function of the controller can be written as:

\[ G_c(s) = K_p \left(1 + \frac{1}{T_i} \cdot s + T_d \cdot s\right) \]

where \( U(s) = G_c(s) \cdot E(s) \).

Due to the mentioned problems with the integral and derivative parts, real controllers in industry utilise slightly modified versions of the presented classical algorithm, where filters are used.

For the controller to be effective in the control loop, the gain values must be tuned. This means selecting appropriate values for \( K_p \), \( K_i \) and \( K_d \) so the process behaves in the desired manner.

The controller does not strictly have to use all the settings of a PID. P only, PI, PD or ID control are all valid versions of the PID method. There are various methods that have been developed for the tuning of PID controllers. One of the most well known methods is the Ziegler-Nichols tuning rules (Ziegler and Nichols, 1942), which is a simple way of tuning a controller in closed loop when the controller is switched on. This is achieved by turning off the derivative and the integral parts, and increasing the proportional gain until a sustained oscillation is obtained. The value of the gain when the oscillations are obtained at, is called the ultimate gain, \( K_u \) and the period of the oscillation is called the critical period, \( T_u \). These values are then multiplied by the predetermined values of the Ziegler – Nichols method to form P, PI or PID structures.

However, the Ziegler – Nichols tuning methods do not always produce good controllers (Yuwana and Seborg, 1982). Firstly this method assumes that all the plants can be represented as first order with a time delay model, as shown in 2-12:

\[ G_p(s) = \frac{K}{\alpha + s} e^{-\alpha t} \]

This representation may not be true for all systems and the resulting controller may not be suitable due to this mismatch. Additionally, the main design centres around a quarter amplitude damping criteria to reduce oscillations against load disturbances, which gives a damping ratio of 0.2 (Cominos and Munro, 2002). However, this criterion does not suit all systems as they may behave in an oscillatory manner depending on the frequency of the input. The maximum sensitivity of the control systems implemented using Ziegler – Nichols is reported to be high. This means the resulting controller makes the closed loop system too sensitive to variations in the system parameters. Depending on the system, the Ziegler – Nichols method could be ideal or provide a starting point for the tuning parameters (Åström and Hägglund, 1988). The Ziegler - Nichols method was developed in 1942 and since then more advanced methods have emerged, which take into account the differences between various systems (Panda et al., 2004), (Yuwana and Seborg, 1982), (Astrom and Persson, 1993).
One of the alternatives to Ziegler – Nichols is gain and phase margin tuning (Ho et al., 1998), which is a frequency response based approach. This method ensures that the controller is not affected by the changes in the input frequency and process gain, thus making the process more robust. Robust means that the controller can work within a range of variations in system parameters and prevent instability. The phase margin is the measure of phase shift required for the system to reach unity gain, meaning instability in closed loop. The gain margin is the measure of change in the open loop gain required for a 180° degree phase shift, again meaning instability in closed loop. The gain and phase margin values that define instability can be obtained from Bode or Nyquist plots. The data for these plots is collected through testing the system in open loop with a sinusoidal input signal that spans a range of frequencies. By changing the values of the derivative and the integral terms of the controller, the frequency response of the open loop system can be influenced. In addition, design using frequency analysis has the advantage of predicting whether the controller will be stable or unstable when it is switched to closed loop. Some examples of other methods that are used for tuning PID controllers are; Coheen and Coon (1953), pole placement, D partitioning, and Yuawana-Seborg (1982) tuning.

Many of the processes used in industry employ regulatory controllers such as the PID. However, one PID controller can only control one plant at a time. In some applications, more than one PID controller will be used to regulate one process, such as the supply of a liquid at a required temperature, flow rate and pressure. Or when the application has multiple outputs that affect each other, such as the top and bottom products of a distillation column, control loops may interact with each other and deteriorate the performance of the plant. Methods such as loop decoupling and feed-forward control may be utilised to improve performance. However, having perfect regulation does not necessarily mean the best economical returns (Lee and Weekman, 1976). For this, optimal control is needed, where the process can be regulated for different operating conditions. Although it must be pointed out that before implementing optimum control strategies, good quality regulatory control is necessary. Examples of optimising a plant for specific outcomes may include, minimum actuator wear or maximum throughput in the shortest time.

One of the main challenges PID controllers face is controlling processes with higher order dynamics and large time delays. To overcome these difficulties further modifications have been suggested to the standard feedback PID regulator structure, where almost all the proposed methods rely on a model based predictive approach. The model of the processes are determined either using mathematical means or the graphical methods mentioned previously.

One of the first of these modifications was proposed by Smith (1957) known as the Smith predictor. The main aim of this method was to overcome the time delay problem when the process is in closed loop. Systems with large time delays can make the control loop unstable, as the existence of the time
delay becomes an exponential in the frequency domain, meaning instability. The Smith predictor uses the model of the plant and the time delay to tackle this issue. The overview of the Smith predictor is given in Figure 2.10.

The Smith predictor calculates the predicted output at that point in time as well as eliminating the delay on the output, before it is fed back to the controller. This enables the application of delay free PID design methods. Around the same era of process control research, other methods have been proposed similar to that of the Smith predictor to deal with systems with large time delays and an integrator, such as the optimal regulator by Fuller (1968) and Kleinman (1969). Watanabe (1981) proposed a modified Smith predictor with a PID controller. In this work it was shown how the regular Smith predictor cannot achieve a steady state error of zero if the time delay is integrated and how some of the alternative methods do not have good transient response. An example transfer function of a system with integrator and time delay is given below:

\[
G_p(s) = \frac{K}{s \cdot (\alpha + s)} \cdot e^{-\alpha s}
\]

Astrom et al. (1994) proposed a further modification to the Smith predictor, where the disturbance is estimated. The new structure gave a better response when compared with the regular Smith and Watanabe methods.

Another alternative approach to the Smith predictor based techniques is the Internal Model Control (IMC) approach. IMC falls under the general scope of model based predictive control, which is presented in more detail in the further sections. An overview of IMC is given in Figure 2.11. Similar to the Smith predictor, IMC uses a model to cancel the plant dynamics and the disturbances by subtracting the predicted output at that point in time from the actual output.
IMC can, in theory, provide perfect control if the plant model matches the real system exactly, as the controller is the inverse of the model, and if the model is identical to the plant, then control can be achieved in open loop. Open loop is when there is no measurement fed back in to the controller, and in theory if the model is perfect, feedback is not necessary for IMC. IMC is implemented by setting the controller to be the inverse of the process model. However, for real systems due to disturbances and noise, having a perfect model is often not possible. Also, representing a high order process with a low order approximation can cause differences in the real plant and the model. This is termed \emph{plant-model mismatch}.

In the formulation of IMC, assuming the model is \( \hat{G}_p(s) = \frac{K}{(\alpha + s)} \cdot e^{-\tau s} \), the model is factorised into two parts. One that is invertible \( \hat{G}_p^+(s) = \frac{K}{(\alpha + s)} \) and one that is non-invertible \( \hat{G}_p^-(s) = e^{-\tau s} \).

The controller is set as the inverse of \( \hat{G}_p(s)^+ \) and in series with a low pass filter where:

\[
G_f(s) = \frac{1}{(1 + \tau_f s)^n}.
\]

Giving \( G_c(s) = G_f(s) \cdot \left( \hat{G}_p^+(s) \right)^{-1} = \frac{(\alpha + s)}{(1 + \tau_f s)^n} \cdot K \)

\[
Y(s) = \frac{G_c(s) \cdot G_p(s) \cdot R(s) + \left( 1 - G_c(s) \cdot \hat{G}_p(s) \right) \cdot D(s)}{G_p(s) - \hat{G}_p(s)} \cdot G_c(s) + 1
\]

If \( G_p(s) = \hat{G}_p(s) = \hat{G}_p(s)^+ \cdot \hat{G}_p(s)^- \)

Then expression 2-14 becomes:
This configuration eliminates the effect of the delay at that current time and decides on the next control move. The filter in the IMC can improve disturbance rejection by tuning for desired bandwidths and help customise controller behaviour for specific problems. The time constant of the filter can also be adjusted to change the response speed of the controller. The addition of the filter was suggested by Brosilow (1979) to compensate for the plant model mismatch. Also in the presence of the perfect controller it is possible to have zero steady state error. The importance of the model must be pointed out again, as if the model is not a close representation of the plant, good control will not be possible.

Implementing the IMC can either be done using a bespoke controller or a PID controller. Chien and Fruehuauf (1990) developed a PI representation of IMC, using the closed loop time constant as the tuning parameter. This work was followed by Luyben and Tyreus (1992) improving the proposed PI controller tuning and eliminating the poor control under transient inputs. The PID formulation of IMC can be found in Appendix 1.

IMC was more of an academic research topic rather than an applied method in industry. However successful pilot plant applications in academia have been reported. Similar in context, a model based predictive control methodology was developed in industry around 1978 – 1979 (Richalet et al. 1978), (Cutler and Ramaker, 1979), based on process models and aimed at handling constraints, which had similarities to IMC in terms of using a model of the process. This is presented next, in the following section.

### 2.3.2 Model Predictive Control

The methods explained up to now presented the system to be controlled as a time invariant and linear Single Input Single Output (SISO) structure. In industry, systems range from being linear SISO to nonlinear, Multi Input Multi Output (MIMO) and time variant. When systems such as electrospinning, become difficult for standard feedback controllers to regulate due to fast system dynamics, nonlinearities or simply when better performance from the control system is needed, model based approaches, such as MPC, can be sought.

MPC can be briefly described as a control methodology that uses the plant model to predict future responses of the plant, usually multiple steps ahead and with numerical optimisation, and decides on the control moves to manipulate the input variables. The optimisation is carried out to minimise a
cost function. The first of the future predicted control moves is applied at every sample and the calculation is repeated for the next. This is to minimise steady state error in the case of changing system dynamics such as disturbances. The moving horizon approach is shown in Figure 2.12, where $y$ is the current and past output and $\hat{y}$ is the future predicted output. $P$ and $M$ are the prediction and control horizons respectively, and the red square wave is the control input. As time moves on and new data becomes available for the MPC, the future outputs are predicted alongside the future control moves.

The backbone of MPC can be tracked down to the work of Zadeh and Whalen (1962) as pointed out by (García et al., 1989), and to the moving horizon controller as proposed by Propoi (1963). The work by Zadeh and Whalen shows how optimisation can be simplified as a linear programming problem in the presence of constraints, meaning when there is a minimum and maximum limit on a variable or the rate it can be changed. Solving the optimisation in a linear program context makes computing more efficient. The original industry application of MPC was presented by Richalet et al. (1978), where it was applied to over a hundred industrial plants and the economical benefits observed. As the ideas of MPC had been around for several years prior to its industrial application, in literature Richalet’s work is described as the “rediscovery” of MPC. However, the main difference between the theoreticians’ work and the version developed in industry is that it was actually applied to real industrial plants rather than remaining in theory.
Initially the work of Kalman et al. (1960) researched the predictive control methodology with the linear quadratic Gaussian (LQG) controller. Like MPC, the LQG control tries to minimise a cost function. In LQG the system is described in state space form, this is an alternative representation of a system, using the inputs, outputs and its states in time. These are related with first order differential equations and the transfer function of the system can be derived in a straightforward manner.

If the system is represented as a discrete time linear state-space model by equations 2-17 and 2-18:

\[ x[k + 1] = A \cdot x[k] + B \cdot u[k] + d[k] \]  \hspace{1cm} 2-17

\[ y[k] = C \cdot x[k] + e[k] \]  \hspace{1cm} 2-18

Where:
- \( x \) is the state vector,
- \( u \) is the input vector,
- \( d \) is the disturbance vector, which is Gaussian white noise with zero mean,
- \( y \) is the output vector,
- \( e \) is the Gaussian white noise vector with zero mean,
- \( A \) is the state matrix,
- \( B \) is the input matrix,
- \( C \) is the output matrix,

The cost function that is required to be minimised is:

\[ J = \sum_{j=1}^{\infty} (x[k + j]^T Q \cdot x[k + j] + u[k + j]^T R \cdot u[k + j]) \]  \hspace{1cm} 2-19

This means the input effort and the state is to be minimised. \( Q \) and \( R \) are weighting matrices and enable fine tuning of the cost function in terms of tradeoffs. The next predicted input is described as:

\[ \hat{x}[k + 1] = A \cdot \hat{x}[k] + B \cdot u[k] + K \cdot (y[k] - C \cdot \hat{x}[k]) \]  \hspace{1cm} 2-20

And the control action is calculated as:

\[ u[k] = -K_C \cdot \hat{x}[k] \]  \hspace{1cm} 2-21

Where \( K \) is the Kalman filter gain and \( K_C \) is the controller gain. The Kalman filter is an estimator of state variables, and works by using the data up to that time point. The LQG method has an infinite horizon, which provides stability and was shown to stabilise linear plants (Qin and Badgwell, 2003).

The LQG theory was then extended by Kwakernaak and Sivan (1973) with the application of the multistep ahead LQG controller and the development of other practical uses, such as offset free...
control and calculating steady state targets (Qin and Badgwell, 2003). LQG methodology is used extensively in the aerospace industry, where physical models of electric motors and other parts can be defined accurately by mathematical models. The reason for this is the LQG controller does not cope well when there are slight modelling inaccuracies and is sensitive to errors. Additionally, PID adaptations of LQG also exist (Grimble and Johnson, 1999). García et al. (1989) outline the limitations of LQG when the process is dependent on conditions that change with time, nonlinear, multivariable and have constraints. It is also pointed out that some plants may have their own performance criterion and different operating modes throughout one production cycle that is not possible to apply through LQG.

Similar to IMC, in its early days LQG was restricted to academic research. García et al. (1989) outline the reason for the academically advanced methods not being readily accepted to industry as, the mindset at the time and unfamiliarity of the engineers with these concepts. Additionally LQG and traditional feedback control does not address all of the issues industrial plants face, which are:

- Constraints
- Nonlinear behaviour
- Multivariable processes
- Custom performance criteria
- Smooth shutdown during plant failure or breakdowns

The points made above are explained in more detail in the following paragraphs.

Constraints are limits on how much a variable can be changed, such as certain temperature and pressure ranges a process is required to operate at, or a limit an actuator can physically operate at. Constraints are categorised as hard and soft. Hard constraints are strict limits that cannot be violated, as they may have dangerous consequences or have negative effects on product quality. Soft constraints are ranges a process can operate between and are not ideal to violate. Constraints should be handled in the control algorithm, and soft constraints can be violated depending on the weighting they are given in the cost function.

System nonlinearities can either arise from the process itself, from actuators near saturation limits, or measurements that are nonlinear such as certain material purities. Nonlinear systems are generally controlled by either linearising the system around its operating points or applying nonlinear control techniques. In practice, nonlinear continuous processes are generally dealt with by detuning the linear controllers, linearising the plant model or the feedback path (García et al., 1989). Nonlinearities can become dominant during start ups and shutdowns due to actuator saturation. To deal with the
said issues there are an abundance of methods such as dynamic rescaling using nonlinear methods, gain scheduling and sliding mode control.

Multivariable processes are systems with more than one input or output and are difficult to control using one regulatory controller. Problems of control loop interferences arise if more than one regulatory controller is implemented that are not aware of each other. Sometimes, high collinearities can exist between the variables, challenging conventional methods. Mainly regression methods are used to tackle the collinearities. These are explained in more detail in the further sections of this thesis.

Different industrial plants can need different performance criteria. They may have more than one mode of operation needing additional logic or time dependent output weights to deal with nonlinearities or to meet specific setpoints. Plants also require smooth shutdown procedures in the case of sensor failures or mechanical breakdowns. Conventional control techniques can react harshly if parts of the plant unexpectedly fail.

MPC was developed through the need to address these issues. An overview of MPC is shown in Figure 2.13 against the traditional control structure, adapted from Qin and Badgwell (2003). In both of the structures the plant optimiser supplies the optimum steady state settings to a local optimiser. This reconsiders the optimisation in terms of specific operating units. Next is the dynamic constraint control section, this is where the main difference between the conventional control and the MPC structure lies.
The dynamic constraint control approach of the conventional method is done with the use of digital logic, a regulatory controller that may be of any type, and a lead-lag compensator. A lead-lag compensator is a frequency domain based correction tool. This combination of different blocks is a complex approach and as the system gets larger, calculating the control requirements becomes more challenging. The aim of the dynamic constraint handling section is to move the process through different steady states, and the way the MPC structure does this is already inherent in its algorithm. The input and output constraints are already included in the formulation of the control problem. MPC uses its model to calculate the process inputs where the aim is to optimise the plant over the prediction horizon. The prediction horizon is a finite time length where the future outputs of the process are predicted, for example 10 samples ahead. As the constraints are already included in the problem formulation, future violations can be predicted in a multi-step ahead manner and prevented. The basic steps of MPC are (Seborg and Mellichamp, 2006):

1. Calculate the future outputs at every sampling instant over the prediction horizon using the model and the input signals.
2. With the main objective being keeping the output at or as close as possible to steady state, calculate future control signals over a control horizon at every sampling instant. The control horizon is a finite number of steps-ahead prediction of the future control moves.
3. Send the first calculated control move in the control horizon to the local controller and recalculate (back to Step 1).
The cost function is represented as:

\[
J = \sum_{j=0}^{P-1} \|\hat{y}(k+j) - y_{ref}(k+j)\|^2 + Q \sum_{j=0}^{M-1} \|\Delta u(k+j)\|^2 R
\]

Where control move \(\Delta u(k)\) is determined so that \(\min(J)\) is achieved where:

- \(P\) is the prediction horizon,
- \(M\) is the control horizon,
- \(y_{ref}\) is the reference trajectory,
- \(\hat{y}\) is the predicted output through the model,
- \(\Delta u\) is what is to be added to the current control move \(u(k)\) at the next sampling instant where:
  \[
  \Delta u(k+1) = u(k+1) - u(k)
  \]
- \(Q\) and \(R\) are weighting matrices, which are selected to add or lower the weight of the important variables of the process. The minimisation is subject to the following constraints:

\[
\Delta u_{min} \leq \Delta u(k+j) \leq \Delta u_{max}, \text{ where } j=0, 1, 2, 3... M-1
\]
\[
u_{min} \leq u(k+j) \leq u_{max}, \text{ where } j=0, 1, 2, 3... M-1
\]
\[
y_{min} \leq \hat{y}(k+j) \leq y_{max}, \text{ where } j=0, 1, 2, 3... P-1
\]

The first developed MPC application was IDCOM (Richalet et al., 1978). IDCOM had the following properties:

- The plant was modelled as an impulse response.
- The cost function was minimised as a quadratic objective over a prediction horizon.
- The future plant output behaviour could be set with a reference trajectory.
- Input and output constraints were handled inherently.
- Optimal control moves were computed using an iterative algorithm.

Using the impulse response model means that the approach is only valid for plants that are open loop stable. One of the applied plants was a nonlinear power plant steam generator (Richalet 1993). The nonlinearity was overcome by introducing a variable sampling time. Overall, significant economical benefits were reported.

Around the same era, Shell also developed their in house MPC technology (Cutler and Ramaker 1979), and although the aims were similar to IDCOM the implementation was different. This method
was called Dynamic Matrix Control (DMC) and did not handle constraints. The properties of DMC were:

- The plant was modelled as a linear step response.
- Cost function was minimised as a quadratic objective over a prediction horizon.
- The future plant output behaviour could be set, where it follows a setpoint as close as possible.
- Optimal inputs were computed with least squares, a regression method.

One of main differences between the two methods is the model that is used to represent the system. DMC uses a step response model where the predicted future output can be represented in terms of the linear combination of the inputs and outputs.

If $S$ is the matrix of step responses, also called the dynamic matrix of size $p \times m$, where $p$ is the number of steps ahead future prediction and $m$ is the number of future control moves:

$$
S = \begin{bmatrix}
    s_1 & 0 & \cdots & 0 \\
    s_2 & s_1 & \ddots & \vdots \\
    s_3 & s_2 & \ddots & 0 \\
    \vdots & \vdots & \ddots & s_1 \\
    \vdots & \vdots & \cdots & \vdots \\
    s_p & s_{p-1} & \cdots & s_{p-m+1}
\end{bmatrix}
$$

DMC can be formulated as shown in equation 2-23, where the $P$ step ahead prediction of the output can be presented as:

$$
\hat{y}(t+k) = \hat{y}(t) + \sum_{i=k-1}^{k} s_i \cdot \Delta u(t-k+1) + \sum_{i=k-1}^{k} (s_i - s_{i-1}) \cdot \Delta u(t-k+1)
$$

For $k = 1, 2, \ldots, P$, and where $t$ is current time, $\Delta u$ is the control action, and $s_i$ is the $i$'th coefficient obtained from the step responses.

The control move is computed as:

$$
\Delta u = (A^T A)^{-1} \cdot A^T \cdot (y_{set \ point} - \hat{y})
$$
Similar to IDCOM’s reference trajectory approach, the aim of the DMC controller is to manipulate the input variables so that the output is as close to the setpoint as possible. With this approach the changes in the input variable magnitudes are small and the outputs are subject to less aggressive responses.

One of the results presented on DMC by Cutler and Ramaker (1979) was temperature control of a furnace. When compared to a standard PID controller with a lead-lag compensator the DMC responded much smoother to the inlet temperature changes, resulting in a more efficient process.

IDCOM and DMC controller are classed as the first generation MPC software. The initial programs IDCOM and DMC worked effectively in comparison to more conventional control methodologies. Although, the constraint handling methods of IDCOM was described as ad-hoc, meaning the disturbances were not fully integrated into the program. DMC did not have any constraint handling properties in its formulation, but constraints were handled separately. The limitations of DMC are explored in the work of Lundström et al. (1995). The conclusion is that limitations come from how the predictor is formed, and in certain scenarios, for good control a high number of step response models may be required. This is not ideal as it is time consuming to gather all the required data from a working plant in production.

To address some of the issues, modifications of these products took place. Different methods were developed, such as QDMC (Garcia and Morshedi, 1986) which introduced a quadratic regulator where the input and output constraints were actually implemented into the algorithm rather than being handled, when they got close to their limits. IDCOM was updated to IDCOM-M (Grosdidier, Froisy and Hammann, 1988), (Froisy and Matsko, 1990). IDCOM-M used two objective functions, one for the outputs and one for the inputs. Similar to QDMC a quadratic objective function for the outputs was minimised subject to hard constrains, which were weighted in orders of priority. Other MPC programs were developed, such as the Shell Multivariable Optimising Controller (SMOC) where state space modelling and MPC was combined (Marquis and Broustail, 1998), (Yousfi and Tournier, 1991). State space modelling was chosen to model the full range of system dynamics and MPC was chosen to handle constraints. Additionally the theoretical work of Di Ruscio (1997) showed state space and MPC working together.

Although the initial analysis of MPC showed insufficiencies, their contribution to the process industries were significant in terms of efficiency and control. In time, commercial software and controller packages became available for general industry use rather than the in-house developed custom systems. Current MPC software packages come with various methods of model identification and model types. The new generation of controllers still utilise the impulse and step response models but also have the option of modelling the disturbances by using other time series approaches such as
Box-Jenkins. Some examples of commercially available packages are DMC-plus from Aspen Tech, which uses a modified version of the least squares algorithm. RMPCT from Honeywell Hi-Spec uses finite impulse response models, Box Jenkins models and autoregressive exogenous (ARX) models with least squares or prediction error methods (Qin and Badgwell, 2003). These methods are explained extensively in the existing literature Söderström and Stoica (1989), Åström and Wittenmark (2008) and are presented in Chapter 5 of this thesis.

The overview of a typical MPC loop is shown in Figure 2.14, where the noisy output and the disturbances are fed back into the controller. The multiple steps ahead predictions of the outputs would be determined and the control moves calculated and applied accordingly.

![Figure 2.14; MPC Loop](image)

In summary the benefits of MPC can be listed as:

- Multivariable processes can be controlled once the model is identified.
- MPC allows for complex dynamics such as nonlinearities, large time delays and inverse response.
- Constraints are handled inherently in the algorithms and future predictions lead to more careful handling. Additionally limitations with actuators are accounted for.
- The tuning of the system is straightforward by adjusting the weightings of the variables.
- If there is a fault, the controller will still carry on and try to meet the control objectives that have higher priorities and sacrifice lower priority objectives. This limits losses which could be higher using a traditional scheme.
When compared with traditional control techniques, MPC can provide improved control and improved efficiency. However, a cost benefit study is important to justify the extra time and resources needed to implement the technology.

Today, MPC can be found in a wide range of applications including the food processing, chemical, pharmaceutical, aerospace, and automotive industries. The following section presents the biomaterials manufacturing methods in the MPC context and explains how they fit in to the framework of this project.

2.3.3 Electrospinning, Solvent Casting and MPC

2.3.3.1 Electrospinning

When the benefits of MPC are considered in the context of the electrospinning process, it lends itself naturally to be an MPC problem. The electrospinning process has not been approached as a control problem before. The main aim of electrospinning depending on the specific application is to collect fibres of constant or varying diameters in either an ordered or a random manner. Many studies exist on the effect of its parameters on final fibres, such as solution properties and environmental conditions, but no conclusive method of controlling fibre diameters has been discussed. The studies have mainly focused on rigorous mathematical models of the jet or the effect of parameters that actually cannot be changed during the running of the plant, such as solution density or solvent type. PID type regulatory controllers could be used to control local actuators but the whole process is multivariable and constraints exist. A model of the process to determine future outputs and regulate the system is a more feasible approach when the range of solvents, polymers and other variables that can be changed is considered.

The variables of the electrospinning process that are possible to manipulate with the aid of control and automation are:

- Voltage
- Electric field shape
- Solution pump flow rate
- Spinning distance
- Rotating collector axle speed and orientation
- Temperature of the solution
- Temperature of the spinning environment
- Humidity of the spinning environment
All of these variables contribute to the final fibre diameters, as they either affect evaporation, stretching of the fibres, drying time of the fibres or any other effects presented in Section 2.2. The electrospinning process also has constraints on the following variables, these are:

- Pump flow rate
- Voltage
- Spinning distance
- Temperature of the solution
- Spinning environmental conditions

It is not possible to electrospin at any given flow rate for the whole range of voltages. Depending on flow rate and voltage, the spinning distance needs to be sufficient for the fibres to have adequate flight time so they can fully dry. If they do not dry, then the fibres join together in clumps of wet polymer and the mat loses its fibrous structure. The temperature of the solution needs to be sufficient to provide enough viscosity to allow spinning to take place. The environmental conditions of electrospinning also need to be within certain ranges depending on the solution temperature and solution type. The temperature of the air affects evaporation and humidity affects fibre diameters, too much humidity depending on the solution can slow evaporation down. In between these constraints various combinations of parameters may exist that provide similar diameter fibres, but the fibre mechanical properties may differ.

Considering the electrospinning process from a control systems modelling perspective is a novel approach. As pointed out in Section 2.2 the variables affect each other, so it is important to know these effects and establish how much they influence the final product. A model of the process would be ultimately useful in applying control and also help choose optimum operating parameters for a certain fibre diameter and material property. However there are challenges when modelling similar highly correlated processes, and how to deal with these is presented in the further sections of this work.

2.3.3.2 Solvent Casting

Solvent casting can be viewed as a batch style process as the film quality is not known until after the evaporation run. Batch processes rely on the input data to predict the final output. It is not possible to apply a standard regulatory controller to batch processes, as there is no feedback from a sensor that is directly related to the output. In practice, a model created using historical data is used for the output prediction and implementation of control. These methods will be discussed in detail in Section 2.3.6.

The variables of solvent casting that can potentially be manipulated during a run are:
• Air temperature
• Temperature of the platform the solution is evaporated from
• Humidity
• Airflow

Solvent casting has soft constraints, such as the temperature, which needs to be high enough for evaporation to take place. The effect of the rate of evaporation is an area that needs further investigation. The humidity has been reported to affect the pits and pores formed on the surface. Airflow affects evaporation and could be an influencing factor on the final film morphology. The combined effect of these three elements has an impact on the final film morphologies. With the use of a model, combinations of variable values can be calculated if a specific pore size is required. The model could also be used for monitoring the process in the presence of a fault causing a change in the environmental conditions.

The importance of models once again must be highlighted. For an MPC scheme to work effectively an accurate model is needed where plant-model mismatch is minimal. Obtaining models from first principles may be straightforward for simple aircraft actuators, but for complex processes with reactions or many variables such as solvent casting, it is not a feasible method. However, there are other ways to identify models. This is addressed by the mature field of system identification. Details of modelling and system identification methods are presented in the following section.

2.3.4 System Identification, Multivariable Systems and Modelling Methods

This section introduces the concept of system identification and presents the difficulties encountered when the systems have more than one input and one output. The methods to model multivariable and correlated systems are also presented. The study of determining models is an active research field and this is introduced first.

2.3.4.1 System Identification and Identification Signals

Identification of the model is the most important part of MPC, as without the correct model the embedded controller will not benefit the control system. As presented previously, there are two types of models; graphical and mathematical.

This section of the thesis presents the mathematical modelling approach. Electrospinning and solvent casting are complex processes and the graphical methods are not suitable on their own. As solvent casting is a batch process, the output is only available at the end of the run. Therefore, an impulse or a step change in the system will not be physically possible to observe. In the electrospinning case, step responses can be used. However, mathematical modelling methods will still be necessary to overcome the correlation amongst the variables of the process.
In mathematical modelling two approaches can be taken. The analytical approach method is to determine a system model purely from laws of physics. This is generally acceptable for simple systems such as servo actuators. However, the method is not suitable for processes where the number of inputs and outputs are high and the disturbances that are inherent within systems cannot be determined and described as an equation in order to complete the model. The alternative to this method is using input signals to excite the process, combined with routine process data and building time series models to describe the dynamics of the process. This approach is called system identification.

System identification is used to understand complex, multivariate systems. The main idea is to gather enough experimental data to build a time series model of the system. When determining a good working model of a process \textit{a priori} knowledge is useful. The modeller’s knowledge and experience of the system is important in determining the useful data for modelling and identifying the unexpected behaviour in the data. The next step following data collection is to determine the best model out of the model set. The main criterion for model selection is to identify a model that minimises the value of sum of the squared error during validation and testing (Söderström and Stoica, 1989).

Identification signals are used to acquire a more accurate picture of the process. These signals help the user collect data by exciting the system, which will describe the dynamics of the process. There are specific signals for the identification exercise. The signals need to be applied to the process, but should typically have minimal impact on the final quality of the product, as it may not be possible to stop production for the purposes of testing in a real industrial process. The main signals that are used for identification in academia and industry are Pseudo Random Binary Signal (PRBS), Generalised Random Binary signal (GBN), white noise and the sum of sinusoids.

The Pseudo Random Binary Signal (PRBS) is a signal proposed by Eykhoff (1974). The signal is generated using a shift register, a clock, and modular two addition and it takes two fixed signal values. A shift register is a mechanism that shifts an array of data by one position, removing the last element of data in the array and inserting a new one to the first place. The registers are given values of 1's and 0's and at every clock pulse the value of the state $n$ becomes the value of state $(n+1)$. The PRBS is deterministic and its autocorrelation function is similar to a random white noise signal meaning the frequency spectrums of the signals are similar for some special cases (Zhu, 2001). The PRBS is a useful signal, since it allows testing of certain frequencies of the process with equal power. This is desirable as the effect of the test signal needs to be minimal on product quality. The PRBS signal loses some of its appeal when the aim is to derive a model that will be used for multistep ahead prediction. The future predictors “overweigh” the fast frequency components within the data therefore giving an inaccurate model of the low frequencies of the system.
The Generalised Random Binary (GBN) signal, presented by Tulleken (1990), is designed to overcome the future predictor weighting bias of the PRBS, and presents a more elegant and control relevant signal. The GBN signal has a slightly modified switching probability when compared to the PRBS. This creates distributions which are not dependent of each other with zero mean.

Although they look similar to the eye, as shown in Figure 2.15, the difference between the GBN and PRBS lies in the power spectrum of the signals. The following Figure 2.16 shows the power spectrum of a GBN and a PRBS signal. The GBN signal does not have any dips at certain frequencies and is smoother than the PRBS, thus making it a more suitable identification signal as it maintains its power through all the frequency ranges.

Figure 2.15; PRBS and GBN Signals
Persistent excitation of the signal is of equal importance to the type of test used for identification. This means the test signal allows for unique solutions for the estimation algorithms to be created. The persistent excitation criterion applies to GBN and coloured noise signals, whereas the PRBS signal depends on the period.

White noise signals are exciting throughout the whole spectrum. Although they may be suitable for some processes, white noise identification signals are not the best type of signal for processes with electro-mechanical actuators. This is due to the frequent changes in magnitude leading to premature wear and tear. Raw white noise over emphasizes high frequencies, which is not desirable for a successful identification test. A filtered white noise, also called coloured noise, can be used for identification when faced with the frequency limitations of PRBS and GBN. Raw and low pass filtered white noise are shown in Figure 2.17. By using the appropriate filter, any frequency content can be realised with coloured noise. Although initially they appear similar in Figure 2.17, Figure 2.18 shows the difference in power spectrums of raw and filtered white noise signals. The filtered signal has a low pass characteristic and this is desirable for system identification, as high frequency components can cause instability and unnecessary wear on the plant.
Figure 2.17; Raw and Low Pass Filtered White Noise

Figure 2.18; Raw and Low Pass Filtered White Noise Power Spectrum
Another identification signal is the sum of sinusoids, where any desired frequency content can be achieved by setting the frequency and the amplitude of the sinusoidal input. Different frequency signals can also be combined and used for identification. The sinusoidal approach and the white noise approach have a stronger case when the systems are nonlinear and the binary inputs of PRBS and GBN are not sufficient to detect the nonlinearity (Zhu, 2001).

In summary the mentioned identification signals are all suitable for their purposes, although GBN has advantages over PRBS. Depending on the process itself, one or both of these signal types may be suitable to implement during a process run that has minimal effect on final product quality. Throughout this thesis the GBN signal is employed, as the simulated models are linear examples and it is more advantageous over PRBS. Once identification data is collected, various methods can be used to build system models, which are presented in the following sections. To begin with, multivariable systems are explained.

2.3.5 Multivariable Systems

Multivariable systems are systems with more than one input or output and the solvent casting and electrospinning methods are examples of this. The solvent casting process is a multivariable batch like process, where the final aim is to ensure that the final quality measures consistently meet specifications. A batch style process is where the product is made following a recipe, for example varying temperature setpoints or mixture levels from start to finish without any outside intervention. The difference with batch processes is that the product quality can only be known at the end of the processing run. There could be many data points logged for the input variables and only one measurement for the output. This requires a model where the output is predicted only using the input variables as they follow the batch recipe.

The electrospinning process is a multivariable continuous process, where the final fibre structure can potentially be adjusted as the process is running, to create fibre bundles with desired characteristics. The different forms multivariable processes can take are presented in the representations in Figure 2.19, where $X$ and $Y$ are the input and output data matrices respectively. A batch like process has $n$ inputs that have been sampled $m$ times and only one sampled output which is obtainable at the end of the batch. The Multi Input Single Output (MISO) system has $n$ inputs that have been sampled $m$ times and one output sampled $m$ times. The MIMO system has $n$ inputs that have been sampled $m$ times and $p$ outputs that have been sampled at the same time as the inputs. Sometimes, correlations that exist amongst the inputs and the outputs that can make modelling difficult, as identifying which variable has what effect may not be straightforward.
Monitoring and modelling methods that are applied to multivariable systems have evolved over time. Starting with quality monitoring charts, multivariable systems have been monitored with statistical methods (Kourti and MacGregor, 1995). Today, both continuous and batch process control systems rely on predictions of the future values of variables in order to achieve optimum results, where the future predictions are used within MPC. There are two ways to model multivariable processes with large data matrices; time series and statistical (Barceló et al., 2011). Methods for collecting relevant data and the statistical approaches that can be implemented to gain an understanding of this data are presented in the following sections.

2.3.6 Statistical Methods

In multivariable systems, controlling the quality of the final product is dependent on the variables of the process. Initially, quality monitoring was applied to processes using Statistical Process Control (SPC) charts, which indicated whether the system remained within specified limits (Kourti, 2003a). The charts were used to monitor product quality variables and certain important process variables, rather than to apply control to the process. When the process control chart shows the quality variable
as out of limits or at an unusual level, it may mean an abnormal event or a fault may have occurred. When this happens, many of the charts will show anomalies due to the correlation between the input variables, which needs to be investigated and recorded for future reference. The anomalies are a result of the system operating out of its control limits. Finding out exactly what caused the fault can be a time consuming process and sometimes a clear conclusion is not reached. However, this approach rather than being preventative and corrective only acts as a data logging exercise, so when it happens again the user knows what the fault may be.

Monitoring only the quality variables of a process is not adequate to meet the ever increasing need to improve quality and make the process more efficient. Therefore SPC alone is rarely sufficient to provide accurate monitoring of a process. To address this issue, multivariate extensions to SPC, such as Principal Components Analysis (PCA), Principal Component Regression (PCR) and Partial Least Squares (PLS) are used. For these methods to be effective they need to describe the process accurately, and for this to be possible, sufficient data is required. The data from the process is acquired with the aid of sensors and data logging.

Advances in technology means that large amounts of data are now regularly recorded from processes, which enables the user to gather information at high sampling intervals and over long time periods (MacGregor and Kourt, 1995), (Kaspar and Ray, 1992). With the developments in sensory technologies, the amount of process data collected today is considerably more than the limited measurements that were recorded several years ago, which may have included quality data only. For example, on an injection moulding process the quality variable monitored was the shrinkage of the final product from the desired mean value. Other possible measureable variables are flow, temperature, pressure, speed and time. Monitoring the quality variables only is considered inadequate in today’s industry, as frequent measurements of pressures, flow rates, temperatures and concentrations are readily available through sensors. However, many quality variables, such as biomass density in fermentation processes, are only available infrequently when laboratory samples are taken or when the batch is terminated. The variables in processes are usually correlated and the events that cause changes to the process can be identified and modelled by using multivariate statistical methods. Trends and relations that cannot be identified by visually looking at the data alone can be discovered using multivariate statistical projection methods.

The large amounts of data obtained from processes need to be analysed. This is difficult due to the size of the data matrices and the correlations that exist between the variables. To help analyse these data matrices several methods have been proposed which compress the data, without losing information. The method for doing this is Principal Components Analysis (PCA). PCA reduces the data to a smaller number of variables called principal components.
PCA decomposes the eigenvectors of the covariance matrix of the measured process variables (Wise and Gallagher, 1996). The measurement data matrix is represented by $X$ which is an $m \times n$ matrix, where $m$ is the number of samples which is also the number of rows of the $X$ matrix. The notation $n$ is the number of measured variables and also the number of columns of the $X$ matrix. The covariance matrix is described as:

$$\text{cov}(X) = \frac{X^T X}{m-1}$$  \hspace{1cm} 2-25

PCA initially finds a direction that represents the data with the greatest variance. Following this the next greatest variation direction is found, and this continues until all of the variance is captured. The identified directions are all orthogonal to each other.

The dimension reduction with PCA is achieved by decomposing the matrix into the sum of the outer product of two vectors, which are referred to as scores and loadings and a residual matrix as presented below:

$$X = T \cdot P^T + \epsilon = \sum_{k=1}^{k} t_k \cdot p_k^T + E = t_1 \cdot p_1^T + t_2 \cdot p_2^T + t_k \cdot p_k^T + \epsilon$$  \hspace{1cm} 2-26

Where $k \leq \min(m, n)$. The $P$ matrix consists of the eigenvectors of the covariance matrix, also named the loadings. This means $P$ is the coordinates of this new matrix and its elements are the individual orthogonal directions. The $T$ matrix is the projection of $X$ on to $P$, and its elements are named scores, and finally $\epsilon$ is the residual, which is the part of the data that could not be described.

The number of principal components used in the PCA model determines the quality of the model. There are methods for determining the number of principal components to be used, such as cross validation (S. Wold, 1978), (Krzanowski, 1987). Cross validation involves dividing the data matrix into several subsections. The PCA model is identified using one or more of the data sections and tested with the remaining sections. This procedure is repeated several times with the number of components determined by that which give the smallest error.

PCA has been further extended by Wold (1987) and later by Nomikos and MacGregor (1994) so that it can be used with three dimensional matrices, which are collected from batch processes. This algorithm is referred to as Multiway Principal Component Analysis (MPCA). The application of MPCA is shown extensively in Kourti and MacGregor (1995), where the three dimensional batch data matrix is converted to a large two dimensional one through a process called unfolding and conventional PCA is applied to the resulting two dimensional matrix. Unfolding is explained further in the following sections of the thesis.
PCA has been applied to many different fields in science and engineering. Examples of this include; the classification of materials when analysing spectroscopic data for medicine production (Krämer and Ebel, 2000), monitoring of batch processes (Nomikos and MacGregor, 1994) (Lu et al., 2004), and finding similarities between covariant behaviours of certain variables on a large data set, that is used to relate hospital admissions for people with respiratory conditions to winter air mass types (McGregor et al., 1999).

The first control relevant applications of PCA simply monitored the principal components, in a similar way to how variables are monitored using SPC. An important statistic related to PCA is the Hotelling’s $T^2$ statistic, which describes the statistical square distance of the process from its mean trajectory. Another important statistic is the Squared Prediction Error (SPE), which indicates the squared difference between the actual data point and the predicted data point. There are also statistics available from the PCA model, which provide information on how each input fits the PCA model. These statistics enable the calculation of control limits for the process, which are beneficial for process monitoring. When the system is out of these limits, it means an event or a fault has occurred. It is possible to determine which variable has contributed the most to the fault by analysing the residuals of the PCA model’s prediction (Qin et al., 1998). Kourtì et al. (1996) and Miller et al. (1998) present applications where PCA was used for monitoring and control of batch processes.

Although very useful, PCA does not provide a cause to effect model, which is vital for incorporation in to a model predictive control system. The method to identify a cause to effect model is the ordinary least squares method, which is a linear regression method where a regression vector is calculated by minimising the error between the predicted and the actual output values. However, this method is problematic if there are collinearities between the input variables, such as will be present with the electrospinning process. This approach also fails if there are fewer samples than the number of inputs. This creates a wide and short data matrix, which is termed ill conditioned, as the inverse of the transpose $(X^T X)^{-1}$ does not exist, where $X$ is the input data matrix. However, this inverse is required to calculate the regression vector, which is described as:

$$Y = \beta \cdot X + r$$  \hspace{1cm} 2-27

Where $\beta$ is the parameter vector that is unknown, $Y$ is the output row vector and $r$ is the residual or the unexplained part of the data caused by noise or disturbances. If we assume $r$ is negligible then $\beta$ is defined as a set of values which minimises the sum of the squared error between the actual output and the predicted output;

$$\epsilon^2 = (Y - \beta \cdot X)^T (Y - \beta \cdot X)$$  \hspace{1cm} 2-28
If 2-28 is differentiated to minimise:

$$\frac{\partial e^2}{\partial \beta} = 2 \cdot X^T X \cdot \beta - 2 \cdot X^T Y = 0$$  \hspace{1cm} 2-29

Then $\beta$ can be estimated as:

$$\hat{\beta} = (X^T X)^{-1} \cdot X^T Y$$  \hspace{1cm} 2-30

When $(X^T X)^{-1}$ does not exist due to ill conditioned matrices, Principal Component Regression (PCR) has been proposed (Næs and Martens, 1988), (Martens and Næs, 1992), (Esbensen and Geladi, 1992), (Høskuldsson, 1996) as an alternative method. This method regresses the output variables over the principal components identified with PCA, and cross validation is used to determine the number of principal components to be used in the PCR model. The output $Y$ is defined as:

$$Y = \hat{R} T + \varepsilon$$  \hspace{1cm} 2-31

Where $T$ is the principal component scores matrix and $\hat{R}$ is the regression vector and $\varepsilon$ is the residual.

$$\hat{R} = (T^T T)^{-1} \cdot T^T Y$$  \hspace{1cm} 2-32

Over time the PCR method has been used extensively, for example Yang et al. (2006) report using PCR for building a predictive model for applying control to an injection moulding plant. The weight of the final product was predicted online with PCR during the manufacturing process and the input variables were adjusted accordingly to meet the tolerances. Al-Alawi et.al. (2008) used PCR in conjunction with neural networks to improve predictions of ozone concentrations and demonstrated that the root mean square error of the prediction was improved over using either PCR or neural networks independently. Although PCR is frequently used, it does have some drawbacks. The examples given where PCR has been successfully used had only a single output variable. In many control systems there can be more than one output variable. PCR will only relate one output variable at a time to all of the input variables. However, it is important to point out that the output variables could also be correlated amongst themselves and PCR does not address this. The correlations amongst the outputs can carry important process information in terms of applying control and it is beneficial to describe them. Using PCR will also result in the loss of this information, as it is possible that very small variations in the input matrix will have a significant effect on the output variable. However, it is possible for the small variations to be discarded in the PCR model (Geladi and Kowalski, 1986), (Jackson, 1991). A more attractive and powerful approach to modelling highly correlated data sets is Partial Least Squares (PLS) also referred to as Projection to Latent Structures (PLS).

The PLS method, mainly pioneered by Wold (1966), aims to maximise covariance by capturing the maximum variance between input variables, and the maximum correlation between the output
variables. The number of factors used to describe the data is termed *latent variables*, these are not directly observed but are deduced from the measured variables. The number of latent variables can be determined using cross validation.

There are various algorithms which can be used to calculate PLS models. The most commonly used methods are SIMPLS (de Jong, 1993) and Nonlinear Iterative PLS (NIPALS) (H. Wold and Lyttkens, 1969). Both algorithms perform at the same level when there is a single output, but results differ slightly when there are multiple outputs. The NIPALS algorithm has been the main focus of research over the years, and starts by decomposing the input and output matrices that are represented as the sum of latent variables. An initial weight is selected from the non-zero rows of the input matrix, and then a score for each of the input and the output matrices is generated for every iteration. Following the first iteration the input matrix and the output score is used to calculate a new weight, and once this new weight converges to the initial weight, the iteration stops and a new loading vector is calculated. This is used to calculate the regression coefficient and is shown in detail in Chapter 5 of this thesis.

The alternative SIMPLS approach is a more statistical approach where the weights are calculated from the original data matrices using the Krylov subspace method (de Jong, 1993). Helland (1988) demonstrated through geometrical means how the predicted output is the projection of the actual output vector on to a subspace described by a particular Krylov sequence. The SIMPLS method avoids the steps of having to deflate the matrices as in NIPALS, and is shown to be computationally faster (de Jong, 1993). Additionally, due to the computational intensity of the NIPALS algorithm when the data matrices are large, more efficient *kernel* versions that only use the covariance matrices have been developed. Dayal and MacGregor (1997) improved the kernel algorithm proposed by Lindgren et al. (1993) by using the covariance matrices, which considerably improved the efficiency of the algorithm. Further details of PLS and its associated algorithms are presented in Chapter 5.

In initial PLS applications, it was used for modelling and condition monitoring of processes. Following the application of PCA in continuous and batch scenarios, PLS was similarly applied to industrial processes. For example Piovoso et al. (1992) applied PLS to monitor a polymerisation process and Wise et al. (1991) applied it to a ceramic melting process, where some benefits compared with PCA were reported. More recently, there have been applications of PLS within control systems. Flores-Cerrillo and MacGregor (2004), for example developed a control strategy that regulated the plant within the latent variable space of the PLS model. This had the effect of reducing the complexity of the problem, making the control calculations more efficient. Kesavan et al. (2000) applied PLS to a batch pulp digester where the mixture measurements taken during production, were related to the final product quality. The PLS model was then used to adapt the parameters of the empirical model in order to overcome the batch to batch variation the process suffered from. The performance of this
controller was compared to that of the controller using the simple empirical model, which used a single chemical measurement taken at the beginning of the production run to compensate for the batch to batch variations. The use of a PLS model to adapt the empirical model parameters according to the variations provided improvements on the final product quality.

PLS has also been extended for use with nonlinear systems. Early works to introduce nonlinearities to PLS included formulating it as a quadratic problem, where the output and the output latent variable relationship is described as nonlinear (S. Wold et al., 1989). A further suggestion was to use a spline function, to describe the nonlinear relationship between the latent variables (S. Wold 1992). However, this technique assumes the nonlinear relationship can be described as a polynomial, and the specific order of the polynomial may be difficult to select. Qin and McAvoy (1992), with a view to avoid the polynomial selection process, proposed an artificial neural network approach to describe the nonlinearities. The disadvantage with this method was that only small nonlinearities could be represented. An error based weighting updated version of the neural network strategy was proposed by Baffi et al. (1999), who were able to improve the nonlinear modelling capabilities of PLS further.

Dynamic extensions to the PLS algorithm have been proposed to meet the requirement to model dynamically changing systems. In time varying dynamic processes, the model should be able adapt to the changes and keep the system under a controlled state. Ricker (1988) proposed using time lagged input variables in the input data matrix $X$, and Qin and McAvoy (1993) proposed using lagged values of both of the input and output variables in the data matrix. Both of these approaches cater for dynamic changes, at the expense of having to compute large matrices due to the addition of the extra lagged variables. Kaspar and Ray (1993) suggested the use of a dynamic filter where a linear algebraic model is found, relating the input data matrix to the output data matrix. This reduces the matrix size when compared with the methods that use purely lagged variables. One disadvantage however is the specification of the filter order, as in all the examples specified, first order filters were utilised due to the first order dynamics as the authors had *a priori* knowledge of the system. The suggested method to determine the order of the filter is to minimise the sum of the squared error of the output residuals.

Lakshminarayanan et al. (1997) further extended this approach to be used with nonlinear MIMO systems, where the modelling of the nonlinearity was handled with a Hammerstein model, a well known dynamic nonlinear modelling approach. The use of this method relates the input and output scores dynamically rather than using a static linear filter. However this approach requires careful selection of identification signals with low frequency contents for model determination.

PLS has been shown to have advantages over other regression methods such as Principal Component Regression (PCR). The extensions of PLS have followed on from the initial NIPALS algorithm to make
it more computer processor friendly, to cope with nonlinearities, and systems with dynamic behaviour. Another method used to deal with time varying dynamics in systems is Recursive Least Squares (RLS) which is the extension of the linear regression method Ordinary Least Squares (OLS).

RLS is the recursive implementation of the OLS method, and aims to minimise the mean prediction squared error. RLS computes the next least squares estimate based on the old estimate, the covariance matrix and the new data coming from the system. RLS is a naturally adaptive algorithm, which can incorporate a forgetting factor that will reduce the weight of the older measurements and hence capture the most recent dynamics. For example, it will adapt to processes that are slowly varying with time. The OLS method will produce a biased estimate in the presence of noise or disturbances. This is also carried through to the RLS method as the output becomes correlated with the noise and is included in the calculation of the regression coefficient. Sandoz and Swanik (1972) proposed an alternative approach to solve this problem, by including the previous estimated outputs and the new incoming output in the calculation of the regression coefficient. This was termed an unbiased estimator and eliminated the effect of noise. The bias caused by noise and ill conditioned data sets and the effect it has on the accuracy of the model is detailed in Chapter 5.

Various research studies have been conducted over the years on RLS using real data, to investigate the strength of the algorithm (James et al., 1974), (Isermann et al., 1974), (Söderström et al., 1978). Sandoz (VanDoren, 2003) pointed out that the RLS algorithm incorporating Bierman's Upper Diagonal (UD) filter (Bierman, 1976) is the most applied adaptive controller algorithm utilised in industry at that time. The UD filter ensures that the covariance matrix is always positive definite by ensuring the symmetrical matrices are always upper diagonal. The UD filter, unlike a Kalman filter, does not have numerical deterioration caused by negative elements in the covariance matrix.

The Bierman filter version of the RLS algorithm is widely used due to its reliable performance and the ease of tuning. In certain scenarios where a forgetting factor has been implemented, the algorithm will slowly erase the stored data if the plant dynamics are not changing, which can be dangerous. The use of a forgetting factor is not encouraged, and an alternative method is suggested to reconfigure the plant. The covariance matrix is reset by “kicking” the model by setting the covariance matrix to a very large number. This has an equal effect to a forgetting factor as the model recalculates the parameters to adapt to the changes (Åström, 2008), (VanDoren, 2003). This method of updating is used in the commercial Connoisseur MPC software supplied by Invensys.

The PLS method also suffers from the same bias problem when predicting forward through time, which is caused by the disturbances being correlated with the output data used for the regression calculation (Awad, 2008). This is a problem not often addressed, due to the benefits of PLS when used in multivariable system identification and in the presence of high correlations. The suggested
solution is to use the unbiased estimates of the unbiased RLS (uRLS) model in the PLS algorithm. This suggested approach forms one of the main starting points for the theoretical modelling part of the research in this thesis. The advantages of unbiased PLS (uPLS) over standard PLS are shown in the examples in Chapter 5.

An alternative modelling method is the Finite Impulse Response (FIR) modelling. FIR provides unbiased estimates, as long as the data gathering exercise was conducted in open loop. This is not always feasible as operating naturally unstable plants in open loop can have unwanted consequences, such as becoming unstable, which can be dangerous and costly if equipment is damaged. The FIR model describes the system as a sum of impulse responses:

\[ y(t) = \sum_{k=1}^{N} h_k \cdot u(t - k) \]  

The FIR method’s most important advantage is that it is not necessary to have previous knowledge of the process or of its order to create working models. The estimate is unbiased if the process is identified in open loop and the number of data points used tend towards infinity. In theory this method is ideal. However, from an application perspective, it requires a large number of parameters to describe the system, which results in a high computational expense. Zhu (2001) also highlights that an extra step of model reduction is necessary to apply linear control methods using FIR.

The least squares methods can be used to build rational transfer function models. One model type is the Autoregressive Exogenous (ARX) model, which is a more compact model form. If the system is described by the difference equations:

\[ y(t) + a_1 y(t-1) + \ldots + a_n y(t-n) = b_1 u(t-1) + \ldots + b_n u(t-n) \]  

Where:

\[ A(q) = 1 + a_1 q^{-1} + \ldots + a_n q^{-n} \]  

and

\[ B(q) = b_1 q^{-1} + \ldots + b_n q^{-n} \]  

The transfer function is:

\[ G(q) = \frac{B(q)}{A(q)} \]
Then the system can also be described as:

\[ A(q) \cdot y(t) = B(q) \cdot u(t) + \epsilon(t) \]

Here \( \epsilon(t) \) is described as the residual or the *equation error*. The parameters of the system are linearly related to the equation error, and the use of a cost function to minimise this is straightforward. Problems with this method arise when there is coloured noise present in the system, which creates a biased estimator. This problem can be partially overcome by increasing the order of the system. The effect of the bias is more dominant in the frequency domain where it causes errors at the low and middle ranges of the frequency spectrum (Ljung, 1987), (Zhu, 2001).

The effect of the noise in the frequency spectrum can be countered by using pre-filters. The pre-filter is used to cancel the effect of bias, which in the frequency domain can be represented as a weighting multiplier affecting the frequency response. This bias can in theory be countered by designing the input signal power spectrum around the weighting constant, so the identification has a cancelling effect and the correct parameters can be obtained. This is not simple, as the parameters of the system are unknown, and repeating an identification test in an iterative manner is inefficient and not financially viable from an industry point of view. Designing a pre-filter can be achieved iteratively by applying filters to the already obtained data from a single test. The error is then minimised in an iterative manner proposed by Steiglitz and McBride (1965), (Stoica and Soderstrom, 1981). Rivera et al. (1992) suggest that the bias is reduced if a pre-filter is used prior to applying PLS. This is caused by a reduction in the degrees of the estimation problem, as PLS represents the data in terms of latent variables. The same trade-off between bias and variance still exists in PLS but at a reduced level. In existing literature it is pointed out that the pre-filtering method does not work for narrow band noise signals and alternative approaches such as the Output Error (OE) or prediction error methods (Ljung, 1978) and recursive methods are suggested (Zhu, 2001). These are presented in the following section.

### 2.3.7 Output Error and Prediction Error Methods

The output and prediction error methods have been suggested as a solution to the inherent bias of the equation error found with linear regression methods. Both are mature methods in the model identification research field.

Output and prediction error methods are alternative approaches to linear regression methods for modelling multivariable systems. They are parameter estimation techniques, where the loss function is formulated so that the output error or the prediction error is minimised. The minimisation can be undertaken using various techniques including least squares, PLS, RLS or numerical search algorithms. The main differences of the output error and prediction error methods from the
regression methods are their formulation of their cost functions. The main approach of these techniques is to describe the model of the system as the next predicted output, which can be further extended to multi step ahead prediction. These methods traditionally rely on building models following a system identification exercise rather than process data or data collected during routine runs with added external coloured noise (Söderström and Stoica, 1989). This gives flexibility when modelling the noise and system dynamics separately, which is not possible using the least squares methods. However, the model orders need to be correctly determined before being used in the equations, in order to obtain accurate results from the output and prediction error methods. This can be a time consuming process when compared with the least squares methods.

The main difference between output error and prediction error is that the output error method does not model disturbances, whereas the prediction error method is able to model them. In some cases, not modelling the disturbances will save time and effort for the control engineer. In others the disturbances will need to be modelled to get the best out of the control system and to increase its disturbance rejection capabilities. Whether this is required or not will depend on the process, and if there is a high safety risk then disturbance rejection may become more important. The output error method is formulated slightly differently to the previously introduced equation error method of the least squares identification methods. This can be shown if the transfer function of the model is assumed to be as expression 2-37:

\[
G(q) = \frac{B(q)}{A(q)}
\]

Where:

\[
A(q) = 1 + a_1 q^{-1} + \ldots + a_n q^{-n} \quad \text{and} \quad B(q) = b_1 q^{-1} + \ldots + b_n q^{-n}
\]

Then the output error is described as:

\[
\varepsilon_{oe}(t) = y(t) - \hat{y}(t) = y(t) - \frac{B(q)}{A(q)} \cdot u(t)
\]

However, this approach offers no analytical solution for the minimisation of its cost function, although local minimum points may be calculated, a global minimum is not possible (Söderström and Stoica, 1982). To minimise the cost function a numerical search routine is required, Gauss-Newton or Levenberg–Marquardt for example. This approach is also more expensive in terms of computation when compared with the linear regression methods.

Prediction error methods fill the gap left by the previously presented methods. They can also be used when the data is collected under closed loop conditions, rather than open loop. This overcomes the
inconvenience and the risk of putting the system under manual control during identification. Additionally, they can obtain less variance in the predictions in comparison with the output error method (Zhu, 2001). This is achieved through a similar approach to pre-filtering, where the disturbance is estimated using shaping filters. Different filter types are used depending on the type of prediction error that is employed such as the Generalised Least Squares (GLS), Auto Regressive Moving Average Exogenous (ARMAX) and Box Jenkins Method (BJ). The prediction error is described as:

\[
\varepsilon_{pem}(t) = y(t) - \hat{y}(t)
\]

Where, the equation error \( \varepsilon_{pem} \) is treated as the prediction error. This is shown in more detail in Chapter 5. Different filter types used in the prediction error methods create different cost functions. The GLS method (like OE) has no analytical solutions, but a simplified representation of the cost function can be obtained and minimised iteratively using the least squares method. However this method does not guarantee global minimum convergence. Söderström (1989) suggests using the least squares approach to estimate the polynomials that form the loss function and perform a model reduction step.

An alternative model type is the ARMAX model (Astrom and Bohlin, 1966). The ARMAX method assumes that the noise or disturbance is of moving average structure, and is widely used in controller design routines due to its similarity in structure to discrete transfer function models. (Linko et al., 1992) have modelled a fermentation process successfully without extensive knowledge of it and recommend the model could be used for monitoring and implementing within model based control systems, such as MPC. Bercu (1992) used the ARMAX model structure and calculated the parameters with weighted least squares to provide an adaptive controller that was able to track a reference trajectory.

Another prediction error model type is the Box-Jenkins method (Box and Jenkins, 1970), and it is classed as an extension of the OE method and is able to model disturbance dynamics. Due to its more complex structure, this method takes longer to compute. However, in today’s computing technology the processing times of any of the algorithms mentioned here are not a significant problem. The ability to model disturbances increases a control systems’ disturbance rejection, which is important in terms of optimal production and safety. Alongside control, the Box- Jenkins method is also widely used for economic modelling, such as the economic forecasting work of Bray (1971). The Box-Jenkins model structure is employed in the PEM method of system identification work of this thesis, presented in Chapter 5.
2.3.8 Design of Experiments

Methods known as Design of Experiments (DEO) were developed as a means of better understanding and optimising processes using the minimal amount of time and effort. The obvious approach to understanding a process is to vary each cause variable on a process independently and then analysing the response of the system to identify the specific effects of each cause variable on the system response. In contrast, DOE methods define a series of tests that can be undertaken such that combinations of cause variables are manipulated simultaneously in a specific pattern, such that the response of the process to each individual cause variable can be determined. DOE methods can be separated into two main methods:

- Full factorial analysis
- Fractional factorial analysis

First the full factorial analysis is explained.

2.3.8.1 Full Factorial Analysis

Full factorial design method allows the investigation of a specific region of a process to discover the cause–effect relationships amongst the variables. If three parameters of a process are considered; Property 1 (P1), Property 2 (P2) and Property 3 (P3) a full factorial design map of this can be illustrated as in Figure 2.20.

![Figure 2.20; Full Factorial Map for Three Variables](image)

The eight points represent individual experiments where the variable value has been set to either a low level (-) or a high level (+). Three variables require eight experiments in a full factorial study, and when the number of variables increases the number of experiments increases drastically. For example 5 variables require $2^5 = 32$ experiments, 8 variables require $2^8 = 256$ experiments. Repeat experiments are also recommended to account for natural variation in the process and when these are considered the number of experiments can reach extremely high numbers, which can be
impractical. The high number of experiments can be a disadvantage for the full factorial method, however it allows the estimation of all main effects and all the interaction effects between variables. Additionally a centre test; a test completed under conditions in the centre of the cube shown in Figure 2.20 should also be undertaken to test the linearity of the system. If a full factorial study is infeasible or it is not necessary to identify the effects of all the process parameters then fractional factorial analysis can be applied. This is explained in Section 2.3.8.2.

The effects are split into two: the first one is the main effects and the other the interaction effect. The main effect is the linear variation of the response over the full range of that property, for example P1. The significance of an effect can be tested with the use of probability plots or significance tests (Esbensen et al., 2002). Interaction effects are the cross-terms between the two variables and can be extended to three factor or more orders and are assessed the same way as the main effects.

The main effects are calculated as follows:

\[
\text{Mean Effect of Variable P1} = \text{Mean Response at High P1} - \text{Mean Response at Low P1}
\]

(Esbensen et al., 2002)

The two-factor interaction effects, I are calculated as follows, for example between P1 and P2:

\[
I = \frac{1}{2} \cdot (\text{Effect of Design variable P1 at high level of P2} - \text{Effect of Design variable P1 at low level of P2})
\]

When the data collection is finalised a table similar that to shown in Table 2.2 will be obtained. This table represents a typical data table for three variables (P1, P2 and P3) and one quality variable (Q1), where n1 to n8 represent numerical values.

<table>
<thead>
<tr>
<th>Test</th>
<th>P1</th>
<th>P2</th>
<th>P3</th>
<th>P1 &amp; P2</th>
<th>P1 &amp; P3</th>
<th>P2 &amp; P3</th>
<th>P1 &amp; P2 &amp; P3</th>
<th>Q1</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>n1</td>
</tr>
<tr>
<td>2</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>n2</td>
</tr>
<tr>
<td>3</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>n3</td>
</tr>
<tr>
<td>4</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>n4</td>
</tr>
<tr>
<td>5</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>n5</td>
</tr>
<tr>
<td>6</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>n6</td>
</tr>
<tr>
<td>7</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>n7</td>
</tr>
<tr>
<td>8</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>n8</td>
</tr>
</tbody>
</table>

Table 2.2; Three Variable Factorial Study Table
As highlighted previously, if the number of variables in the process is high then the number of experiments and the data gathered increases significantly. If a large table of data is obtained then some of the methods reviewed in the previous section can be utilised to find the effects of each of the variables, such as regression analysis, PCA or PLS. The regression coefficients will provide an indication of the impact that the cause variables have on the system. Expression 2-41 represents the regression equation for the data in Table 2.2.

\[ Q_1 = b_0 + b_1P1 + b_2P2 + b_3P3 + b_4P1P2 + b_5P1P3 + b_6P2P3 + b_7P1P2P3 \quad 2-41 \]

In expression 2-41 \( b_0 \) is the mean of the effects of the experiments. The effect of variable \( P1 \) can be estimated by multiplying its coefficient \( b_1 \) by two and the interaction effects can be estimated multiplying the coefficients from \( b_4 \) to \( b_7 \) by two.

### 2.3.8.2 Fractional Design

The number of experiments that need conducting for full factorial studies is often prohibitive, both in terms of time and economics. The main effects and interactions of a process can be established without the need for testing all the variables and possible combinations, using fractional designs (Diamond, 2001).

The determination of variables for the fractional design can be achieved by setting up the test schedule as with the full factorial design, with the high (+) and low (-) values for only two variables and multiplying their signs to determine the level of the third test variable. This can be represented in Table 2.3.

<table>
<thead>
<tr>
<th>Test</th>
<th>P1</th>
<th>P2</th>
<th>P3</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>-</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td>2</td>
<td>-</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>3</td>
<td>+</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>4</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
</tbody>
</table>

**Table 2.3; Fractional Design Table**

Although there are benefits with the fractional approach, there are also limitations, most notably *confounding*, which can be summarised as follows:

- Due to the cross multiplications it can be difficult to tell whether an effect has occurred due to one variable or because of an interaction.
- The fractional method is not suitable if the variables have more than two levels.
- Two variable interactions may not be distinguishable between each other.
Today with the help of software it is straightforward to identify cases of confounding where the risks of misinterpretation is reduced (Barrentine 1999). However this does not solve this problem.

In addition to Fractional designs there are other screening design methods, such as the Plackett-Burman Design (Plackett and Burman, 1946). This method is advantageous if there are many variables and a short space of time to conduct tests. This method works using a mathematical formula and the number of experiments need to be in the multiples of 4. In terms of application Plackett – Burman method can be used as a first screening method.

2.3.8.3 Analysis

The DOE methods introduced above can show how the variables effect the end product and how they interact with each other. There are also other data analysis methods that provide more insight in to the data.

The effects of a large data set can be determined by plotting a Normal probability plot, also termed the Normal-B plot. For analysis, the $b$ coefficients obtained from the data are plotted (x-axis) against their probabilities (y-axis), which is obtained using the F-test, explained below. The values that lie to the top right and bottom left of the line of best fit are the coefficients that effect the process significantly.

To determine whether the effect of a potential cause variable is significant or not is to use the F-Test. For this test, the F value is calculated using the relative standard deviation $R$, which is calculated as shown in equation 2-42 (Esbensen et al., 2002).

\[
R = \sqrt{\frac{\sum_{i=1}^{n} (y_i - \bar{y})^2}{(n-1)}}
\]  

The F value is then calculated as shown in 2-43 (Esbensen et al. 2002).

\[
F = \frac{e^2}{(m-1) - \frac{R^2}{(n-1)}}
\]  

Where $m$ and $n$ are the number of experiments and the number of repeats, respectively.
The effect of a cause variable is then defined as being significant or not by comparing the F-value calculated with the F distribution, using three parameter values which are \((m-1), (n-1)\) and the confidence limit (a value between 95% - 99%). If the calculated F-value lies outside the confidence limits then it suggests that the cause variable has a significant effect on the system.

2.4 Summary

Starting with the tissue scaffold manufacturing techniques, an analysis of the manufacturing methods has been provided, and the reasons for the use of electrospinning and solvent casting techniques have been outlined. Tissue engineering has progressed significantly since its introduction, and today, researchers are trying to replicate tissue by creating scaffolds using synthetic polymers that degrade in the body. There are different methods that can be used to manufacture these scaffolds, and the two that are of interest to this study are electrospinning and solvent casting. The background work concludes that although biologists and materials scientists have discovered useful potential products and ranges for creating scaffolds, they lack understanding of the underlying effects of variables and need a reliable way of manufacturing these products. The use of designed experiments and a control systems approach should be able to shed light on to some of the specified issues.

Electrospinning lacks understanding of the underlying effects of its variables and solvent casting is not repeatable enough to create the same scaffold multiple times. Both systems are multivariable, and the suitability of introducing control is investigated in order to improve repeatability and enhance understanding. These methods have never been considered from a control systems perspective with the intent to apply data based and empirical modelling techniques. The full understanding of these manufacturing methods is vital for improvements in product quality and repeatability.

From the various control systems methods that have been analysed and the model based prediction methods presented, MPC is assessed as the most suitable candidate for implementing control to the electrospinning process. The main reason for this is that the system is multivariable, and correlations exist amongst the input variables as well as soft constraints. The predictive capabilities of a model will be important when trying to control or calculate operating conditions for producing fibres of a desired structure, strength and alignment. Local PID controllers would be suitable for controlling the speed of the rotating axle, the voltage of the power supply and other actuators that could be introduced into the system in the future.

The solvent casting method would benefit from a regression based method such as PLS, as the quality of the final product is not known until the end of the run. None of the time series modelling methods would be suitable for this process due to the lack of output measurements, unless a soft-sensor is used to predict the output variables. The model could be used to calculate the steady state conditions that the evaporation needs to take place, or to provide monitoring and control to the process if deemed necessary.
Potential models of the processes can be identified using data collected during specific tests. The approach known as DOE can be used to define these tests and this approach was described in this chapter. Examples of factorial studies performed on the processes investigated in the work described in this thesis are presented in Chapters 3 and 4.

In this chapter, various modelling methods have been presented with a view to analyse both of the manufacturing processes for control systems implementation purposes. From the analysis, the ease of use of the PLS approach and the power of MPC stand out, and there are recent publications with a view to use PLS in MPC applications (Laurí et al., 2010), (Rossiter, 2010). In many publications the need for PLS is justified by the presence of highly correlated, ill conditioned data matrices. However, PLS is not always reliable as the sole modelling tool as a bias is always present, which will be shown in the later chapters of this thesis. The work of Zhu (Zhu, 2001) and Sandoz (1972) identifies the bias present in the least squares algorithm, which also carries through to the PLS method with less effect due to the reduced variable space. The output error method, prediction error method and recursive parameter estimation methods are recommended to overcome this bias. The use of the unbiased version of RLS for an MPC application has been shown to reduce the bias (Awad, 2008). The OE and PEM methods however do not perform well when the data is highly correlated, which highlights the need for an unbiased method that is suitable for MPC applications, and that can also cope with highly correlated data. Additionally, in Chapter 5, novel unbiased hybrid PLS methods are proposed.
3 Solvent Casting

This chapter presents the work undertaken towards improving the solvent casting technique used by the Biomaterials Research Group (BRG). Solvent casting is an evaporation based process that is used for making thin films. The films are created by dispensing a required quantity of polymer-solvent mixture on to a glass substrate and letting this completely evaporate. Following this, the films are rolled up in to a tube to form a conduit used for healing nerve injuries. This is currently the preferred method used by the BRG to manufacture artificial nerve conduits.

3.1 Properties of Solvent Cast Films and Manufacturing Issues

3.1.1 Solvent Cast Films

As explained in Chapter 2, the solvent casting method allows rapid manufacturing of thin films for use as nerve conduits. The capability to accommodate cells is one of the most important features of the nerve conduit, where the conduit needs to provide mechanical support as well as the correct environment for neural regeneration. The existence of motor neuron cells and Schwann cells are two of the main components in forming a growth promoting environment for the nerve cells. For the cells to attach to the created films, a specific surface morphology is desired. It has been shown that cells prefer attaching to rough surfaces, which can be obtained by creating pits and pores on the surface (Dalby et al., 2007), (Sun et al., 2010).

Creating a surface with pits and pores is possible using the solvent casting method. The solvent cast films have two surfaces; one is the surface that is in contact with the glass substrate, which the solvent is originally dispensed on to and the other is the surface that is in contact with the environment. From these two surfaces, the side that is in contact with the environment is the pitted surface and the side that is in contact with the glass substrate is the smoother side.

The pits and pores on the surface are created by the evaporated solvent. Once the solvent mixture is dispensed on to the glass substrate, it immediately begins to evaporate and different concentrations of polymer solvent mixtures are generated at different locations of the substrate. In some areas, the polymer molecules are more heavily concentrated than others. These different concentrations give rise to the pits and pores observed on the surface of the films.

The formation of a film is presented in Figure 3.1, with the photos taken at various times during the formation of the film. Horizontally, the first three pictures were taken at ten second intervals, the following three at thirty seconds and the remainder at two minute intervals. This shows that the solvent evaporation is a fast process and the surface structure is almost pre determined depending on the local concentration levels of the solvent and the internal flow within the liquid sample. However, analysis of existing studies presented in Chapter 2, indicate that environmental factors have an effect
on solvent evaporation. Although the polymers and solvents used in this study reported in Chapter 2 vary from other research publications, the environmental analysis conducted in this research study is expected to be valid for similar non-aqueous solutions as they will be affected by the environment in a similar way.

![Figure 3.1; Time Lapse Picture of Solvent Casting](image)

One of the other main desired properties of an ideal nerve conduit is that it should be permeable enough to allow nutrients from the body to reach the injury site, as well as provide mechanical support. Permeability and stiffness are both related to the intrinsic properties of the materials the conduit is manufactured from. The PCL/PLA blend, discussed in Chapter 2 gives satisfactory permeability and is strong and flexible enough for longer nerve injuries, or where there could be a bend in the injury site (Sun et al., 2010), potentially minimising local irritation. The PCL/PLA blended films are proposed by the BRG to be used as nerve conduits due to their superior mechanical and permeability properties over single polymer films. However, the current method of manufacture does not produce films with consistent properties. The problems with the current manufacturing methods and the proposed solutions are presented in the following section.
3.1.2 Manufacturing Issues

Prior to this project, solvent casting in the BRG took place on the laboratory bench top or under a fume hood. The laboratory conditions can vary in temperature, humidity and even air speed when a person walks past or works nearby. Some of the films produced previously were identified as suitable platforms to grow nerve cells on, but were not always reproducible as the conditions were unknown. The problems encountered can be summarised as follows:

- The manufacturing technique was unreliable. Variation in the environment affects the final product, which varies from day to day.
- There was a lack of sufficient data, caused by the absence of data logging. This means that the conditions required to produce high quality films were unknown.
- There was no measure of film quality. The lack of a quality variable to assess the film properties in a quantitative manner meant that there was no meaningful way to compare the films that were made under different conditions.
- Although cells have been reported to attach to pitted films, no work had taken place to determine what pit sizes they may prefer, if any, and under what conditions could these films be produced.
- The surface morphology of the films, the pits and pores can affect its mechanical properties; the full effect of this has not been investigated.

To eliminate the identified problems it was proposed that the films were produced under the following conditions:

- The films should be created under a known controlled environment where the temperature of the solution, the syringe and the glass substrate are all constant.
- The humidity should be known and regulated.
- The airflow should be at a known rate and regulated.

The main aims of the solvent casting study from this project’s perspective were:

- Introduce concepts of data logging and repeatable manufacturing conditions.
- Confirm the effects of environmental conditions on the films.
- Produce films with known diameters where the pit and pore sizes are repeatable.
- Identify a suitable pit diameter where successful cell growth could be observed.

The first step in this study was to determine a way to control the environment. Initially, the design and build of an environmental chamber was considered; the basic working principle would be to use a humidifier, a heater, and a cooler with refrigerant to condition the air with feedback control using
temperature and humidity sensors. The conditioned air would then be moved to the actual workspace where solvent casting will take place. Additionally, the purchase or custom manufacture of an environmental chamber was taken in to consideration. Following a thorough justification study, the decision to purchase an environmental chamber was made, where the required specifications were identified as:

- Computer programmable temperature, humidity and air flow control.
- Computer link to log humidity and temperature from the built in sensors of the chamber and other external sensors.
- Suitable size to fit the stirrer, the syringe and have ample room for the movement of hands for dispensing.
- Having an inner door to keep the pre-set environment in the chamber when the main door is opened to dispense the solvent.

Following the creation of the specification and a review of the available products, the MMM CLIMACELL 111 environmental chamber was purchased. However, the chamber required modifications as it was not suitable for solvent casting in its standard configuration. The modifications are explained in the following section.

### 3.2 Development of the Experimental Setup

The MMM CLIMACELL 111 appliance is intended for use as an incubator in pharmaceutical laboratories and meets the desired specifications determined by this project. To make the chamber useable for solvent casting, a number of modifications were implemented. These can be summarised as:

- Design and build of a suitable height stand, so the user can insert their hands in to the chamber whilst standing upright.
- Removal of the inner glass door, and replacement of this with an in house made Perspex door containing glove ports. This is to prevent opening of the door and disturbing the stabilised internal environment when solvent casting.
- Wiring a data logger with humidity and temperature sensors.
- Providing a way to evacuate the solvent once the process is finished.
Figure 3.2; Inner Door and Glove Ports

Figure 3.3; CLIMACELL Chamber
The final chamber is shown in the images presented in Figure 3.2 and Figure 3.3. The air extraction port was also used for the routing of the cables for the data logger and the USB communication for the scales used for the solvent evaporation study. The air was removed using an air extractor connected to a valve with chemically resistant tubing. The completion of the modifications meant that the planning of the experimental work could begin. These are explained in the following section.

### 3.3 Experimental Methods & Results

Before conducting any experiments, the initial experimental approach needed to be outlined. Referring back to the identified problems, a measure of quality or a comparable property of the produced film needed to be determined. Since the ultimate aim of this investigation is to find suitable surface morphologies to grow nerve cells successfully, the following reverse approach was proposed for determining variables that should be considered for measuring.

- The surface area of the film covered by cells.
- Cell proliferation.
- Long term health of the cells, i.e. biocompatibility.
- Nerve cell conductivity.
- The diameters of the pits and pores on the surface of the film.
- The area of the film covered with pits and pores.
- Mechanical properties of the film, such as tensile strength.
- Temperature.
- Humidity.
- Airflow.
- Solvent – Polymer mixture levels.
- Density.
- Thermodynamic properties.
- Heat transfer properties.

The proposed measurements above are useful when comparing films and when measuring the degree of success in terms of cell growth. If viewed from a modelling perspective, the process variables would be the input, or cause variables and the film surface properties would be the output, or response variables that are only available once the film has been manufactured, much like typical batch processes. Finally, the cell success measurements would ultimately be the quality variables. The data would help to identify the parameters that can be used to create films of desired pit and pore sizes, which will then be related to cell attachment success.

The main aim of the outlining and planning exercise was to minimise the number of tests conducted in the limited time frame of this project. For this, a Design of Experiments (DOE) approach was used. In experimental design, rather than changing one variable at a time through its complete operating
range, whilst keeping the others constant and observing a response, many variables are changed simultaneously at selected key operating ranges. This results in a much reduced number of experiments that need to be conducted, when compared with the traditional approach of changing one variable at a time. Utilising existing knowledge and the results from a previously conducted pilot study by the BRG (Mobasseri, 2012), a working range of steady state conditions, where films that could be potentially usable for cell studies would be created, was identified.

Before starting actual experimentation, the environmental chamber was tested at one of the identified test conditions of 35°C 80% RH to observe the response of the controller when the temperature was stepped down to 25°C. A short, closed loop step test was applied to the temperature set-point (black line) and the temperature response (red line) was observed, shown in Figure 3.4. The purple and blue lines represent the humidity setpoint and actual humidity values respectively. The response shows the temperature settles with a first order response over a 20 minute period. This also had an effect on humidity (blue line) which dropped until the temperature became close to its setpoint and then was ramped up by the controller to the required value.

![Figure 3.4; Environmental Chamber Controller Inverse Response 35°C 80% RH](image)

Initial test runs of the chamber showed that trying to change internal conditions such as ramping the temperature or the humidity caused a large disturbance in the airflow, which is known to generate undesirable surface morphologies from the previous fume hood studies. Additionally, the temperature and humidity response times are slow when compared with the rapid evaporation of the films, which is presented later. These findings suggested that the tests need to be conducted under steady state conditions and that the introduction of a GBN or PRBS like excitation signal would be unsuitable. The methods outlined in the following section describe the tests and results in more detail.
3.3.1 Temperature and Humidity Test

Using *a priori* knowledge from the pilot study (Mobasseri, 2012), the minimum working temperature was specified to be 15°C and the maximum 35°C. The minimum and maximum humidity levels were selected as 20% RH and 80% RH respectively. In addition to these, a medium test condition was also selected at 25°C and 50% RH. The relative humidity levels were selected so that the water content of each test was different from the others. The water content values are presented in Table 3.1, where the water holding capacity of air increases with temperature. For this test a small factorial study was undertaken and the test conditions for the nine experiments are presented in Figure 3.5.

<table>
<thead>
<tr>
<th>Environmental Conditions</th>
<th>Water Content (gr of water per kg of air)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Temperature</td>
<td>Humidity</td>
</tr>
<tr>
<td>15°C</td>
<td>20% RH</td>
</tr>
<tr>
<td>15°C</td>
<td>50% RH</td>
</tr>
<tr>
<td>15°C</td>
<td>80% RH</td>
</tr>
<tr>
<td>25°C</td>
<td>20% RH</td>
</tr>
<tr>
<td>25°C</td>
<td>50% RH</td>
</tr>
<tr>
<td>25°C</td>
<td>80% RH</td>
</tr>
<tr>
<td>35°C</td>
<td>20% RH</td>
</tr>
<tr>
<td>35°C</td>
<td>50% RH</td>
</tr>
<tr>
<td>35°C</td>
<td>80% RH</td>
</tr>
</tbody>
</table>

**Table 3.1; Water Content of Applied Tests**

The main aim of this test was to create films with constant pore sizes and to establish the effects of temperature and humidity settings on the film surfaces. The films produced under the nine different conditions were studied using an SEM and the captured images were analysed with an image and pattern recognition software called ImageJ. This software automatically detected the pits and pores and calculated their individual areas. For this test, three films were produced for each batch, creating...
27 films. Three tests were used as this is common practice when implementing DOE (Esbensen et al., 2002) and having multiple samples is beneficial to determine repeatability.

The presented numerical results in the following sections are the mean values of the film properties from the overall batches for respective temperature and humidity levels. The next series of images shown in Figure 3.6, Figure 3.7 and Figure 3.8 present a selection of results from these tests for the 15°C condition. Only one section of the film is shown, as with the high level of magnification used the whole film is difficult to present. In these images the effect of humidity can be seen. The low humidity at 20% RH for 15°C did not create any pits or pores and the film was smooth, but with the increasing humidity pits and pores are formed on the surfaces. The 15°C 50% RH film has a considerable number of small pits and some larger ones. The 15°C 80% RH film has more of the larger pits.

As expected, the higher humidity levels aid in the creation of pits and pores. However, a range of pits and pores are created rather than a consistent size, and where the larger holes are created the vein-like structure was observed in Figure 3.1. The larger holes are formed around areas where higher solvent concentrations are present, which evaporates quicker therefore drawing more condensation, hence creating larger holes.

![Figure 3.6; 15°C - 20% RH PCL-PLA (Mobasseri, 2012)](image-url)
Figure 3.7; 15°C - 50% RH PCL-PLA (Mobasseri, 2012)

Figure 3.8; 15°C - 80% RH PCL-PLA (Mobasseri, 2012)

The better film in this instance is the one shown Figure 3.7, as Figure 3.8 has more pores than pits, which is a disadvantage as cells can fall through large pores. The difference between pits and pores is shown in Figure 3.9. The two large holes to the right in Figure 3.9 are pores which are not desirable for cells. The two large shapes to the left in the image are pores and all the smaller diameter holes are pits, and not through-holes like pores.
The next analysis to be undertaken was comparison of pit diameters for the batches of films; the results are shown in Table 3.2. As observed from the images there is a large variation caused by a wide range of diameters, which reflect on to the results presented in Table 3.2 where the standard deviations are large. All the films that were produced at 20% relative humidity are smooth and no results are available. From this it can be concluded that, for the current ratio of the PCL/PLA blend solution, using a water content below 10 gr/kg of air does not create any pits or pores and this should be set as a non working range. It can be observed that as the humidity increases the mean pit area and the range of size of the pits created increases. Temperature appears not to have a significant effect on the pit diameters in comparison to humidity.

<table>
<thead>
<tr>
<th>Environmental Conditions</th>
<th>Water Content (gr of water per kg of air)</th>
<th>Mean Pit Area (µm²)</th>
<th>Standard Deviation (µm²)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Temperature</td>
<td>Humidity</td>
<td></td>
<td></td>
</tr>
<tr>
<td>15°C</td>
<td>50% RH</td>
<td>4.9</td>
<td>15</td>
</tr>
<tr>
<td>15°C</td>
<td>80% RH</td>
<td>7.9</td>
<td>22</td>
</tr>
<tr>
<td>25°C</td>
<td>50% RH</td>
<td>10.2</td>
<td>13</td>
</tr>
<tr>
<td>25°C</td>
<td>80% RH</td>
<td>16.3</td>
<td>30</td>
</tr>
<tr>
<td>35°C</td>
<td>50% RH</td>
<td>19.2</td>
<td>16</td>
</tr>
<tr>
<td>35°C</td>
<td>80% RH</td>
<td>29.2</td>
<td>82</td>
</tr>
</tbody>
</table>

**Table 3.2; Pit Diameter Study Results**
The histogram in Figure 3.10 presents the percentage distributions of pit areas on the created films. Confirming the results in Table 3.2, with the increase in humidity more pits with larger diameters are formed. The largest pits are in the 35°C temperature range. This is a result of the increased air temperature, as with increasing air temperature the air’s water holding capacity increases. As the solvent evaporates from the film it creates colder spots and the moisture in the air condensates around these points. When there is more moisture in the air the water condensates as larger droplets, creating larger pits and pores.

Although one of the aims was to create films with similar pit sizes at different environmental settings, from the results it is clear that only using the environmental factors as the manipulated variables is not sufficient. Even with minimal airflow and controlled environmental conditions, pit formation is highly variable. However, the wide range of pit sizes can be considered as an advantage, as it provides an ideal platform to test cell growth. Rather than making many films with different diameter pits, one film which has a wide range of diameters would allow for rapid testing of the cell growth. In addition having a range of pits and pores would determine if the cells prefer a certain size or a range of various size pits to grow on. This would be evident in the way they populate the film, which is presented in the cell attachment study under section 3.3.3.

![Figure 3.10; Pit Area Distributions (Mobasseri, 2012)](image)

In Figure 3.10 the first of the four presented samples are the 25°C at 50% RH condition, shown in green, where the smoother side of the film that is facing the glass substrate is analysed. This gives the highest distribution in the 5 µm² range, and there are smaller diameter pits in the larger ranges. The second sample, shown in red, is the rougher side of the film created using the 25°C at 50% RH condition. This is the side that faced the humid air and as expected has a higher distribution of pits.
and pores in the larger ranges. The third condition, shown in blue, is the 25°C at 80% RH where larger diameter pits and pores would be expected, and most of the pits are formed around the 50 µm² area range which is higher than for the films produced under the lower humid condition. The last condition presented, shown in purple, is the 35°C at 80% RH film. As with the other films, this film has pits and pores in all of the ranges, but has a significantly higher distribution in the higher 100 µm², 500 µm², 1000 µm² and 5000 µm² ranges. This confirms once again that as the humidity increases the number of samples with larger pits and pores also increase.

Figure 3.10 shows that the distribution of the pit sizes does not follow a Gaussian distribution. This is caused by the batch to batch variations and the inconsistency caused by only relying on the environmental variables to regulate the formation of the pits and pores. This variability suggests solutions such as the use of master substrates with a predetermined orientation consisting of pillars or grooves. Although for this the required diameters of the pillars or the grooves will still need to be established, as this may influence cell growth. The following section presents the effect of the environmental conditions on solvent evaporation.

### 3.3.2 Evaporation Rate Study

This test was conducted in order to determine the evaporation profile of the solvent and to see how temperature and humidity affected this. All the tests in this section were carried out using the same initial procedure. Two biodegradable polymers (PCL and PLA) were mixed by a set ratio and dissolved in a predetermined amount of dichloromethane, which mixed on a magnetic stirrer for 24 hours. Before solvent casting, the dispensing pipette, the solution and the glass substrates were placed in the environmental chamber for at least an hour to bring everything to the same temperature.

The first test was to log the solvent evaporation rates at different conditions. The evaporation rate was measured by monitoring the weight of the sample until the solvent completely evaporated. This was achieved by using an Ohaus SP123 digital balance with 0.0001 grams sensitivity with a USB output to the computer. The temperature and humidity levels were logged using the same computer.
The results of this test are presented in Figure 3.11. In this figure the 0 point indicates the unloaded scale, the sharp increase is the dispensing action, and the negative slope is the reducing weight of the solution due to evaporation. The effect of the trialled conditions on the evaporation rate can be seen in Table 3.3, where the sample weight is presented throughout the evaporation process. The slowest evaporation took place during the 15°C condition, and the difference between 25°C and 35°C can be observed in the gradient of the lines. From these results it can be determined that the changes in temperature have more of an effect on the evaporation rate compared with changes in humidity. As the solvent Dichloromethane is non-aqueous (not water based), humidity does not have a significant effect on the evaporation.

<table>
<thead>
<tr>
<th>Environmental Condition</th>
<th>Evaporation Rate ( gr/sec )</th>
</tr>
</thead>
<tbody>
<tr>
<td>15°C 20% RH</td>
<td>0.0033</td>
</tr>
<tr>
<td>25°C 50% RH</td>
<td>0.0050</td>
</tr>
<tr>
<td>25°C 50% RH</td>
<td>0.0049</td>
</tr>
<tr>
<td>25°C 80% RH</td>
<td>0.0047</td>
</tr>
<tr>
<td>35°C 50% RH</td>
<td>0.0058</td>
</tr>
<tr>
<td>35°C 50% RH</td>
<td>0.0061</td>
</tr>
<tr>
<td>35°C 80% RH</td>
<td>0.0065</td>
</tr>
<tr>
<td>35°C 80% RH</td>
<td>0.0066</td>
</tr>
</tbody>
</table>

Table 3.3; Evaporation Test Results
This test shows that monitoring the evaporation rate during film production could help identify the effects of the evaporation rate on the surface morphology and mechanical properties of the film. Changing the temperature within the chamber during the production of a film is not suitable as the resulting high airflow will have an impact on the resulting quality of the film. However, if it was found necessary to change the temperature during the formation of the film, the glass substrate could be placed on a heat controlled, thin metal block. Rapid changes to temperature of this block would result in varying evaporation rates. This would in turn affect the surface morphology and the mechanical properties of the film.

This test highlighted a further problem. Although the pipette is set to load 600 μl of liquid, there was found to be a weight difference in each sample, with the mean percentage error 14.02%. This suggests that the pipette was not fully accurate and introduces a slight variability into the samples. This variation will not affect the evaporation rate but will determine how much polymer is leftover in the final film, thus affecting the film thickness and its mechanical properties. Additionally the internal flow of the liquid once it is dispensed onto the glass substrate, varies depending on how fast the user pushes the pipette plunger. These inconsistencies are not desirable if repeatability and up-scaling is considered, an automatic dispensing mechanism would overcome this problem where equal volumes will be delivered at the same flow rate to the same location of the substrate.

Although the mechanical properties are vital, the ultimate aim is attaching cells to the films. The following study presents the investigation conducted to determine the sizes of the pits the nerve cells prefer.

3.3.3 Cell Attachment Analysis

The next study was the cell attachment study to investigate if the cells have a preferred surface morphology and how they attach. This was undertaken in collaboration with other members of the BRG, where nerve cells were grown and analysed on the substrates created in the earlier experiments. The success of the cells were compared to those grown on tissue culture plastic which is considered as the benchmark for cell growth. Tissue culture plastic is an ideal film that has extracellular matrix coated on its surface. Cells naturally attach themselves to an extracellular matrix in real tissue, which are made from fibrous proteins and polysaccharides. The films that were selected from the 27 batches for this test and their corresponding average pit diameters are shown in Table 3.4. These films were selected as they cover a wide range of pit sizes to test the cells on.
Table 3.4; Cell Test Samples

<table>
<thead>
<tr>
<th>Film</th>
<th>Mean Pit Diameter (µm)</th>
</tr>
</thead>
<tbody>
<tr>
<td>15°C at 20% RH</td>
<td>0</td>
</tr>
<tr>
<td>25°C at 50% RH (Substrate Side)</td>
<td>3</td>
</tr>
<tr>
<td>25°C at 50% RH</td>
<td>13</td>
</tr>
<tr>
<td>25°C at 80% RH</td>
<td>30</td>
</tr>
<tr>
<td>35°C at 80% RH</td>
<td>82</td>
</tr>
<tr>
<td>Tissue Culture Plastic</td>
<td>Not Applicable</td>
</tr>
</tbody>
</table>

For this study NG108-15 nerve cells were used. The cells were planted to the selected PCL/PLA films at 5000 cells/ml density per sample. The cell media was replenished at days 3, 5, and 10, and SEM images of the cells were taken during these intervals. In addition, 100 µl Alamar blue dye solution was added to each well plate where the cells are kept. After a 2.5 hour incubation time the measurement of the fluorescent emissions were taken using a plate reader at 530-510 nm wavelength excitation. The Alamar blue method is commonly used for cell essays by biologists. The fluorescent emissions help assess cells’ metabolic activity, which in this study was found to be constant on all the samples. In addition to SEM imaging for assessing how the cells spread, a technique called cytoskeleton staining was used to visualise the cell nuclei and the cell body. The cytoskeleton can be described as the internals of the cell that is contained within the cell body.
The presented images in Figure 3.12, show the cytoskeleton staining result of the 15°C at 20% RH smooth film. The images were taken 3, 5 and 10 days after cell implantation. The images show that cells (blue and green shapes) are alive and healthy, but have not spread around on the film. This is not ideal as the nerve conduit should be able to accommodate cells and promote cell growth. The cytoskeleton images only show the morphology of the cells and do not represent their spread accurately. For this, normal SEM images are used which are presented in the following sections.

The following image in Figure 3.13 shows the SEM image of the film with 13 µm mean diameter pits, created under 25°C at 50% RH conditions. The lighter gray clusters with white boundaries are the cells, and the black circles are the pits and pores on the surface of the film. In comparison to the previous image, the cells appear to have spread significantly more from the first sample on Day 3 and have continued to spread over the forthcoming days, almost covering the whole film.
These time lapse images are useful in terms of observing how the cells spread, and show that the cells prefer to attach to certain sized pits. They, then spread along certain routes on the film. These routes follow the location of the pits with the diameters they prefer. For example, the images for days three and five show that the cells spread quickly across the areas where the smaller sized pits exist. The tenth day growth images are of most importance as the conduit would remain in the body for long periods of time and eventually degrade. Therefore its behaviour after several days is important to analyse. The tenth day growth images are presented in Figure 3.14 for the best two conditions alongside the tissue culture plastic. The film with 82 µm mean pit diameters shows strong cell growth. However, it must be noted that cells do not populate the whole of the film. When analysed closer, as speculated earlier, it is evident that the cells prefer a certain size pit, for example the cells appear more concentrated in the areas where small pits are located and avoid the areas where large pits or pores exist.
A bar chart representing the cells attachment preference is presented in Figure 3.15. The cells prefer to attach to the edges of the pits at roughly 10 µm to 30 µm diameter ranges.
The findings of this test can be summarised as:

- Cells preferred to attach to the edges of 10 µm to 30 µm diameter pits, then spread further on the film.
- In comparison to the smooth films, as the surface got rougher with the films created under higher humidities, strong cell distributions were observed.

### 3.3.4 Discussion

For the ideal nerve conduit, in addition to healthy cells, another requirement is the alignment of the cells. The method of only using the environmental variables is unsuitable for creating consistent films with a set range of diameters before beginning to consider cell alignment. The batch to batch variations in the film morphologies found in this study occur due to the following reasons:

- Following dispensing, the evaporation process starts immediately. Whilst the solvent is evaporating it is still flowing around on the substrate. Different areas of solvent concentration form and these evaporate further, forming small pits. The larger pits and pores are a product of humidity and water condensation. It is impossible to control where the water will condensate and in what quantities. While the film is drying it starts taking on a more rugged shape, causing water droplets to move around the film.
- As pointed out in the evaporation rate study, the dispensing method is not reliable and each film varies in terms of the volume of solution used for making it. This means that each time, a different amount of solvent and polymer mixture is used, making the films different in terms of mechanical properties. An automated accurate dispenser will provide more repeatable dispensing to eliminate this error, with more consistent internal liquid flow dynamics.
- It is known that air flow is necessary to keep the cabinet environment at constant conditions. Even at its lowest level, which was not possible to measure due to the lack of an accurate enough air flow meter, light turbulences of air affect evaporation. Using a temperature controlled metal plate would be more suitable for this study rather than using the air temperature to manipulate solvent evaporation.

Analysing the points above with the aim of providing films that will accommodate healthy cells, the need for a more reliable method of creating films is apparent. A major outcome of this study was that following the identification of the preferred pit size for the cells, the use of a custom made aligned grooved substrate was proposed as a novel form of solvent casting (Mobasser, 2012). This would provide two benefits, one is that the substrate could provide a consistent 20 µm gap, and secondly the grooves are already aligned forcing the cells to grow in an aligned manner, which is the same direction as the two injury points. Based on the outcomes of this study, conditions were set to manufacture these films in a consistent manner, such that batch to batch variability would be minimal.
and cell growth maximised. The proposed substrate is shown in the representative diagram in Figure 3.16 where the cross sectional representative diagram of the custom substrate. Using the new method the side that cells are planted on becomes the side that faces the substrate. The pits and pores created on the side facing the environment, only influence the mechanical properties of the film. From the evaporation study it was found that evaporating films slowly resulted in a more homogenous structure. The initial batch of films manufactured with this method used the 15°C at 20% RH condition with a low water content which did not create any pits or pores on the outside. However, other conditions still need to be evaluated and their material properties tested.

![Figure 3.16; Parallel Grooved Substrate](image)

Using this master grooved substrate method of film production, the effects of humidity is minimised. Unless a high water content is present, pores do not form in the films and a lower water content creates pits, which have minimal effect on the side of the film that is facing the substrate. The temperature can be used to speed up the evaporation.

![Figure 3.17; Overview of Nerve Repair using Conduit with Aligned Grooves](image)

The cells grown on the grooved films, illustrated in Figure 3.17, have better cell responses, proving this method is more suitable for creating films that can be used as nerve conduits. Images of these
are not available at this current time but the initial cell assessments show alignment and healthy cell growth (Mobasseri, 2012).

### 3.3.5 Summary

This section presented the work accomplished on creating nerve conduits using the solvent casting method. The main challenge with this project was creating repeatable films whilst working in parallel to determine the pit sizes the nerve cells preferred. The investigation of the effects of environmental conditions on the evaporation of the polymer blend and the solvent mixture concluded that temperature had a more significant effect on the evaporation rate than humidity, and humidity had more of an effect on the surface morphology. Using low temperatures and a slower evaporation rate created films with more homogenous surfaces. An additional study needs to be conducted on the effects of the environmental conditions on the material properties of the final conduit.

The environmental chamber, with the implemented modifications enabled repeatable environmental conditions and air flow, which helped to determine a range of pit sizes for successful cell growth. A short term maximum and minimum parameter test was conducted, and a small factorial study was undertaken in a predetermined range. Although the creation of films with the same size pits in a repeatable manner was not possible with this method, the wide range of pits created on the individual films led to more rapid testing. The results showed that for pit sizes of approximately 10 µm to 30 µm, the nerve cells were most likely to grow and develop, thus promoting neural regeneration. Following these findings, custom substrates were outsourced to a specialist fabrication company to be manufactured with aligned grooves. The latest results suggest that by using this method with aligned grooves, strong networks of nerve cells can be grown.

This research study could potentially be improved with the use of MSPC methods as control of it using the environmental method is not possible due to the slow dynamics of the actuators and the fast evaporation of the solvent. However, MSPC methods can be used to improve the product quality. In addition a temperature controlled hot plate together with the environmental chamber or a pressure controlled vessel could be utilised in the future. With this configuration, rapid solution property tests could be undertaken and included in the model. At the same time the evaporation rate could be measured, and at the end of the batch the final film material properties such as tensile strength and elasticity can be used as the output variables. Additionally, an automatic solution dispensing system would need to be implemented to eliminate the problems identified during the evaporation test, where the dispensed volume using a manual pipette is not always consistent. The formulation of a data matrix where excitation was applied to the cause variables would enable predictive models to be identified, alongside helping to identify optimal operation parameters and film properties. A batch like system example is presented in Chapter 6 of this thesis, where the structure of the simulation is similar to the solvent casting process where a measurement of output quality is only available at the
end of the batch. The modelling methods applied to this system should be applicable to the solvent casting process when the necessary data is made available.
4 Electrospinning

This chapter presents the work conducted on the electrospinning process, which aims to discover how the process variables affect the final product and how they can be manipulated to control the properties of the produced material. The work consists of mechanical, electrical and software design, small scale design of experiments exercises to reduce test times, SEM analysis and Finite Element Analysis (FEA). The first section briefly explains the development of the experimental setup which is then followed by the experimental sections and summary.

The application of electrospinning is a relatively new method within the biomaterials manufacturing field. It is used to produce fibres which fall into the nano and micro scale, that are of interest for cell growth and support when engineering artificial tissue. The process itself involves applying high voltages in the 5000 Volts – 30000 Volts range to electrically charge a polymer solvent solution. The charged liquid is then forced out in fine fibres to the nearest target with the lowest potential, such as a metal plate connected to ground.

The main aims of this work were to:

- Fully understand the process and find out which factors affect fibre diameters, their alignment and the material properties.

- Identify the conditions necessary to create fibres of constant or varying diameters in a controlled and repeatable manner as a continuous process. This will aid the manufacturing of layered surfaces with different morphologies, for example muscle to bone adapter tissues.

From the beginning of the project, as briefly discussed in Chapter 1, it was identified that the existing electrospinning equipment was not capable of producing repeatable results. Electrospinning for biomaterials applications is globally ad-hoc, with no thorough analysis of how consistent products should be produced. Similarly to the solvent casting method, electrospinning at the BRG took place under a fume hood without environmental control, and day to day temperature and humidity changes affected the process. Furthermore, the high voltage wiring was unsuitable for use with the required voltage ranges and created unwanted disturbances, such as the fibres landing only to one side of the collector. The rotating axle of the collector also had visible shaft whirl even at low speeds, acting as a disturbance on the alignment of the fibres. Conditions similar to these appear to be standard across different research groups, where the equipment is either not suitable or the experimental methods neglect some of the important variables.

There are various publications presenting attempts at discovering electrospinning parameters in research institutions around the world. De Vrieze et al. (2009) present their work on analysing the
effects of the environment on the fibre diameters, where the solutions used to dissolve the polymers
are both non-aqueous. They use a closed system with a fan in a temperature controlled laboratory
and salt baths to influence the humidity, where the tested environmental conditions are 10°C, 20°C
and 30°C at 20%, 30% and 50% RH. In their study there is no record of temperature and humidity
and how it changes during the process, as the fan to circulate the air is switched off. The results
reported by De Vrieze et al. (2009) are not repeatedly tested and a measure of variance has not been
given. The fibre diameter findings between the conditions are small and if repeated they may vary
and contradict the first findings. Although the environmental effects are limited, this study still lacks
full environmental stability.

Jeun et al. (2007) conducted a similar test in a closed cabinet, without any environmental control and
without any repeat testing. They utilised a humidifier to add water vapour to the air without any
temperature measurements. The amount of water that can be held by the air strongly depends on
temperature, so the results presented in this study are of limited value. Additionally, the temperature
also affects solution viscosity, and since there is no record of this, it further undermines the
consistency of their findings.

You et al. (2006) presented their investigation into the effect of solution properties on electrospun
nanofibres, where the solution viscosity was altered using different polymer-solvent concentrations.
However, none of the experiments have taken place in a controlled environment. Similar to other
research examples, this again compromises the value of their findings.

Deitzel et al. (2001) present a comprehensive study analysing the effects of certain electrospinning
processing variables, such as voltage, solution concentration and viscosity. They concluded that
increasing solution concentration increased fibre diameter. However, no significant effects on the
diameters were observed with the changing voltage and instead a phenomenon called beading was
reported. Beading was observed at voltages higher than 9 kV and became more intense with the
increase in voltage. The study used a gravity fed pump which did not guarantee a constant flow rate
of the solution, which could lead to unreliable results.

Beading is a commonly reported problem that is related to the chosen solvent and the voltage used
during electrospinning, and was also observed in the original electrospinning setup at the BRG. At the
beginning of this project the solvent used by the Biomaterials Research Group was changed from
Acetone to hexa-iso-fluoropropanol (HFIP) and the beading problem did not re-occur. Acetone and
HFIP have different conductive properties, Acetone has a conductivity of 0.02 µS (micro-Siemens) at
18°C and HFIP with a non conductive polymer was measured to be 0.22 µS at 18°C. From this it can
be suggested that beading occurs with poor conducting polymers at high voltages when the flow is
unchanged, the electrostatic forces cause the electrospinning jet to stutter and to start *spraying* rather than continually drawing fibres.

None of the studies mentioned above, or other similar studies, such as the works of Milleret et al. (2011), Zong et al. (2002), Boland et al. (2005) that attempt to analyse or compare differences between electrospinning variables, consider the high correlations amongst the variables. In all of the cases at least one of the variables is not controlled, or not measured. None of the publications have analysed the process from a control systems perspective or proposed solutions to controlling fibre diameter whilst the process is running. This is far from straight forward, as the electrospinning process is highly complex. For example, changing the solution properties by adding more polymer is not possible as the polymer and the solvent need to be mixed until homogenous for at least 24 hours. Higher voltages mean increased electric force is induced on the fibres, which stretches them more and influences the fibre diameters. However increasing the electric force induces more instability to the jet and actually controlling it becomes difficult.

There has been no previous research that has applied computer-based feedback control to regulate the electrospinning process. Without control, keeping the variables constant at all times is not always guaranteed. For example, if spinning for a sustained period, an insulating layer of polymer can build up on the collector, affecting the electric field. Ideally this needs to be detected and the electric force adjusted accordingly.

Following the assessment of the existing equipment and the incomplete attempts presented in existing literature, the specifications for the electrospinning research equipment were determined as:

- The control of the environment as temperature and humidity are contributors to the evaporation rate of the solvent. The temperature also influences the solution properties such as viscosity and many tests do not take these into consideration.
- Minimal electrical disturbances from wires and other equipment must be ensured. This is necessary, as on the previous set up at the BRG the fibres would land significantly biased to one side of the collector and sometimes the high voltage would spark to the syringe pump or the fume hood wall, causing disturbances in the electric field that affected the spinning process.
- Data logging of process variables is required as it is important to know the exact conditions and settings used in the process. Recording as many variables as possible can help in highlighting correlations and calculating the effects the variables have on the final product.
- Computer control of voltage, collector rotation speed and syringe pump is necessary, as these are the main variables of the process. From a control systems perspective, being able to control and change these values without stopping the process, enables the potential
application of step tests and other identification signals. This also provides the potential to bring the system under closed loop control, where a model of the process would set the variable values to meet the required fibre specifications entered by the user.

Without the described functionality it is not possible to study the process, and the results will not be repeatable or consistent. The next section explains in detail the development of the experimental setup.

4.1 Development of the Experimental Setup

This section describes the development of the electrospinning test chamber, which was used to produce materials under repeatable conditions. In the scope of this project, it is used specifically to investigate the system’s suitability for the application of control systems, and the effect of the process variables. The development has been divided in to hardware and software, and will be described in detail in the following two sections.

4.1.1 Hardware

The hardware development for electrospinning is divided in to two sub categories, mechanical and electrical. Firstly the mechanical part is explained.

4.1.1.1 Mechanical

Following the purchase of an environmental chamber from Vindon Scientific, the electrospinning equipment needed to be incorporated inside the cabinet. The unit was designed to be interchangeable due to the use of various collector types and spinning methods preferred by individual users, such as vertical or horizontal spinning. The next major design consideration for the mechanical components inside the cabinet, excluding the collector and the needle, were their non conductive properties. The main supporting structure for the collector was constructed using Polyvinyl Chloride (PVC) plastic. The main axle used to mount the aluminium collector plates of different sizes was manufactured using Tufnel, a hard phenolic resin material. The design of all the components was undertaken in SolidWorks, a Computer Aided Design (CAD) package.
The axle on the original electrospinning system suffered from considerable shaft whirl. To overcome this in the new setup, two bearings were used to mount the axle on to the support frame. Figure 4.1 shows a representative section view of the designed frame, where the bearings are push fitted into the main walls and the axle carrying the collector is fed through these.

There are three different collectors with different widths and diameters that can be attached to the system illustrated in Figure 4.1, and these can be interchanged in a short period of time. The motor chosen for the drive was an Allied Motion Premotec 24V DC Servomotor. An adapter – pulley pair was designed and manufactured for the motor to accommodate a belt drive, which keeps the motor away from the electric field of the experiment, minimising disturbances. In addition, an adapter was also designed and manufactured to couple the air filter to the extraction fan, along with a clamp to attach a plastic guard to a rotary solenoid to shield the collector during start-up. The shield prevents droplets of solution from falling on to the connector. These droplets form when the syringe pump is started before the electric field and will continue to form until consistent flow is achieved at the needle tip. Due to the silicone tubing between the syringe and the needle, there can be time delays until the pressure stabilises in the tube during start up. Once the flow rate is stable, the shield is removed and the voltage switched on simultaneously.

An image of the electrospinning kit is presented in Figure 4.2, where the various components of the system are labelled. The “Bell”, which has not been mentioned previously, is a piece of machined brass that fits tightly around the needle and is used to influence the electric field. This will be explained in more detail in the further sections of this chapter.
The original electrospinning set-up suffered from electrical arcing to the syringe pump, which on several occasions broke the pump and also affected the electric field. This was overcome by using high pressure silicone tubing and keeping the pump away from the high voltage, as presented in Figure 4.3. The top image illustrates the problem of having the high voltage wire too close to the syringe pump, where metallic edges are present. Depending on humidity levels, at high voltages the air breaks down and the current rushes to the nearest conductive spot. The bottom illustration shows how this problem was solved with the use of a long silicone tube, which allowed the syringe pump to be located away from the high voltage connector.
4.1.1.2 Electrical

The first part of the electrical work was to develop a safety circuit, to prevent the user touching anything in the cabinet when the high voltage power supply is switched on. As shown in Figure 4.4, the system simply cuts off an AC relay that provides power to the high voltage supply. There are further safety interlocks on the glove ports of the main inner glass doors, the glass door itself and an emergency stop switch on the control panel to cut off the power.

Figure 4.3; Arcing Problem and the Solution
The next addressed issue was the replacement of the wire used to conduct the high voltage from the power supply to the needle. The original setup often arced to the fume hood interior and to the syringe pump, which affected the electric field. This was caused by a lack of dielectric isolation surrounding the conductor of the cable, and the interference resulted in the fibres landing with a significant bias to one side of the collector. This has been rectified by using high voltage cabling supplied by Lemo Cables, rated at 60,000 V breakdown voltage.

Following these measures and the health and safety checks by the department, the chamber was considered ready for use. The next part of the work formed the development of the instrumentation, measurement and actuation systems. There are three actuators in the system, the electric motor driving the collector, the syringe pump and the rotary solenoid, which is used to catch drips from the needle tip prior to electrospinning and providing more reliable start-up and shut down routines.

Data Acquisition (DAQ) cards were purchased from National Instruments to interface the physical equipment with a computer. Two cards were used, one for actuation and one for measurements. This separation was done to prevent electrical noise in the measurement signals. The cards have analogue inputs and outputs, alongside digital input-output channels. The rotary solenoid was interfaced using a transistor circuit controlled by the digital output of the DAQ card.

The electric motor has built-in electronics that require a 0 - 5V signal to set the speed. However, the different drive gear sizes created by the pulley means the collector spins at a different speed to the motor shaft. To overcome this, a tachometer was integrated into system and the collector speed controlled, this will be explained in more detail in the next section. Additionally, due to the slow response times of temperature and humidity in the environmental chamber, mainly caused by its large volume of over 1000 litres, it was not feasible to manipulate these during a run. The chamber was therefore used as a steady state environment provider.
The most sensitive measurement to be taken during the electrospinning research was the current at the positive supply and ground sides of the circuit. The currents are very low, in the nano amperes range, and are prone to be lost in the electrical noise of the circuitry. To prevent this, a point2point optically decoupled measurement system was used, where the transmitters were battery powered to separate the high voltage side of the circuit away from the sensitive electronics. The overview schematic of how this works is shown in Figure 4.5. A 1 MΩ resistor is connected in series on to both the live and ground sides of the wires, going to the needle and collector respectively. Initially, smaller resistors were trialled but as the current measurement was found to be in the nA range, a larger resistor was used to detect the changes in current more accurately. This setup uses Ohm’s Law, where the current is determined by measuring the voltage across the resistors. The point2point opto transmitter receives this voltage value, and transmits it via the fibre optic link to the receiver that is outside the cabinet. The output voltage of the receiver is then read with the DAQ card into the computer. In the main software the voltage value is converted to a current value, by taking the 5:1 ratio between the transmitter and the receiver in to account.

![Figure 4.5; The Opto-Isolated Current Measurement System](image)

The syringe pump did not need any external electrical circuitry, as the serial RS232 communication port enabled it to be connected and controlled by a computer. However, to do this a communications protocol needed to be programmed in the LabVIEW programming environment, which is compatible with the DAQ cards. To cover this in more detail the software section of the development is presented next. More photographs of the hardware described in this section can be found in Appendix 2.
4.1.2 Software

The National Instruments DAQ cards are compatible with the National Instruments LabVIEW software, and the site licence of the University of Manchester for LabVIEW made this software the ideal choice. After developing the operating procedures of the new equipment, the individual sections of the software were developed. These were the axle speed controller, the custom serial communications platform for the syringe pump, the collector shield on-off controller, the voltage setpoint, current limit and voltage cut control for the high voltage power supply and the data logging of the measured variables of the process.

The axle speed controller was a PI controller, and the feedback was from the laser tachometer. This outputs frequency pulses on each rotation of the collector, which are then counted by the DAQ and converted to revolutions per second (rpm). The syringe pump controller responds to setpoint changes of the flow rate. Once the volume of liquid and syringe diameters are entered the flow rate setpoint is sent to the pump’s internal processor over the serial link. The voltage control and current limit control of the high voltage power supply is a 0 – 10 V analogue voltage signal which relates to 0 – 30 kV range, and the 0 - 4 mA current limit. The screenshot presented in Figure 4.6 shows part of the syringe pump communication platform and the speed controller created in LabVIEW.

Figure 4.6; Syringe Communications Platform and Speed Control
The following screenshot presented in Figure 4.7 shows the speed controller’s graphical user interface, which is part of the overall electrospinning operating program. The tabs labelled Pump, Collector, Voltage, Data Logger allow navigation between the respective parts of the program.

![Figure 4.7; Speed Control - Graphical User Interface](image)

Once all the individual parts of the software were completed a main user panel was developed to integrate all the pieces of software. Further screenshots of the program can be found Appendix 3. The software is written so that the whole system can be put under closed loop when required. However, to enable this, extra sensing would be required. The reasons for this are explained in the following sections. The overview schematic of the developed set up is shown in Figure 4.8, where the squares represent the controlled elements and the circles represent the measured elements of the process. The following section explains the conducted experiments.
4.2 Experimental Methods and Results

The electrospinning process can be split into three sections for analysis as shown in Figure 4.9.

These three identified areas are interconnected and affect each other strongly. The jet splitting and thinning section should be viewed as a result of the Taylor cone formation, and the separation
distance of the needle to the collector. Taylor cone formation is considered as the main influence in jet formation and is analysed in detail. For this reason the jet splitting section is not explained as a separate section and is covered within the other two sections.

To explore the process, short designed experiments and factorial studies were implemented. The main aim of these experiments was to highlight which variables affected the final fibres the most, and how. The experiments and the results are presented by outlining the constant parameters, the manipulated variables and then the results. The results are presented with the aid of images, histograms, bar charts and numerical values of mean fibre diameters. The fibre data presented in the results was obtained by using Fibremetric software, which automatically detects, counts and calculates diameters of the fibres. The software results were also checked with manual measurements to eliminate any outliers. The software is part of the Phenom Scanning Electron Microscope (SEM) system. The following section explains the collection part of the electrospinning process.

4.2.1 Collection

From a tissue engineering perspective the ability to collect random, ordered or mixed layers of fibres is crucial, as different morphologies may be needed depending on the type of tissue. According to literature, alignment happens when the rotational speed of the collector matches or exceeds the flight speed of the jet (Boland et al., 2001), (Matthews et al., 2002). This section presents the test results of the experiments that were conducted to determine the effect of the collector speed on jet collection. The collector is where the final product from the process is accumulated and is the lowest potential point that is closest to the needle. This is generally 0 V, but can also be at negative potential if desired, making the electric field even stronger. 0 V was chosen throughout all of the experiments, as this is the typical value used in this research field.

4.2.1.1 Effect of Speed Control on Fibre Alignment

This experiment was performed to test and compare results with existing literature regarding fibre alignment on the collector. The manipulated variable was the collector speed. The settings for the experiment are shown in Table 4.1.

<table>
<thead>
<tr>
<th>Constant Parameters</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Voltage</td>
<td>10 kV and 20 kV</td>
</tr>
<tr>
<td>Flow Rate</td>
<td>2 ml/h</td>
</tr>
<tr>
<td>Separation Distance</td>
<td>15 cm</td>
</tr>
<tr>
<td>Solution concentration</td>
<td>8% w.v. PCL - HFIP</td>
</tr>
<tr>
<td>Temperature</td>
<td>23°C</td>
</tr>
<tr>
<td>Humidity</td>
<td>45% Relative Humidity (R.H.)</td>
</tr>
</tbody>
</table>

Table 4.1; Experiment Settings
For this experiment a fixed temperature of 23°C and 45% RH was used. These values were chosen as they were closest to the conditions on the day of the experiment. This meant that the environmental chamber would reach the humidity and temperature setpoints quickly and maintain the environment with minimal effort. The separation distance between the needle and the collector was set at 15 cm. If the distance is any closer than 15 cm for the solvent and voltage ranges used, the fibres do not dry properly on the collector, forming clumps of unusable fibre that are not suitable for tissue engineering purposes.

The speed of the collector was varied from 100 rpm to 700 rpm in 100 rpm steps. Beyond 700 rpm the alignment was unchanged for this range of settings, as the collector’s ground speed was equal or more than the fibre jet speeds. Each condition was repeated three times, and the presented numerical values are the mean values from these experiments.

The following images in Figure 4.10, Figure 4.11, Figure 4.12 present a section of the fibre mats at different rotational speeds. The accompanying polar plots, split in to 18 bins, show the amount of fibres at the corresponding rotational degrees in the whole of the fibre mat samples. Polar plots represent the data over the polar coordinate system, showing the degrees of rotation a sample has and its quantity, similar to a histogram.

![Figure 4.10; 200 rpm Orientation Plot](image)

At 200 rpm the fibre distribution was random, and the polar plot in Figure 4.10 gives an almost round shape, meaning the fibres are spread relatively evenly over the 360° range.
At 450 rpm the fibres appear more aligned. There are large distributions at 90° and 240° pointing towards an increase in alignment.

At 700 rpm the distributions were mostly around 270° indicating alignment. However, there was a significant 340° - 360° distribution, but this was smaller than the vertically aligned fibres.

The standard deviation of the fibre orientation for the whole experiment is presented in Figure 4.13. This confirms the downward trend observed in the images, as the speed increases, the alignment of fibres increases with it.
This experiment confirms that fibre orientation can be controlled up to a certain level and the results are consistent with similar studies presented existing literature (Sundaray et al., 2004), (Yang et al., 2005b), (Pan et al., 2006b).

4.2.1.2 Current as a Measured Variable and a Process Monitor

In this project, the current at the collector side of the circuit was proposed as a measured variable throughout the process to monitor the status of the jet and the formation of the Taylor cone. The polymer solvent solution carries electrical charges to the collector. More flow carries more charge, which means it makes more solvent available to be dragged by the electric field at any point in time. For every flow rate, there is a minimum voltage level that electrospinning starts at. Figure 4.14 shows the low side current at different flow rates through the voltage ranges of the power supply, 1 kV to 30 kV in 1 kV steps.

Placing a resistor in series on the ground path of the collector and measuring the voltage across this enables the current being carried by the solution to be determined.
In Figure 4.14 the three flow rates follow the same pattern, but with increased flow rate more charge is carried and a higher current value was observed. The current measurement can also be added to the data matrix as one of the measured variables during a modelling exercise.

There are three distinct regions of electrospinning, shown more clearly in Figure 4.15. This shows the low side current data of the 5 ml/h flow rate test. The three regions are labelled as 1, 2 and 3 in the plot. The initial 1 kV to 5 kV section can be ignored, as electrospinning does not take place at these levels. This is because the electric field is not strong enough to overcome the surface tension of the solution.
Region 1 is where electrospinning begins, and the electric force is relatively low. The fibres gathered in this region are thick in comparison to the ones where higher voltages are used. Region 2 is where some splitting starts to occur and finer fibres are obtained as the electric force starts to increase and stretch them. Region 3 is where the jet splitting happens close to the needle and even finer fibres are collected. These effects will be explained in more detail in the Taylor Cone Formation and Jet Splitting section.

The next plot shows the low current during a normal electrospinning run. The voltage is pre set and steps to 6 kV from 0 kV on start up. The instant the fibres hit the collector is visible in the data, presented in Figure 4.16.
Figure 4.16; 16 kV, Low Side Current

The low current measurement can be used to observe the state of the electrospinning jet and detect disturbances in the electric field and flow. Figure 4.17 shows a situation where the system is almost stuttering due to the fibres changing their ground path momentarily to different regions of the collector, or when the fibres are not landing continuously. This is sometimes caused by air bubbles in the system, or the syringe pump having to push against minor blockages.
Figure 4.17; 16 kV, Low Side Current and Disturbance

The sensitivity of the current measurement is advantageous and has the additional advantage of being used in safety situations, such as when the pump runs out of solution to deliver, or when there is a fault with the system resulting in fibres not being collected. In situations like these the voltage can be cut or an alarm can be raised.

Another situation would be during a long electrospinning run if a significant thickness of material is being collected; the mat can act as an insulator and a slight increase in the voltage may be required to keep the electric field strength constant. This would be straightforward to implement in to the closed loop control system.

4.2.2 Taylor Cone Formation and Jet Splitting

The Taylor cone, as explained in the Chapter 2, is where the electrospinning process starts. The jet that overcomes the surface tension of the liquid then emanates out of the Taylor cone to form fine fibres on the collector.

To explore the Taylor cone formation, different voltages were used at a fixed temperature of 23°C and 35% RH. These values were chosen, like the previous test, as they were closest to the conditions on the day of the experiment. The separation distance was maintained at 15 cm, as this allows the whole voltage range of the power supply to be used for electrospinning the polymer – solvent mixture.
ranges. The fixed temperature means the viscosity of the solution flowing in the syringe is always constant for that concentration level.

Three voltages were tried for the solution concentration levels of 8%, 10% and 12% weight volume (w.v.) of polymer mixture to solvent. This concentration range was chosen as it is known to work (Bosworth et al., 2008), (Robb, 2012) and it is possible to electrospin due to the viscosity and the surface tension properties of the mixtures. The following corresponding conductivities were measured as 0.22µS, 0.20µS and 0.19µS. There were no obvious effects that could be observed with the eye on the splitting distances and Taylor cone shapes with the different concentrations. In literature it is reported that increasing the concentration of the polymer in the solvent increases the fibre dimension. However, as pointed out earlier, changing the concentrations is not a viable option when the process is running.

The variables that are reported to affect the Taylor cone are the solution properties, needle diameter and the voltage. However this study proposes that the intensity of the electric field at the needle tip, rather than just the voltage is the main contributor to the Taylor cone shape. This effect was identified in this study, but is explained in detail in the next section. The shorter the Taylor cone, the earlier the jet splits and the larger the Taylor cone, the later it splits. When the jet splits fine fibres are created, and theoretically when splitting happens earlier, the split jet has to travel further creating thinner fibres. Due to the lack of equipment there are no accurate numerical results of this experiment on the Taylor cone dimensions. However the observations are presented below in Table 4.2.

<table>
<thead>
<tr>
<th>Manipulated Variable</th>
<th>Output</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Voltage (kV)</strong></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>Long</td>
</tr>
<tr>
<td>15</td>
<td>Short / Medium</td>
</tr>
<tr>
<td>25</td>
<td>Short / Receding into the Needle</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Taylor Cone Shape</strong></td>
<td></td>
</tr>
<tr>
<td>3-4 cm above the collector</td>
<td></td>
</tr>
<tr>
<td>6-8 cm above the collector</td>
<td></td>
</tr>
<tr>
<td>13-14 cm above the collector</td>
<td></td>
</tr>
</tbody>
</table>

Table 4.2; Taylor Cone Observations

The Taylor cone elongates and shortens because of the change in the electric field intensity on the needle tip. This is achieved by changing the voltage. More voltage means more electric force and therefore the surface tension is overcome quicker and more material is carried over at a faster rate. The lower the voltage, the later the Taylor cone breaks, which produces thicker fibres at a slower rate.
The photographs presented in Figure 4.18 were taken at a later date near the end of this project, after a camera was purchased. The same environmental conditions, flow rates and voltages were used for the 8% polymer – solvent mixture.

The images confirm the observations presented in Table 4.2. Starting from the left, the first image shows the long Taylor cone at the 8 kV setting and the jet appears thicker than the other two images. The middle image is the 15 kV setting and the Taylor cone has reduced in size and the jet appears thinner. The last image shows the 25 kV condition where the Taylor cone has almost receded into the needle and the jet has split into two. A thick jet in the middle and a secondary split jet shooting off to the left can be observed. These images confirm the effect of field intensity at the needle tip on the Taylor cone formations.

![Figure 4.18; Taylor Cone Shapes, 8 kV, 15 kV, 25 kV respectively (Robb, 2012)](image)

### 4.2.2.1 Effect of Voltage and Flow Rate on Fibre Diameters

The next study tested the effects of voltage and flow rate on fibre diameters. The main aim was to identify which of these variables contributes the most. A simple experiment was conducted to investigate the effects, and the settings are presented in Table 4.3, where the variables that were held constant are shown.

<table>
<thead>
<tr>
<th>Constant Parameters</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Separation Distance</td>
<td>15 cm</td>
</tr>
<tr>
<td>Solution concentration</td>
<td>8% w.v. PCL - HFIP</td>
</tr>
<tr>
<td>Temperature</td>
<td>23°C</td>
</tr>
<tr>
<td>Humidity</td>
<td>45% Relative Humidity (R.H.)</td>
</tr>
<tr>
<td>Collector Speed</td>
<td>150 rpm</td>
</tr>
</tbody>
</table>

*Table 4.3; Experiment Settings*
The environmental variables were selected as close to the conditions of the day to minimise the work load of the environmental chamber. The collector speed was set at a rate where random fibres were collected. Fibre orientation is not of importance to this study and operating the collector at a high speed was deemed unnecessary. Furthermore, the separation distance was kept the same as in the previous studies. As this is a fibre diameter and concentration study, the volume dispensed for each sample was kept constant for all of the produced films. This is straightforward to implement, as the syringe pump can be programmed to deliver a desired set volume of liquid.

![Figure 4.19; Manipulated Variables](image)

Figure 4.19 shows the six configurations that were used in this study. High, low and medium voltage settings of 10 kV, 8 kV and 6 kV were selected respectively. High and low values were also selected for the flow rates, which were 4 ml/h and 2 ml/h respectively.

For this experiment three films were created for each of the six conditions. The films were then analysed with the SEM and the fibre measurements taken. Following the data collection the mean values of the fibre diameters were calculated alongside the mean standard deviations. The mean values of the fibres produced in each experiment are given in Table 4.4. It is clear from the results that voltage has a more significant effect on diameter than the flow rate. The standard deviation suggests that the variances between the samples are low and that the fibres are being created in a repeatable manner.

<table>
<thead>
<tr>
<th>Voltage</th>
<th>Flow Rate</th>
<th>Mean Diameter</th>
<th>Standard Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>6 kV</td>
<td>2 ml/h</td>
<td>1.3320 µm</td>
<td>0.1408 µm</td>
</tr>
<tr>
<td>8 kV</td>
<td>2 ml/h</td>
<td>0.9034 µm</td>
<td>0.1103 µm</td>
</tr>
<tr>
<td>10 kV</td>
<td>2 ml/h</td>
<td>0.6012 µm</td>
<td>0.06012 µm</td>
</tr>
<tr>
<td>6 kV</td>
<td>4 ml/h</td>
<td>1.3572 µm</td>
<td>0.0860 µm</td>
</tr>
<tr>
<td>8 kV</td>
<td>4 ml/h</td>
<td>0.9641 µm</td>
<td>0.0640 µm</td>
</tr>
<tr>
<td>10 kV</td>
<td>4 ml/h</td>
<td>0.7173 µm</td>
<td>0.0565 µm</td>
</tr>
</tbody>
</table>

Table 4.4; Results of the Experiment
To find out which of the variables most affects the fibre diameters, the mean differences can be calculated. The independent effect of a variable can be calculated using the equation given below, where $A$ in this case is either voltage or flow rate and the response is the mean fibre diameters.

$$\text{Mean Effect of Variable } A = \text{Mean Response at High } A - \text{Mean Response at Low } A$$  \hspace{1cm} (Esbensen et al., 2002)  

If we analyse the 6 kV and 8 kV results at 2 ml/h and 4 ml/h, the effect of flow and voltage can be calculated. To make this easier to present, the results are shown in the experimental framework for the said cases.

For the 6 kV case, when the flow rate is changed from 2 ml/h to 4 ml/h, the mean fibre diameter increases from 1.3320 µm to 1.3572 µm respectively. The mean response at the 6 kV case, which is the low voltage case is:

$$\overline{V}_{low} = \frac{(1.3572 + 1.3320)}{2} = 1.3446 \text{ µm}$$  \hspace{1cm} (4-2)

For the 8 kV case with the same change in the flow rates, the mean fibre diameter increases from 0.9034 µm to 0.9641 µm, so the mean response for the high voltage case is:

$$\overline{V}_{high} = \frac{(0.9641 + 0.9034)}{2} = 0.9337 \text{ µm}$$  \hspace{1cm} (4-3)

The main effect of flow rate is:

$$\overline{V}_{high} - \overline{V}_{low} = 1.3446 - 0.9337 = 0.4109 \text{ µm}$$
If the main effect of voltage for two cases is analysed:

$$
\overline{F}_{\text{low}} = \frac{(0.9034 + 1.3320)}{2} = 1.1177 \ \mu\text{m} \quad 4-4
$$

$$
\overline{F}_{\text{high}} = \frac{(0.9641 + 1.3572)}{2} = 1.1607 \ \mu\text{m} \quad 4-5
$$

The main effect of voltage is:

$$
\overline{F}_{\text{high}} - \overline{F}_{\text{low}} = 1.1607 - 1.1177 = 0.0430 \ \mu\text{m} \quad 4-6
$$

Additionally, the effect of interaction can also be calculated with the following formula:

Interaction =

$$
\frac{1}{2} \cdot (\text{Effect of Design variable } A \text{ at high level of } B - \text{Effect of Design variable } A \text{ at low level of } B )
$$

In this case the $A$ is voltage and $B$ is flow. For the high level of flow at 4 ml/h the effect of an increase in voltage from 6 kV to 8 kV on the fibre diameters is:

$$
\overline{V}_{\text{eff high}} = (0.9641 - 1.3572) = -0.3931 \ \mu\text{m} \quad 4-7
$$

For the low level of flow at 4 ml/h the effect of an increase in voltage from 6 kV to 8 kV on the fibre diameters is:

$$
\overline{V}_{\text{eff low}} = (0.9034 - 1.3320) = -0.4286 \ \mu\text{m} \quad 4-8
$$

The interaction can be calculated as:

$$
I = \frac{1}{2} \cdot (-0.3931 - (-0.4286)) = 0.0177 \ \mu\text{m} \quad 4-9
$$

This shows that the level of interaction tested at this range is quite low between the voltage and the flow rate, as an increase on one variable does not depend highly on the level of another. The interaction calculation is repeated for the larger range change from 6 kV to 10 kV.
\[ \bar{V}_{\text{eff high}} = (0.7173 - 1.3572) = -0.6399 \mu m \]  
\[ \bar{V}_{\text{eff low}} = (0.6012 - 1.3320) = -0.7308 \mu m \]

The interaction can be calculated as:

\[ I = \frac{1}{2} \cdot (-0.6399 - (-0.7308)) = 0.0454 \mu m \]

This result once again shows the effect of interaction between the two variables is low. This confirms the earlier findings that the main effect influencing the fibre diameter is voltage. However it must be noted that the flow rate must be sufficient for electrospinning to take place, as it is not possible to electrospin if the flow is not providing solution to the needle tip as quickly as the electric force is removing it.

The fibre distribution histograms for this study are shown in Figure 4.21, which also agree with the numerical indications. The first two images show the fibre distributions at the 6 kV voltage condition, where the image on the left is for the flow rate of 2 ml/h and the image on the right is for 4 ml/h. There is a slight change in the distributions, but the range of fibres created is similar for both flow rates. When compared with the second row of images depicting the increased voltage level of 8 kV, two effects can be observed. Firstly, the number of fibres increases in the samples, this is due to a thinner jet electrospinning the same volume solution. Secondly, the range of fibre diameters shifts towards the smaller values from 0.5 µm to 2.3 µm and for the 6 kV 2 ml/h it shifts to 0.4 µm to 1.7 µm. The third row of images shows the same trend and the ranges shift to 0.2 µm to 1.5 µm range, with most of the fibres being concentrated on the smaller end of the scale.
This experiment proves that the voltage has a more significant effect on the fibre diameters than the flow rate.

### 4.2.2.2 The Effect of Field Intensity on the Needle Tip and a New Electrospinning Setup

This section reports the study conducted in investigating the effect the electric field shape and its intensity has on the fibre diameters. When voltages above 15 kV were used, with the needle only set-up, the fibres on occasion were seen to change ground path and momentarily not go to the collector. This is caused by an increase in electric force on the fibre jet, where the repulsions between the fibres themselves also increase. Sometimes a proportion of fibres would collect on the environmental chamber walls due to static charging, which indicated that the jet is prone to disturbances. This highly unsteady behaviour is also caused by the fibres repelling each other during flight. This appears to be a more complex physics problem and could not be covered in the scope of this research.
However, to minimise the effects of this problem the shape of the electric field was considered. The presence of an extra electric field has been reported previously, where the jet is steered with the aid of switching mini electromagnets to create custom patterns (Bellan and Craighead, 2011), but has not been used in a fibre diameter control context prior to this study.

This section of the work proposes the use of a new configuration on the setup, whereby the electric field intensity at the needle tip and the shape of the electric field are manipulated to influence fibre diameters. To influence the electric field shape, two different size brass cylindrical shapes, termed *bells*, were manufactured. The larger bell had a 7cm radius and the smaller one had a 3.5 cm radius. Both were made from the same material and were of the same height. One of the bells can be seen in use in the photo presented previously in Figure 4.2. The use of these bells changes the electric field shape and dilutes the charge concentration at the tip of the needle.

An overview of the two different configurations with and without the bell is illustrated in Figure 4.22, where the effect of the bell on the electric field is shown on the right hand side drawing. The drawing on the left shows the needle only configuration, where the electric force concentrates solely on the needle tip. The new configuration with the bell means there is less of an electrical charge at the needle tip and therefore a higher voltage is required to start electrospinning. This is shown in further detail utilising finite element analysis.

![Figure 4.22; Needle Only and Bell – Needle Configurations](image-url)
The conditions for electric field shape tests are given below in Table 4.5.

<table>
<thead>
<tr>
<th>Constant Parameters</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Voltage</td>
<td>25 kV</td>
</tr>
<tr>
<td>Flow Rate</td>
<td>4 ml/h</td>
</tr>
<tr>
<td>Separation Distance</td>
<td>15 cm</td>
</tr>
<tr>
<td>Solution concentration</td>
<td>8% w.v. PCL – HFIP</td>
</tr>
<tr>
<td>Temperature</td>
<td>23°C</td>
</tr>
<tr>
<td>Humidity</td>
<td>45% Relative Humidity (R.H.)</td>
</tr>
<tr>
<td>Collector Speed</td>
<td>150 rpm</td>
</tr>
</tbody>
</table>

Table 4.5; Experiment Settings for the Modified Electric Field Tests

The parameter that was varied in this instance is the length of needle protruding from the bell. Three protrusion lengths were tested, 2.5 mm, 5 mm and 7.5 mm. The 2.5 mm protrusion distance was selected so the concentration of the field was minimal on the needle tip, but electrospinning could still take place. When the needle is flush with the bell, voltages above 27 kV are required and this is close to the 30 kV limit of the power supply. The 5 mm and 7.5 mm distances were selected to show how a small increment can influence fibre diameters and their distributions. For this study three samples were created for each of the conditions, totalling 9 samples. The fibres were gold coated before being viewed under the SEM and analysed. The values presented are the mean values of the three samples for each of the protrusion conditions.

The effects of the modified electric field and the protrusion distance can be seen in the collected fibres. Figure 4.23 shows the results of the 2.5 mm protrusion. The range of fibres is between 1.3 µm to 2 µm. There are very few fibres with diameters less than 1.6 µm and the majority are concentrated between 1.6 µm and 1.8 µm.
Figure 4.24 shows the distributions for the 5 mm protrusion case. As the electric field intensity on the tip of the needle has increased, the range of fibre diameters has decreased as expected. The fibre diameters now vary from 1.2 µm to 1.8 µm, with the majority of the fibres having diameters in the region of 1.4 µm to 1.5 µm. In the films made with the 5 mm protrusion setting, there were no fibres produced with diameters below 1.2 µm.

Figure 4.25 shows the mean fibre distributions taken from three samples where the protrusion was set at 7.5 mm. The range of collected fibres has decreased once again, with the range now being 0.5 µm to 1.5 µm and the majority of the fibres having diameters in the range of 1.2 µm to 1.4 µm.
Table 4.6 presents the mean and standard deviations of the measured fibres. The mean diameter of the fibres decreases as the intensity of the field on the tip of the needle increases. Although when compared with the previous studies the standard deviation has not changed significantly, which suggests the fibres are being electrospun more consistently. However, it was observed that all of the fibres landed on the collector without a change in their ground path.

<table>
<thead>
<tr>
<th>Protrusion Distance</th>
<th>Mean Diameter</th>
<th>Standard Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.5 mm</td>
<td>1.7163 μm</td>
<td>0.1615 μm</td>
</tr>
<tr>
<td>5 mm</td>
<td>1.4786 μm</td>
<td>0.1752 μm</td>
</tr>
<tr>
<td>7.5 mm</td>
<td>1.1808 μm</td>
<td>0.1923 μm</td>
</tr>
</tbody>
</table>

Table 4.6; Protrusion Test Results

To fully understand the effects of the protrusion of the needle, two dimensional finite element models were developed and analysed using the Finite Element Analysis (FEA) software FEMM. Hohman et al. (2001) reported a similar study where the fringe lines of the needle relative to the spinning distance was investigated. They showed how the electric field effected the Taylor cone shape and performed theoretical analysis on the solution, where the solution's charge carrying characteristics were explained. The contribution of the investigation presented in this thesis is that the full electric field was taken in to consideration and an investigation was made to how the fibre collection process could be improved by the changing the electric field properties, and how the diameters could be regulated by manipulating the protrusion distance.
The differential equations used by the FEMM software is presented in the next section alongside the
electric field simulation study results. The first equation is the differential from of Gauss’s Law. In this
equation explains the conservation of charge, and shows that the flux coming out of a closed surface
is equal to the charge within that surface (Meeker, 2009).

\[ \nabla \cdot \mathbf{D} = \rho \]  
\[ \text{4-13} \]

Where \( \rho \) is charge density and \( \mathbf{D} \) is flux density.

Ampere’s Law is also considered, shown in equation 4-14:

\[ \nabla \times \mathbf{E} = 0 \]  
\[ \text{4-14} \]

Where \( \mathbf{E} \) is the electric field intensity.

The flux density \( \mathbf{D} \) is related to the electric field via the electrical permittivity of the material. This is
described in equation 4-15:

\[ \mathbf{D} = \varepsilon \mathbf{E} \]  
\[ \text{4-15} \]

The FEMM program uses the electric potential \( V \) to calculate the electric field intensity, this is shown
in expression 4-16:

\[ \mathbf{E} = -\nabla V \]  
\[ \text{4-16} \]

The partial differential equation that the software solves is shown in expression 4-17.

\[ \varepsilon \nabla^2 V = \rho \]  
\[ \text{4-17} \]

FEMM solves equation 4-17 over a constant \( \varepsilon \) and for a given voltage \( V \) with defined boundary
conditions and materials. The FEMM software allows the user to enter sizes, distances, material
properties and electric potential values for evaluation through an intuitive graphical user interface. For
this test the needle was set at 30kV and the collector at 0kV.

The first simulated scenario was the needle only configuration and the second was where the large
bell was used with a 2.5 mm needle protrusion.
Figure 4.26 shows the FEA results of the needle only scenario. The electric field intensity is at maximum on the needle tip shown by the red colour, and is less intense through the flight path of the jet. This possibly encourages a higher incidence of fibre to fibre repulsion events. When a bell is used, the maximum intensity points move to the corners of the bell instead of the needle tip, as shown in Figure 4.27. This explains why more voltage is needed to electrospin. In addition, the central field intensity values are overall higher in the flight path of the fibres.
The tests and the FEA results show the effect of the intensity on the needle tip and the field shape on fibre formations. This creates an additional manipulated variable which can be used in future investigations, which is the protrusion distance of the needle through the bell. This new configuration can be easily automated where the protrusion distance is changed on demand, or automatically by a control loop when the desired fibre diameters are requested.

The results presented in this section have demonstrated how it is not possible to collect fibres with a single desired diameter. Instead, a variety of fibres in a particular diameter range are created. This could be due to the variation of the forces amongst the fibres where they repel each other, or due to small disturbances that could be caused by minor blockages on the solution line. The syringe pump is a stepper motor and operates in high frequency steps, and these minor pulses may also be contributing to the variation seen amongst the fibres. In the previous tests the environmental variables were kept constant and continuously monitored. The next set of experiments present the work investigating the effect of the environmental variables on the fibre diameters and distributions.

### 4.2.2.3 Environmental Effects

This study presents a short three step designed experiment where the environmental variables were tested. Temperature and humidity are the two environmental variables that can be changed in the process. However, the slow response of the variables makes their use as manipulated variables in a real time control situation unfeasible. Nonetheless, the effect that they have on fibre properties is important. The response of the system is slow because the volume of the cabinet is large, at approximately 1000 litres, and the air that needs to be conditioned to change the internal conditions takes a significant amount of time to reach its setpoint.

The parameters of this experiment are shown in Table 4.7. These settings were selected as the electrospinning takes place at these ranges and the fibres created are suitable for tissue engineering purposes. The collector speed was set to a low value as there was no requirement for alignment in this test.

<table>
<thead>
<tr>
<th>Constant Parameters</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Voltage</td>
<td>15 kV</td>
</tr>
<tr>
<td>Flow Rate</td>
<td>2 ml/h</td>
</tr>
<tr>
<td>Separation Distance</td>
<td>15 cm</td>
</tr>
<tr>
<td>Solution concentration</td>
<td>8% w.v. PCL – HFIP</td>
</tr>
<tr>
<td>Collector Speed</td>
<td>150 rpm</td>
</tr>
</tbody>
</table>

*Table 4.7; Experiment Settings*
Figure 4.28 shows the values of the manipulated variables over the three experiments.

![Diagram showing temperatures and humidity levels](image)

**Figure 4.28; Manipulated Variables**

The temperature and humidity values used in this experiment were specified by consideration of the calculated water content in the air. If temperatures or humidity values are selected at random, then the air could have the same water content in different experiments, meaning the effect of humidity will not be observed correctly. 27°C at 41% RH contains 9.5 gr of water in 1 kg of air and at 55% RH it contains 12.7 gr of water. 32°C at 41% RH contains 12.7 gr of water per kg of air, the same as 27°C at 55% RH condition. The effect of temperature can be analysed by comparing the results with the same water content. Similarly the effect of humidity can be observed by comparing the results at the same temperature. The mean diameters of the fibres resulting from this experiment are shown in Table 4.8.

<table>
<thead>
<tr>
<th>Temperature</th>
<th>Humidity</th>
<th>Water Content</th>
<th>Mean Fibre Diameter</th>
<th>Standard Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>27°C</td>
<td>41% RH</td>
<td>9.5 gr/kg</td>
<td>1.6102 µm</td>
<td>0.2437 µm</td>
</tr>
<tr>
<td>27°C</td>
<td>55% RH</td>
<td>12.7 gr/kg</td>
<td>1.5178 µm</td>
<td>0.2173 µm</td>
</tr>
<tr>
<td>32°C</td>
<td>41% RH</td>
<td>12.7 gr/kg</td>
<td>1.5642 µm</td>
<td>0.2576 µm</td>
</tr>
</tbody>
</table>

**Table 4.8; Conditions and Fibre Diameter Results**

The rise in temperature together with humidity, whilst keeping the water content the same, creates a 0.05 µm increase in fibre diameters. The increase in water content causes the fibre diameters to decrease by approximately 0.1 µm. When only the temperature is increased by keeping the humidity the same at 41% RH, the fibre diameters decrease by 0.05 µm.

The results suggest that humidity may have more of an effect on this particular solvent polymer mixture at 8% w.v. concentration. However different results would be observed when different solvents are used, because of the chemical properties of the solvents and whether they absorb water or not (Casper et al., 2003).
With the presented standard deviations observed among the fibre diameters in Table 4.8, it is not possible to come to a definite conclusion as to which variable affects the final fibres the most. For the determination of the relative effects, additional tests are required, where the changes in the water content levels are larger.

Similar to solvent casting, to identify the effects of viscosity and surface tension when they are changed due to temperature, a quick response syringe heater/cooler would be required. Additionally, keeping the relative humidity constant and only changing the temperature can be misleading, a point ignored in other research studies. Reneker et al. (2002) showed that when temperature and humidity are increased together, which is effectively increasing the water content, then the fibre diameters decreased. Although the solvents used are different, this result also contradicts the evaporation rate study conducted for the solvent casting project with the accurate scales, where the temperature influenced the rate of evaporation more significantly than humidity. This reiterates that it is important to discover the properties of the solvents before they are used for evaporation based processes. Furthermore, for electrospinning, the solution conductivity, viscosity and surface tension would need to be measured over the range of solution temperatures to be tested. A mathematical model of this would help estimate the values at the different mixture levels and temperatures.

4.3 Summary

This study conducted all the experiments under measured conditions and none of the variables varied significantly from their setpoints. Everything was data logged by a computer and with the help of actuators and computer control, human interference was minimised. This approach was far more advanced when compared with the equipment and methods of other research groups around the world.

The tests and the observations conducted in the presented studies prove the existence of correlation between the variables of the electrospinning process. A manipulated variable map and their effects on the process and final products determined during this study are shown in Figure 4.29. This map is important when trying to establish what is affecting a specific electrospinning process. Most research studies attempt changing one variable at a time and reporting findings. Considering the interactions, the presented map is an important contribution to understanding of the electrospinning process and how the process can be improved.
The environmental variables are slow to change and are therefore not suitable as the main manipulated variables. The most suitable way to change and influence fibre diameter is through the control of the electric field shape, intensity and the shape of the Taylor cone. These can be controlled by manipulating the voltage, the needle protrusion distance using the bell and the flow rate through the syringe pump. Using the low side current as a jet stability monitor and relating the current values to the Taylor cone shapes, would benefit the process control aspect of electrospinning in terms of modelling. The first step in applying this would be to relate the Taylor cone shapes to the fibre diameters and distributions, whilst all the other variables are held constant apart from the protrusion distance of the needle from the bell. This can be tried for different voltage levels to build up a model between voltage, flow, protrusion distance and the Taylor cone. In addition, the initial jet diameter from the Taylor cone also needs to be investigated. Although the Taylor cone shapes may be similar, the initial jet diameters may change under different electric field strengths and intensities, which could in return influence the diameter of the fibres.

For the above approach to be implemented, several improvements to the current system are required. Ideally a camera with a live feed and fast image processing would be used. This would provide rapid testing with the manipulated variables. Once the Taylor cone shape and initial jet diameter is related to the final fibres, a control system for controlling the cone shape can be implemented. To do this the construction of an electric field control actuator system is necessary, where the bell can be moved up or down along the needle autonomously. Additionally, the introduction of electrodes could be considered to steer the jet to the desired areas of the collector for three dimensional scaffold manufacturing.
To further investigate the effect of temperature on solution properties and the knock-on effect of this on the Taylor cone shape, a syringe heater/cooler should be utilised. This would eliminate the reliance on the environmental chamber’s internal temperature, as this could be a limiting factor in the solution temperatures that can be tested. This will enable the solution to heat up and cool down quickly to test a range of conditions. A thermoelectric module could be used for rapid heating and cooling, which would make the solution temperature an appropriate manipulated variable for the control system.

To fully understand the effects of the bells on the electric field shapes, a full three dimensional finite element analysis should be conducted. The two-dimensional study presented here gives a broad overview of what is happening. However, a three-dimensional study would provide more detail on the effect that the bell has. The next natural step in this project is to collect a large set of routine operating data and build models using the techniques introduced in Chapter 2, however, as discussed, for this to happen further modifications to the equipment would be required. These models would then be used to predict fibre diameters, or help determine the necessary conditions to produce fibres with specific diameters.

In summary, this part of the project has completed essential preliminary work towards the development of a control system for continuous fibre production using electrospinning, and has worked towards understanding the effects that several manipulated variables have on the final product. The effect of the collector on orientation has been confirmed and the changes introduced to manipulate the electric field shape also affected the charge intensity at the needle tip. These effects meant that slightly higher voltages were needed to electrospin, but a more stable jet was achieved. The potential of moving the bell up and down the needle, as well as being able to control the voltage gives the system flexibility and a wider range for fibre production. The monitoring of the low side current is also an important step towards understanding the process. With the aid of data logging and some of the suggested further modifications, an example of the input data matrix that is potentially possible to obtain, is shown in Table 4.9 and Table 4.10, where “n” is the final number of samples.
The output data matrix would be as shown in Table 4.11. This would be used in combination with the input data to identify a model of the process.

Table 4.10; Example Input Data Matrix- Part 2

<table>
<thead>
<tr>
<th>Samples</th>
<th>Electric Field</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Voltage</td>
</tr>
<tr>
<td></td>
<td>Spinning Distance</td>
</tr>
<tr>
<td></td>
<td>Protrusion</td>
</tr>
<tr>
<td></td>
<td>Current</td>
</tr>
<tr>
<td>1</td>
<td>.</td>
</tr>
<tr>
<td>2</td>
<td>.</td>
</tr>
<tr>
<td>.</td>
<td>.</td>
</tr>
<tr>
<td>.</td>
<td>.</td>
</tr>
<tr>
<td>n</td>
<td>.</td>
</tr>
</tbody>
</table>

Table 4.11; Example Output Data Matrix

<table>
<thead>
<tr>
<th>Samples</th>
<th>Fibre</th>
<th>Taylor Cone</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Diameter</td>
<td>Volume</td>
</tr>
<tr>
<td>1</td>
<td>.</td>
<td>.</td>
</tr>
<tr>
<td>2</td>
<td>.</td>
<td>.</td>
</tr>
<tr>
<td>.</td>
<td>.</td>
<td>.</td>
</tr>
<tr>
<td>.</td>
<td>.</td>
<td>.</td>
</tr>
<tr>
<td>n</td>
<td>.</td>
<td>.</td>
</tr>
</tbody>
</table>

The electrospinning process contains fast dynamics and changes in the cause variables has an almost instantaneous effect on the Taylor cone and therefore also in the fibre diameters and morphologies. For this process a fast acting MPC algorithm would be beneficial as standard MPC algorithms have high computational requirements.

With traditional MPC, at each decision point or sample an optimisation problem is solved over a finite horizon, which can take time. Kouvaritakis et al. (2002) present an alternative approach to MPC, which removes the requirement to solve a quadratic programming problem by computing the control moves using ellipsoidal approximations of the control space. Although this approach allows the control calculations to be performed very quickly in milliseconds, the complexity of the calculation does increase with the size of the system and the complexity of the process dynamics.

An alternative strategy for systems with fast dynamics is to pre-define the MPC and calculate the control moves offline, where the online controller uses a look up table. This method can be a feasible alternative as it is shown to work well with systems that have a small number of inputs. For example, Wang and Boyd (2010) developed an offline MPC system for an application with 12 states, 3 control variables and a 30 step ahead prediction horizon. The problem contained 450 quadratic programming variables and 1284 constraints, which were solved in 5 milliseconds. This speed was achieved by
terminating the quadratic program early by fixing the length of the primal barrier method for the quadratic solver (Nocedal and Wright, 1999). It was shown that even though the quadratic program produced poor results due to its early termination, the quality of the control was good when compared with the control objective.

Another suggested approach for fast MPC is to reformulate the quadratic problem in terms of control inputs, however this means the problem structure is lost in the equation (Wright, 1997). This method also suffers when the prediction horizon is large and can be slower to converge than traditional MPC optimisers (Y. Wang and Boyd 2010). For electrospinning the ellipsoidal approximation offline MPC approach by Kouvaritakis et al. (2002 and 2005) would be suitable, as this method reduces the computational burden on a systems with fast dynamics, such as electrospinning.

Using identification signals with persistent excitation such as GBN, it would be possible to collect sufficient data for modelling this continuous system. The determination of a suitable model will benefit the system in three main aspects; firstly, the model would be able to calculate optimum parameters for desired fibre production. Secondly, the process understanding would be significantly enhanced, and thirdly a Model Predictive Control scheme could be implemented for real-time control. However, this would require a standalone FPGA based computer fast enough to cope with the multiple processes. Potential methods that could be used to identify models for this type of continuous process, which has highly correlated variables, are presented in the following chapter. Additionally, the methods used for determining models from batch type data, such as the solvent casting process are also presented. Chapter 6, following the next presented chapter, demonstrates through simulated correlated systems how these models are determined and proposes additional novel modelling methods for highly correlated data sets.
5 Numerical Methods

This chapter presents the background theory for the numerical methods used in this thesis. The modelling methods presented are proposed to be used as part of an MPC control scheme, where an accurate model of the process is necessary for the control system to work. The modelling methods are explained with the aid of equations and the derivations of the cost functions are shown. The bias problem in the least squares methods is presented with the aid of a simulation, and novel hybrid modelling methods that eliminate this problem are described.

5.1 Theory of the Numerical Methods

This section introduces the concept of multivariable statistical process control which has the potential to be applied to the solvent casting method to improve the final product quality. In addition, the theory of the methods applied to the simulation studies in Chapter 6 are presented and analysed in terms of their predictive qualities for use in MPC applications.

To begin Principal Components Analysis (PCA) is presented as this technique provides a stepping stone to understanding the PLS technique applied in this work.

5.1.1.1 Principal Components Analysis

Principal Components Analysis, as discussed in Chapter 2, has been demonstrated as a useful tool for monitoring multivariate processes. The method identifies major trends in a data set by projecting pairs of variables on to a reduced space. Before applying PCA the data should be mean centred. This sets the mean to zero, and gives the data unit variance. The formula is given in equation 5-1, where \( X \) is the data matrix and \( \bar{X} \) is the mean.

\[
X_{normalised} = \frac{(X - \bar{X})}{\text{var}(X)}
\]

As PCA is scale dependent, if required, the data should also be scaled as incorrect scaling can yield false predictions (MacGregor and Kourti, 1995). This is only required if there are measurements that have different values and units, such as flows and temperatures. Scaling ensures that each variable has the same weighting rather than one outweighing the other based on its magnitude.

The data matrix containing the measurements is defined as \( X \), which has \( m \) rows representing the number of samples and \( n \) columns representing the number of variables. The amount of data captured by the principal components is calculated through the percentage variance. When PCA is applied to a matrix, a certain amount of data is captured based on the greatest variance in the first principal component pair. Following the calculation of the first pair, the next pair is calculated. To
ensure maximum variance, the principal components are orthogonal to each other. The mathematical
description of the PCA method is given in equation 5-2, where $j < n$:

$$X = T \cdot P^T + \varepsilon = \sum_{k=1}^{j} t_k \cdot p_k^T + E = t_1 \cdot p_1^T + t_2 \cdot p_2^T + t_k \cdot p_k^T + \varepsilon$$

Singular value decomposition is used to identify the score and loading pairs, where the score matrix
$T$ and the loading matrix $P$ describe the data matrix $X$. The $p$ vectors represent the orthogonal
directions of the coordinate system described by $P$, which the original data is projected on to. The
loading vectors are the eigenvalues of the covariance matrix, which is:

$$\text{cov}(X) = \frac{X^T X}{m-1}$$

Where \text{cov}(X) $p_k = \lambda_k$ $p_k$ and $\lambda_k$ is the eigenvector. The percentage variance captured is
calculated as:

$$\% \text{ Var} = \frac{\lambda_1 + \ldots + \lambda_{n_p}}{\lambda_1 + \ldots + \lambda_n} \cdot 100\%$$

where $n_p$ is the number of principal components used and $n$ is the total number of variables in the data matrix $X$.

It is important that a suitable number of components are specified in the model. Several techniques
have been proposed for doing this, but the most prevalent and arguably robust is Cross Validation,
where the data is divided into training and testing data sets. The ideal number of components to
give the minimum error is then determined by cross testing all of the data sets at least once against
the models developed with the training data.

PCA can also be extended into the Multiway PCA (MPCA) (S. Wold et al., 1987) method, where
multidimensional matrices of data can be analysed. In its normal form PCA works on static data.
However, if application to dynamic data is required then Dynamic PCA (DPCA) method can be used
(Ku et al., 1995). This method introduces a time shift operator which introduces dynamics in to the
model.

From a control systems perspective PCA is useful for analysing the input data. The ability of PCA to
detect abnormal conditions makes monitoring systems with large numbers of variables less complex.
However, to apply control it is necessary to explicitly capture the relationship between the input and
output variables. The most common way to do this is by using regression methods. Least squares
regression using an Autoregressive Exogenous (ARX) model is presented next.
5.1.2 Least Squares Estimation and the ARX Model

Least squares estimation is used to calculate the parameters of a system so that the squared sum of the chosen error criterion is minimised. This is also described as the cost function of the estimators.

If a system is described as:

\[ y(t) = u_1(t) \cdot \theta_1 + u_2(t) \cdot \theta_2 + \ldots + u_n(t) \cdot \theta_n \]  

5-4

Where \( t = 1, 2, \ldots, N \)

\[
y = \begin{bmatrix} y(1) \\ y(2) \\ \vdots \\ y(m) \end{bmatrix}, \quad X = \begin{bmatrix} x(1) \\ x(2) \\ \vdots \\ x(m) \end{bmatrix}, \quad u = \begin{bmatrix} u_1(1) & u_2(1) & \cdots & u_n(m) \\ u_1(2) & u_2(2) & \cdots & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ u_1(m) & u_2(m) & \cdots & u_n(m) \end{bmatrix} \quad \text{and} \quad \theta = [\theta_1 \theta_2 \cdots \theta_n]^T
\]

Where \( y = X \cdot \theta \) and if the \( X \) matrix is square, then the parameters can be estimated as, \( \hat{\theta} = X^{-1} \cdot y \). However, in real scenarios this condition is rare and to calculate \( \hat{\theta} \), the equation error is introduced:

\[ \varepsilon(t) = y(t) - \hat{y}(t), \quad \text{where} \quad \hat{y}(t) = x(t) \cdot \theta \]

And \( \varepsilon(t) \) is minimised as shown in equations 5-5 and 5-6:

\[ J_{LS} = \frac{1}{m} \sum_{t=n+1}^{m} \varepsilon(t)^2 = \frac{1}{m} \sum_{t=n+1}^{m} \left[ y(t) - x(t) \cdot \theta \right] \cdot \varepsilon^2 = \frac{1}{m} \cdot \varepsilon^T \varepsilon \]

5-5

\[ J_{LS} = \frac{1}{m} (y - X\theta)^T (y - X\theta) = \frac{1}{m} (y^T y - \theta^T X^T y - y^T X\theta - \theta^T X^T X\theta) \]

5-6

By taking the differential and equating it to zero, the minimum can be found:

\[
\frac{\partial V_{LS}(\theta)}{\partial \theta} \bigg|_{\theta = \hat{\theta}} = \frac{1}{m} (-2 X^T y + 2 X^T X \hat{\theta}) = 0
\]

5-7

Which becomes:

\[ X^T \hat{X} = X^T y \]

5-8
Then the estimated parameter vector is:

$$\hat{\theta} = X^T y \cdot (X^T X)^{-1}$$  \hspace{1cm} (5-9)

For this solution to exist, as pointed out in Chapter 2, the $X^T X$ matrix needs to have an inverse. In situations where this matrix is singular or close to being singular, then principal component based approaches are recommended to determine $\hat{\theta}$. However, if the identification signal is persistently exciting and its order is at least twice that of the system then $X^T X$ will be non-singular, meaning it will have an inverse.

The ARX model can be written as:

$$y(t) = G(q) \cdot u(t) = \frac{B(q)}{A(q)} \cdot u(t) = \frac{b_2q^{-1} + b_2q^{-2} + \ldots + b_nq^{-n}}{1 + a_1q^{-1} + a_2q^{-2} + \ldots + a_nq^{-n}} \cdot u(t)$$  \hspace{1cm} (5-10)

Where $a$ and $b$ are the system parameters. Then the transfer function can be re-written as:

$$A(q) \cdot y(t) = B(q) \cdot u(t) + \varepsilon(t)$$  \hspace{1cm} (5-11)

The error $\varepsilon$ introduced here is the equation error and will be minimised when calculating the parameters. It represents the error of the fit between the actual data and the estimated data through the regression. When the transfer function is re-written as a least squares representation, it becomes:

$$y(t) = x(t) \cdot \theta + \varepsilon(t)$$  \hspace{1cm} (5-12)

Where the input data sequence is:

$$x(t) = [-y(t-1) \ldots y(t-n) \ldots u(t-1) \ldots u(t-n)]$$  \hspace{1cm} (5-13)

$$\theta = [a_1 \ldots a_n b_1 \ldots b_n]^T$$, is the vector of parameters.

If written in a general form, the whole system can be described as $y = X \cdot \theta + \varepsilon$.

Where, $y = \begin{bmatrix} y(n+1) \\ y(n+2) \\ \vdots \\ y(m) \end{bmatrix}$, $\varepsilon = \begin{bmatrix} \varepsilon(n+1) \\ \varepsilon(n+2) \\ \vdots \\ \varepsilon(m) \end{bmatrix}$ and
\[
X = \begin{bmatrix}
  x(n+1) \\
  x(n+2) \\
  \vdots \\
  x(m)
\end{bmatrix} = \begin{bmatrix}
  -y(n) & \ldots & -y(1) & -u(n) & \ldots & -u(1) \\
  -y(n+1) & \ddots & \vdots & -u(n+1) & \ddots & \vdots \\
  \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\
  -y(m-1) & \ldots & -y(m-n) & -u(m-1) & \ldots & -u(m-n)
\end{bmatrix}  \quad 5-14
\]

From equation 5-12, the error \( E \) can be written as:
\[
\varepsilon(t) = y(t) - x(t) \cdot \theta \quad 5-15
\]

The minimisation algorithm then becomes:
\[
J_{ARX} = \sum_{i=n+1}^{m} \varepsilon(t)^2 = \sum_{i=n+1}^{m} [y(t) - x(t) \cdot \theta] \quad 5-15
\]

The system parameters are estimated from the least squares principle as:
\[
\hat{\theta} = [X^T X]^{-1} X^T y \quad 5-16
\]

\[
\hat{\theta} = \left[ \sum_{i=n+1}^{m} x_i^T (t) \cdot x(t) \right]^{-1} \left[ \sum_{i=n+1}^{m} x_i^T (t) \cdot y(t) \right] \quad 5-17
\]

### 5.1.2.1 Effect of Bias

As highlighted in Chapter 2, the mathematical simplicity of the least squares method makes it attractive to use for modelling. However, it is biased and this can be shown if we assume the system is represented as:
\[
y(t) = G(q) \cdot u(t) + r(t) = \frac{B(q)}{A(q)} \cdot u(t) + r(t) = \frac{b_1 q^{-1} + b_2 q^{-2} + \cdots + b_n q^{-n}}{1 + a_1 q^{-1} + a_2 q^{-2} + \cdots + a_n q^{-n}} \cdot u(t) + r(t) \quad 5-18
\]

Where, \( r(t) \) is a disturbance or noise uncorrelated with the input \( u(t) \).

The system can be re-written as an equation error form:
\[
y(t) = x(t) \cdot \theta + \varepsilon(t) \quad 5-19
\]

Where,
\[
x(t) = [-y(t-1) \ldots -y(t-n) \quad u(t-1) \ldots u(t-n)] \quad \text{and} \quad \theta = [a_1 \ldots a_n \ b_1 \ldots b_n]^T
\]

From this it can be seen that \( y(t) \) is partially dependent on the disturbance \( r(t) \), which shows that \( y(t) \) is correlated with the equation error \( \varepsilon(t) \). Then it can be shown as:
\[
\varepsilon(t) = A(q) \cdot r(t) \quad 5-20
\]
To show what affect this has, the least squares estimator is calculated as:

\[
\hat{\theta} = \left[ \sum_{t=n+1}^{m} x^T(t) \cdot x(t) \right]^{-1} \cdot \left[ \sum_{t=n+1}^{m} x^T(t) \cdot y(t) \right]
\]  

(5-21)

If equation 5-19 is written in to equation 5-21:

\[
\hat{\theta} = \left[ \sum_{t=n+1}^{m} x^T(t) \cdot x(t) \right]^{-1} \cdot \left[ \sum_{t=n+1}^{m} x^T(t) \cdot (x(t) \cdot \theta + \varepsilon(t)) \right]
\]  

(5-22)

\[
\hat{\theta} = \left[ \sum_{t=n+1}^{m} x^T(t) \cdot x(t) \right]^{-1} \cdot \left[ \sum_{t=n+1}^{m} x^T(t) \cdot x(t) \cdot \theta + \sum_{t=n+1}^{m} x^T(t) \cdot \varepsilon(t) \right]
\]  

(5-23)

When equation 5-23 is multiplied out and cancellations have taken place we are left with:

\[
\hat{\theta} - \theta = \left[ \sum_{t=n+1}^{m} x^T(t) \cdot \varphi(t) \right]^{-1} \cdot \left[ \sum_{t=n+1}^{m} x^T(t) \cdot \varepsilon(t) \right]
\]  

(5-24)

This represents the difference between the actual parameters and the estimated parameters, where the difference is the bias. Due to the correlation identified earlier between \( y(t) \) and \( r(t) \), \( x^T(t) \) will be correlated with \( r(t) \). If equation 5-24 is expanded as below:

\[
\hat{\theta} - \theta = \left[ \sum_{t=n+1}^{m} x^T(t) \cdot \varphi(t) \right]^{-1} \cdot \left[ \sum_{t=n+1}^{m} y(t) - y(t-1) \cdot u(t-1) \right] \cdot \varepsilon(t) + \left[ \sum_{t=n+1}^{m} x^T(t) \cdot x(t) \right]^{-1} \cdot \left[ \sum_{t=n+1}^{m} -r(t-1) \right] \cdot \varepsilon(t)
\]  

(5-25)

As \( m \to \infty \) the left hand side of the equation becomes zero. However, the right hand side will only be zero if \( \varepsilon(t) \) was white noise with zero mean, which is unrealistic in real life situations. Zhu (2001) points out that when the system is represented as an ARX model, even if the noise is white noise it is still filtered through \( A(q) \), as shown previously, and hence a bias will still exist.

The following section presents the effect of bias in the frequency domain, and explains why frequency weighted pre-filtering has been suggested as a solution.
5.1.2.2 Effect of Bias in the Frequency Domain and Pre-filtering

If the least squares cost function is written as:

\[
J_{LS} = \frac{1}{m-1} \sum_{i=n+1}^{m} e(t)^2 = \frac{1}{m-1} \sum_{i=n+1}^{m} \left( A(q) \cdot y(t) - B(q) \cdot u(t) \right)^2
\]

Parseval's formula (Zhu, 2001) is used to represent this in the frequency domain and calculate the energy, also known as the power spectrum, of the signals. Perseval’s formula can be written as:

\[
\frac{1}{2\pi} \int_{-\pi}^{\pi} |f(x)|^2 \, dx = \sum_{i=1}^{\infty} |u(t)|^2
\]

When applied to the cost function, this becomes:

\[
J_{LS} = \frac{1}{2\pi} \int_{-\pi}^{\pi} \left( G(e^{j\omega}) - \frac{B(e^{j\omega})}{A(e^{j\omega})} \right)^2 \cdot A^2(e^{j\omega}) \cdot \Phi_u(\omega) \, d\omega + \frac{1}{2\pi} \int_{-\pi}^{\pi} A^2(e^{j\omega}) \cdot \Phi_r(\omega) \, d\omega
\]

Here \( \Phi_u \) and \( \Phi_r \) are the signal power spectra of the input and the disturbance. In this equation the transfer function error \( \left( G(e^{j\omega}) - \frac{B(e^{j\omega})}{A(e^{j\omega})} \right)^2 \) is weighted by \( A^2(e^{j\omega}) \cdot \Phi_u(\omega) \).

If the system's frequency weighting is assumed to be \( F^2(e^{j\omega}) \), the identification input signal can potentially be tailored to suit this. This can be achieved by using a pre-filter. The ideal input signal's power spectrum for identification can be described as:

\[
\Phi_u(\omega) = \frac{F^2(e^{j\omega})}{A^2(e^{j\omega})}
\]

Although this appears convenient in theory, in practice \( A(q) \) is unknown. Therefore the ideal input signal is not as straightforward to obtain. As discussed in Chapter 2, a trial and error approach to obtain the pre-filter is generally financially unfeasible and an iterative method is used in practice, which can be described as:

\[
u_{filt} = L(q) \cdot u(t) \quad \text{and} \quad y_{filt} = L(q) \cdot y(t),\]

where \( L(q) \) is a stable filter, filtering the input and the output data. These pre-filters can be designed using the least squares method. However this method of pre-filtering proposed by Steiglitz and McBride (1965) is not a real minimisation technique as it is an iterative procedure where every iteration is not guaranteed to minimise the loss function (Zhu, 2001), therefore it is not used in this study.

As discussed earlier, in situations where the covariance matrix is singular or close to singular, the ordinary least squares method is not suitable. In these situations, Partial Least Squares (PLS) is recommended and this algorithm is described in the following section.
**5.1.3 Partial Least Squares - PLS**

Partial Least Squares also known as Projection to Latent Structures is a regression method that is closely related to PCR and OLS, as presented in Chapter 2. The PLS method aims to maximise the covariance between the predictor variables, which means trying to find the covariance and correlation at the same time. This is accomplished by projecting the predicted variables in matrix $Y$, on to the measured variables in $X$. PLS was proposed as a method to deal with data containing high levels of co-linearity, such as short and wide data matrices, when there are very few samples available. PLS has also been applied in control and monitoring of batch processes, where the output or the quality variable is only measured at the end of a process run.

The two most common ways to identify a PLS model are to use the Nonlinear Iterative Partial Least Squares (NIPALS) algorithm and the SIMPLS algorithm. The NIPALS algorithm calculates loadings and scores similarly to the PCR method, but in addition calculates a vector of weights (Wise et al., 2004). In contrast, the SIMPLS algorithm calculates the weights slightly differently, which makes the calculation faster in comparison to the original NIPALS algorithm introduced by Wold (de Jong, 1993).

The difference between the loadings and the scores of PLS and other methods such as PCA and PCR, is that they are rotated for the identification of the variables to predict $y$. This means PLS tries to capture both variance and correlation amongst the variables.

The steps of the NIPALS algorithm are given below, where the first principal component weighting, scores and loadings are described.

1. Pick an initial value for the output score estimate from any column of $Y$ and all its samples, $u_i = Y(:, r)$.
2. In the $X$ data calculate, the weighting vector, $w_i = \frac{X^T u_i}{u_i^T u_i}$ then normalise, $w_i = \frac{w_i}{\|w_i\|}$.
3. Calculate the score vector, $t_i = Xw_i$.
4. In the $Y$ data calculate the loading vector, $q_i = \frac{Y^T t_i}{t_i^T t_i}$.
5. Scale the loading vector $q_i$ to unit length, $q_i = \frac{q_i}{\|q_i\|}$.
6. Calculate a new score vector, $t_{new} = Yq_i$ and check for convergence, $|t_{new} - t_i| < \epsilon$. If this condition is not satisfied, re-calculate by iterating one step, from the second point onwards of this
list, where \( w_1 \) will become \( w_2 \). The convergence is checked against an error tolerance, such as \( e << 0.000001 \).

7. For the \( X \) data, calculate the loadings, \( p_i = \frac{X^T t_i}{t_i^T t_i} \) and rescale \( p_i = \frac{p_i}{\|p_i\|} \).

8. Calculate the regression coefficient, \( b_i = \frac{u_i t_i}{t_i^T t_i} \).

9. Calculate the residuals of the \( Y \) and \( X \) data blocks as \( E_i = X_i - t_i p_i^T \) and \( F_i = Y_i - b_i t_i q_i^T \).

This sequence continues \( k \) times until the desired number of latent variables is obtained. Choosing \( k \), the number of latent variables, to represent the model is the most important aspect of PLS modelling and, as with PCA, is typically obtained using cross validation methods.

For a more straightforward representation, the PLS scores \( t_k \) can be represented as a product of the original data matrix, where \( r_k = X \cdot r_k \). In a matrix representation this expression will become \( T = XR \) for the \( k \) number of latent variables, where \( R = [r_1, r_2, \ldots, r_k] \). \( R \) is calculated by regressing the pseudo-inverse of \( X \) on to \( T \), shown in equation 5-30:

\[
R = X^+ T = (X^T X)^{-1} X^T T = (X^T X)^{-1} P (T^T T)^{-1}
\]

The predicted outputs can now be calculated through \( \hat{Y} = XC \) and \( C = RBQ^T \) where;

\[
B = [b_1, b_2, \ldots, b_k] \quad \text{and} \quad Q = [q_1, q_2, \ldots, q_k].
\]

The SIMPLS algorithm, which is computationally faster than the original NIPALS algorithm operates as follows:

1. Calculate the cross product, \( S = X^T Y \).
2. Compute the singular valued decomposition of \( S \) to obtain the weights and normalise.
3. Calculate the weights \( r \), which are the first left singular values of \( S \).
4. Calculate the score vector, \( t = X r \).
5. Calculate the loadings, \( p = \frac{X^T t}{t^T t} \).
This sequence is repeated until the desired number of latent variables, $k$, is obtained. Then the regression coefficients are obtained as $B = RT^{-1}Y$, where $T = [t_1, t_2, \ldots, t_k]$ and $R = [r_1, r_2, \ldots, r_k]$. Output prediction is obtained by $\hat{Y} = XB$.

Additionally the SIMPLS method’s formulation meets the following conditions (De Jong, 1993):

1. The covariance is truly maximised, $MAX = u_k^T t_k = q_k^T (Y^T X) \cdot r_k$.
2. As with NIPALS, weights and loadings are normalised.
3. Like NIPALS, orthogonality of the scores is ensured.

The SIMPLS algorithm is the PLS method of choice in this study, as it has been shown to have slightly improved accuracy when the data is multivariate when compared with the NIPALS algorithm. This improvement in accuracy is due to maximising the covariance.

Kernel representations of the PLS method have been presented to further the calculation efficiency. Rather than calculating the PLS parameters using the full size data matrices of $X$ and $Y$, Lindgren et al. (1993) proposed using smaller "kernel" matrices. Dayal and MacGregor (1997) further simplified this approach by only using the covariance matrices $X^T X$ and $X^T Y$ in the calculation of the PLS model, where $X^T Y \cdot Y^T X$ is used. This was again further extended so the covariance matrices were updated recursively, and this was termed Recursive PLS (Dayal and MacGregor, 1997a).

The covariance matrices used to calculate the weights and the loadings for the regression vector calculation are updated as shown below as shown in equations 5-31 and 5-32, where $\lambda$ is a forgetting factor. This is a constant that is used to reduce the weighting of the old data so the system is more responsive to dynamic changes as it gives new data higher priority.

$$
\left(X^T X\right)_{r+1} = \lambda \cdot \left(X^T X\right)_r + x_{r+1}^T x_{r+1} \tag{5-31}
$$

$$
\left(X^T Y\right)_{r+1} = \lambda \cdot \left(X^T Y\right)_r + x_{r+1}^T y_{r+1} \tag{5-32}
$$

Although widely used, PLS is still a biased estimator, and the kernel matrix is the most common form of PLS used in computer software for calculating PLS models. In this work the covariance matrices are calculated and then the SIMPLS algorithm is applied. When there is noise or disturbances present in the system, as with the examples shown earlier, the $Y$ in $X^T Y$ and $Y^T X$ will be correlated with any noise, which will result in a biased PLS estimator. This problem is shown with the aid of an example in the further sections of this chapter.
As PLS is being used with increasing regularity for modelling dynamic systems with the intention of applying control, it is worth re-iterating the bias problem. The work of Kaspar and Ray (1993) utilises PLS to compensate for the constantly changing process dynamics when identifying models for process control. Similarly to this work, Lakshminarayanan et al. (1997) added a nonlinear extension to their dynamic PLS method to apply control to nonlinear systems, where PLS is used dynamically to model the linear parts of the process. MacGregor et al. (1994) and Kourt (2003b) used PLS to model and apply control to batch processes. Mc Avoy et al. (2001) used PLS modelling to optimise and shorten the batch times of a polymer process. Lauri et al. (2010) and Rossiter et al. (2010) propose the use of PLS models in MPC controllers. More recently, Prívara et al. (2013) propose using PLS in MPC control schemes for climate management in buildings. All the mentioned studies overlook the bias existent in the PLS method in the presence of noise. This thesis proposes further improvements to the standard PLS algorithm by incorporating more traditional system identification methods into the model identification exercise. The next section presents the recursive least squares and the unbiased version of the Recursive Least Squares (RLS).

5.1.4 Recursive Least Squares, Unbiased Recursive Least Squares and Unbiased PLS

Recursive computations offer significant benefits in real applications, as they require less computing power and can adapt to changing process characteristics. The unbiased RLS algorithm is the main choice for adaptive controllers in industry today and is also used for the identification of fixed models in static controllers. First, the conventional RLS is presented.

Assuming the system can be described as:
\[ y(t) = x(t) \cdot \theta + \varepsilon(t), \] where \( \varepsilon(t) \) is the equation error.

\[ x(t) = [-y(t-1) - y(t-2)]... - y(t-n) \ u(t-1) \ u(t-2)... \ u(t-n)] \]

and \( \theta = [a_1 \ ... \ a_n \ b_1 \ ... \ b_n]^T \) is the vector of parameters.

The cost function is formulated as:
\[ J_{RLS} = \frac{1}{T} \sum_{k=1}^{T} \varepsilon(k)^2 = \frac{1}{T} \sum_{k=1}^{T} (y(k) - x(k) \cdot \theta)^2 \]

Where \( T \) represents the dependence on time of \( J \) and \( \hat{\theta} \).

The coefficients are then estimated as:
\[ \hat{\theta} = \left[ \sum_{k=1}^{T} x^T(k) \cdot x(k) \right]^{-1} \left[ \sum_{k=1}^{T} x^T(k) \cdot y(k) \right] \]

To introduce recursive calculations, the parameter vector estimate can be written as:
\[ \hat{\theta}(t) = P(t) \cdot \left[ \sum_{k=1}^{t} x(k)^T y(k) \right] \] 5-35

Where \( P(t) \) is assumed to be:

\[ P(t) = \left[ \sum_{k=1}^{t} x(k)^T x(k) \right]^{-1} \] 5-36

\( P(t) \) is iterated as follows:

\[ P(t + 1)^{-1} = P(t)^{-1} + x(t + 1) \cdot x^T(t + 1) \quad \text{(Sandoz and Swanik, 1972)} \] 5-37

When expanded, equation 5-37 becomes:

\[ P(t + 1)^{-1} = \left[ \sum_{k=1}^{t+1} x(k)^T y(k) \right] = \left[ \sum_{k=1}^{t} x(k)^T y(k) \right] + x(t + 1)^T x(t + 1) \] 5-38

For the iteration of \( \hat{\theta}(t) \) at time \( (t+1) \), \( \hat{\theta}(t+1) \) is written as:

\[ \hat{\theta}(t + 1) = P(t + 1) \cdot \left[ \sum_{k=1}^{t} x(k)^T y(k) + x(t + 1)^T y(t + 1) \right] \] 5-39

Now introducing a \( P(t)^{-1} \cdot P(t) \) operator and using equation 5-35, the next iteration becomes:

\[ \hat{\theta}(t + 1) = P(t + 1) \cdot \left[ (P(t)^{-1} \cdot P(t)) \cdot \sum_{k=1}^{t} x(k)^T y(k) + x(t + 1)^T y(t + 1) \right] \] 5-40

Further details of this algorithm can be found in Åström and Wittenmark (1995).

5.1.4.1 Recursive Least Squares and Unbiased Recursive Least Squares

The derivation of the unbiased RLS algorithm is outlined below:

First \( \hat{\theta}(t+1) \) is rewritten as:

\[ \hat{\theta}(t + 1) = P(t + 1) \cdot \left( P(t)^{-1} \hat{\theta}(t) + x(t + 1) \cdot y^T(t + 1) \right) \] 5-41

Then using equation 5-37, equation 5-41 becomes:

\[ \hat{\theta}(t + 1) = P(t + 1) \cdot \left[ \left( P(t + 1)^{-1} - x(t + 1)^T \cdot y^T(t + 1) \right) \cdot \hat{\theta}(t) + x(t + 1)^T \cdot y(t + 1) \right] \] 5-42

Which then can be written as:

\[ \hat{\theta}(t + 1) = \hat{\theta}(t) + P(t + 1) \cdot x(t + 1)^T \left( y(t + 1) - x(t + 1) \hat{\theta}(t) \right) \] 5-43
Now $\hat{\theta}(t+1)$ can be represented as a series of recursive representations:

$$\hat{\theta}(t+1) = \hat{\theta}(t) + K(t+1) \cdot \varepsilon(t+1)$$  \hfill 5-44

$$\varepsilon(t+1) = y(t+1) - x(t+1) \cdot \hat{\theta}(t)$$  \hfill 5-45

$$K(t+1) = P(t+1) \cdot x(t+1)^T$$  \hfill 5-46

For the next iterative calculation, expression 5-37 is used. The matrices can be inversed using the matrix inversion lemma:

$$[A + BCD]^{-1} = A^{-1} - A^{-1} B \left[C^{-1} + DA^{-1} B\right]^{-1} DA^{-1}$$  \hfill 5-47

Inverting $P(t+1)^{-1}$ gives the following expression:

$$P(t+1) = \left(\sum_{s=1}^{t} x(s)^T x(s) + x(t+1)^T x(t+1)\right)^{-1}$$  \hfill 5-48

$$P(t+1) = P(t) - P(t) \cdot x(t+1)^T x(t+1) \cdot P(t) \cdot \left(1 + x(t+1) \cdot P(t) \cdot x(t+1)^T\right)^{-1}$$  \hfill 5-49

This can then be substituted into 5-46, giving:

$$K(t+1) = P(t+1) \cdot x(t+1)^T = P(t) x(t+1)^T \cdot \left(1 + x(t+1) \cdot P(t) \cdot x(t+1)^T\right)^{-1}$$  \hfill 5-50

Using this result, the recursive formulas can be rewritten without the need for matrix inversions in the calculations:

$$\hat{\theta}(t+1) = \hat{\theta}(t) + K(t+1) \cdot \varepsilon(t+1)$$  \hfill 5-51

$$\varepsilon(t+1) = y(t+1) - x(t+1) \cdot \hat{\theta}(t)$$  \hfill 5-52

$$K(t+1) = P(t+1) \cdot x(t+1)^T = P(t) x(t+1)^T \cdot \left(1 + x(t+1) \cdot P(t) \cdot x(t+1)^T\right)^{-1}$$  \hfill 5-53

$$P(t+1) = P(t) - P(t) \cdot x(t+1)^T x(t+1) \cdot P(t) \cdot \left(1 + x(t+1) \cdot P(t) \cdot x(t+1)^T\right)^{-1}$$  \hfill 5-54

The inclusion of a forgetting factor will slightly modify expression 5-54 to:

$$P(t+1) = \frac{P(t) - P(t) \cdot x(t+1)^T x(t+1) \cdot P(t) \cdot \left(1 + x(t+1) \cdot P(t) \cdot x(t+1)^T\right)^{-1}}{\lambda}$$  \hfill 5-55

Here $\varepsilon(t+1)$ is treated as a prediction error, which is different from the traditional least squares approach where the error to be minimised is the equation error. If $\varepsilon(t+1)$ is small, the estimated parameter vector $\hat{\theta}(t+1)$ should provide an accurate model. $K(t+1)$ is the weighting vector modifying each parameter on every iteration to minimise the error.
The Unbiased RLS (uRLS) method has a slight tweak to the presented RLS method. Instead of using the next real value of \( y(t) \) when updating \( x(t) \), the predicted \( \hat{y}(t) \) is used, as shown in expression 5-56:

\[
x(t) = [-\hat{y}(t-1) - \hat{y}(t-2) \ldots - \hat{y}(t-n) \ u(t-1) \ u(t-2) \ldots \ u(t-n)]
\]

This approach ensures that \( x(t) \) contains data that is uncorrelated with noise. Further improvements to the uRLS algorithm can be implemented using Biermann’s work on UD covariance factorisation (Bierman 1976). This algorithm can also be updated by inflating the diagonal elements of the \( D \) matrix to very high values (Åström and Wittenmark, 1995).

This is formulated as follows:

\[
Y = X \cdot \theta + \varepsilon
\]

The joint covariance matrix for \( y(t) \) and \( x(t) \) is written as:

\[
R = \begin{bmatrix} X^T \cdot P \cdot X & X^T \cdot P \\ P \cdot X & P \end{bmatrix} + \begin{bmatrix} \sigma^2 & 0 \\ 0 & 0 \end{bmatrix}
\]

The \( P \) matrix can be decomposed as \( P = L \cdot D \cdot L^T \), where \( L \) is a lower triangular matrix and \( D \) a diagonal matrix. The joint covariance matrix is then re-written:

\[
R = \begin{bmatrix} X^T \cdot L \cdot D \cdot L^T \cdot X & X^T \cdot L \cdot D \cdot L^T \\ L \cdot D \cdot L^T \cdot X & L \cdot D \cdot L^T \end{bmatrix} + \begin{bmatrix} \sigma^2 & 0 \\ 0 & 0 \end{bmatrix}
\]

This can now be decomposed as:

\[
R = \begin{bmatrix} 1 \ X^T L \\ 0 \ L \end{bmatrix} \begin{bmatrix} \sigma^2 & 0 \\ 0 & D \end{bmatrix} \begin{bmatrix} 1 \ 0 \\ L^T X \ L^T \end{bmatrix}
\]

Then for recursive updates the estimated version can be written as:

\[
R = \begin{bmatrix} 1 & 0 \\ K & \hat{L} \end{bmatrix} \begin{bmatrix} \hat{\sigma}^2 & 0 \\ 0 & \hat{D} \end{bmatrix} \begin{bmatrix} 1 \ K^T \\ 0 \ \hat{L}^T \end{bmatrix}
\]

The recursive estimate of the parameter vector can be written as:

\[
\hat{\theta} = \theta + K \ (Y - X^T \hat{\theta})
\]

Again here, \( K \) matrix is the weights, which are also described as the gains, used to update the calculated coefficients in each iteration.

5.1.4.2 Unbiased PLS

This method employs a similar idea to the uRLS method, where the predicted output is fed back into the input data matrix and used in the recursive calculations, thus providing an uncorrelated output estimate that does not contain noise.
The same approach is used where the output calculated through the uRLS method is fed into the PLS algorithm. The kernel version of PLS covariance matrices is used to represent this. Instead of updating $X^TY$ as follows:

$$
(X^TY)_{k+1} = (X^TY)_k + X^T_{k+1}Y_{k+1}
$$

The estimate from uRLS $\hat{Y}_{r+1}$ is used:

$$
(X^TY)_{r+1} = (X^TY)_r + X^T_{r+1}\hat{Y}_{r+1}
$$

Where $\hat{Y}_{r+1}$ is the next predicted value of $Y$. Once the covariance matrices are obtained the SIMPLS algorithm is applied. The predicted output $\hat{Y}_{r+1}$, obtained by using the RLS algorithm is similar to that of the one obtained with the PLS algorithm when using training data. However the improvements with the PLS algorithm are observed when the predictions are obtained using unseen data. Examples of the presented methods are given in the following section.

### 5.1.5 OLS, PLS and RLS Modelling Examples

This section presents an example of a Single Input Single Output (SISO) system, where an identification study is carried out and OLS, PLS and RLS methods are applied. This simulated system is a second order model with low pass filtered noise as disturbance. The system consists of the transfer function $G(q)$ and a disturbance signal $v(t)$.

$$
y(t) = G(q) \cdot u(t) + v(t) = \frac{q^{-1} + 0.5q^{-2}}{1 - 1.5q^{-1} + 0.7q^{-2}} \cdot u(t) + v(t)
$$

Where the disturbance $v(t)$ is low pass filtered white noise;

$$
v(t) = \frac{1}{1 - 0.9q^{-1}} \cdot e(t)
$$

The aims of using this example are:

- Initial testing of the biased and unbiased least squares methods.
- Investigating the effects of noise.

The first example given is when there is no noise present. The identification input and the system’s output is shown in Figure 5.1. The signal used is a GBN signal of clock period 8 with 500 samples. The test signal is a step input at the 50th sample out of 100 samples. The compared methods are
OLS, RLS, uRLS, uPLS and PLS, where PLS will be referred to as bPLS (biased PLS). The number of latent variables used to describe the data for the PLS models was determined by using cross validation, where the data was split into several subgroups and with half of the groups, models were built using a range of latent variables. Each model was then tested with the remaining data sets to determine the most suitable number of latent variables.

![Identification Signal](image1.png)

![Output Signal](image2.png)

**Figure 5.1; Input and Output Signals**

![Step Test](image3.png)

**Figure 5.2; Step Response of the Actual System and Models of the System**
Figure 5.2 shows the step response of the actual system and the models identified using the three biased methods. As there is no noise, the identified models are perfect and there are no observable errors. All the bRLS and the bPLS methods converge to the OLS result as expected. This is because the system is SISO and there are no correlations to consider, therefore all the least squares methods calculate the same solution.

The next example is when a low pass filtered noise signal is introduced at the output, where the signal to noise ratio (SNR) is kept at 10dB for the length of the identification samples. The input and output is shown in Figure 5.3 alongside the noise signal. The number of latent variables was selected in a similar manner to the no noise test, and four latent variables were used.

Figure 5.3; Identification and Noisy Output Signals

Figure 5.4 shows the step response of the actual system and the prediction of the biased models. This figure shows that there is now a small error on the biased methods caused by the noise.
Figure 5.4; Step Response of the Actual System and Models of the System

Figure 5.5; Step Response of the Actual System and Models of the System

Figure 5.5 shows the actual output along with the estimates made by the unbiased methods of uRLS and uPLS. Similar to the biased methods, uPLS converges to uRLS as there are no correlations to deal with. However in this case, both of the identified models show almost zero error, due to the unbiased nature of their predictors. The next example is when the SNR is increased to 1dB. The input, output is shown in Figure 5.6, where the increased noise can be observed.
The step responses of the models generated with the biased methods is shown in Figure 5.7, where the increase in noise has increased the error on the predicted models.

Figure 5.8 shows the response of the models determined by the unbiased methods. Whilst a small error does exist in the predictions, the size of this error is significantly less than that obtained using
the biased methods. This small error can be eliminated by increasing the number of samples used to identify the models.

![Step Test](image)

**Figure 5.8; Step Response of the Actual System and Models of the System**

This short example has demonstrated the necessity of unbiased methods and how unwanted noise can have a negative effect on the system identification exercise. In addition, the application of the uRLS prediction in the PLS algorithm to create an unbiased PLS model provides the justification of this powerful concept, although the full benefits of using PLS, rather than other identification methods, has not been demonstrated in this example. In addition to recursive estimation techniques, output error and prediction error methods have also been developed for eliminating model bias. These techniques are presented in the following section.

### 5.1.6 Output Error Method and Prediction Error Methods

#### 5.1.6.1 Output Error Method

The output error method is slightly different from the equation error method presented in the ARX minimisation algorithm in terms of formulation of its cost function. The output error approach is chosen as the model formation is more relevant for control systems applications and the least squares methods are chosen as they are mathematically easy to implement, but suffer from bias. If a system is represented by:

\[
G(q) = \frac{B(q)}{A(q)} = \frac{b_1 q^{-1} + b_2 q^{-2} + \ldots b_n q^{-n}}{1 + a_1 q^{-1} + a_2 q^{-2} + \ldots a_n q^{-n}}
\]

5-65
Then the output error is defined as:

\[ \varepsilon_{OE}(t) = y(t) - \frac{B(q)}{A(q)} u(t) = y(t) - \hat{y}(t) \]  \hspace{1cm} 5-66

The parameters are determined by minimising the following cost function:

\[ J_{OE} = \frac{1}{N-n} \sum_{t=n+1}^{N} \varepsilon_{OE}(t)^2 \]  \hspace{1cm} 5-67

Where \( N \) is the length of data and matrix and \( n \) is the model order. Providing the order \( n \) is correct, and the excitation signal order is at least twice the model order and the cost function converges to a global minimum, then the output error method will be stable (Zhu, 2001). Assuming the true process is given by:

\[ y(t) = \frac{B(q)}{A(q)} \cdot u(t) + r(t) \]  \hspace{1cm} 5-68

When Parseval’s formula is applied to the cost function, the following description is obtained (Zhu, 2001):

\[ J_{OE} = \frac{1}{2\pi} \int_{-\pi}^{\pi} \left( G(e^{j\omega}) - \frac{B(e^{j\omega})}{A(e^{j\omega})} \right)^2 \Phi_u(\omega) d\omega + \frac{1}{2\pi} \int_{-\pi}^{\pi} \Phi_r(\omega) d\omega \]  \hspace{1cm} 5-69

This shows the difference between the output error and equation error least squares methods where the bias error \( G(e^{j\omega}) - \frac{B(e^{j\omega})}{A(e^{j\omega})} \) was weighted by \( A^2(e^{j\omega}) \Phi_u(\omega) \). In the output error method this is only weighted by the identification signal spectrum \( \Phi_u(\omega) \), as shown in expression 5-69.

This minimisation of the cost function is undertaken with a numerical search algorithm as \( \varepsilon(t) \) is nonlinear in the \( A(q) \) parameters. The search algorithm can be chosen between various methods, such as the steepest descent, Gauss – Newton, Adaptive Gauss – Newton or Levenberg – Marquardt methods. Although any other optimisation method can be used, these are the methods employed within the System Identification Toolbox of MATLAB, which was utilised in this work. The Gauss – Newton method linearises the nonlinear relationship with a first order Taylor expansion and forms a quadratic problem. This simplifies the optimisation as the quadratic method is computationally easy to solve and does not require the second derivatives. However, this method has a disadvantage in terms of convergence, as it is not guaranteed.

The alternative is the steepest descent method, which minimises the sum of the squared error with iterative moves to update the parameters. The iterative steps are taken with regards to the inverse of the gradient of the function, which corresponds to the greatest reduction in error. However, as the
gradient gets smaller and the solution gets closer to the minimum, more and more iterations are needed and after the first few, this method is slow. The Levenberg–Marquardt method (Marquardt 1963) is an alternative to tackle these issues. It is viewed as combination of the Gauss Newton and the steepest descent methods. The way it works is by using a damping parameter. When the value of the parameter is small the Gauss – Newton method is used, and when it is large the steepest descent method is used. The System Identification Toolbox automatically tries all of the methods and uses the most optimal solution.

One disadvantage of the OE method is that it will not model the disturbances, unless the disturbance is white noise. Prediction error methods with suitable model structures such as Box – Jenkins and ARMAX can be used to model non-white noise disturbances. These methods are presented in the following section.

5.1.6.2 Prediction Error Methods

The prediction error methods are used to fully describe a system along with its disturbance dynamics. This method addresses the problems of disturbance modelling, closed loop identification and obtaining a minimum variance estimate (Zhu, 2001).

The prediction error methods follow the methodology listed below (Söderström and Stoica, 1989):

1. Analysis of the output data following the identification test, to determine trends and comparison of autocorrelation tests. This is to decide whether to include moving average or autoregressive moving average components of the system.
2. Estimation of identified models to fit the data by using optimisation methods.
3. Checking of model accuracy by testing the residuals and auto correlations.

The prediction error method starts by representing the cost function to be minimised in terms of a prediction error, as opposed to an equation error or an output error. Assuming the system can be represented as an ARX model, the prediction error can be described as follows:

\[ y(t) = \frac{B(q)}{A(q)} \cdot u(t) + \varepsilon(t) \]  

This can be re-written as:

\[ y(t) = [1 - A(q)] \cdot y(t) + B(q) \cdot u(t) + \varepsilon_{eq}(t) \]  

Then neglecting the equation error \( \varepsilon_{eq} \), the predicted output can be written as:

\[ \hat{y}(t) = x(t) \cdot \theta = [1 - A(q)] \cdot y(t) + B(q) \cdot u(t) \]
And the prediction error is represented as:
\[ e_{pred}(t) = y(t) - \hat{y}(t) \]

As mentioned in Chapter 2, the prediction error methods use various filters to model the disturbances and determine an unbiased estimate. The different filter types create different methods. In this study the PEM method with a Box Jenkins model and the ARMAX method are used. These are presented in the following section.

### 5.1.6.3 ARMAX Method

The first prediction error method presented is the ARMAX method (Astrom and Bohlin, 1966) which has the following structure:

\[
y(t) = \frac{B(q)}{A(q)} \cdot u(t) + \frac{C(q)}{A(q)} \cdot e(t)
\]

Where:
\[
A(q) = 1 + a_1 q^{-1} + a_2 q^{-2} + \ldots + a_n q^{-n}, \quad B(q) = b_1 q^{-1} + b_2 q^{-2} + \ldots + b_n q^{-n} \quad \text{and}
\]
\[
C(q) = 1 + c_1 q^{-1} + c_2 q^{-2} + \ldots + c_n q^{-n} \quad \text{and} \quad e(t) \quad \text{is white noise. The ARMAX method describes the error as a moving average noise sequence.}
\]

The predictor can be written as:
\[
\hat{y}(t) = \frac{B(q)}{A(q)} \cdot u(t) + \left[ 1 - \frac{C(q)}{A(q)} \right] y(t)
\]

Then the loss function to be minimised is:
\[
V_{ARMAX} = \frac{1}{N} \sum_{i=1}^{N} \epsilon(t)^2 = \frac{1}{N} \sum_{i=1}^{N} \left( \frac{1}{C(q)} \right) \left[ A(q) \cdot y(t) - B(q) \cdot u(t) \right]^2
\]

Where, \( N \) is the length of the data matrix. The search algorithm for this method is one of the numerical optimisation methods mentioned previously, such as Gauss-Newton. The main aim of this modelling approach is to model the equation disturbance, which gives rise to the use of the \( A(q) \) polynomial as part of the transfer function as well as the disturbance. For systems with slow disturbances, the ARMAX method can be extended with an integrator to become an ARIMAX model (Ljung, 1987). An alternative to ARMAX is the PEM method, where the plant and the disturbances are modelled separately.
5.1.6.4 PEM

This method uses the Box-Jenkins structure (Box, Jenkins, and Reinsel, 1970), representing the system as:

\[ y(t) = \frac{B(q)}{A(q)} \cdot u(t) + \frac{C(q)}{D(q)} \cdot e(t) \]

where:

\[ A(q) = 1 + a_1q^{-1} + a_2q^{-2} + \ldots + a_nq^n \]
\[ B(q) = b_1q^{-1} + b_2q^{-2} + \ldots + b_nq^n \]
\[ C(q) = 1 + c_1q^{-1} + c_2q^{-2} + \ldots + c_nq^n \]
\[ D(q) = 1 + d_1q^{-1} + d_2q^{-2} + \ldots + d_nq^n \]

The predictor can be written as follows:

\[ \hat{y}(t) = \frac{D(q)B(q)}{C(q)A(q)} \cdot u(t) + \left[ \frac{1 - D(q)}{C(q)} \cdot y(t) \right] \]

The parameters are determined by minimising the following cost function:

\[ J_{BJ} = \frac{1}{N} \sum_{i=1}^{N} e(t)^2 = \frac{1}{N} \sum_{i=1}^{N} \left( \frac{D(q)}{C(q)} \cdot \left[ y(t) - \frac{B(q)}{A(q)} \cdot u(t) \right] \right)^2 \]

where \( N \) is the length of the data matrix. The search algorithm for this method is numerical optimisation. The part provides a disturbance model, which is beneficial when the disturbance is coloured. This is more realistic in terms of application to industrial plants, where inherent disturbances exist that cannot be represented as an output error and will not be white noise. The extra complexity of the PEM method’s cost function means it is calculated slower in comparison with the other methods.

5.1.7 Hybrid PLS Methods

In addition to the standard methods, new hybrid PLS methods have been proposed in this thesis with a view to be used in MPC schemes. The main aim of these hybrid methods is to improve the performance of the standard algorithms in the cases of high noise, correlated data and short data sets. As presented, the standard PLS method is biased, and in the presence of noise creates inaccurate models. However, PLS has advantages as extensively presented in literature when used with highly correlated data sets, where traditional empirical methods are said to fail.

The hybrid methods aim to combine the benefits of both system identification empirical modelling techniques, such as disturbance modelling, and the advantages of the PLS method where the covariance between the input and output is maximised. Providing the most accurate output prediction to the PLS model should in theory give the best unbiased model, when the data set is both noisy and
correlated. As all the hybrid methods are potentially unbiased, presenting them individually will help in distinguishing their differences. First the unbiased RLS-PLS method is presented.

5.1.7.1 Unbiased RLS – PLS

The Unbiased RLS-PLS (uRLS-PLS) method has been proposed recently as an alternative to eliminate the bias in PLS (Awad, 2008). When building this PLS model, one step ahead predicted values of the output data matrix \( \hat{Y} \) obtained using the uRLS method are used instead of the real output data matrix \( Y \). The model has the following structure when expressed in terms of the PLS kernel algorithm, where \( (X^T Y)_{r+1} = (X^T Y)_{r} + x^T_{r+1} \hat{y}_{r+1} \) is the traditional representation with the next value of the \( Y \) output matrix.

\[
(X^T Y)_{r+1} = (X^T Y)_{r} + x^T_{r+1} \hat{y}_{r+1}
\]

is the uRLS-PLS update where \( \hat{y}_{r+1} \) is estimated using uRLS.

This algorithm was demonstrated comparatively by Awad (2008) against traditional PLS and was shown to work in an MPC scheme in closed loop, when short data sets were used to adapt the model. However, it was not tested with a highly correlated data set.

5.1.7.2 OE – PLS, PEM-PLS, ARX-PLS and ARMAX-PLS

The OE-PLS, PEM-PLS, ARX-PLS and ARMAX-PLS are the proposed new alternatives to the uRLS-PLS presented previously. These new methods are investigated in this thesis alongside the more traditional methods.

The new methods work by using the one step ahead predictions of the traditional methods and then building the hybrid PLS models. The PLS models are then used for multistep ahead predictions with a view to be applied to MPC schemes. The multistep ahead calculations form the all important prediction horizon of the MPC controller are important for the controller’s performance.

If \( \theta \) is the parameter vector of the identified system, then we can represent the system as:

\[
y(t) = x(t) \cdot \theta + \varepsilon(t), \text{ where } \varepsilon(t) \text{ is the equation error.}
\]

\[
x(t) = [-y(t-1) - y(t-2) \cdots - y(t-n) \ u(t-1) \ u(t-2) \cdots u(t-n)]
\]

Where \( \hat{\theta} \) is the estimated parameter vector the one step ahead prediction is:

\[
\hat{y}(t+1) = x(t+1) \cdot \hat{\theta} + \varepsilon(t), \text{ where}
\]

\[
x(t+1) = [-y(t) - y(t-1) \cdots - y(t+1-n) \ u(t) \ u(t-1) \cdots u(t+1-n)]
\]

Then the multistep ahead prediction can be shown as:

\[
\hat{y}(t+p) = x(t+p) \cdot \hat{\theta} + \varepsilon(t), \text{ where } p \text{ is the prediction horizon, and } x(t+p) \text{ is formed of:}
\]

\[
x(t+p) = [-y(t+p-1) \cdots - y(t+p-n) \ u(t+p-1) \cdots u(t+p-n)].
\]
Again represented in terms of the PLS kernel form, the updated covariance matrices of each of the hybrid models are described as:

\[
\begin{align*}
(X^T Y)_{t+1} &= (X^T Y)_t + x^T_{t+1} \hat{\eta}_{ARX_{t+1}} \quad \text{for ARX-PLS.} \\
(X^T Y)_{t+1} &= (X^T Y)_t + x^T_{t+1} \hat{\eta}_{ARMAX_{t+1}} \quad \text{for ARMAX-PLS.} \\
(X^T Y)_{t+1} &= (X^T Y)_t + x^T_{t+1} \hat{\eta}_{OE_{t+1}} \quad \text{for OE-PLS.} \\
(X^T Y)_{t+1} &= (X^T Y)_t + x^T_{t+1} \hat{\eta}_{PEM_{t+1}} \quad \text{for PEM-PLS.}
\end{align*}
\]

Here, \( \hat{\eta}_{t+1} \) is estimated using the previously presented methods, such as ARX, ARMAX, OE and PEM. The benefits of the hybrid models are that they combine the disturbance modelling properties of methods such as ARMAX and PEM and the computation simplicity of PLS to give more accurate models when the traditional methods fail to do so. The next section shows the application of the recently presented OE and PEM methods and the PEM-PLS hybrid method to the SISO system.

### 5.1.8 Further Modelling Examples

This section presents the extension of the SISO test presented previously in section 5.15 using the OLS, RLS and PLS methods. Additionally, the one step ahead predictions of the OE, PEM and one of the hybrid methods, PEM-PLS are presented with the aim to justify the need for unbiased estimates, and to show the performance of OE and PEM methods against the least squares methods. The first example is when the disturbance is only white noise. If the system is re-written:

\[
y(t) = G(q) \cdot u(t) + v(t) = \frac{q^{-1} + 0.5q^{-2}}{1 - 1.5q^{-1} + 0.7q^{-2}} \cdot u(t) + \epsilon(t)
\]

Where, \( \epsilon(t) \) is white noise with zero mean. The input and output signals are shown in Figure 5.9 alongside the disturbance signal. The identification signal is again a GBN signal with clock period of 8.
The predicted models are then tested with a step test and the responses are shown in Figure 5.10. The OLS, PLS and bRLS methods again give a biased prediction. The uRLS and uPLS-RLS methods converge to each other, and the error is smaller than the biased methods. The OE, PEM and PEM-PLS methods identify models with almost zero error. PEM-PLS converges to the PEM solution for the same reasons explained previously, as there are no correlations to capture the PLS model converges to the PEM model.
The next example is when the noise is filtered to create coloured noise. This adds dynamics to the noise and the expected result is for the OE method to identify a biased model, as it cannot model the noise dynamics. The PEM and the PEM-PLS methods should identify models with lower errors. The input output sequence and the SNR are the same as presented in Figure 5.6. The estimated step responses are shown in Figure 5.11. As expected, the PEM and PEM-PLS method has modelled the disturbance, created unbiased estimates and OE has shown to have a small error.

Analysing only the step responses gives an insight into the predictive capabilities of the models. However, the results may vary if the experiment is repeated, as there will be slight changes in the noise as this relies on random number generation. To get a more accurate picture, the tests presented were repeated 100 times for different identification signal lengths, with two different SNR levels, one at 10dB and the other at 2.5dB. Table 5.1 presents the signals used to identify the models. The system modelled is second order, so a signal with excitation order of four or above should be sufficient for identification. The GBN signal is a highly exciting signal and is of order 50. The mean condition number of the data matrices used for building models is presented in the final column. The 2-norm condition number is used where the largest singular value is divided by the smallest singular value of the data matrix. The lower this number is, the better conditioned the data set. As the number gets higher, the more likely the data is to be ill conditioned, which will make obtaining working models difficult.
The results presented from this point forward are based on the multistep ahead prediction capabilities of the models. A 30-step prediction horizon is used in this study, as a long prediction horizon is important in terms of application to MPC schemes, where the one-step ahead prediction accuracy is of limited importance.

<table>
<thead>
<tr>
<th>Signal Type</th>
<th>Samples</th>
<th>Signal to Noise Ratio (SNR)</th>
<th>Excitation Order</th>
<th>Mean Condition Number</th>
</tr>
</thead>
<tbody>
<tr>
<td>GBN(8)</td>
<td>500</td>
<td>10</td>
<td>50</td>
<td>24.7</td>
</tr>
<tr>
<td>GBN(8)</td>
<td>2500</td>
<td>10</td>
<td>50</td>
<td>24.5</td>
</tr>
<tr>
<td>GBN(64)</td>
<td>500</td>
<td>10</td>
<td>50</td>
<td>63.2</td>
</tr>
<tr>
<td>GBN(64)</td>
<td>500</td>
<td>2.5</td>
<td>50</td>
<td>71.1</td>
</tr>
<tr>
<td>GBN(64)</td>
<td>2500</td>
<td>10</td>
<td>50</td>
<td>66.5</td>
</tr>
<tr>
<td>GBN(64)</td>
<td>2500</td>
<td>2.5</td>
<td>50</td>
<td>75.7</td>
</tr>
</tbody>
</table>

Table 5.1; Excitation Signal Properties
Table 5.2 shows the results of the 100 repeated runs, and presents them as the root mean squared error at the 30-step ahead prediction horizon point. The models are compared using Root Mean Squared Error (RMSE) of prediction, and the coefficient of determination over the multistep ahead prediction range. In addition, the variance between each prediction is compared between tests, as the models are identified using different methods more than once, to guarantee consistency in the results.

The RMSE is calculated as:

$$RMSE = \sqrt{\frac{1}{s} \sum_{k=1}^{s} (y(k) - \hat{y}(k))^2}$$

Here, \(s\) is the length of the test input, \(n\) is the number of repeat runs, \(y(k)\) is the actual output and \(\hat{y}(k)\) is the predicted output at the 30 steps ahead prediction point.

<table>
<thead>
<tr>
<th>Test</th>
<th>Signal Type</th>
<th>RGS(Pass band)</th>
<th>GBN(Time)</th>
<th>SNR</th>
<th>Samples</th>
<th>PLS</th>
<th>uRLS-PLS</th>
<th>RLS</th>
<th>uRLS</th>
<th>PEM</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>GBN(8)</td>
<td>10</td>
<td>500</td>
<td>43.3</td>
<td>9.2</td>
<td>42.3</td>
<td>7.1</td>
<td>1.8</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>GBN(8)</td>
<td>10</td>
<td>2500</td>
<td>31.9</td>
<td>2.3</td>
<td>31.1</td>
<td>1.4</td>
<td>0.3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>GBN(64)</td>
<td>10</td>
<td>500</td>
<td>31.2</td>
<td>10.2</td>
<td>31.1</td>
<td>10.7</td>
<td>3.9</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>GBN(64)</td>
<td>2.5</td>
<td>500</td>
<td>130.4</td>
<td>56.1</td>
<td>125.4</td>
<td>58.5</td>
<td>24.5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>GBN(64)</td>
<td>10</td>
<td>2500</td>
<td>30.5</td>
<td>2.9</td>
<td>30.1</td>
<td>2.9</td>
<td>0.9</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>GBN(64)</td>
<td>2.5</td>
<td>2500</td>
<td>108.5</td>
<td>13.5</td>
<td>101.9</td>
<td>15.0</td>
<td>5.4</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Table 5.2; Mean Squared Error of the Repeated Tests**

The PEM method produces models that have the smallest prediction error at 30 steps ahead. The PLS and the RLS methods, which are both biased, show a similar performance with significantly larger errors in comparison with other methods. The effect of noise can be observed across all the methods where the increase in the SNR creates larger errors in tests 4 and 6. The results are in agreement with the mean condition number and the data length. For example, an increase can be observed on the overall errors for the unbiased methods and the PEM method, between Test 1 and Test 3, and also between Test 2 and Test 6, due to the increase in mean condition number of the matrices. The PLS and RLS methods are not significantly affected by the mean condition number of the matrix, as they rely on finding correlations between the variables. In this test it should be noted that the hybrid uRLS-PLS method provides smaller errors in comparison to the PLS and RLS methods, and the errors
are similar to that of the uRLS. Their results do not converge exactly to each other as there is some variation between the repeated tests. This test shows that hybrid methods have the potential to provide improved predictions when compared with the standard PLS algorithm.

To demonstrate the deterioration of the models’ predictive capabilities, the coefficient of determination $R^2$ is calculated as follows:

$$R^2 = 1 - \frac{\sum_{k=1}^{n}(y(k) - \hat{y}(k))^2}{\sum_{k=1}^{n}(y(k) - \bar{y}(k))^2}$$

Where $y(k)$ is the actual output, $\hat{y}(k)$ is the predicted output and $\bar{y}(k)$ is the mean value of the actual output. $R^2$ is a commonly used indicator of performance in literature.

Figure 5.12 shows the performance of several models identified using the various identification techniques, using a GBN signal with a clock period of 8 and a length of 2500 samples. The mean of 30 runs has been taken for this exercise. The tested methods were, uRLS, bRLS, OE, Recursive OE (ROE), bPLS, PEM and the proposed hybrid methods uRLS-PLS, bPLS, PEM-PLS, OE-PLS. From the mean $R^2$ plot it can be shown that some methods such as PEM, and uRLS suffer from model deterioration, although the deterioration is not significant. All of the hybrid methods performed without significant enough deterioration over the prediction horizon to validate their justification for use in modelling. However, just by looking at the $R^2$ plot it is difficult to judge the true performance of the models.
Figure 5.13 shows the mean squared error at the 30 step ahead prediction point, when the models are identified using a matrix containing between 300 to 9900 samples. The bPLS and RLS methods stand out with larger mean squared errors. The other methods, including the hybrids, start at a high error value when the sample lengths used for identifying models are short. This means there is less information to build the model with. Their accuracy then improves almost to zero error, as the length of samples the models are trained with increases. All the hybrid models converge to their base models, for example, PEM-PLS converges to PEM and OE-PLS converges to OE, with the exception of bPLS and bRLS.

![Graph showing mean squared error vs number of samples](image)

**Figure 5.13; Mean Squared Error at 30 Steps Ahead Prediction Point**

5.2 Batch Process Methods

This section presents the common methods used to monitor, diagnose faults and control batch processes using the reduce dimension latent variable space. As previously mentioned, batch processes, such as solvent casting, typically only have the measurements of the output quality available at the termination of the process. The following section describes the methodologies used in multivariable statistical process control.

5.2.1 Multivariable Statistical Process Control

The properties of a control system that is aimed at improving product quality in multivariable batch systems needs to satisfy certain requirements. Firstly, it must be able to utilise the process measurements taken during the batch to provide a prediction of the final product quality as the batch
progresses. This can be achieved with the aid of an empirical or numerical model used as a soft sensor. Secondly, the system should be able to determine changes in the operating conditions that will affect the product quality, which can be achieved by utilising the soft sensor to optimise the process parameters. As a final requirement the system should be able to determine faults such as malfunctioning sensors. These requirements are met through the use of the Multivariable Statistical Process Control (MSPC) techniques, such as PCA and PLS, that were first discussed in Chapter 2.

Developments in computing and sensing technology have led to the introduction of data acquisition and advanced sensory technologies that enable the user to gather information at very high sampling intervals and over long periods (Nomikos and MacGregor 1995). The quantity of process data available is considerably more than the quality data recorded in the process control charts mentioned in Chapter 2. Recently, multivariate statistical methods have been proposed as a method for analysing the large amounts of data routinely collected from industrial processes, with the intention of improving the final product quality.

The most common statistical methods used in MSPC are PCA and PLS. The utilisation of PCA and PLS with process control charts was proposed by Kresta et al. (1991) and was later applied MacGregor and Kourti (1995) amongst many others. The need to incorporate data reduction methods in to process control charts emerged from the inadequacies of the existing charts such as the Cumulative Sum (CUSUM), Exponentially Weighted Moving Average (EWMA) or Shewhart charts. The traditional process monitoring charts were mainly used with univariate data sets or to monitor independent variables. The increased amount of data collected from a process and the existence of correlations between the process variables meant that these charts were not always reliable and alternative methods were necessary. Although multivariate extensions of the traditional charts were developed, MacGregor and Kourti (1995) point out that more reliable multivariate statistical methods can be obtained by incorporating all the data in to the analysis rather than the most recent observation used in the Shewhart chart. Jackson (1973) investigated the use of the $T^2$ statistic in principal component based control charts, where $T^2$ is the measure of variation inside the model obtained using PCA. Following this Jackson & Mudholkar (1979) proposed the use of residual, or Squared Prediction Error (SPE), analysis using PCA. Residual analysis and $T^2$ is explained in the further sections.

Before attempting any modelling and analysis with multivariate data matrices, the data needs to be pre-treated, this is explained in the following section.

5.2.2 Pre-Treatment of Data and Unfolding

The pre-treatment of data is necessary as it removes the systematic variance from the measurements. This avoids variables with large ranges outweighing the other variables during
identification. The pre-treatment is accomplished by removing the mean from the actual data set and dividing by its variance as shown in equation 5-2 in Chapter 5.

When dealing with batch data, it is useful for the $X$ data matrix to go through a process called batch-wise unfolding before it can be normalised and used for modelling. The unfolding process is shown in the diagram below, where the batches of data are stacked next to each other to create a two dimensional matrix. This method of unfolding is preferred in batch processes as the structure allows the time varying characteristics of a batch process to be approximated (MacGregor and Kourti, 1995).

![Unfolding Batch Data](image)

5.2.3 $T^2$ and Squared Prediction Error (SPE)

As discussed previously PCA projects the original data matrix on to a reduced space, represented by principal components. The principal component scores can then be used to derive the $T^2$ statistic, which is also termed the Hotelling’s statistic. $T^2$ is calculated as shown below:

$$T^2 = \sum_{i=1}^{n} \frac{t_i^2}{\hat{\sigma}_i^2}$$  \hspace{1cm} 5-87

Where $t_i$ is the $i$'th score and $n$ is the number of principle components used. $\hat{\sigma}_i^2$ is the estimated variance of the $i$'th score. $T^2$ is the optimum statistical indicator that can be used to detect any general shifts in the process. It must be highlighted that the way $T^2$ is calculated means that each principle component has an equal effect regardless of the variance it captures. This causes a problem when the data set is highly correlated or ill conditioned. MacGregor et al. (1995) point out that dividing the scores with their small variances can lead to an “out-of-control” signal in the $T^2$ monitoring chart. As $T^2$ alone does not provide a good enough solution for monitoring batch processes, only the first $k$ number of principal components are reliable to test the deviations from product quality (MacGregor and Kourti 1995).
Furthermore, using only $T_k^2$ alone is not always reliable for detecting unexpected events or faults and to overcome this the SPE is recommended. SPE represents the shortest distance between the new predicted data from the projected variable plane. In the event of an unexpected fault, new correlation patterns can appear and these can be detected with SPE as the predicted data will deviate from the projected plane.

SPE, also termed the $Q$ - statistic (Jackson, 1991) is calculated as follows:

$$SPE = \sum_{i=1}^{k} (y_i - \hat{y}_i)^2$$

Where $y_i$ is the $i$th new observation, $\hat{y}_i$ is the predicted observation and $k$ is the number of process variables used in the SPE calculation.

With sufficient data collected during a batch, the $T^2$ and SPE values can be calculated and violations of their confidence limits observed and used to identify faults and changes in the operating conditions. When an unusual event occurs the covariance of the data changes, and this is detected through the high values of the SPE. Goulding et al. (2000) highlight that faults such as sensory faults are detected on the SPE chart and changes in operating conditions are detected on the $T^2$ chart. The correct selection of confidence limits is also critical in ensuring successful batch completion. In practice these limits vary for each process and are often set between 95% and 99%.

Control limits for each individual score with a significance level of $\alpha$, is calculated as follows:

$$t_{n-1,\alpha/2} s_r \left( 1 + \frac{1}{n} \right)^{1/2}$$

Where $t_{n-1,\alpha/2}$ is the critical value of the Student’s T distribution with significance $\alpha/2$, $s_r$ is the estimated standard deviation of the score, $n$ is the number of samples at sampling rate $k$. Further information on the Student’s T distribution can be found in Esbensen et al. (2002). The $T^2$ statistic for a new score vector can be calculated as using the F-distribution (Lennox et al., 2001):

$$T^2 = \frac{np(k-1)}{k-np} F_{np,k-1,\alpha}$$

Where $k$ is the number of samples and $np$ is the number of principal components used for modelling. The control limit for the SPE chart can be calculated by assuming this value follows a chi-squared distribution (Jackson and Mudholkar 1979). The limit $Q$ at the $\alpha\%$ level is shown in expression 5-91.
\[ Q = \theta_1 \left[ 1 - \frac{\theta_2 h_0 (1 - h_0)}{\theta_1^2} + z_a \frac{\sqrt{2 \theta_2 h_0^2}}{\theta_1} \right]^{1/2} \]

In 5-91 \( \theta_1 \) and \( \theta_2 \) are the trace of the residual covariance matrix to the relevant power, i.e. \( \theta_2 = \text{trace} \left( \text{cov}(X)^2 \right) \). \( h_0 \) is the function of the traces of the covariance of the residual matrix, and \( z \) is the new process measurements.

The following sections discusses how MSPC can be applied to processes and how the final product quality can be improved.

### 5.2.4 Product Quality Improvement with MSPC

As presented in the previous section, it is possible to detect faults and monitor process quality with MSPC, however the full potential of this technology is only realised if MSPC is used to improve product quality. Batch processes have inherent challenges when implementing control. The main issue is that product quality is often only available at the end of the batch and different variables can have varying sampling rates. Additionally changes in raw material quality introduce further complications. The presented methods in the previous section, PCA and PLS and the statistical indicators relevant to these methods provide a convenient method for monitoring and controlling batch processes.

The ease by which latent variable methods can be applied has made them a popular tool in real world applications by control engineers, where rigorous mathematical modelling methods can be time consuming. The inputs and outputs of batch processes are usually highly correlated, which justifies the use of latent variable methods where the data matrix dimension is significantly reduced.

Applications of MSPC techniques for active process control has only been investigated relatively recently. Jaeckle and MacGregor (1998) proposed the use of PCA and PLS to calculate new conditions for the recipe in a batch process during product grade transitions. In this study all of the data from the process was projected on to a reduced dimension space and the necessary control moves were determined in the latent variable space. This reduced space, eased the demand on computing and saved significant time on control move calculations. Product quality improvements were made in this example by using the model to estimate product quality, which was then analysed to determine whether the process was within the “in-control” region or not. If control action was required the control move was estimated in the latent variable space and the model was then inverted to establish the manipulated variable trajectory changes. The regression vector used in this approach can be represented as follows (Jaeckle and MacGregor, 1998):
Kano and Nakagawa (2008) outline a “data driven quality improvement” scheme where MSPC and offline optimisation were utilised to improve the process from batch to batch. Following the modelling of the system using MSPC techniques, the proposed methodology optimised the operating conditions and the operation profile. The optimisation of the operating conditions was similar to that used by Jaeckle and Macgregor (1998), however it considered the desired quality and the constraints of the system.

Similar MSPC quality improvement methods exist throughout literature (Keats et al., 1994), (Yang and Sheu, 2006), (Ferrer 2007). In addition to these, alternative methods have been proposed that use MSPC techniques and then integrate them with more advanced approaches, such as artificial neural networks or Hammerstein models. Kramer (1992) showed that artificial neural networks could be used in a method similar to PLS to overcome the non-linearities associated with industrial processes. Cheng (1995) proposed a multilayer neural network to detect changes in the process mean, Pham and Oztçemel (1994) used learning vector quantisation networks alongside MSPC and Barghash and Santarisi (2004) developed a pattern recognising neural network to learn the control charts and relate them to causes. MSPC techniques have also been proposed to be used within MPC schemes, where Chen, McAvoy, and Piovoso (1998) proposed a model predictive controller that applied control in the latent variable space. Laurí et al., (2010) also used PLS in their model based controller.

In addition to fault detection and quality monitoring, soft sensing is also an important aspect of quality improvement in conjunction with MSPC. Soft sensors can be estimated using an array of methods such as neural networks and state space methods, however the use of PLS is common as it is relatively easy to implement as shown by Kamohara et al. (2004). Soft sensors have advantages in the absence of a physical sensor or during a sensor fault to estimate the actual measured variable value. Furthermore, physical sensors can be costly in terms of maintenance, calibration, noise and they may contain large time delays. During sensor failures, such as a line blockage a software sensor would be of importance to keep the process under control by preventing missing and inaccurate data from entering the controller.

In the case of solvent casting the mentioned methods are highly relevant. As outlined in Chapter 3 control of the solvent casting process is difficult and optimising the operating conditions for desired material qualities is a more feasible option and the use of MSPC methods is recommended. In the presence of relevant data the prediction of material properties would be possible. PLS could be used as a soft sensor to estimate the type of film that will be produced, as well as calculating modified operating conditions in the presence of raw polymer quality variations. The use of the PCA method
would highlight the most significant contributors to the process and the level of correlation that exists within the variables. The solvent casting method is likely to suffer from batch to batch variations due to slight changes in raw material properties, and the changes in mixture properties over time due to variations in ambient temperature and time. Using MSPC techniques the scores from the previous batches can be incorporated into the data matrix for modelling and monitoring. Flores-Cerrillo and MacGregor (2004a) outline an iterative method to track the batch to batch variations in the model. With the aid of historical data, and data collected from a DOE, the end point quality of the film could be predicted, and by using the $T^2$ and SPE indicators it would be possible to monitor the variations in the batch and be able to assign causes to some of the variations in quality.

To track the batch to batch variations, iterative learning techniques can also be applied. In standard MSPC the next batch is operated by tracking the desired reference trajectory which may be subtly altered depending on the monitoring statistics. In iterative learning the previous batch inputs as well as the tracking error are utilised, but the tracking error can be utilised differently depending on the system, such as normal gain block or a weighted filter block imitating the system dynamics and the forgetting factor. The main principle is the use of past iterations to calculate the best range of inputs for the next batch (Moore et al., 1992).

5.3 Summary

This section has explained the theories behind the modelling methods most widely used in existing literature and also in this study. These methods are used to build models within model predictive control schemes, increase product quality and provide process monitoring through MSPC methods. The tests conducted on the SISO system were presented as a proof of concept of the hybrid methods, the effect of noise on the model qualities and the effect of multistep ahead prediction. The findings can be summarised as:

- White noise has an unwanted effect on bPLS, OLS and RLS model qualities.
- The uRLS, uRLS-PLS and OE methods can be used to create unbiased models.
- Coloured noise has a negative effect on all the methods, but most significantly in bPLS, OLS, RLS and OE.
- Using PEM to model the disturbances creates models with lower errors in comparison with the biased methods and the OE method. Furthermore, the OE method produces models which contain a bias when the noise is not purely white.
- Sample length has an effect on empirical modelling methods such as OE and PEM, and as the length of the data increases the modelling error decreases. This also valid for the bPLS and bRLS methods but the effect is not so noticeable when using these techniques.
- MSPC has the potential to offer an elegant solution to monitoring and improving the quality of the solvent casting process with the aid of DOE and control-quality monitoring charts.
• Alternative methods such as neural networks or iterative learning could be used alongside MSPC to apply a quality improvement system to the process and to make the best use of the historical batch to batch variation.

Although the proof of concepts and the effects of various parameters affecting model quality were analysed in this section, the methods need to be tested on multiple input systems with correlations. The next chapter presents more realistic systems that are representative of the solvent casting and electrospinning processes in structure and discusses how these methods can be applied to these two real physical systems.
6 Application of Numerical Methods to MIMO Systems

This chapter presents the work undertaken on the system identification and modelling part of this thesis on Multiple Input Multiple Output (MIMO) systems. Firstly, an example of the benefits of PLS is demonstrated on the PENSIM batch process simulator (Birol et al., 2002), which operates similar to the solvent casting process in that the output quality measurement is only available at the end of the batch. Following this, the modelling methods and identification signals are tested and justified on two MIMO systems, one being the benchmark MIMO Wood and Berry Distillation Column simulation (Wood and Berry, 1973) and the other a highly correlated system with ten inputs and a single output. The examples selected here are used due to their similarities with the solvent casting and electrospinning processes. The application of system identification techniques to these simulations is used to demonstrate the proof of principles, that these methods can be applied to aid in the production of materials for nerve and tendon regeneration.

6.1.1 Application of PLS to PENSIM

This study shows the application of PLS modelling to a batch process called PENSIM. PENSIM is a penicillin production simulator that simulates a batch/fed-batch process. The simulator was developed by Birol et al. (2002) to test batch process monitoring, application of advanced control techniques, multivariate techniques and fault diagnosis. This process is used due to its similarities with solvent casting at a systems level. Both have multiple inputs, and they undergo a reaction; in solvent casting evaporation takes place, and in PENSIM the penicillin is produced as a by product from the biomass, and the output for both is only measured at the end of the batch. The input and output variables within this process are listed in Figure 6.1.

Penicillin is used in antibiotics and is produced by fermenting filamentous microorganisms. The produced penicillin is not directly correlated with the growth of the microorganisms. In industrial practice, the microorganisms are first grown through a controlled batch stage until their biomass concentration reaches a suitable level for penicillin production. When the penicillin starts to
The substrate feed rate is slowed down and added at a lower rate, which is the fed-batch stage. The aim during this stage of operation is to keep the biomass levels at a low rate while the penicillin synthesis takes place. PENSIM simulates this process by mathematically modelling the penicillin growth process. In addition to the core penicillin algorithm, the fermenter temperature controllers for the hot and cold water feeds are implemented as individual PID controllers. The simulation allows the user to adjust the operating conditions, introduce disturbances and determines the outputs of the process by taking these effects into account.

Figure 6.2 shows the batch and the fed-batch stage of the penicillin fermentation process. For the first 60 hours the biomass of the microorganism increases rapidly, then after this at a more controlled rate until the end point. The variation in the biomass concentration is caused by a small amount of white noise, between 0 and 0.005, being injected to the biomass growth factor constant in the growth equation. This creates a natural variance similar to what would be observed in a real process.

The penicillin concentration is shown in Figure 6.3, where until the 60th hour of the batch the concentration level is 0 g/L. After this, the biomass begins to synthesise and glucose is slowly fed to control the reproduction of the biomass until the end of the batch.
The overall system can be represented as illustrated in Figure 6.4. The pH sensor sends feedback to the flow controller (FC) which controls the levels of acid and base being pumped into the process. These regulate the pH levels of the substrate so the environment is suitable for bacterial growth. Similarly, the temperature sensor (T) provides feedback to the flow controller (FC), which adjusts the hot and cold water mixture by controlling the valve levels. Again, this is to ensure the temperature of the substrate is at a suitable level to encourage a desired growth rate. Too hot or too acidic conditions could result in the death of the biomass.
The functional relationship between the process variables are given as follows:

\[
X = f(X, S, C_L, H^+, T), \quad S = f(X, S, C_L, H^+, T), \quad C_L = f(X, S, C_L, H^+, T)
\]
\[
P = f(X, S, C_L, H^+, T, P), \quad CO_2 = f(X, H^+, T), \quad H = f(X, H^+, T)
\]

Where \(X\) is the biomass concentration, \(S\) is the substrate concentration, \(P\) is the penicillin concentration, \(C_L\) is the dissolved oxygen concentration, \(CO_2\) is the carbon-dioxide concentration, \(H^+\) is the hydrogen ion concentration of the pH and \(T\) is the temperature of the solution in the fermenter.

The measured variables of the process are the biomass concentration, penicillin concentration, culture volume in the fermenter, \(CO_2\) level, pH level, temperature of the biomass, and the heat generated by the reaction. The biomass and the penicillin measurements are only available at the end of the batch and the penicillin concentration is the controlled variable.

The manipulated inputs of the process are aeration rate, agitator power, substrate feed rate and the substrate feed temperature. For identification purposes, a small external GBN signal is applied to the aeration rate and the agitator power input variables to enrich the data during the building of the PLS model. The water temperature and the pH values were excited with a smooth filtered PRBS sequence.
that was already inherent in the simulation. The smooth PRBS is advantageous in terms of limiting drastic changes to the set-points of the already slow temperature and pH processes. The initial set-points of the input variables were:

Aeration rate: 8 L/h  
Agitator power: 30 W  
Substrate feed rate: 0.045 L/h  
Substrate temperature: 296 K

This small amount of smooth PRBS variation is inherent in the simulation to represent the unmeasured disturbances acting on the system. These are shown in Figure 6.5.

![Figure 6.5; PENSIM Inputs](image)

The final products, biomass and penicillin of five consecutive batches are shown below in Figure 6.6. The differences in the biomass concentration are caused by the small variation in the biomass growth factor, which in turn affects penicillin production.
The four inputs were excited with a GBN sequence as shown in Figure 6.7, as a small excitation on the inputs is necessary for capturing the dynamics of the system for predictive models. The modelling method used in this study was the unfolded-PLS method. This was selected as the final output variables (penicillin and biomass) on a real process are typically only available at the end of the batch. In this study the biomass and penicillin concentrations were modelled. The other measured variables used in the input data matrix for modelling were the substrate concentration, dissolved oxygen, pH, substrate temperature, generated heat, acid flow rate, base flow rate, and cooling/heating water flow rate. The reason for considering these measurements was that they are easily measurable using standard industrial sensors, unlike the biomass and penicillin concentrations.
The number of latent variables used in the model was obtained using cross validation. Before this, one of the training batches was tested on the developed model to predict the end point output value of the biomass, the RSMSE of prediction against the latent variables used is shown in Figure 6.8. This model was obtained using 10 training batches with GBN excitation. It is apparent that after the 8th latent variable, almost all of the data is described and the error approaches zero. It is also apparent that the overall error of the model is relatively low, which is expected for a simulated system such as this.
Figure 6.9 shows the prediction error when a test data set, which was not used to train the model was utilised to predict the biomass. The minimum error was obtained at the 3rd latent variable. Using more latent variables results in the model over-fitting the data, which is caused by including too many latent variables, increasing the bias in the predictions. It must be noted that bias can be higher when the data is under-fitted, or in other words not enough latent variables are used.

The selection of latent variables was conducted by using a 10 batch long data set and dividing this into equal length groups. This was undertaken for both biomass and penicillin using the same data sets. Then five different models were identified using half of the data from each group which were cross validated with the remaining data sets. The latent variables were selected according to the minimum root mean squared error of cross validation. Based on Figure 6.10, four latent variables were selected. Although the error corresponding to the fifth latent variable might be marginally smaller, it does not provide more than a 2% improvement and is not selected. This is a recommended approach applied in industry (Wise et al., 2004). Additionally the variance of the input and output data matrices captured was also taken into account.
Following the selection of latent variables, the number of training batches used in identifying the models was investigated. This was to demonstrate the effects of using more historical data and how more of the process behaviour can be captured. For this test 1000 batches were created. The first model was trained using 10 batches and tested with the remaining 990. The end point prediction of penicillin is shown in Figure 6.11. The concentration values are the raw mean centred values obtained from the prediction.
As shown in Figure 6.11, the 10 batch trained model does not predict many of the actual end points. Figure 6.12 shows the accuracy of the model when more training batches were used, in this case 100 batches were used for training and the remaining 900 batches were used for testing the model. The model has provided improved predictions over the 10 batch trained model as expected and illustrates that a large amount of data is required when identifying a batch process model.

Figure 6.12; Penicillin End Point Prediction

Presenting the predictions in the above format makes analysis difficult when the accuracy of the models increases. As an alternative, the end point predictions are plotted against the real values. For the 100 batch trained model presented in Figure 6.12, this is shown in Figure 6.13. As the predictions improve with more accurate models, the plot should tend towards a straight line with a function of $y = mx$, where $m = 1$.

Figure 6.13; Actual -vs- Predicted End Point, 100 Batch Trained
Figure 6.14 and Figure 6.15 show the prediction against actual outputs when 250 and 500 batches respectively were used to identify the models. The prediction plots are shown to have narrowed in width, meaning less variation, with Figure 6.15 showing the greatest prediction accuracy. However, the model accuracy is limited because of the bias. This is mainly caused by the random fluctuations in the growth rate, which essentially creates noise on the output measurements.

This study shows that it is possible to predict the end point of a batch process using the PLS modelling method. Although the method provides biased predictions, with sufficient training data, working models can be obtained. This method is a viable option for predicting, and subsequently controlling the solvent casting process, where the quality variable is only available at the end-point of the batch or the electrospinning process in its current state, where the fibre measurements are only
available after SEM analysis. Following on from the conclusions of the solvent casting process, the potential capability of the PLS model successfully predicting the material properties and pore sizes could be utilised in two ways:

1. The model could be used to determine the air temperature, base heater temperature and humidity levels that are required to produce a film with desired properties.

2. The PLS model could be used as a soft sensor to provide a continuous prediction of the output quality, such as the pore sizes. An empirical model such as this could then be used within an MPC scheme.

Both of the points made above are potential improvements that could be implemented to the solvent casting and the electrospinning method. The availability of the model would maximise the potential of the equipment and the range of materials that can be produced. The following example shows the application of PLS to the electrospinning data, where the final fibre diameters, which are only available at the end of the process, are predicted.

6.1.1.1 Application of PLS to Electrospinning Data

This example utilises data from eleven separate runs of the electrospinning process, collected during the preliminary runs and the further experiments presented in Chapter 4. The PLS method was applied to the data presented in Table 6.1.

<table>
<thead>
<tr>
<th>Test</th>
<th>Voltage (kV)</th>
<th>Flow Rate (ml/h)</th>
<th>Solution Concentration (% w/w)</th>
<th>Air and Solution Temperature</th>
<th>Water Content (gr water / kg air)</th>
<th>Low Side Current</th>
<th>Fibre Diameter</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>6</td>
<td>2</td>
<td>0.08</td>
<td>23</td>
<td>8.07</td>
<td>2.46e-05</td>
<td>1.332</td>
</tr>
<tr>
<td>2</td>
<td>8</td>
<td>2</td>
<td>0.08</td>
<td>23</td>
<td>8.07</td>
<td>3.91e-05</td>
<td>0.903</td>
</tr>
<tr>
<td>3</td>
<td>10</td>
<td>2</td>
<td>0.08</td>
<td>23</td>
<td>8.07</td>
<td>4.95e-05</td>
<td>0.601</td>
</tr>
<tr>
<td>4</td>
<td>6</td>
<td>4</td>
<td>0.08</td>
<td>23</td>
<td>8.07</td>
<td>2.76e-05</td>
<td>1.357</td>
</tr>
<tr>
<td>5</td>
<td>8</td>
<td>4</td>
<td>0.08</td>
<td>23</td>
<td>8.07</td>
<td>4.88e-05</td>
<td>0.964</td>
</tr>
<tr>
<td>6</td>
<td>6</td>
<td>0.5</td>
<td>0.10</td>
<td>22.5</td>
<td>6.28</td>
<td>2.27e-05</td>
<td>1.219</td>
</tr>
<tr>
<td>7</td>
<td>6.5</td>
<td>1</td>
<td>0.10</td>
<td>22.5</td>
<td>6.28</td>
<td>2.99e-05</td>
<td>1.211</td>
</tr>
<tr>
<td>8</td>
<td>10</td>
<td>4</td>
<td>0.08</td>
<td>23</td>
<td>6.28</td>
<td>7.08e-05</td>
<td>0.717</td>
</tr>
<tr>
<td>9</td>
<td>7.5</td>
<td>2</td>
<td>0.10</td>
<td>22.5</td>
<td>6.28</td>
<td>3.85e-05</td>
<td>1.203</td>
</tr>
<tr>
<td>10</td>
<td>9.5</td>
<td>4</td>
<td>0.10</td>
<td>22.5</td>
<td>6.28</td>
<td>6.91e-05</td>
<td>1.308</td>
</tr>
<tr>
<td>11</td>
<td>13</td>
<td>8</td>
<td>0.10</td>
<td>22.5</td>
<td>6.28</td>
<td>9.14e-05</td>
<td>1.081</td>
</tr>
</tbody>
</table>

Table 6.1; Electrospinning PLS Data
The input variables were not changed during any of the experiments, and the fibre diameter was measured afterwards using the SEM following gold coating. The data was pre-treated as shown in section 6.1.1 of this chapter. Leave one out cross validation was undertaken to determine the number of latent variables. Leave one out cross validation works by keeping some of the data apart from the training data and using it to test the model accuracy, meaning predicted values against real values. The data was used to predict the end point average diameters of the fibres, where the model was identified using 10 experiments and one test was left out for validation. This was repeated with each of the experiments, and the predicted diameter against the actual fibre diameter is presented in Figure 6.16. In this case the predictions appear accurate and demonstrate it could possible to identify models with short data sets using the PLS technique. However, it must be pointed out that the data presented in Table 6.1 does not contain high amounts of excitation. This is not ideal as the data is not able to capture the full dynamics of the system and can only predict in this narrow data range. If a new, unseen data set was tested the results would be inaccurate due to the defective model.

![Figure 6.16; End Point Prediction of Electrospinning Fibres using PLS](image)

In the future, using a larger data set that has more variation especially in the water content and solution mixture levels, a more accurate model could be identified. This can be used to produce fibres of specific average diameters and desired mechanical properties once the mechanical measurements are taken and recorded as additional quality measurements. Although the ultimate aim would be the cell growth success, electrospun fibres are used in many other applications in industry, such as specialist filters, and these models would be invaluable for further research in manufacturing such products. This is an important result, as no such final fibre diameter predictions have been published using data that is relatively easy to obtain. Instead, the previous attempts at modelling have focused
on complex equations of motion, fluid dynamics and solving rigorous mathematical equations (Fridrikh et al., 2003). The next section presents the application of the modelling methods to continuous MIMO systems, similar to the electrospinning process where the jet diameter can be measured continuously, or the solvent casting process when a soft sensor approach is assumed.

### 6.1.2 Multi Input Multi Output System

The first of the continuous process examples presented is the Wood and Berry distillation column simulation, which represents a binary distillation column (Wood and Berry, 1973). Distillation columns are used to extract different materials from one fluid feedstock. The incoming fluid is evaporated in stages, leaving the correct materials in different locations of the column. The process inputs are the re-boiler steam flow rate and the reflux. The top and bottom products are the process outputs. This simulation is considered as one of the benchmark MIMO systems in control systems research (Lakshminarayanan et al., 1997), (Laurí, Martínez, et al. 2010), (Seborg and Wood, 1978).

This example was chosen as it is considered to be similar to electrospinning in terms of its structure. Like electrospinning, this process is continuous, it has multiple inputs and multiple outputs with complex disturbances and both of the inputs are correlated with these. The structure of the system is shown in equations 6-1 and 6-2, the correlations can be observed on the full transfer function in equations 6-3 and 6-4.

\[
y_1(k) = G_{11}(q) \cdot u_1(k) + G_{12}(q) \cdot u_2(k) + D_{11}(q) \cdot \xi_1(k) + D_{12}(q) \cdot \xi_2(k)
\]

6-1

The disturbances in the Wood and Berry simulation have complex dynamics when compared to a typical first order low pass filter. The disturbances also have large delays, which are estimated with a second order Padé approximation. The dynamics of the process are described as follows:

\[
\begin{bmatrix}
y_1(k) \\
y_2(k)
\end{bmatrix} = \begin{bmatrix} G_{11}(q) & G_{12}(q) \\ G_{21}(q) & G_{22}(q) \end{bmatrix} \cdot \begin{bmatrix} u_1(k) \\
u_2(k) \end{bmatrix} + \begin{bmatrix} D_{11}(q) & D_{12}(q) \\ D_{21}(q) & D_{22}(q) \end{bmatrix} \cdot \begin{bmatrix} \xi_1(k) \\
\xi_2(k) \end{bmatrix}
\]

6-2

Where, \( \xi_1(k) \) and \( \xi_2(k) \) are white noise with zero mean, \( u_1(k) \) and \( u_2(k) \) are input signals and \( G(q) \) and \( D(q) \) are the transfer functions of the plant and the disturbances respectively, which are given as:

\[
G_{11}(q) = \frac{0.06q^{-1} + 0.63q^{-2} + 0.072q^{-3}}{1 - 0.93q^{-1} - 0.013q^{-2} - 0.0023q^{-3}}, \quad G_{12}(q) = \frac{0.033q^{-1} + 0.087q^{-2} - 0.58q^{-3}}{1 - 1.57q^{-1} + 0.72q^{-2} - 0.13q^{-3}}
\]

6-3

\[
G_{21}(q) = \frac{0.20q^{-1} - 0.063q^{-2} + 0.052q^{-3}}{1 - 2.18q^{-1} + 1.58q^{-2} - 0.39q^{-3}}, \quad G_{22}(q) = \frac{0.051q^{-1} + 0.12q^{-2} - 0.85q^{-3}}{1 - 1.55q^{-1} + 0.71q^{-2} - 0.13q^{-3}}
\]
In this continuous MIMO study, GBN and Random Gaussian Signals (RGS) with lengths ranging from 200 to 9500 samples were used to excite the system for the identification data collection exercise. Different identification signal lengths were implemented to observe the effect of the amount of data on modelling, and to determine how accurate the models were when estimated with short data sets. These tests are important as the electrospinning system model may need frequent updates, and building accurate models from short data sets would be necessary. In addition to the usual GBN identification signal, the RGS signal was also used. This was only implemented for comparison purposes with recent literature (Laurí et al., 2010), where the Wood and Berry column is excited with the low order RGS signal and PLS is applied and recommended for model identification.

The system identification and modelling exercise conducted here, is similar to the SISO example shown in Chapter 5. The order of the identification signals and condition number of data matrices were analysed in addition to the spectral analysis of the signals. The input signal analysis was undertaken to establish how the changes in spectrum and excitation affect the model accuracy.

The four different signals used in this study were RGS with a pass band of 0.01, RGS with a pass band of 0.04, GBN with a clock period of 64, and GBN with a clock period of 8. The pass band of 0.01 means that the signal is expressed in fractions of the Nyquist frequency, in this case the low limit is 0 and high is 0.01. The power spectrum of the signals is shown in Figure 6.17. All of the input spectrums are of a low pass characteristic, as this is important for an identification signal. Signals that are of high power in the high frequency ranges risk making the tested system unstable, depending on the system’s frequency response characteristics (Zhu, 2001). The RGS signals have the lowest power over the frequency ranges and almost drop to zero power immediately. The RGS with 0.04 pass band has more power over the 0.01 pass band. The GBN 64 signal contains more power and the GBN 8 even more, never becoming zero over the whole frequency range. These power spectrums can be backed up with the excitation order data from the signal properties table, presented in Table 6.2.

\[
D_{11}(q) = \frac{0.10q^{-1} - 0.3q^{-2} + 0.23q^{-3}}{1-2.29q^{-1} + 1.74q^{-2} - 0.45q^{-3}} , D_{12}(q) = \frac{0.0037q^{-1} - 0.011q^{-2} + 0.0087q^{-3}}{1-2.28q^{-1} + 1.72q^{-2} - 0.44q^{-3}}
\]
\[
D_{21}(q) = \frac{0.003q^{-1} - 0.095q^{-2} + 0.25q^{-3}}{1-1.65q^{-1} + 0.84q^{-2} - 0.16q^{-3}} , D_{22}(q) = \frac{0.0052q^{-1} - 0.014q^{-2} + 0.01q^{-3}}{1-2.34q^{-1} + 1.83q^{-2} - 0.48q^{-3}}
\]

In this continuous MIMO study, GBN and Random Gaussian Signals (RGS) with lengths ranging from 200 to 9500 samples were used to excite the system for the identification data collection exercise. Different identification signal lengths were implemented to observe the effect of the amount of data on modelling, and to determine how accurate the models were when estimated with short data sets. These tests are important as the electrospinning system model may need frequent updates, and building accurate models from short data sets would be necessary. In addition to the usual GBN identification signal, the RGS signal was also used. This was only implemented for comparison purposes with recent literature (Laurí et al., 2010), where the Wood and Berry column is excited with the low order RGS signal and PLS is applied and recommended for model identification.

The system identification and modelling exercise conducted here, is similar to the SISO example shown in Chapter 5. The order of the identification signals and condition number of data matrices were analysed in addition to the spectral analysis of the signals. The input signal analysis was undertaken to establish how the changes in spectrum and excitation affect the model accuracy.

The four different signals used in this study were RGS with a pass band of 0.01, RGS with a pass band of 0.04, GBN with a clock period of 64, and GBN with a clock period of 8. The pass band of 0.01 means that the signal is expressed in fractions of the Nyquist frequency, in this case the low limit is 0 and high is 0.01. The power spectrum of the signals is shown in Figure 6.17. All of the input spectrums are of a low pass characteristic, as this is important for an identification signal. Signals that are of high power in the high frequency ranges risk making the tested system unstable, depending on the system’s frequency response characteristics (Zhu, 2001). The RGS signals have the lowest power over the frequency ranges and almost drop to zero power immediately. The GBS with 0.04 pass band has more power over the 0.01 pass band. The GBN 64 signal contains more power and the GBN 8 even more, never becoming zero over the whole frequency range. These power spectrums can be backed up with the excitation order data from the signal properties table, presented in Table 6.2.
The excitation orders of the signals follow their spectral properties, the higher the excitation the more power the signal contains in the frequency domain. A signal with non-zero power over the whole frequency range would be expected to capture more frequency dynamics of the system in comparison to a low exciting signal, such as the RGS.

<table>
<thead>
<tr>
<th>Signal Type</th>
<th>Samples</th>
<th>Signal to Noise Ratio</th>
<th>Excitation Order</th>
<th>Mean Condition Number $X_1$</th>
<th>Mean Condition Number $X_2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>RGS (0.01)</td>
<td>500</td>
<td>10</td>
<td>11</td>
<td>66</td>
<td>24</td>
</tr>
<tr>
<td>RGS (0.01)</td>
<td>2500</td>
<td>10</td>
<td>11</td>
<td>54</td>
<td>22</td>
</tr>
<tr>
<td>RGS (0.01)</td>
<td>9500</td>
<td>10</td>
<td>11</td>
<td>45</td>
<td>22</td>
</tr>
<tr>
<td>RGS (0.04)</td>
<td>500</td>
<td>10</td>
<td>26</td>
<td>50</td>
<td>10</td>
</tr>
<tr>
<td>GBN(64)</td>
<td>500</td>
<td>10</td>
<td>50</td>
<td>25</td>
<td>13</td>
</tr>
<tr>
<td>GBN(64)</td>
<td>2500</td>
<td>10</td>
<td>50</td>
<td>26</td>
<td>13</td>
</tr>
<tr>
<td>GBN(64)</td>
<td>9500</td>
<td>10</td>
<td>50</td>
<td>25</td>
<td>12</td>
</tr>
<tr>
<td>GBN(8)</td>
<td>500</td>
<td>10</td>
<td>50</td>
<td>12</td>
<td>4</td>
</tr>
<tr>
<td>GBN(8)</td>
<td>9500</td>
<td>10</td>
<td>50</td>
<td>9</td>
<td>4</td>
</tr>
</tbody>
</table>

**Table 6.2; Identification Signal Properties**
Table 6.2 presents the properties of the input signals of certain data lengths that were used in the identification study. This is to give an insight into how different excitation orders affect the data in terms of mean condition number. This is the ratio of the highest and the lowest singular value of the data matrix, and represents the condition of the matrix. Matrices with very high values of this are considered *ill-conditioned*. The condition number of the matrices improves as the identification signal increases in excitation. The condition number of the $X_2$ data matrix is always smaller than the $X_1$ data matrix. This is caused by the relatively large gain of 0.10 in the first disturbance transfer function of $D_{11}$. As expected from the first study, the larger overall condition number in the first output’s data matrix leads to higher modelling errors.

In this study, the signal to noise ratio was kept at a constant 10 dB throughout the simulations to ensure consistency between the tests. The results are presented differently from the previously shown SISO and PENSIM examples. The results from this point forward are presented using the overall Root Mean Squared Error (RMSE) rather than the prediction against the actual value of the output. The root mean squared error gives an overall measure of the modelling error and simplifies analysis. During the tests, model orders were assumed to be known. The recommended way of establishing the appropriate model order is to use the Output Error method to identify a number of models of varying orders and cross validate them (Zhu, 2001).

The latent variables for the PLS based methods used in this study were selected using cross validation. The PLS based methods applied were bPLS, uRLS-PLS, ARX-PLS, ARMAX-PLS, OE-PLS and PEM-PLS. Similar to the PENSIM test, a data set containing 2000 samples was collected and split into ten subsets. Five of these subsets were used to identify the models, and the remaining five were used for testing. The Root Mean Squared Error of Cross Validation (RMSECV) was then used to select the number of latent variables. As examples of this, Figure 6.18 and Figure 6.19 show the cross validation error plot for bPLS and the hybrid ARMAX-PLS methods respectively, when the models were identified using the RGS signal of pass band 0.01. For the bPLS result shown in Figure 6.18 four latent variables were selected, as the fifth latent variable did not offer an improvement over 2%. 

For the ARMAX-PLS method shown in Figure 6.19, three latent variables were selected. The latent variables were selected using the same method for all the tests reported in this chapter.

The identification techniques were also tested with the length of samples used for obtaining the models. An example plot is shown in Figure 6.20 for the ARX-PLS method. As expected with the increasing sample number the overall prediction error decreases. The latent variable responses
appear similar, with the most accurate models using 4 or 5 latent variables. Increasing the number of latent variables further had very little impact on model accuracy.

![Figure 6.20; Sample and Latent Variable Map](image)

Following the latent variable selection exercise, the first test consisted of identifying the system using an RGS signal with a pass band of 0.01 and where the disturbance SNR is 10 dB. The models were tested with a step signal, as the response to a step is particularly important for a model that is to be used in a model based control system.

The results of this test are presented as the RMSE rather than the $R^2$ statistic. As discussed previously, although model deterioration can be observed over the prediction horizon the $R^2$ statistic does not give an indication of the magnitude of the error. The RMSE over the prediction horizon provides both pieces of information. This is also more useful in terms of observing the value of actual deviation from the real system. In Figure 6.21 and Figure 6.22 the errors for Output 1 and Output 2 are shown respectively through the 1 to 30 steps ahead prediction horizon. The overall error of the second output prediction is lower than that of the first output, which is due to the noise structure and the system parameters of the first input. These factors mean that the condition number for Input 1 is higher than that for Input 2, as previously shown in Table 6.2, which results in the modelling error for Output 1 being higher than Output 2.

In the second part of the test, the results presented are only the RMSE at the 30 step ahead prediction point, as this is more relevant for MPC applications, where long prediction horizons are typically used. The tests were each repeated 100 times, where each time a new identification data set
was created and tested with a step signal. The tests were repeated multiple times as the model predictions can vary due to the small amounts of data used.

In Figure 6.21 and Figure 6.22, the method with the lowest error is the bPLS method for both outputs. This is due to the maximisation of the covariance property of the PLS algorithm, and illustrates that PLS models offer benefits when identifying models from data sets containing low excitation. The uRLS-PLS method performs better than its base method (uRLS), as again the PLS method helps accommodate the lack of excitation in the data. It must be noted that the errors for the bPLS, OE-PLS and uRLS-PLS are all low, and all of the hybrid methods have outperformed their base methods of OE, ARX, ARMAX, PEM and uRLS. It can also be seen that the variance in the PEM prediction is too high for it to be considered as a valid method in this test instance. The results for the OE method are not displayed, as the error was too large to display in the plots with the current scale. This once again demonstrates the advantages of PLS and the hybrid PLS methods when identifying models using data containing limited excitation.

The results observed in this test are consistent with the results found in the work of Laurí, et al. (2010), where the bPLS method was found to produce low error models, although in the mentioned work the tests were not conducted in a systematic repeatable way. The next set of results presents a more through test run, where the models were again identified with an RGS signal with pass band 0.01, but repeated 100 times. The models are identified using a range of identification signals from 200 to 9500 samples and validated with a step test.
The following section presents the results when the simulation was run multiple times, and only the 30 step ahead prediction points were estimated. The results table presenting the RMSE for Output 1 is shown in Table 6.3 and for Output 2 in Table 6.4. For Output 1, the OE method in both cases fails to compute usable models due to the inherent noise dynamics in the data. The OE method assumes the disturbances are only white noise and formulates the cost function as the output error. When the noise dynamics are complex it creates biased estimates. The PEM method performs similar to OE in both cases, calculating unusable models with large errors. This is because the RGS signal with the 0.01 pass band is not sufficiently exciting to provide rich enough data for these methods. As expected the alternative hybrid PLS methods give significantly lower errors. The uRLS-PLS, ARX-PLS and ARMAX-PLS models outperform the bPLS method when low numbers of samples are available for identifying the models. In the case of the second output the bPLS and the uRLS-PLS methods obtain the models with the lowest errors. PEM-PLS and OE-PLS produced smaller errors when compared with their base methods, which is also the case for ARX-PLS and ARMAX-PLS. This is the result of creating unbiased PLS models with the base model predictions and taking advantage of the PLS method’s ability to capture the covariance properties in limited size matrices.
In the case of Output 2, presented in Table 6.4, the results pattern is similar, with the overall error being smaller due to the condition number of the data matrix. This also affects the accuracy of the first prediction of the base models used by the hybrid techniques. The bPLS and uRLS-PLS methods identify the models with the lowest prediction error. The hybrid PEM-PLS and OE-PLS show high variance between each run, and are not reliable choices. The ARX, ARMAX, OE and PEM methods are also unable to identify models with low errors, due to the excitation signal not being sufficient for these methods.

Table 6.3; Output 1

<table>
<thead>
<tr>
<th>Sample Length</th>
<th>bPLS</th>
<th>uRLS-PLS</th>
<th>ARX-PLS</th>
<th>ARMAX-PLS</th>
<th>PEM-PLS</th>
<th>OE-PLS</th>
<th>ARX</th>
<th>ARMAX</th>
<th>OE</th>
<th>PEM</th>
</tr>
</thead>
<tbody>
<tr>
<td>200</td>
<td>6.88</td>
<td>3.33</td>
<td>2.45</td>
<td>2.45</td>
<td>2.82</td>
<td>290.14</td>
<td>1.6E+4</td>
<td>1.9E+4</td>
<td>1.3E+5</td>
<td>8.9E+4</td>
</tr>
<tr>
<td>500</td>
<td>4.60</td>
<td>3.33</td>
<td>3.13</td>
<td>3.16</td>
<td>3.23</td>
<td>25.32</td>
<td>279.79</td>
<td>252.16</td>
<td>1.3E+4</td>
<td>6.8E+4</td>
</tr>
<tr>
<td>1000</td>
<td>3.85</td>
<td>3.14</td>
<td>3.55</td>
<td>3.56</td>
<td>5.29</td>
<td>40.80</td>
<td>81.15</td>
<td>76.47</td>
<td>2.3E+6</td>
<td>5.1E+15</td>
</tr>
<tr>
<td>2500</td>
<td>3.35</td>
<td>3.11</td>
<td>3.62</td>
<td>3.68</td>
<td>3.38</td>
<td>984.97</td>
<td>58.20</td>
<td>55.95</td>
<td>9.7E+4</td>
<td>7.0E+9</td>
</tr>
<tr>
<td>5000</td>
<td>3.21</td>
<td>3.20</td>
<td>3.62</td>
<td>3.67</td>
<td>3.58</td>
<td>7.01</td>
<td>54.35</td>
<td>52.21</td>
<td>6.2E+3</td>
<td>2.0E+5</td>
</tr>
<tr>
<td>7500</td>
<td>3.11</td>
<td>2.93</td>
<td>3.58</td>
<td>3.66</td>
<td>3.62</td>
<td>10.09</td>
<td>53.78</td>
<td>51.63</td>
<td>7.7E+3</td>
<td>1.4E+8</td>
</tr>
<tr>
<td>9500</td>
<td>3.05</td>
<td>2.96</td>
<td>3.57</td>
<td>3.66</td>
<td>6.74</td>
<td>9.53</td>
<td>53.49</td>
<td>51.38</td>
<td>1.1E+4</td>
<td>2.1E+17</td>
</tr>
</tbody>
</table>

Table 6.4; Output 2

<table>
<thead>
<tr>
<th>Sample Length</th>
<th>bPLS</th>
<th>uRLS-PLS</th>
<th>ARX-PLS</th>
<th>ARMAX-PLS</th>
<th>PEM-PLS</th>
<th>OE-PLS</th>
<th>ARX</th>
<th>ARMAX</th>
<th>OE</th>
<th>PEM</th>
</tr>
</thead>
<tbody>
<tr>
<td>200</td>
<td>9.92</td>
<td>3.25</td>
<td>4.82</td>
<td>4.82</td>
<td>3.11E+4</td>
<td>487.36</td>
<td>3.0E+4</td>
<td>1.1E+5</td>
<td>2.5E+5</td>
<td></td>
</tr>
<tr>
<td>500</td>
<td>3.05</td>
<td>2.92</td>
<td>4.80</td>
<td>4.78</td>
<td>497.89</td>
<td>14.39</td>
<td>167.42</td>
<td>275.97</td>
<td>1.5E+5</td>
<td>1.1E+5</td>
</tr>
<tr>
<td>1000</td>
<td>1.70</td>
<td>2.46</td>
<td>4.84</td>
<td>4.82</td>
<td>252.31</td>
<td>10.31</td>
<td>70.16</td>
<td>114.41</td>
<td>8.6E+4</td>
<td>4.1E+4</td>
</tr>
<tr>
<td>2500</td>
<td>1.21</td>
<td>1.78</td>
<td>4.60</td>
<td>4.76</td>
<td>2.8E+3</td>
<td>67.35</td>
<td>35.04</td>
<td>54.72</td>
<td>2.8E+5</td>
<td>4.0E+5</td>
</tr>
<tr>
<td>5000</td>
<td>1.01</td>
<td>1.37</td>
<td>4.29</td>
<td>4.67</td>
<td>264.68</td>
<td>728.19</td>
<td>22.43</td>
<td>32.80</td>
<td>4.1E+5</td>
<td>1.5E+4</td>
</tr>
<tr>
<td>7500</td>
<td>0.92</td>
<td>0.98</td>
<td>4.12</td>
<td>4.61</td>
<td>188.65</td>
<td>3.77</td>
<td>18.03</td>
<td>25.62</td>
<td>3.6E+3</td>
<td>1.8E+4</td>
</tr>
<tr>
<td>9500</td>
<td>0.89</td>
<td>0.98</td>
<td>4.00</td>
<td>4.59</td>
<td>116.56</td>
<td>6.54</td>
<td>16.51</td>
<td>21.69</td>
<td>1.8E+4</td>
<td>4.1E+4</td>
</tr>
</tbody>
</table>

Although these results demonstrate that the compared methods can provide accurate models when the data has low excitation, the PLS based methods should be the method of choice. The hybrid methods provide a small improvement over the traditional bPLS method and significant improvements over their base models.
In this next test, the system was identified with an RGS signal with a pass band of 0.04, which is a signal with a higher excitation order than RGS 0.01, meaning it is exciting over a broader range of frequencies. This should mean the overall modelling errors will be lower, and the non hybrid methods will also be able to identify models with small errors from a low number of samples. Table 6.5 and Table 6.6 present the RMSE at the 30 step ahead prediction horizon for Output 1 and Output 2 respectively. The PEM and OE once again result in models with high and variable errors, which also affects their hybrid versions, PEM-PLS and OE-PLS. There is a significant improvement in the modelling error when their PLS hybrids are used. The other PLS based methods all perform similarly, identifying models with low errors.

<table>
<thead>
<tr>
<th>Sample Length</th>
<th>bPLS</th>
<th>uRLS-PLS</th>
<th>ARX-PLS</th>
<th>ARMAX-PLS</th>
<th>PEM-PLS</th>
<th>OE-PLS</th>
<th>ARX</th>
<th>ARMAX</th>
<th>OE</th>
<th>PEM</th>
</tr>
</thead>
<tbody>
<tr>
<td>200</td>
<td>4.60</td>
<td>2.92</td>
<td>4.80</td>
<td>5.63</td>
<td>4.57</td>
<td>8.94</td>
<td>12.56</td>
<td>12.10</td>
<td>3.4E+2</td>
<td>1.2E+16</td>
</tr>
<tr>
<td>500</td>
<td>3.29</td>
<td>2.90</td>
<td>3.25</td>
<td>3.62</td>
<td>3.76</td>
<td>8.52</td>
<td>7.10</td>
<td>7.02</td>
<td>5.6E+2</td>
<td>1.1E+4</td>
</tr>
<tr>
<td>1000</td>
<td>2.28</td>
<td>2.60</td>
<td>2.22</td>
<td>2.95</td>
<td>4.52</td>
<td>7.57</td>
<td>6.25</td>
<td>6.23</td>
<td>9.8E+2</td>
<td>2.3E+6</td>
</tr>
<tr>
<td>2500</td>
<td>1.70</td>
<td>2.50</td>
<td>1.68</td>
<td>2.07</td>
<td>3.28</td>
<td>11.09</td>
<td>5.91</td>
<td>5.91</td>
<td>2.2E+3</td>
<td>9.1E+10</td>
</tr>
<tr>
<td>5000</td>
<td>1.48</td>
<td>2.12</td>
<td>1.47</td>
<td>1.65</td>
<td>3.20</td>
<td>63.20</td>
<td>5.86</td>
<td>5.87</td>
<td>8.5E+2</td>
<td>4.7E+11</td>
</tr>
<tr>
<td>7500</td>
<td>1.41</td>
<td>1.93</td>
<td>1.40</td>
<td>1.48</td>
<td>2.80</td>
<td>69.14</td>
<td>5.84</td>
<td>5.84</td>
<td>8.9E+2</td>
<td>2.8E+8</td>
</tr>
<tr>
<td>9500</td>
<td>1.39</td>
<td>2.12</td>
<td>1.38</td>
<td>1.43</td>
<td>2.65</td>
<td>43.36</td>
<td>5.81</td>
<td>5.81</td>
<td>1.0E+3</td>
<td>6.6E+8</td>
</tr>
</tbody>
</table>

Table 6.5; Output 1

The overall error for Output 2 shown in Table 6.6 is lower as expected. This identification signal provides a richer data set as a whole, and this can be observed in the OE method’s prediction for the second output. The models obtained using 2500 samples or more, have lower error in comparison with those identified using fewer samples. For both of the outputs, ARMAX has performed better than PEM. Although they are both prediction error methods and use a very similar algorithm, the PEM method formulates the prediction errors slightly differently, by taking the full disturbance in to account. When the data is not sufficiently exciting, the disturbances are not predicted accurately leading to a larger prediction error.
The fluctuations in the results as the sample numbers increase is caused by model to model variance between each repeated test. Occasionally, a model with an exceedingly high error is obtained as a result of convergence problems, resulting from the data not being sufficiently rich. This affects the mean error values in the presented tables where it does not monotonically decrease as the number of samples increase. When the identification signals used on an individual test that created large errors were analysed and compared with those that produced normal error values, no significant differences were found to exist in the power spectrums or the mean condition number of the data matrices. The same issue was also observed on the hybrid methods. This type of variation on the repeated tests is only seen when the signals are not highly exciting, therefore some of the methods are unable to identify satisfactory initial models.

During the numerical minimisation of the cost functions, one of the expressions used is the gradient of prediction. For ARMAX the regression vector is filtered through $\frac{1}{C}$ which is of a low pass structure.

This could mean in situations where data is not sufficient, there is more chance of obtaining a stable disturbance as only $C$ needs to be estimated. PEM uses $\frac{D}{C}$ where two factors need to be estimated, increasing the possibility of non-convergence. To illustrate this, if the identified parameters for the first output are compared from a single run, when using the 500 sample long RGS 0.01 signal where a high number of repeated variations were observed for PEM, where:

$$y_1(t) = \frac{B_{11}(q)}{A_{11}(q)} \cdot u_1(t) + \frac{B_{12}(q)}{A_{12}(q)} \cdot u_2(t) + \frac{C_1(q)}{D_1(q)} \cdot e(t)$$

6-5
\[ B_{11}(q) = -4563q^{-1} + 9030q^{-2} - 4477q^{-3} \]  
\[ B_{12}(q) = -7694^{-1}q^{-1} + 1.531q^{-2} - 7619q^{-3} \]

The parameters, given in equation 6-6, within the model are very large, which can result in very large prediction errors (Levine, 1996).

\[ B_{12}(q) = -7694^{-1}q^{-1} + 1.531q^{-2} - 7619q^{-3} \]

The problem in 6-7 is the same. The rest of the calculated parameters are listed as below:

\[ A_{11}(q) = 1 + 0.813q^{-1} + 0.889q^{-2} - 0.714q^{-3} \]  
\[ A_{12}(q) = 1 + 0.884q^{-1} - 0.911q^{-2} - 0.945q^{-3} \]  
\[ C_{1}(q) = 1 + 0.290q^{-1} - 0.763q^{-2} - 0.089q^{-3} \]  
\[ D_{1}(q) = 1 - 0.709q^{-1} + 0.937q^{-2} - 0.737q^{-3} \]

When the parameters predicted with ARMAX are calculated as:

\[ y_1(t) = \frac{B_{11}(q)}{A_1(q)} \cdot u_1(t) + \frac{B_{12}(q)}{A_1(q)} \cdot u_2(t) + \frac{C_1(q)}{A_1(q)} \cdot e(t) \]  
\[ A_1(q) = 1 - 0.9429 \cdot q^{-1} \]  
\[ B_{11}(q) = -0.0008 \cdot q^{-3} + 0.00001 \cdot q^{-4} + 0.0008 \cdot q^{-5} \]  
\[ B_{12}(q) = -0.0006 \cdot q^{-3} - 0.00003 \cdot q^{-4} + 0.0006 \cdot q^{-5} \]

In 6-14 the parameters are relatively small compared to that of in 6-6.

\[ C_1(q) = 1 - 0.034 \cdot q^{-3} + 0.059 \cdot q^{-4} - 0.023 \cdot q^{-5} \]

The one step ahead predictions of these models are shown in Figure 6.23, and are used to calculate the unbiased hybrid models. The identified model will also be used to predict the 30 step ahead PEM and ARMAX responses. The instability will get worse as shown in the previous examples, where the model deterioration is evident as the prediction horizon increases.
Figure 6.23 shows that the resonant response caused by the estimated parameters of the PEM prediction. The red plot is the PEM estimation, the green is the actual output value that the models are trying to predict, and the blue plot is the ARMAX estimation which cannot be observed without increasing the scale further. Figure 6.24 provides a closer look at the ARMAX prediction. In Figure 6.24 it can be seen that the ARMAX method calculates an accurate prediction of the output. This confirms the large variations observed in the PEM models, where some of the RMSE values were exceedingly high. These findings also apply to the ARMAX and OE methods.
The third example is when a GBN signal with a clock period of 64 was used to identify the models. This signal has a higher excitation order than the previously used RGS 0.04 and a power spectrum that is effective over more of the frequencies. As expected, the overall error is higher for Output 1 than Output 2. The overall error of OE and PEM methods is lower than the models identified using the less exciting RGS signal with pass band 0.04. In this example all of the models predict well, as the signal provides rich data. The variations between the repeated tests are minimal, with the exception of OE and OE-PLS when 5000 samples are used. The large errors in this example are caused by the biased estimation problems that result from the relatively complex disturbance dynamics.

![ARMAX - PEM One Step Ahead Comparison - RGS 0.01](image)

**Figure 6.24; Larger Scale plot of the ARMAX - PEM One Step Ahead Prediction**

<table>
<thead>
<tr>
<th>Sample Length</th>
<th>bPLS</th>
<th>uRLS-PLS</th>
<th>ARX-PLS</th>
<th>ARMAX-PLS</th>
<th>PEM-PLS</th>
<th>OE-PLS</th>
<th>ARX</th>
<th>ARMAX</th>
<th>OE</th>
<th>PEM</th>
</tr>
</thead>
<tbody>
<tr>
<td>200</td>
<td>3.97</td>
<td>3.01</td>
<td>4.40</td>
<td>4.41</td>
<td>4.01</td>
<td>4.90</td>
<td>3.45</td>
<td>3.42</td>
<td>5.74</td>
<td>50.78</td>
</tr>
<tr>
<td>500</td>
<td>2.58</td>
<td>3.07</td>
<td>2.55</td>
<td>2.56</td>
<td>2.61</td>
<td>9.09</td>
<td>2.65</td>
<td>2.70</td>
<td>10.45</td>
<td>2.5E+3</td>
</tr>
<tr>
<td>1000</td>
<td>1.62</td>
<td>2.57</td>
<td>1.60</td>
<td>1.60</td>
<td>1.62</td>
<td>8.32</td>
<td>1.56</td>
<td>1.62</td>
<td>8.67</td>
<td>1.3E+3</td>
</tr>
<tr>
<td>2500</td>
<td>2.09</td>
<td>2.69</td>
<td>2.29</td>
<td>2.28</td>
<td>2.12</td>
<td>3.93</td>
<td>1.60</td>
<td>1.62</td>
<td>3.74</td>
<td>105</td>
</tr>
<tr>
<td>5000</td>
<td>2.26</td>
<td>2.31</td>
<td>2.51</td>
<td>2.50</td>
<td>2.27</td>
<td>65.29</td>
<td>1.67</td>
<td>1.65</td>
<td>48.89</td>
<td>0.67</td>
</tr>
<tr>
<td>7500</td>
<td>2.30</td>
<td>1.49</td>
<td>2.55</td>
<td>2.54</td>
<td>2.29</td>
<td>3.97</td>
<td>1.68</td>
<td>1.66</td>
<td>3.98</td>
<td>1.20</td>
</tr>
<tr>
<td>9500</td>
<td>2.11</td>
<td>2.13</td>
<td>2.34</td>
<td>2.33</td>
<td>2.12</td>
<td>4.31</td>
<td>1.57</td>
<td>1.57</td>
<td>4.17</td>
<td>0.49</td>
</tr>
</tbody>
</table>

**Table 6.7; Output 1**
The results for Output 2 shown in Table 6.8 are all of a low magnitude, again meaning all the models have predicted well. As expected the overall errors are reduced due to the lower condition number of this data matrix.

<table>
<thead>
<tr>
<th>Sample Length</th>
<th>bPLS</th>
<th>URLS-PLS</th>
<th>ARX-PLS</th>
<th>ARMAX-PLS</th>
<th>PEM-PLS</th>
<th>OE-PLS</th>
<th>ARX</th>
<th>ARMAX</th>
<th>OE</th>
<th>PEM</th>
</tr>
</thead>
<tbody>
<tr>
<td>200</td>
<td>1.52</td>
<td>1.90</td>
<td>1.54</td>
<td>1.55</td>
<td>1.57</td>
<td>2.70</td>
<td>1.38</td>
<td>1.40</td>
<td>3.14</td>
<td>2.95</td>
</tr>
<tr>
<td>500</td>
<td>1.27</td>
<td>1.88</td>
<td>1.28</td>
<td>1.30</td>
<td>1.33</td>
<td>1.97</td>
<td>1.03</td>
<td>1.05</td>
<td>1.91</td>
<td>1.56</td>
</tr>
<tr>
<td>1000</td>
<td>0.92</td>
<td>0.91</td>
<td>0.91</td>
<td>0.92</td>
<td>0.94</td>
<td>3.22</td>
<td>0.74</td>
<td>0.77</td>
<td>3.98</td>
<td>1.56</td>
</tr>
<tr>
<td>2500</td>
<td>0.79</td>
<td>0.73</td>
<td>0.76</td>
<td>0.77</td>
<td>0.73</td>
<td>1.98</td>
<td>0.53</td>
<td>0.60</td>
<td>1.51</td>
<td>0.67</td>
</tr>
<tr>
<td>5000</td>
<td>0.70</td>
<td>0.61</td>
<td>0.66</td>
<td>0.67</td>
<td>0.65</td>
<td>1.06</td>
<td>0.39</td>
<td>0.49</td>
<td>0.66</td>
<td>0.51</td>
</tr>
<tr>
<td>7500</td>
<td>0.76</td>
<td>0.77</td>
<td>0.71</td>
<td>0.72</td>
<td>0.75</td>
<td>1.04</td>
<td>0.37</td>
<td>0.50</td>
<td>0.36</td>
<td>0.28</td>
</tr>
<tr>
<td>9500</td>
<td>0.71</td>
<td>0.69</td>
<td>0.67</td>
<td>0.68</td>
<td>0.70</td>
<td>1.00</td>
<td>0.34</td>
<td>0.46</td>
<td>0.31</td>
<td>0.24</td>
</tr>
</tbody>
</table>

Table 6.8; Output 2

The following example is when the model was excited with a GBN signal with clock period of 8. This signal has the most power over the whole of the frequency ranges from the four excitation signals used in this study. This signal was applied to confirm that when the data was sufficiently exciting, all of the methods produced models with small errors and with limited variation between the repeat tests. The errors for Output 1 are presented in Table 6.9, where the OE and PEM methods provide low errors even when the models were identified using 200 samples. The same applies to Output 2 presented in Table 6.10 where all the methods, in particular OE and PEM, have identified models with lower errors.

<table>
<thead>
<tr>
<th>Sample Length</th>
<th>bPLS</th>
<th>URLS-PLS</th>
<th>ARX-PLS</th>
<th>ARMAX-PLS</th>
<th>PEM-PLS</th>
<th>OE-PLS</th>
<th>ARX</th>
<th>ARMAX</th>
<th>OE</th>
<th>PEM</th>
</tr>
</thead>
<tbody>
<tr>
<td>200</td>
<td>2.07</td>
<td>2.94</td>
<td>3.92</td>
<td>4.02</td>
<td>2.94</td>
<td>4.20</td>
<td>6.09</td>
<td>7.35</td>
<td>8.53</td>
<td>6.34</td>
</tr>
<tr>
<td>500</td>
<td>2.12</td>
<td>2.28</td>
<td>4.72</td>
<td>4.68</td>
<td>2.73</td>
<td>4.71</td>
<td>3.60</td>
<td>4.76</td>
<td>3.73</td>
<td>2.22</td>
</tr>
<tr>
<td>1000</td>
<td>1.95</td>
<td>2.39</td>
<td>4.61</td>
<td>4.60</td>
<td>2.50</td>
<td>4.13</td>
<td>3.18</td>
<td>4.15</td>
<td>3.04</td>
<td>1.30</td>
</tr>
<tr>
<td>2500</td>
<td>2.39</td>
<td>1.61</td>
<td>5.09</td>
<td>5.06</td>
<td>2.90</td>
<td>4.47</td>
<td>2.39</td>
<td>3.53</td>
<td>3.25</td>
<td>0.66</td>
</tr>
<tr>
<td>5000</td>
<td>2.47</td>
<td>1.95</td>
<td>5.05</td>
<td>5.01</td>
<td>2.83</td>
<td>4.45</td>
<td>2.15</td>
<td>3.29</td>
<td>2.44</td>
<td>0.65</td>
</tr>
<tr>
<td>7500</td>
<td>2.59</td>
<td>1.67</td>
<td>5.20</td>
<td>5.16</td>
<td>2.93</td>
<td>4.61</td>
<td>2.03</td>
<td>3.17</td>
<td>1.85</td>
<td>0.63</td>
</tr>
<tr>
<td>9500</td>
<td>2.53</td>
<td>1.83</td>
<td>5.12</td>
<td>5.08</td>
<td>2.87</td>
<td>4.35</td>
<td>1.98</td>
<td>3.12</td>
<td>2.06</td>
<td>0.55</td>
</tr>
</tbody>
</table>

Table 6.9; Output 1
<table>
<thead>
<tr>
<th>Sample Length</th>
<th>bPLS</th>
<th>uRLS-PLS</th>
<th>ARX-PLS</th>
<th>ARMAX-PLS</th>
<th>PEM-PLS</th>
<th>OE-PLS</th>
<th>ARX</th>
<th>ARMAX</th>
<th>OE</th>
<th>PEM</th>
</tr>
</thead>
<tbody>
<tr>
<td>200</td>
<td>3.33</td>
<td>1.73</td>
<td>3.12</td>
<td>2.56</td>
<td>2.97</td>
<td>5.07</td>
<td>2.49</td>
<td>2.19</td>
<td>3.91</td>
<td>3.08</td>
</tr>
<tr>
<td>500</td>
<td>3.39</td>
<td>2.46</td>
<td>3.72</td>
<td>3.48</td>
<td>3.21</td>
<td>3.99</td>
<td>2.52</td>
<td>2.44</td>
<td>1.38</td>
<td>1.08</td>
</tr>
<tr>
<td>1000</td>
<td>3.55</td>
<td>2.67</td>
<td>4.10</td>
<td>3.90</td>
<td>3.43</td>
<td>3.91</td>
<td>2.70</td>
<td>2.67</td>
<td>0.82</td>
<td>0.69</td>
</tr>
<tr>
<td>2500</td>
<td>3.24</td>
<td>3.63</td>
<td>3.59</td>
<td>3.48</td>
<td>3.07</td>
<td>3.40</td>
<td>2.42</td>
<td>2.48</td>
<td>0.47</td>
<td>0.39</td>
</tr>
<tr>
<td>5000</td>
<td>3.25</td>
<td>2.79</td>
<td>3.59</td>
<td>3.47</td>
<td>3.11</td>
<td>3.43</td>
<td>2.33</td>
<td>2.42</td>
<td>0.34</td>
<td>0.28</td>
</tr>
<tr>
<td>7500</td>
<td>3.20</td>
<td>2.68</td>
<td>3.59</td>
<td>3.49</td>
<td>3.07</td>
<td>3.42</td>
<td>2.29</td>
<td>2.38</td>
<td>0.28</td>
<td>0.23</td>
</tr>
<tr>
<td>9500</td>
<td>3.09</td>
<td>2.96</td>
<td>3.46</td>
<td>3.36</td>
<td>2.96</td>
<td>3.29</td>
<td>2.24</td>
<td>2.31</td>
<td>0.26</td>
<td>0.21</td>
</tr>
</tbody>
</table>

Table 6.10; Output 2

To summarise the four examples presented in this section, it can be concluded that the PLS based methods work better than traditional methods when the data has low excitation. The hybrid methods offer a slight improvement over bPLS and should always be considered when identifying unbiased PLS models. When the excitation is increased, as shown in the RGS pass band 0.04 example, the ARX and ARMAX methods’ performances increase, but not to the extent where they outperform the PLS based methods at low sample lengths. When using a GBN signal with high excitation the OE and PEM methods start producing estimations with low errors. The ARMAX prediction error method produced models with low errors even when the data sets used were short, with the exception of the very low exciting RGS 001 signal. Additionally, the ARX least squares regression method can also work well in most cases when there are sufficient amounts of data. This also shows that relying on one modelling method is not good practice, as varying input data may lead to large errors. Analysing the data frequency spectrum or the condition number of the data matrices could help in choosing between a PLS based method and a traditional approach such as ARX. Ideally, all methods would be tried and tested at every necessary model update.

These results also undermine the usefulness of PLS based techniques when used with systems that have sufficiently exciting data. Another argument for the utilisation of PLS, excluding batch situations, is the presence of a high number of correlated variables. The following example will analyse a similar scenario, to investigate exactly where PLS and the hybrid methods can provide benefits. These results also show that it is possible to build models of MIMO systems such as the electrospinning process, where some correlations and disturbances exist. This example acts as proof of concept and is an important ground study for the further modelling work associated with the electrospinning and solvent process.
6.1.3 Highly Correlated Multi Input Multi Output System

This example shows an extreme case where strong correlations exist amongst the input variables. The aim is to determine when the use of PLS is necessary, when considered as a small scale process rather than an industrial data set with hundreds of variables.

Initially a model with ten independent inputs and one output was tested, where the output is defined as:

\[ y(t) = G_1(q) \cdot u_1(t) + G_2(q) \cdot u_2(t) + \ldots + G_{10}(q) \cdot u_{10}(t) + v(t) \]  

Where \( G_i(q) = \frac{1.1q^{-1} + 0.2q^{-2}}{1-1.1q^{-1} + 0.8q^{-2}} \) and the other transfer functions are of similar order and structure, see Appendix 4 for full details. The noise signal \( \varepsilon(t) \) is a white noise signal of zero variance, and the disturbance, \( v(t) \) is low pass filtered, where \( v(t) = \frac{1}{1-0.9q^{-1}} \cdot \varepsilon(t) \).

The prediction error in this example, like the previous cases, is analysed in terms of the Root Mean Squared Error (RMSE) of 100 repeated tests. In addition the mean variance of the models is calculated.

The first example in this section shows the results of the tests where the system was excited with ten inputs which were independent to one another, and the second example presents the findings when the system was excited with ten inputs that were highly correlated. The correlation was achieved by relating the first two inputs to the remaining ones through linear equations. Details of the data sets generated for this test are provided in Table 6.11. The data sets generated using independent input variables have a low mean condition number, whereas those generated using the highly correlated variables have much larger condition numbers, indicating the data is ill conditioned. The number of latent variables was determined the same way as the previous examples, where the data set was split into groups and cross validated.

<table>
<thead>
<tr>
<th>Signal Type</th>
<th>Signal to Noise Ratio (dB)</th>
<th>Excitation Order</th>
<th>Mean Condition Number of Inputs</th>
</tr>
</thead>
<tbody>
<tr>
<td>RGS (0.01) Correlated</td>
<td>10</td>
<td>11</td>
<td>&gt; 1000</td>
</tr>
<tr>
<td>GBN (64) Not Correlated</td>
<td>10</td>
<td>50</td>
<td>0.7</td>
</tr>
<tr>
<td>GBN (64) Correlated</td>
<td>10</td>
<td>50</td>
<td>&gt; 1000</td>
</tr>
<tr>
<td>GBN (64) Not Correlated</td>
<td>0.9</td>
<td>50</td>
<td>0.7</td>
</tr>
<tr>
<td>GBN (64) Correlated</td>
<td>0.5</td>
<td>50</td>
<td>&gt; 1000</td>
</tr>
</tbody>
</table>

Table 6.11; Identification Signal Properties
The first example looks at the predictions when all inputs are independently excited with a GBN 64 signal, and where the validation signal is a step test. Figure 6.25 shows the RMSE of the models, where the OE prediction error is not shown as the error is very large. The OE result was caused by the data being insufficiently exciting for this particular method and the disturbance not being white noise. For models identified using fewer than 1000 samples, most methods provided predictions with low errors with the exception of PEM, and for models identified using more than 1000 samples PEM produced the lowest error. As all the inputs are independently exciting, all of the methods except OE worked well. It must be noted that although OE does not perform well and the error is too large to show in this plot, the hybrid OE-PLS performs similar to the other hybrid methods providing a significantly lower error over its base model.

![Figure 6.25; RMSE of the Un-correlated System Identified with RGS 0.01](image)

Figure 6.26 shows the variance in the errors produced by each of the models when they were identified 100 times. PEM, OE-PLS and ARX show high variations until the size of the data sets become sufficient to identify models, which occurs after when 5000 or more samples are used. All of the PLS based methods show low variance, and demonstrate repeatable performance.
In the following test a model with ten correlated inputs was constructed to investigate the advantages of PLS in highly correlated systems. The inputs $u_1$ and $u_2$ are correlated with the other inputs through the linear relationships shown between equations 6-19 and 6-26. The data for this exercise was generated with the input variables defined as follows:

$$y(t) = G_1(q) \cdot u_1(t) + G_2(q) \cdot u_2(t) + \cdots + G_{10}(q) \cdot u_{10}(t) + v(t) \tag{6-18}$$

$$u_3 = 0.15 \cdot u_1(t) + 0.7 \cdot u_2(t) + v_1(t) \tag{6-19}$$

$$u_4 = 0.61 \cdot u_1(t) + 0.53 \cdot u_2(t) \tag{6-20}$$

$$u_5 = 0.7 \cdot u_1(t) - 0.3 \cdot u_2(t) \tag{6-21}$$

$$u_6 = 0.46 \cdot u_1(t) + 2 \cdot u_2(t) + v_2(t) \tag{6-22}$$

$$u_7 = 1.15 \cdot u_1(t) - 1 \cdot u_2(t) \tag{6-23}$$

$$u_8 = 2 \cdot u_1(t) + 0.3 \cdot u_2(t) \tag{6-24}$$

$$u_9 = 1 \cdot u_1(t) - 0.92 \cdot u_2(t) \tag{6-25}$$

$$u_{10} = 1.12 \cdot u_2(t) \tag{6-26}$$
Where \( G_1(q) = \frac{1.1q^{-1} + 0.2q^{-2}}{1-1.1q^{-1} + 0.8q^{-2}} \) and the other transfer functions were of similar order and structure. The noise signals \( \varepsilon_1(t) \) and \( \varepsilon_2(t) \) were white noise signals with zero variance, and the disturbances \( v_1(t) \) and \( v_2(t) \) were low pass filtered with \( v_1(t) = \frac{1}{1-0.9q^{-1}} \cdot \varepsilon_1(t) \) and \( v_2(t) = \frac{1}{1-0.7q^{-1}} \cdot \varepsilon_2(t) \) respectively. The noise structure of this process is not complex, and is only correlated with two of the inputs. The full system used for this test is provided in Appendix 5.

The first example uses a low exciting RGS signal of pass band 0.01 for identification, and the test signal is a step test. In this test, OE and PEM methods do not converge in many of the runs and give large errors, which affect their respective hybrid methods, OE-PLS and PEM-PLS. When the one step ahead prediction provided by the base model is large, the hybrid methods were found to fail to converge. The error values of these four modelling methods are not presented in this section as their magnitudes were so high.

The RMSE for the models identified for this test are shown in Figure 6.27, where all of the identified models provide errors with relatively low magnitudes. This shows that a low excitation signal was sufficient to excite this system, and that in practice not all systems need high excitation signals such as PRBS or GBN to provide satisfactory models. Additionally, the ARX and the ARMAX methods are not affected by the high correlation. This suggests that PLS is not strictly necessary in highly correlated data sets, which is an assumption often used in other research studies. The ARX-PLS and ARMAX-PLS hybrid methods offer a slight improvement over their base methods of ARX and ARMAX respectively.
Figure 6.27; RMSE of the Correlated System Identified with RGS 0.01

The variance between the methods is compared in Figure 6.28. Similarly to the uncorrelated example, higher variance was observed for the models identified using fewer of samples.

Figure 6.28; Variance of the Correlated System Identified with RGS 0.01
Although the low excitation signal can identify satisfactory models it is shown in the following examples that the higher signal power of the GBN signal provides models with lower errors at the 30 step ahead prediction horizon. The next example involved modelling the correlated system when a data set obtained using a GBN signal of clock period 64, which was then tested using a step signal. The overall error was found to be lower when compared with the models identified with RGS. The hybrid methods perform marginally better in this example, which is due to the existence of correlations, which are modelled using the capabilities of the PLS technique.

Figure 6.29; RMSE of the Correlated System Identified with GBN 64

The variance of the methods is presented in Figure 6.30 where all the methods are shown to produce models with very similar modelling accuracy.
The next example is where the value of the correlation multipliers for inputs $u_i$ and $u_j$ were changed from their original values at each test run. This example aims to show how the methods would respond to a sudden change in the system, requiring the models to be re-identified. Changes in the dynamics of the system can be caused by a variety of reasons, such as part of the plant being isolated, or when there is a mechanical failure, such as a pump or a compressor shutdown. The changes to the dynamics of the system are defined in equations 6-27 and 6-28.

$$u_3 = 0.15 \cdot u_1(t) \cdot ar(1) + 0.7 \cdot u_2(t) \cdot ar(2)$$  \hspace{2cm} 6-27

$$u_4 = 0.61 \cdot u_1(t) \cdot br(1) + 0.53 \cdot u_2(t) \cdot br(2)$$  \hspace{2cm} 6-28

Where $ar$ and $br$ are set to be random numbers between 0 and 0.1.
The RMSE for this test is shown in Figure 6.31. All of the methods show similar performance and the overall error of the model identified using uRLS-PLS method is the lowest.

To understand the results for this test, one of the runs in this example is considered in detail. In this example, the system with the previously described input parameter changes is identified using 5000 samples. Equations 6-29, 6-30 and 6-31 compare the dynamics of the real process, with those obtained using bPLS and uRLS-PLS. The actual plant is:

\[ G_2(q) = \frac{1.1q^{-1} + 0.2q^{-2}}{1-1.1q^{-1} + 0.8q^{-2}} \quad \text{and} \quad G_3(q) = \frac{0.1q^{-1} + 0.6q^{-2}}{1-1.4q^{-1} + 0.6q^{-2}} \]

\[ 6-29 \]

The estimated transfer function using bPLS are:

\[ G_2(q) = \frac{0.3q^{-1} + 0.02q^{-2}}{1-0.9q^{-1} + 0.1q^{-2}} \quad \text{and} \quad G_3(q) = \frac{0.4q^{-1} + 0.08q^{-2}}{1-0.9q^{-1} + 0.1q^{-2}} \]

\[ 6-30 \]

For uRLS-PLS the transfer functions are:

\[ G_2(q) = \frac{0.7q^{-1} + 0.4q^{-2}}{1-1.3q^{-1} + 0.5q^{-2}} \quad \text{and} \quad G_3(q) = \frac{0.1q^{-1} + 0.4q^{-2}}{1-1.3q^{-1} + 0.5q^{-2}} \]

\[ 6-31 \]
Equation 6-31 shows that the calculated parameters of the uRLS-PLS method are closer to those of the actual system, shown in equation 6-29. The reason for the differences in the model is due to the presence of disturbances. This means uRLS-PLS method is providing a model closer to the actual one, thus a lower prediction error.

Figure 6.32; Variance, GBN Correlated Step Dynamic 10%

In this test all of the models showed similar errors, with uRLS-PLS standing out as the one with the lowest RMSE. This is due to its recursive properties. Although a forgetting factor is not utilised the use of the previous values in the computation becomes an advantage when the model has to adapt. The variance of the model predictions for this test is shown in Figure 6.32. The ARX and ARMAX methods provide the most consistent prediction, where the consistency increases as the number of samples used to identify the models is increased. The remaining methods show slightly higher variations, caused by the constant changes in the model parameters on each run, which has an effect on the predicted parameters.

This again highlights the need to use several modelling methods and compare them, rather than just one. As shown, when the system dynamics change the method producing the model with the smallest error could also change. Some industrial MPC controllers only use one type of model, but in situations like these a suitable model may take a longer time to identify. The hybrid methods, in cases where the data is not rich enough for the base method to identify a model, offer a significant improvement in modelling performance.
The next example shows how the modelling methods perform when the structure of the disturbance was changed to an integrated moving average sequence $v(t) = \frac{1 + 0.5q^{-1}}{1 - 0.9q^{-1}} \cdot e(t)$. This test was applied to observe the responses of the models if the type of disturbance is changed at the start of the test from the previously used low pass structure to an Integrated Moving Average (INMA) structure. This model structure suits the ARMAX and PEM methods, as they can fully model the disturbance. However, PEM cannot predict well with highly correlated data sets. As a result ARMAX and ARMAX-PLS perform well with the addition of uRLS-PLS, where the RMSE plot at the 30 steps ahead point is shown in Figure 6.33. The traditional least squares methods bPLS and ARX models have the largest error due to the more complex dynamics of the noise. Although it should be noted, that all of the methods have predicted relatively well.

![Figure 6.33; RMSE, GBN Correlated INMA Disturbance](image)

The variance of these models is shown in Figure 6.34, and the ARX-PLS shows high variance at fixed sample lengths between the repeat tests. The result of this can also be observed in the RMSE plot where the average error is high for this method. This is due to the calculated ARX model providing large coefficients, as the raw variance of the ARX method itself is low.
6.2 Guidance on the Application of MPC to Electrospinning and Solvent Casting

The previous section has shown that in the presence of suitable data, accurate models of complex processes can be obtained using a variety of methods. The identification of an accurate model is a critical aspect in the application of MPC. The previous sections do not describe how in the presence of suitable data, collected from the physical system, MPC can be applied. This section of the thesis aims to provide guidance on how MPC could be applied to the two processes investigated in this work. Unfortunately, application of MPC to the real processes was not possible during this research project because of time constraints. To begin, the solvent casting method is presented.

6.2.1 MPC for Solvent Casting

Solvent casting is a batch style process where a polymer solvent mixture is evaporated from a glass substrate to leave behind an ultra thin film. The current problems associated with this process have been addressed in Chapter 3 and the tests conducted have identified some of the important variables. The most important outcomes of this analysis were the determination of the quality variables, the pit size preferred by the nerve cells and identification of the difficulty in producing films with a low variance pit diameters using flat substrates. The latter problem has been overcome by using grooved substrates and the next stage is to discover the actual material properties of the films and optimal operating conditions required to produce films with desired properties. Additionally an improved dispensation method is required to improve consistency of the produced films.
Three approaches can be taken when applying advanced control to the solvent casting process. The first is to regulate the environmental conditions while the film is forming; the second is the rapid heating/cooling substrate approach where the substrate is placed on a heating/cooling Peltier device; the third is controlling the evaporation in a pressurised vessel by manipulating the air pressure and temperature. The environmental approach is the currently applied method and is based on conditioning the evaporation environment and relying on this to regulate the evaporation of the solvent. Changes in the environmental conditions during the batch mean there will be increased airflow which affects the surface morphology of the films which is undesirable. The input variables of the solvent casting process can be listed as:

- Temperature
- Humidity
- Solvent - Polymer mixture levels
- Mixture density
- Airflow speed
- Thermodynamic properties of the mixture
- Heat transfer properties of the mixture
- Raw polymer properties of the batch

These input variables are all highly correlated and will have an effect on each other as well as the final product, so it is necessary to include all the variables in the model. The properties of the mixture and how this reacts to the environment is the most important area to understand when this control approach is applied.

The MPC strategy for the environmental evaporation method is not as feasible as using MSPC for quality improvement. This is because the evaporation of the solvent takes approximately 5 minutes, after which time the pits and pores are formed on the film. The time taken to make a change to any of the inputs such as the humidity or the temperature will exceed the evaporation time and hence it will not be possible to adjust the pit size using this approach. By using MSPC charts it would be possible to model the inputs, detect out of control situations and determine whether the material quality would satisfy the requirements. An overview schematic of this system is shown in Figure 6.35. The measured variables during operation are temperature, humidity and airflow and they are not changed significantly throughout the batch. The pre-batch measurements of the solution mixture consist of the remaining variables listed above. To build a cause-effect model a suitable output variable is required. The potential quality variables that are essential for cell success for the solvent cast films that have been identified are as follows, assuming a flat substrate is used:

- Diameters of the pits and pores on the surface of the film
• The area covered by the pits and pores
• The mechanical properties of the film, such as tensile strength and stiffness

In the case of grooved films only the mechanical properties of the film is relevant. This is an important property as the film needs to be able to remain intact in the patient’s body long enough to support the new nerve axons before biodegrading.

The second control solution would be applied if the evaporation is not dependant on the air temperature, but the temperature of the substrate itself. In the environmental control case, as the solvent evaporates it has a cooling effect on the substrate as it loses heat. If the substrate was to be placed on a rapid heating/cooling Peltier device, the temperature of the substrate could be kept constant, leading to consistent evaporation conditions. In this case, the temperature of the substrate could be regulated with a local PI controller.

The initial studies presented in Chapter 3 showed that greater uniformity was achieved using cooler temperatures as this leads to slower evaporation. Further effects could be investigated by keeping the air cool, but allowing the substrate to be kept at a higher temperature. The effect of the evaporation profile and the mechanical properties could then be investigated and modelled using PLS or similar techniques.
A full MPC system would be beneficial to this process if the whole procedure was automated from the raw material mixing stage to the drying the films stage. However, as the method is still at laboratory stage the identification of optimal operating parameters is of priority. By considering all the input variables, this could be achieved in the latent variable space of the PLS model using Quadratic programming. The PLS model could then adapt from batch to batch to track dynamic changes in operation.

Using this approach, decision points would be identified during the batch. At these decision points, the optimiser would attempt to move the end-point quality closer to the desired set point by manipulating conditions in the process. The cost function that would need to be minimised is shown in equation 6-32 adapted from (Flores-Cerrillo and MacGregor 2004b):

\[
\min_{\Delta(t|\theta)} (\hat{y} - y_{qsp})^T Q_1 (\hat{y} - y_{qsp}) + \Delta^T Q_2 \Delta t + \lambda T^2
\]

6-32

Where \(\theta_i\) is the decision point at time \(i\), \(\hat{y}\) is the end point quality estimate, \(y_{qsp}\) is the quality set-point, for example, the required stiffness of the film, \(Q_1\) is a diagonal weighting matrix highlighting the importance of the end point quality variables (applicable if more than one variable), \(Q_2\) is a diagonal movement limiting matrix to control the control behaviour. \(T^2\) is the Hotelling’s statistic, \(\lambda\) is a weight which determines how close the solution is constrained to the region of the score space that was determined from the previous batches. Expression 6-32 is valid subject to the expressions 6-33, 6-34, 6-35 and 6-36:

\[
\hat{y}^T = (\Delta t + \hat{t}_{\text{present}})^T Q^T
\]

6-33

\[
T^2 = \sum_{a=1}^{A} \frac{(\Delta t + \hat{t}_{\text{present}})^2_a}{s^2_a}
\]

6-34

\[
\Delta t_{\text{min}} \leq \Delta t \leq \Delta t_{\text{max}}
\]

6-35

\[
\Delta t^T = t^T - \hat{t}_{\text{present}}^T
\]

6-36

The quadratic solution to this minimisation problem and the model inversion, required for the calculation of the optimum manipulated variables is presented in Appendix 6 and Appendix 7 respectively.
The diagram in Figure 6.36 shows how a model based controller identified using historical data can be updated with new batch information collected during the process. Such an approach would be required if the process dynamics change from batch to batch.

![Figure 6.36; Batch MPC for Solvent Casting](image)

Next the electrospinning process is addressed.

### 6.2.2 MPC for Electrospinning

The electrospinning process is a continuous process used by the BRG to make tendon support structures. The process is explained in detail in Chapter 4, where the existing problems, the modifications applied during this project and the factorial study to establish the effect of the process variables were presented. This section aims to provide information on how to set-up a model based control scheme to control and improve the current process.

Model predictive control is a widely used control methodology in today’s industry and the justification for this is presented in detail in Chapter 2. Data based models provide a convenient way to model complex processes such as the electrospinning process. Once such a model is identified, it can be used to provide future predictions of plant behaviour, which can be used to determine necessary changes to the process conditions that are required to regulate quality of the material produced.

Before applying MPC a model of the process is necessary, and the first step in model estimation is data collection. The tests applied in Chapter 4 are not sufficient to build a working model and were only used for establishing the effects of the variables on the final product. A larger factorial study for the whole system or a DOE study for individual solvent-polymer mixture types and concentration levels is required. Tests need to be conducted with sufficient excitation, using PRBS or similar signals.
Such test would enable the dynamics of the process to be accurately identified. However for this, it would be necessary to install a high speed camera to record the state of the Taylor cone and the final fibre diameters. It would not be possible to develop a generic model of the electrospinning process as the solvents and polymers used may change, which affects the process dynamics considerably, in particular changing the intrinsic properties of the electrospun liquid.

If extensive tests are to be conducted then it is essential that a syringe heater/cooler is incorporated in to the electrospinning set-up. An overall illustration of this methodology is presented in Figure 6.37. The solution properties are pre-recorded before electrospinning, then with the aid of DOE, camera and syringe heater/cooler system, a rapid data gathering exercise is applied. Following this a mathematical model is estimated and validated using real test data. A number of models can be identified with this technique and certain patterns may be observed depending on the solution properties and therefore the model selection could potentially be fully automated.

![Figure 6.37; Model Based Approach Overview for Electrospinning](image)

The scheme proposed in Figure 6.37 depends on sufficient data being collected through the DOE methods presented in Chapter 2. The inclusion of solution properties, visual Taylor cone data and fibre diameter data would enable the identification of a reliable process model. The modelling methods presented in Chapter 5 should be considered for identifying such a model, in particular the statistical data based methods, such as PLS and its variants. As previously mentioned the constraints in the process need to be considered as electrospinning is only possible between certain field strengths and at certain solvent conductivity levels.

The discussion in Chapter 4 regarding fast MPC is also relevant to this section. The electrospinning process contains fast dynamics and changes to the input variables, such as voltage or solution
temperature have an immediate impact on the Taylor cone and the fibre diameters. A further complication is that any model of the process will need to adapt to the build-up of an insulating layer of film on the collector when the process is used to produce a thick fibre mat. A complete model based control system will benefit the system threefold:

- It will allow the determination of the optimal operating conditions required to produce a fibre mat with desired properties.
- In the event of failures, such as running out of solvent or an interference in the electric field, the control system will be able to shut the process down safely without spoiling the fibre mat on the collector.
- The manufacture of layered mats with different properties would be possible without the need to stop and reset the experiment.

The MPC could be implemented using a standard PC with real-time software. The experimental chamber currently utilises LabVIEW together with DAQ cards from National Instruments and therefore LabVIEW’s MPC toolbox would be feasible to use. It is also possible to run MATLAB code within LabVIEW and model adaptation could be achieved using custom algorithms created in MATLAB. Additionally if the PC is not sufficiently powerful to perform the required computation in time then an FPGA based controller, such as the National Instruments Compact RIO could be used as a standalone controller. This controller is pre-programmed with LabVIEW, runs as a standalone computer and can communicate with a PC in real time if required.

It is important to note that the electrospinning system has constraints and it is not possible to create continuous fibres unless the electric field strength and the flow rate is at an appropriate level. As presented in Chapter 4 with the map of variables, many of the process inputs affect the flow rate and the electric field which affects the Taylor cone. It would be possible to derive a relationship between the low side current and the flow rate to determine the working region for each solvent-polymer mixture using the methods presented in Chapter 5.

With the model identified and the constraints set, the MPC would be ready for implementation. The cost function would need to be determined alongside the optimisation parameters. A suggested cost function for the process would be to minimise the SPE in the latent variable space shown in expression 6-37 (Laurí et al. 2010).

$$J(t_d) = \sum_{j=1}^{n} \left[ r_j(k) - \hat{y}_j(k) \right]^2 W_y + \lambda_u \left\| x_{def}(k) W_u \right\|^2_2$$

6-37
Where \( t_d \) is the decision variable, \( r_j(k) \) is the reference at time \( k \), \( \lambda_u \) is weighting for the control action, \( W_y \) and \( W_u \) are diagonal output and control variable weightings. Further expressions used to perform the minimisation are shown in Appendix 8.

The model for the electrospinning process could be identified using PLS as the process is complex and deriving it from physical equations is likely to be infeasible.

### 6.3 Summary

This section demonstrated that when sufficient data is collected from a manufacturing processes predictive models can be identified using the methods described in this work. These models would enable the ideal operating parameters for creating fibre mat and nerve film morphologies with desired properties to be calculated. Although more work is required on solvent casting and the electrospinning process, this chapter provides a proof of concept that such an approach is feasible, despite the likelihood of small, highly correlated data sets.

Due to the size of the project and the time limitation not all methods were possible to apply to the processes. Instead a guide and suggestions on how to apply MPC to solvent casting and electrospinning has been presented. The suggestions and methods illustrated in this thesis should be sufficient enough for a future user with a suitable background, to model and then implement MPC on to the processes. The use of DOE, MSPC for monitoring and also other methods for modelling will need to be applied together with a fast MPC algorithm to overcome the fast dynamics of the electrospinning process. For solvent casting, a modelling and optimisation approach is recommended together with the use of modified substrates with grooves.

The numerical results presented in this chapter show how different methods can provide benefits in different scenarios. Following recent publications (Lauri et al., 2010), (Rossiter et al., 2010), (Prívara et al., 2013) the suitability of modelling continuous systems with bPLS is analysed and compared to the more traditional system identification methods. The studies investigated in this chapter, show that it was not possible to determine definitive rules regarding which type of model should be used to identify a specific dynamic system. However, some general conclusions can be drawn from the carried out work.

Firstly, bPLS is a suitable method to identify models to be used for prediction. However, even in instances where there are highly correlated variables, it is not always the best choice. In situations where noise is high, unbiased versions of the PLS provide the smallest errors, but not the most consistent models. Secondly, the bPLS method is only advantageous when applied to data sets that contain low levels of excitation and are short in length. This is a result of the maximising the
covariance amongst the latent variables feature of the SIMPLS algorithm. Thirdly, the OE, PEM and ARMAX methods struggle to identify models when the data has insufficient excitation. However their hybrid PLS versions provide significant improvements.

As the data length used to identify models gets larger, the prediction of the non-PLS based methods improves. This is also observed when the data becomes more exciting, where as shown with the GBN 8 signal in the Wood and Berry simulation, a short sample of data is sufficient to identify models that provide low errors when using the PEM and ARMAX methods. This means that as long as there is enough excitement, regardless of data length, the time series methods can perform as well as and occasionally better than PLS. This shows that PLS is not always necessary when the data availability is limited. This is an important result, as limited data is one of the main justifications for using PLS.

Following on from the Wood and Berry column a further study was presented where two simulations were investigated, one with ten independent inputs and one output, and one with ten inputs with extreme correlations and one output. In the case of the independently excited system, when a data set with low excitation was used to identify the process, all the identification methods studied in this work were able to calculate models that produced low errors. This demonstrated that it is not always necessary to use a signal for identification with high excitation such as PRBS, GBN or white noise, where the actuators and other system components could potentially wear out prematurely due to the high workload during the identification exercises.

The system that contained high correlations was used to investigate how the various identification methods were able to cope with highly correlated data sets. It was shown that with this system, all of the methods produced reasonable models, with the exceptions of OE and PEM. Although on occasions accurate OE and PEM models could be identified, there was significant inconsistency with these modelling methods, which suggests that they are not suitable for use with highly correlated data. The only occasion where bPLS was seen to be necessary was when the system was identified using short sequences of data, that contained little excitation. This is an interesting result, as the main justification for using PLS is that it is able to model highly correlated data sets. This exercise has shown that if there is sufficient data, then any of the other identification methods studied in this work can also be used. However, it should be stressed that the focus of this work has not been on identifying models from “short-fat” matrices, where research has clearly demonstrated the benefits of using PLS.

In situations where PLS is necessary, the unbiased models including the newly proposed hybrid methods, were shown to offer small improvements in the predictions. This is important if process industries such as such as petrochemicals and pharmaceuticals are considered, where a very small improvement in control could mean extra profitability and higher returns in terms of production.
As a final observation, there is not a clear outstanding method when identifying process models, and suggestions can be made depending on the model type, the correlation of the variables and how complex the noise dynamics are. These properties can be analysed using the power spectrum of the data and the condition number of the data matrices. In the cases of ill-conditioned data matrices with very high condition numbers, the PLS based methods should be considered. However, several modelling methods should be trialled alongside this to get the best performance from the model for an MPC controller, as PLS does not always provide the models with the lowest errors.
7 Project Conclusions and Recommended Future Work

This thesis has presented the work conducted on a research project aimed at providing groundwork for the potential application of advanced control systems to electrospinning and solvent casting process. These methods are used at the University of Manchester for artificial tissue scaffold production, which is a key part of cutting edge tissue engineering research. These structures are intended to supersede conventional healing methods. The currently used method of grafting has low success rates, with only a 40% motor function recovery rate when used to treat nerve injuries. Current statistics strongly highlight the need for alternative methods of nerve and tendon healing. Tissue engineering has a far bigger potential than the two areas this thesis focuses on. With further understanding and advancements in research, custom scaffolds and organs could be manufactured for individuals.

From a control systems perspective, which is the main theme of this research, electrospinning and solvent casting are multivariable processes similar to many other systems encountered in the process industries. The solvent casting method is a batch style process where regression methods are necessary for modelling, and electrospinning is a continuous process with correlated variables. This thesis has shown the difficulties encountered when modelling batch-like systems, correlated systems with complex disturbances and complex dynamics. Through the use of simulations, new hybrid methods are shown to provide improvements to some of the problems traditional methods face, by combining the PLS method with conventional methods to predict multiple steps ahead in to the future.

In summary, this chapter reflects on the initial problems and aims of the project, and compares them to the completed work, the results and the contributions. The project is summarised in the first sub section and the potential future work is discussed in the following.

7.1 Conclusions

As explained in Chapters 3 and 4, the electrospinning and solvent casting processes needed considerable engineering work to bring them up to the state of the art research facilities they are today. Analysis of the existing literature and original equipment pointed towards similar issues in both processes that were also encountered globally by other research groups. These can be listed as:

- Lack of repeatability in the conducted tests and manufacturing conditions.
- Lack of knowledge of the effects of the variables on the final products.
- Lack of computer links and data logging.

In this project the identified technological shortcomings, such as data logging, stable environmental conditions and minimisation of disturbances have been resolved. For both manufacturing methods,
the most important variables have been discovered and further suggestions have been made for their improvements. However, there is scope for further development through additional studies in to the effects of all the variables on the final products.

For electrospinning, the main contributions have been the discovery of the effect of electric field intensity on the needle tip, and the effect of the electric field shape on the fibre paths and diameters. The ability to influence the electric field intensity enables control of fibre diameters within certain ranges. This is dependent on the solution concentration and its properties, such as viscosity and conductivity. In this study the importance of determining the properties of the solution has been pointed out, as this is important for modelling the process and selecting working ranges. Another contribution was the proposal to use the current provided to the process, to monitor the quality of the fibre jet. Once the low side current is related to Taylor cone shapes with the aid of a camera, it has the potential to be used as a controlled variable, which will in return govern the fibre diameters. The electrospinning part of the project concludes with how the fibre diameters can be influenced using the electric field intensity. However, further measurements such as Taylor cone shapes are necessary to move to the next stage of model building. Once the necessary data is collected and a model is obtained, the camera monitoring the Taylor cone shapes would no longer be necessary. The proposed method of influencing the fibre diameters and keeping the electric field free from disturbances is an important advancement in electrospinning, and can be exploited to its full potential with future projects in this area. The modelling of the system and application of MPC to this process as presented in Chapter 6 would be the first step in improving the product quality and consistency. This will in addition will give extra flexibility to the manufacturing chamber in terms of the range of fibres it can produce.

For solvent casting, the determination of a preferred pit size for the NG108-15 nerve cells was the main contribution. Previously, the scaffold studies did not research why cells attached or what size pits or gaps they preferred for attachment. The ideal size was determined to be approximately 20 µm, and this has lead on to the use of grooved films with this width. The introduction of the controlled environment and data logging has enabled repeatable tests to determine which variables affect the films’ morphologies. The inconsistency of the solvent dispensing method was also identified, which contributed to the varied results obtained in different batches and caused varying amounts of internal liquid flow during the formation of the film. The evaporation study of the solvent benefits both the electrospinning and solvent casting processes. This study confirmed evaporation of non-aqueous solvents is more dependent on temperature than humidity, but humidity still plays a major role in surface morphology. The solvent casting study concluded that only using the environmental conditions is not sufficient to create consistent films, and a method of physical pattern casting is necessary.
In addition, recommendations have been made on the physical systems and how MSPC and MPC can be applied to them. The guide in Chapter 6 should provide ample support for any future users of the systems to increase product quality, through the methods presented in Chapter 5.

In parallel to the discussed practical work, system identification studies have investigated the challenge of identifying highly correlated systems using simulations. The main objective of this was to analyse and develop methods that could potentially be used for modelling electrospinning and solvent casting. To improve the existing modelling approaches, hybrid methods were proposed and shown to work successfully. When combined with PLS, the identified unbiased models outperformed their base methods for multi step ahead predictions. The theoretical part of this thesis outlined the bias problem present in the least squares estimators, and how it could be overcome with using alternative methods such as ARMAX, PEM and the new hybrid methods. Most importantly, this part of the work tried to identify instances when the frequently used PLS method is necessary. The studies found that PLS is only justified for short data sets with insufficiently exciting data, which is a limited case when the common application of GBN type identification signals is considered. In addition, it is shown that not all systems need to be identified with highly exciting signals such as the GBN or PRBS, and prediction error methods can still identify accurate models of uncorrelated processes in the presence of less exciting RGS signals. For highly correlated systems prediction error methods worked as well as the PLS methods, once again undermining the necessity of PLS. The further shortcoming of PLS was also highlighted when the noise increased in complexity, and the ARMAX and the hybrid methods outperformed it over multiple runs. Additionally the drawback of using just one modelling method was highlighted. Using one method is common practice, and by utilising other methods, more accurate models could potentially be obtained. This work recommends that varieties of identification methods are trialled and compared at every model update.

In summary, the majority of the goals set at the beginning of this project have been completed. The only exception is fully modelling the physical systems, where difficulties were faced due to the complexity of the processes and the time taken to identify some of the important variables. However, essential research has been conducted towards modelling in the time available for the project, and novel modelling methods have been proposed that has a wider application than just electrospinning and solvent casting. In the given time of three years, significant progress has been made in these fields and their scope has been widened, which has laid the foundation for numerous potential future projects. The most promising of these that may be followed with further research are discussed in the following section.

### 7.2 Future Work

Following the end of this thesis, the next steps for the continuation of this research are explained in this section. The suggestions are made in order of their appearance in this thesis, starting with
solvent casting, followed by electrospinning, and finally the multivariable system modelling methods. The areas selected for further exploration were chosen as they hold the greatest potential benefits for this research.

Creating nerve conduits using solvent casting has been shown to be a viable method in the previous studies conducted at the University of Manchester (M. Sun et al., 2010). However, none of the previous studies examined what conditions the cells preferred in terms of surface morphology. The study in this thesis concentrated on discovering surface morphologies of the PCL/PLA blend films for nerve cell growth. With the identification of the ideal size of pits for cell attachment and the use of grooved films, it is shown that ideal surface morphologies can be achieved. From this point onwards, the mechanical properties of the films need to be investigated, including how the environmental variables and solution mixture rates affect these. To help with the application of the control system, MSPC methods could be implemented as presented in Chapter 5. However, initially a DOE study would be necessary to capture more of the system dynamics for modelling. Once the model is complete the operating conditions can be optimised for certain material properties. With this, a PLS model relating the mechanical properties to the environmental conditions could be constructed, which could be used to determine the optimum conditions to produce a film of desired specifications. The MSPC monitoring methods such as the $T^2$ and SPE charts can be used in the control system for implementing improvements in a batch to batch manner.

To implement the said studies further modifications are required to the current solvent casting configuration. As highlighted, the solution temperature is important in terms of evaporation rates, and relying on the temperature of the environment is not necessary. A rapid heating/cooling system where the glass substrate sits would be sufficient to evaporate the solvent in a more controlled manner. The necessity of an automated dispensing system is also shown in the evaporation study results in Chapter 3. This would consist of an accurate pump that can deliver desired volumes of liquid quickly and evenly, so the internal flow of the solution is consistent. More importantly, in addition to the surface characteristics, potential chemical modifications on the film surfaces need to be investigated, which will encourage further alignment and growth of the cells. Using proteins and peptides for laying down chemical paths and chemo-tactical cues for cell growth is an active area of research, and could potentially eliminate the need for specific surface morphologies for cell growth. The next stage after mechanical properties data gathering and optimisation in this process, is the in vitro testing of the developed films on laboratory test animals.

Additionally, the solvent casting set-up can be re-designed as a whole, where the evaporation is controlled in a pressurised chamber. Rather than the solvent evaporating instantly after dispensing the pressure of the vessel could be manipulated to change the evaporation time of the solvent. This
will require the design and build of a new system with new actuators. Again the methods introduced and explained in this thesis would be applicable to this new system.

The electrospinning process requires further engineering work in terms of constructing a moving platform to control the needle protrusion and spinning distance accurately. This is suggested to control the electric field intensity and its shape which influences the fibre diameters. However, in parallel to this a full three dimensional finite element analysis of the electric field shapes would be mandatory where the protrusion distances can be related to the field shape. Following this, the introduction of a camera to capture the behaviour of the Taylor cone is essential along with image analysis software. Relating this to the final fibre diameters, electric current, electric field shape and intensity, will enable the creation of models that can be used for control. The models would be built using the methods presented in Chapters 5 and 6. Once relevant data is gathered and models are obtained, the need for the camera will also be eliminated as predictions will be made depending on the process current and the other measured variables such as solution temperature and viscosity.

As previously mentioned, the solution properties including its conductivity are important for incorporation into the model. Once an accurate data model is obtained the system can be put under closed loop in an MPC scheme, where the controller would pick the optimum conditions depending on the specified cost function to create a desired fibre mat. Furthermore a moving stage can be utilised to create three dimensional objects, or alternatively the spinneret would be moved in the y and z plane, similar to a 3D printer.

The hybrid numerical methods proposed in this thesis have shown their advantages over conventional methods such as RLS, PLS, ARX, ARMAX, OE and PEM. However, they need further testing in a real MPC scheme. Their ease of updating for adaptation needs to be further identified, alongside testing with other model varieties, such as large industrial plant type models with hundreds of variables and systems with faster dynamics. A more thorough study of determining which method provides an accurate model by only analysing the input data is needed to overcome the computational expense, created by using a range of modelling methods at every model update. Using a high number of multiple simulations and different model and disturbance types, a thorough study could be conducted where the methods that are most likely to work in certain scenarios will be determined. Although the initial investigation and data gathering stage for this project would be laborious, the future benefits would outweigh the costs. Although mainly utilised in the field of neural networks, an alternative approach for this could be using the stacking models approach (Breiman, 1996), (Wolpert, 1992) where the final predicted output is the average obtained from all of the methods.
8 Bibliography


Downes, Sandra, Atefeh Mobasseri, Mingzhu Sun, and Giorgio Terenghi. "MEDICAL DEVICE". The University of Manchester.


Formhals, A. "Process and Apparatus for Preparing Artificial Threads". Mainz-on-the-Rhine.


Mobasser, Atfeh. 2012. 2nd Year Transfer Report.


Ramakrishna, Seeram. 2005. An Introduction to Electrospinning and Nanofibers. World Scientific.


Robb, Brendan. 2012. Second Year Transfer Report.


Robb, Brendan. 2012. Second Year Transfer Report.


Xing, Xiaobo, Yuqing Wang, and Baojun Li. 2008. “Nanofibers Drawing and Nanodevices Assembly in Poly(trimethylene Terephthalate).” Optics Express 16 (14) (July 7): 10815–10822.
9 Appendices

Appendix 1

PID derivation of IMC is shown in this section adapted from Tham (2002).

If the IMC loop representation above is re-arranged to the form below:

Now the PID representation can be assumed to be:
Where the transfer function can be described as:

\[
G_c(s) = \frac{G_{IMC}(s)}{1-G_{IMC}(s)\hat{G}_p(s)} = \frac{\hat{G}_p(s)^{-1}G_f(s)}{1-\hat{G}_p(s)G_f(s)}
\]

And if the process is assumed to be:

\[
\hat{G}_p(s) = \frac{K}{1+\tau s} \cdot e^{-\theta s}
\]

Where the time delay is approximated with the Padé approximation:

\[
e^{-\theta s} \approx \frac{1-\frac{\theta}{2}s}{1+\frac{\theta}{2}s} \quad \text{so} \quad \hat{G}_p(s) = \left(\frac{K}{1+\tau s}\right) \cdot \frac{1-\frac{\theta}{2}s}{1+\frac{\theta}{2}s}
\]

And like in Chapter 2 the invertible and non-invertible parts can be separated:

\[
\hat{G}_p^+(s) = \frac{K}{(1+\tau s) \cdot (1+\frac{\theta}{2}s)}
\]

\[
\hat{G}_p^-(s) = 1 - \frac{\theta}{2}s
\]

Now the controller transfer function can be written as:

\[
\hat{G}_c(s) = \frac{\hat{G}_p^+(s)^{-1}G_f(s)}{1-\hat{G}_p^-G_f(s)} = \frac{(1+\tau s) \cdot (1+\frac{\theta}{2}s)}{K} \cdot \frac{\frac{1}{1+\tau_f s}}{1-\frac{\theta}{2}s} \cdot \frac{1}{1+\tau_f s}
\]

\[
\hat{G}_c(s) = \frac{(1+\tau s) \cdot (1+\frac{\theta}{2}s)}{K \cdot (\tau_f + \frac{\theta}{2}) \cdot s}
\]

The conventional PID controller can be written as:

\[
G_{PID}(s) = K_c \left(1 + \frac{1}{T_i s} + sT_d\right)
\]
Then the $\hat{G}_c$ obtained through the IMC algorithm can be partitioned as:

\[
K_c = \frac{(\tau + \frac{\theta}{2})}{K \cdot (\tau + \frac{\theta}{2})}
\]

\[
T_i = \frac{\theta}{2} + \tau
\]

\[
T_d = \frac{\pi\theta}{\theta + 2\tau}
\]
Appendix 2

Further images of the developed electrospinning hardware are presented in the following pages. First the whole setup is shown once again.
The following image shows the large collector.

The following image shows the syringe pump, the needle and the silicone tube going to the spinneret.
The next image shows the two DAQ cards used for data logging and control.

![DAQ cards](image)

The following image shows the computer system used for controlling the electrospinning process.

![Computer system](image)
The next image shows the fibre optic voltage measurement system’s transmitter and its battery.

The next image shows the point2point system’s receiver where the data is sent to the DAQ card.
Appendix 3

Further images of the electrospinning software. The first image shows the voltage control screen of the main user interface. The trap door actuator and the voltage cut buttons can be seen alongside the voltage level and the current limit settings.

The next image shows the syringe pump control interface.
The next screenshot shows the data logging screen, where the start logging button, the file path and the current graphs can be seen. These were taken when the system was not running so only small signal noise is visible.
The next image shows the data logging code, the solenoid and the voltage cut digital signals.

The next two images in the next page show the whole syringe pump communication software.
Appendix 4

The full transfer function of the uncorrelated system in Chapter 6.

\[ y(t) = G_1(q) \cdot u_1(t) + G_2(q) \cdot u_2(t) + G_3(q) \cdot u_3(t) + G_4(q) \cdot u_4(t) + G_5(q) \cdot u_5(t) + \\
G_6(q) \cdot u_6(t) + G_7(q) \cdot u_7(t) + G_8(q) \cdot u_8(t) + G_9(q) \cdot u_9(t) + G_{10}(q) \cdot u_{10}(t) \]

Where:

\[
G_1(q) = \frac{1.1q^{-1} + 0.2q^{-2}}{1-1.1q^{-1} + 0.8q^{-2}} ,
\]
\[
G_2(q) = \frac{1.0q^{-1} + 0.5q^{-2}}{1-1.4q^{-1} + 0.7q^{-2}} ,
\]
\[
G_3(q) = \frac{0.1q^{-1} + 0.6q^{-2}}{1-1.4q^{-1} + 0.6q^{-2}} ,
\]
\[
G_4(q) = \frac{0.8q^{-1} + 0.3q^{-2}}{1-1.5q^{-1} + 0.55q^{-2}} ,
\]
\[
G_5(q) = \frac{1.1q^{-1} + 0.3q^{-2}}{1-1.2q^{-1} + 0.7q^{-2}} ,
\]
\[
G_6(q) = \frac{1.2q^{-1} + 0.6q^{-2}}{1-1.3q^{-1} + 0.8q^{-2}} ,
\]
\[
G_7(q) = \frac{1.0q^{-1} + 0.5q^{-2}}{1-1.5q^{-1} + 0.6q^{-2}} ,
\]
\[
G_8(q) = \frac{0.8q^{-1} + 0.35q^{-2}}{1-0.9q^{-1} + 0.1q^{-2}} ,
\]
\[
G_9(q) = \frac{0.4q^{-1} + 0.8q^{-2}}{1-1.3q^{-1} + 0.8q^{-2}} ,
\]
\[
G_{10}(q) = \frac{1.1q^{-1} + 0.7q^{-2}}{1-1.5q^{-1} + 0.68q^{-2}}
\]

\[
u3=0.15*u1+0.7*u2; \\
u4=0.61*u1+0.35*u2; u5=0.7*u1-0.3*u2; \quad \%u4=0.61*u1+0.5; \\
u6=0.46*u1+2*u2; u7=1.15*u1-u2; u8=2*u1+0.3*u2; \\
u9=u1-0.921*u2; u10=u1*u2+u2*1.12; \\
\]

\[
G_1(q) = \frac{1.1q^{-1} + 0.2q^{-2}}{1-1.1q^{-1} + 0.8q^{-2}} ,
\]
\[
G_2(q) = \frac{1.0q^{-1} + 0.5q^{-2}}{1-1.4q^{-1} + 0.7q^{-2}}
\]
**Appendix 5**

The full transfer function of the correlated system in Chapter 6 is

\[ y(t) = G_1(q) \cdot u_1(t) + G_2(q) \cdot u_2(t) + G_3(q) \cdot (u_1(t) \cdot 0.15 + u_2(t) \cdot 0.7) + G_4(q) \cdot (u_1(t) \cdot 0.61 + u_2(t) \cdot 0.53) + G_5(q) \cdot (u_1(t) \cdot 0.7 - u_2(t) \cdot 0.3) + G_6(q) \cdot (u_1(t) \cdot 0.46 + u_2(t) \cdot 2) + G_7(q) \cdot (u_1(t) \cdot 1.15 - u_2(t)) + G_8(q) \cdot (u_1(t) \cdot 2 + u_2(t) \cdot 0.3) + G_9(q) \cdot (u_1(t) - u_2(t) \cdot 0.92) + G_{10}(q) \cdot (u_1(t) \cdot 0.1 - u_2(t) \cdot 1.12) \]

Where:

\[
G_1(q) = \frac{1.1q^{-1} + 0.2q^{-2}}{1 - 1.1q^{-1} + 0.8q^{-2}}, \quad G_2(q) = \frac{1.0q^{-1} + 0.5q^{-2}}{1 - 1.4q^{-1} + 0.7q^{-2}}, \quad G_3(q) = \frac{0.1q^{-1} + 0.6q^{-2}}{1 - 1.4q^{-1} + 0.6q^{-2}},
\]

\[
G_4(q) = \frac{0.8q^{-1} + 0.3q^{-2}}{1 - 1.5q^{-1} + 0.55q^{-2}}, \quad G_5(q) = \frac{1.1q^{-1} + 0.3q^{-2}}{1 - 1.2q^{-1} + 0.7q^{-2}}, \quad G_6(q) = \frac{1.2q^{-1} + 0.6q^{-2}}{1 - 1.3q^{-1} + 0.8q^{-2}},
\]

\[
G_7(q) = \frac{1.0q^{-1} + 0.5q^{-2}}{1 - 1.5q^{-1} + 0.6q^{-2}}, \quad G_8(q) = \frac{0.8q^{-1} + 0.35q^{-2}}{1 - 0.9q^{-1} + 0.1q^{-2}}, \quad G_9(q) = \frac{0.4q^{-1} + 0.8q^{-2}}{1 - 1.3q^{-1} + 0.8q^{-2}},
\]

\[
G_{10}(q) = \frac{1.1q^{-1} + 0.7q^{-2}}{1 - 1.5q^{-1} + 0.68q^{-2}}.
\]
Appendix 6

Equation 6-32 can be re-written as a quadratic programming problem (Flores-Cerrillo and MacGregor, 2004b)

\[
\min_{\Delta t(\theta)} \frac{1}{2} \Delta t^T H \Delta t + f^T \Delta t
\]

Where

\[
H = Q^T Q_1 Q + Q_2 + Q_3
\]

\[
f^T = (\hat{Q}_{\text{present}} - y_{\text{sp}})^T Q_1 Q + \hat{t}_{\text{present}}^T Q_3
\]

\[
Q_3 = \text{diag}\left[ \frac{\lambda}{s_a^2} \right]
\]

And \( \Delta t_{\text{min}} \leq \Delta t \leq \Delta t_{\text{max}} \) or if there are no hard constraints \( \Delta t^T = -f^T H^{-1} \)

If \( Q_1 \) is chosen to be equal to \( I \), \( Q_2 = 0 \) and \( \lambda = 0 \)

Then at the end of the batch this forces the quality variable prediction \( \hat{y} \) to be equal to the set-point \( y_{\text{sp}} \).

\[
\min_{\Delta t(\theta)} (\hat{y} - y_{\text{sp}})^T Q_1 (\hat{y} - y_{\text{sp}})
\]

Subject to:

\[
\hat{y}^T = (\Delta t + \hat{t}_{\text{present}})^T Q^T
\]
Appendix 7

Following the control move in the score space calculation presented in Appendix 6, this needs to be converted to future control moves and future process variable predictions.

For each control interval $\theta_i$ the $x$ vector trajectory which is:

$$x^T = [x_{m,measured}^T(0;\theta_i), u_{c,implemented}^T(0;\theta_i), x_{m,future}^T(\theta_i;\theta_i), u_{c,future}^T(\theta_i;\theta_i)]$$

consist of measured process variables for interval $0 \leq \theta < \theta_i$.

If $x_i^T = [x_{m,measured}^T(0;\theta_i), u_{c,implemented}^T(0;\theta_i)]$ and $x_2^T = [x_{m,future}^T(0;\theta_i), u_{c,future}^T(0;\theta_i)]$

Then the remaining trajectories are computed as shown:

$$[x_1^T, x_2^T] = [t^TP_1^T, t^TP_2^T]$$

At time $\theta_i$ the computed $t^TP_i^T$ is not equal to the actual observed trajectories, so an assumption is necessary which is shown below:

$$x_2^T = (t^T + \alpha^T)P_2^T$$, here $\alpha^T P_2^T$ is used as an adjustment for the inequality mentioned above.

To complete this calculation in the score space the following equation is used:

$$t^T = [x_1^T, x_2^T] \begin{bmatrix} W_1 \\ W_2 \end{bmatrix} = x_1^TW_1 + x_2^TW_2$$

Which can be re-arranged to $x_2^TW_2 = t^T - x_1^TW_1$ and substituting $x_2^T = (t^T + \alpha^T)P_2^T$ the equation becomes:

$$(t^T + \alpha^T)P_2^TW_2 = (t^T - x_1^TW_1)$$

Which can be re-arranged to:

$$(t^T + \alpha^T) = (t^T - x_1^TW_1)(P_2^TW_2)^{-1}$$

This now can be re-substituted back in to $x_2^TW_2 = t^T - x_1^TW_1$:

$$x_2^T = (t^T - x_1^TW_1)(P_2^TW_2)^{-1} P_2^T$$

This is then repeated at every control interval of $\theta_i$. 

March 2013 280
Appendix 8

The minimisation of the cost function presented in equation 6-37 is shown in this section. In a similar manner to Appendix 7, the minimisation is conducted in the latent variable space (Lauri, et al., 2010).

Here $x_{\text{def}}(k)$ and $\hat{y}_f(k)$ is represented in the score space, $t_d$. At instant $k$:

$$t(k) = x(k)Z,$$ where $Z$ is decomposed.

$$t = [x_p \ x_f \ x_{\text{def}}] \begin{bmatrix} Z_p \\ Z_f \\ Z_{\text{def}} \end{bmatrix}$$

$$t = [x_pZ_p \ x_fZ_f \ x_{\text{def}}Z_{\text{def}}] \begin{bmatrix} \Delta p \\ \Delta f \\ \Delta_{\text{def}} \end{bmatrix}$$

$$t = [x_p + x_f + x_{\text{def}}]$$

If $u_{k+i}$ is set as $u_{k+n_f-1}$ for $i \in [n_u \ n_{f-1}]$ then;

$$x_f = x_{\text{def}} \begin{bmatrix} I_{n_1} \cdots I_{n_i} \\ 0(n_u - 1)n_i \times (n_f - n_u)n_i \end{bmatrix}$$

The decision vector in the latent variable space is set as $t_d = t_{\text{def}}$

Then $x_{\text{def}}$ can be shown as:

$$x_{\text{def}} = t_d(Z_{\text{def}}^T Z_{\text{def}})^{-1} Z_{\text{def}}^T$$

Next the expression of $\hat{y}_f$ is shown

$$\hat{y}_f = tBQ^T = (t_p + t_f + t_{\text{def}})BQ^T$$

$$\hat{y}_f = (x_pZ_p + t_{\text{def}})BQ^T$$

Then

$$x_f = t_d M_{\text{def}} \Gamma$$ and $t_f = t_d M_{\text{def}} \Gamma Z_f$

$$t_{\text{def}} = t_d M_{\text{def}} \Gamma Z_f + t_d(M_{\text{def}} \Gamma Z_f + I_{n_f})$$

$$\hat{y}_f = x_pZ_pBQ^T + t_d M_f BQ^T$$